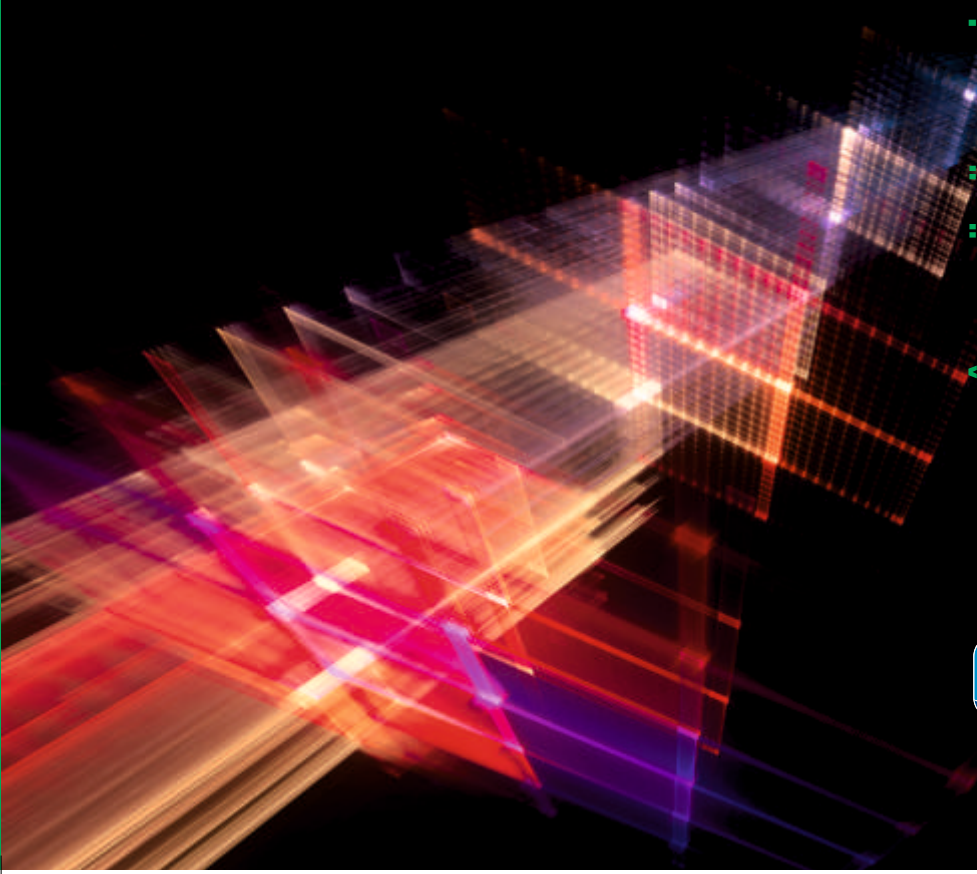
YORUMLANABİLİR VE AÇIKLANABİLİR YAPAY ZEKÂ VE GÜNCEL KONULAR

Editörler
Prof. Dr. Şeref Sağıroğlu
Dr. Mustafa Umut Demirezen



YAPAY ZEKÂ VE BÜYÜK VERİ KİTAP SERİSİ 4

nobel

Yapay Zekâ ve Büyük Veri Kitap Serisi 4:

Yorumlanabilir ve Açıklanabilir Yapay Zekâ ve Güncel Konular

Editörler

Prof. Dr. Şeref SAĞIROĞLU

Dr. Mustafa Umut DEMİREZEN

nobel

YAPAY ZEKÂ VE BÜYÜK VERİ KİTAP SERİSİ 4:
Yorumlanabilir ve Açıklanabilir Yapay Zekâ ve Güncel Konular
Editörler: Prof. Dr. Şeref SAĞIROĞLU - Dr. Mustafa Umut DEMİREZEN

Yayın No.: 4460
Mühendislik/Teknik No.: 442
ISBN: 978-625-427-476-3
E-ISBN: 978-625-427-477-0
Basım Sayısı: 1. Basım, Kasım 2022

© Copyright 2022, NOBEL AKADEMİK YAYINCILIK EĞİTİM DANIŞMANLIK TİC. LTD. ŞTİ. SERTİFİKA NO.: 40340

Nobel Yayın Grubu, 1984 yılından itibaren ulusal ve 2011 yılından itibaren ise uluslararası düzeyde düzenli olarak faaliyet yürütmekte ve yayınladığı kitaplar, ulusal ve uluslararası düzeydeki yükseköğretim kurumları kataloglarında yer almaktadır.

Genel Yayın Yönetmeni: Nevzat Argun -nargun@nobelyayin.com-
Genel Yayın Koordinatörü: Gülfem Dursun -gulfem@nobelyayin.com-
Alan Editörü: Sadık Küçükakman -sadik@nobelyayin.com

Sayfa Tasarım: Tarkan Kara -erdal@nobelyayin.com-
Redaksiyon: Sergen Öz -sergen@nobelyayin.com-
Kapak Tasarım: Sezai Özden -sezai@nobelyayin.com-
Görsel Tasarım Uzmanı: Mehtap Yürümez -mehtap@nobelyayin.com-
Baskı Sorumlusu: Yavuz Şahin -yavuz@nobelyayin.com-

Kütüphane Bilgi Kartı

Sağıroğlu, Şeref., Demirezen, Mustafa Umut.
Yapay Zekâ ve Büyük Veri Kitap Serisi 4: Yorumlanabilir ve Açıklanabilir Yapay Zekâ ve Güncel Konular / Şeref Sağıroğlu, Mustafa Umut Demirezen
1. Basım, XVIII+ 272 s., 16x23,5 cm. Kaynakça ve dizin var.
ISBN: 978-625-427-476-3
E-ISBN: 978-625-427-477-0

1. Yorumlanabilir Makine Öğrenmesi 2. Açıklanabilir Yapay Zekâ Metodolojileri ve Uygulamaları 3. Graf Algoritmaları ve Yapay Zekâya Katkıları 4. Kuantum Derin Öğrenme 5. Robot İşletim Sisteminde (RİS) Yapay Zekâ Temelli Anormal Durum Tespit Yaklaşımları 6. Eğitimde Yapay Zekâ, Sanal Gerçeklik ve Sanal Evren (Metaverse) 7. Büyük Veri ve Yapay Zekâ Tabanlı Kapalı Alan Konum Tespiti: İnavi 8. Güvenli Kodlamada Yapay Zekâ Kullanımı: Kod Karıştırma (Code Obfuscation) Yaklaşımları 9. Yapay Zekâ ile Çocuk Çizimleri Analizi

Genel Dağıtım

ATLAS AKADEMİK BASIM YAYIN DAĞITIM TİC. LTD. ŞTİ.
Adres: Bahçekapı Mh. 2465 Sk. Oto Sanayi Sitesi No:7 Bodrum Kat, Şaşmaz/ANKARA
Telefon: +90 312 278 50 77 - **Faks:** 0 312 278 21 65 - **Sipariş:** siparis@nobelyayin.com-
E-Satış: www.nobelkitap.com - esatis@nobelkitap.com / www.atlaskitap.com - info@atlaskitap.com

Dağıtım ve Satış Noktaları: Alfa Basım Dağıtım, Arasta, Arkadaş Kitabevi, D&R Mağazaları, Dost Dağıtım, Ekip Dağıtım, Kida Dağıtım, Kitapsan, Nezih Kitabevleri, Pandora, Prefix, Remzi Kitabevleri

Baskı ve Cilt

Meteksan Matbaacılık ve Teknik Sanayi Tic. Anonim Şirketi Sertifika No.: 46519
Beytepe Köy Yolu No.: 3 06800 Bilkent-Çankaya/ANKARA

BÖLÜM YAZARLARI

BÖLÜM 1
YORUMLANABİLİR MAKİNE ÖĞRENMESİ
Feyza Yıldırım Okay, Mustafa Yıldırım

BÖLÜM 2
AÇIKLANABİLİR YAPAY ZEKÂ METODOLOJİLERİ VE UYGULAMALARI
Utku Köse

Bölüm 3
GRAF ALGORİTMALARI VE YAPAY ZEKÂYA KATKILARI
Elif Eryaşar, Şerife Büyükköse

BÖLÜM 4
KUANTUM DERİN ÖĞRENME
Hande Çavşi Zaim, Metin Yılmaz, Esra Nergis Yolaçan

BÖLÜM 5
ROBOT İŞLETİM SİSTEMİNDE (RİS) YAPAY ZEKÂ TEMELLİ ANORMAL DURUM TESPİT YAKLAŞIMLARI
Hande Çavşi Zaim, Esra Nergis Yolaçan

BÖLÜM 6
EĞİTİMDE YAPAY ZEKÂ, SANAL GERÇEKLİK VE SANAL EVREN (METAVERSE)
Betül Ersöz, Halil İbrahim Bülbül

BÖLÜM 7
BÜYÜK VERİ VE YAPAY ZEKÂ TABANLI KAPALI ALAN KONUM TESPİTİ: İNAVİ
Murat Akın, Emrah Yılmaz, Tuncay Küçükpehlivan, Şeref Sağıroğlu

BÖLÜM 8
GÜVENLİ KODLAMADA YAPAY ZEKÂ KULLANIMI: KOD KARIŞTIRMA (CODE OBFUSCATION) YAKLAŞIMLARI
Çağrı Şahin

BÖLÜM 9
YAPAY ZEKÂ İLE ÇOCUK ÇİZİMLERİ ANALİZİ
S. Mine Okumuş, Pelin Canbay

ÖNSÖZ

Üniversitelerin ülkelerin bilimsel ve teknolojik gelişmelerine büyük katkılar sağladığı çok iyi bilinmektedir. İnsanlığın teknolojik ilerlemesinde büyük katkı sağlayan bu bilim alanı; bilimsel bilgiye dayalı araştırma, geliştirme, yönetme, eğitme, bilgilendirme, veriye dayalı karar verme, veri temelli üretim, bilgi ekonomisi ve sonuçta bilgi toplumu olmanın temel yapı taşlarından birisidir. Dijital değişim ve dönüşüme katkı sağlayan en önemli alanıdır. İçinde bulunduğumuz küresel salgın döneminde de insanlığa ve toplumlara katkı düzeyinin ne kadar yüksek olduğu alenen görülmüştür. Ayrıca, dijital becerilerin geliştirilmesi, dünyadaki dijital dönüşümün ve ekonomik büyümenin gerçekleşmesi, refah düzeyinin artması ve dijital ekonominin de ön koşulu olarak görülmektedir. Son yıllarda yapay zekâ ve büyük veri; ülkelerin kalkınma planlarını, politikalarını, stratejilerini, yönetimini; eğitimden sağlığa, ticaretten hizmet kalitesini artırmaya, ar-geden teknolojiye, verilerden değer üretme gibi pek çok alanı etkilemekte ve katkılar sunmaktadır.

Gazi Üniversitesi; ülkemizin ihtiyaç duyduğu pek çok alanda nitelikli uzmanlar yetiştiren, akredite olmuş pek çok programı bünyesinde barındıran, disiplinler arası çalışmaları, farklı üniversite ve kurum-kuruluşlarla paydaşlık ilişkilerini geliştirerek birçok üniversiteye rehberlik eden, bünyesinde Türkçe ve İngilizce eğitim veren programları da bulunan, farklı üniversitelerle birlikte lisansüstü eğitim verdiği Aselsan Akademi, yurtdışındaki üniversitelerle ortak diploma programları gibi işbirlikleri de bulunan ve bunu **96 yıldır** yapan bir **araştırma üniversitesidir**.

GÜ bünyesinde "büyük veri ve yapay zekâ" alanı destekleyen Bilgisayar Bilimleri, Bilgisayar Mühendisliği, Bilişim Sistemleri, Bilgi Güvenliği Mühendisliği, Adli Bilişim, Yazılım Mühendisliği gibi bu alanları desteleyen yüksek lisans ve doktora programlarını bünyesinde barındırmakta ve ülke bilimi ve teknolojilerinin geliştirilmesine katkılar sağlamaktadır. Yapay zekâ, büyük veri, siber güvenlik alanlarında ülkemizde en çok bilimsel faaliyet yapan da üniversitelerden birisidir. Ayrıca öncelikli alanlarda da eğitim-öğretim faaliyetlerini sürdüren Gazi Üniversitesi; ülkemizde ilk kez "büyük veri analitiği

laboratuvarı" bünyesinde 2021 yılında açmış; WoS verilerine göre Siber Güvenlik alanın da ülke bilimine katkı sağlamada 1. sırada, Yapay Zekâ konusunda 6. sırada ve büyük veri alanında ise ülke bilimine katkı sağlamada 5. sırada katkı sağlayan; Yapay Zekâ ve Büyük Veri Analitiği Güvenliği ABD (İngilizce) yüksek lisans programına sahip; ve Yapay Zekâ ve Büyük Veri Analitiği Güvenliği Uygulama ve Araştırma Merkezi 2021 yılında hizmete açmıştır. Bunlara ilave olarak bu alanları hem destekleyen hem de gelişimine katkı sağlayan diğer merkez ve programları da bünyesinde barındırmaktadır.

Üniversitemizin kabiliyet ve yetenekleri ile sahip olduğu altyapı ve faaliyetleri yukarıda kısaca özetlenmiştir. Bunun sebebi ise Yapay Zekâ ve Büyük Veri Analitiği Güvenliği Uygulama ve Araştırma Merkezinin belirli bir birikim ile kurulduğunu açıklamak içindir. Üniversitemizde bu konuda büyük bir bilgi birikimi ve potansiyel bulunmaktadır. Bu merkezin kurucu müdürü olarak, büyük veri ve yapay zekâ alanlarında üniversitemiz bundan sonrada bu alanlarda katkılarını sunmaya devam edecektir.

Editörlüğünü de yürüttüğüm "Yapay Zekâ Ve Büyük Veri Kitap Serisi" de bu alana üniversite olarak diğer bir katkımızdır. Gazi Üniversitesi Yapay Zekâ ve Büyük Veri Analitiği ve Güvenliği Uygulama ve Araştırma Merkezi (**Gazi AI R&D Center**) olarak ta bu tür faaliyetleri destekliyoruz. Kitap serisinin "**Yorumlanabilir ve Açıklanabilir Yapay Zekâ ve Güncel Konular**" isimli 4. cildine **Gazi AI R&D Center** olarak sponsorluk yapmaktan da ayrıca çok mutlu olduğumuzu belirtmek isterim.

Yapay Zekâ Ve Büyük Veri Kitap Serisinin hayata geçirilmesinde katkı sağlayan, başta editörlerimiz ve değerli yazarlarımız olmak üzere katkı verenlere teşekkür eder, faydalı olmasını dilerim.

Prof. Dr. Şeref Sağıroğlu

Yapay Zekâ ve Büyük Veri Analitiği Güvenliği Uygulama ve
Araştırma Merkez Müdürü

EDİTÖRLERDEN

Yapay Zekâ ve Büyük Veri Kitap Serisi fikri Cumhurbaşkanlığı Dijital Dönüşüm Ofisi ile yaptığımız Türk Beyin Projesi kapsamında yürütülen işbirliği projesi kapsamında ortaya çıkmıştır. Gerek Gazi Üniversitesinde yaptığımız çalışmalar gerekse ülkemizde bu konuda yapılan çalışmalar incelendiğinde, dünyada büyük veri ve yapay zekâ konusunda çok fazla açık kaynak veri, kod, kitap, proje, doküman var iken ülkemizde maalesef bu konuda kaynaklar olsa da bunların yeteri kadar genele açık olmadığı, yeteri kadar Türkçe kaynak bulunmadığı görüldüğünden, Büyük Veri ve Yapay Zekâ konusunda açık kaynak kitap serisi hazırlama projesi başlatılmış ve bu kapsamda bu kitap serisi çıkarılmaya başlanmıştır.

Açık kaynak yaklaşımların, ülkelerin gelişimine büyük katkı sağladığı, yapılan işlerin ve çalışmaların gerek nitelik gerekse niceliğini arttırdığı, dijital dönüşümün faydaya dönüştürülmesinde çarpan etkisine sahip olduğu ve ülkelerin bilimsel gelişmelerini hızlandırdığı bilinmektedir. Açık kaynak yaklaşımların bilimsel çalışmaların artmasına katkısının %65'lere kadar yükseldiği raporlanmaktadır. Foster Açık Bilim raporunda yer alan bir ankette akademisyenlerin %73'ü yayınlanan araştırma verilerine erişmenin kendi araştırmalarına katkı sağladığını belirtmişlerdir. Açık kaynak yaklaşımların katkılarını daha iyi anlatmak için dünyada önemli pek çok örnek vardır. Bilgisayar işletim sistemlerinde Linux, kelime işlemde OpenOffice, açık kaynak resim işleme tekniklerini yayımlayan OpenCV kütüphaneleri, Google sunduğu yüzlerce açık kaynak ve ücretsiz hizmetler (Google Çeviri, Google Akademik, Gmail, GoogleDrive, vb.), kod geliştirmeye verilen fırsatlar (Github, GitLab veya Bitbucket'te), verilerin tutulması (S3, Dataverse, Figshare, One Drive, Google Drive, Box, Dropbox), alıntılar (Mendeley, Zotero), belgelerin saklanması (OSF wiki alanında) bunlara verilebilecek örneklerden en önemlileridir.

Yukarıda belirtilen çalışmalar, ürünler, yazılımlar ve çıktılar verilerin paylaşımı veya oluşturulan açık kaynak ortamlar ile üretilmiş ve yaygınlaştırılmıştır. Veri paylaşımının artmasıyla çalışmalar çoğalmış, yeni çalışma ortamları oluşturulmuş, farklı çalışmalar desteklenmiş ve yeni başarılı modeller geliştirilmiştir. Ülkemizde de bu felsefenin yaygınlaşmasının ülke

olarak yaşadığımız hızlı değişim ve dönüşüme katkı sağlayacağı değerlendirilmektedir. Ülkemizin, Yapay Zekâ ve Büyük Veri konusunda ilerlemesi bu konuda yapılan yayınlar, farkındalık çalışmaları, ekosistemlerin oluşturulması, teknoloji geliştirme ortamları ve bu konuda üretilen içerikle doğru orantılıdır. Bu kitap serisi ülkemizde bu alanda yapılan, yapılması planlanan veya yapılacak olan çalışmaları bir çatı altında toplamak ve bunu kamuoyuna açık kaynak olarak sunmak için oluşturulmuş, gönüllülük esasına dayalı olarak geliştirilen açık kaynak içerik üretimi olup, üniversiteler başta olmak üzere bu alanlarda bilgi birikimi, deneyimi ve uzmanlığı olan araştırmacıları bir araya getirerek, yerli ve millî teknolojiler geliştirilmesine katkılar sağlamak amacıyla hazırlanmıştır.

Bu kitap serisi konuları belirlenirken, hazırlanırken ve basılırken aşağıdaki şekilde bir yol izlenmiştir.

1. Ülkemizde bu alanda çalışanlar uzmanlar, açık kaynak içerik üretilmesi için başlatılan "Yapay Zekâ Ve Büyük Veri Kitap Serisine" katkı vermeye davet edilmiştir.
2. Her alanda olduğu gibi son yıllarda stratejik konulardan birisi haline gelen Yapay Zekâ ve Büyük Veri konularında "açık erişim" felsefesi temel alınarak geliştirileceği herkese bildirilmiştir. Açık erişim; Libre açık erişim, Budapeşte Açık Erişim Girişimi, Berlin Fen Bilimleri ve İnsani Bilimlerde Bilgiye Açık Erişim Bildirgesindeki açık erişim tanımına uymakta olup "online araştırma çıktılarının erişiminde hiçbir kısıtlama olmamasına ve kullanımda da kısıtlamadan serbest olması anlamındadır".
3. Geliştirilecek olan bu projede hedeflenen hususlar ve kurallar tüm katılımcılarla paylaşılmıştır. Bunlar;
   - Türkçe kaynak oluşturma, yapılan çalışmaları kamuoyu ile paylaşma, farkındalığı arttırma, beraber çalışabilirliği geliştirme, Sektör-Kurum-Üniversite işbirliklerini büyütme,
   - Açık kaynak içerik geliştirme fikrini yaygınlaştırma,
   - Verilerden değer elde etme konusunda örnekleri ve uygulamaları arttırma,
   - Kişisel verilere saygı göstererek araştırmalar yapılmasına katkılar sağlama,
   - Kişisel, kurumsal ve ulusal bilgi varlıklarımızdan daha farklı değer elde etme ve

- Farklı teknolojilerin, fikirlerin, stratejilerin, uygulamaların, çalışmaların veya ortamların geliştirilmesine katkılar sağlamaktır.

4. Yapay Zekâ Ve Büyük Veri Kitap Serisi projesi kapsamında hazırlanacak olan kitap serisine ait her cilt için konu başlıklarına ait planlanan bilgiler aşağıda verilmiştir.

Yapay Zekâ ve Büyük Veri Kitap Serisi

Kitap 1: Yapay Zekâ ve Büyük Veri : Teknolojiler, Yaklaşımlar ve Uygulamalar (Bu kitap yayımlanmıştır.)

Kitap 2: Yapay Zekâ ve Büyük Veri : Endüstri 4.0 ve Yapay Zekâ

(Bu kitap yayımlanmıştır.)

Kitap 3: Yapay Zekâ ve Büyük Veri : Siber Güvenlik, Mahremiyet ve Uygulamalar

(Bu kitap yayımlanmıştır.)

Kitap 4: **Yorumlanabilir ve Açıklanabilir Yapay Zekâ ve Güncel Konular (Bu kitaptır.)**

Kitap 5: Açık Kaynaklar, Kodlar ve Platformlar

Kitap 6: Doğal Dil İşleme

Kitap 7: Bilgisayarlı Görü ve Uzaktan Algılama

Kitap 8: Derin Öğrenme Yapıları ve Algoritmaları

Kitap 9: Makine Öğrenme Yapıları ve Algoritmaları

Kitap 10: Tarım Uygulamaları

Kitap 11: Havacılık Uygulamaları

Kitap 12: Ulaşım Uygulamaları

Kitap 13: Sağlık Uygulamaları

Kitap 14: Endüstriyel Uygulamalar

Kitap 15: Gerçek Zamanlı Uygulamalar

Kitap 16: Sosyal Bilim Uygulamaları

Kitap 17: Eğitim Uygulamaları

Kitap 18: Haberleşme Uygulamaları

5. Yapay Zekâ Ve Büyük Veri Kitap Serisi projesine katkı vermek isteyen yazarlardan aşağıdaki bilgiler istenilmiştir.
   - İsteklilerin; başta alan uzmanlığını belirtmeleri ve daha sonra da hangi cilde veya ciltlere katkı vermek istediklerini belirtmeleri gereklidir.
   - Yukarıda önerilen alanlarda kitap yazılması planlandığından, doğru kitap serisine öneri yapılmalıdır.
   - Kitap bölümlerinin hazırlanmasına ait detay bilgiler yazarlara daha sonra gönderilecektir.
   - Hazırlanacak olan bölümlerde sayfa sınırlaması yoktur.
   - Yapılan önerilerde yazarların uzmanlığı dikkate alınarak seçim yapılır.
   - Ön başvurusu kabul edilen yazarlara, önerdiği bölümlerle ilgili olarak bir içerik hazırlaması istenilir ve uygun görülürse de yazım süreci başlar.
   - Kitap bölümü veya bölümlerini yazmayı isteyenlerin belirlenen takvim dikkate alınarak, en kısa sürede editörlere bilgi vermeleri gerekmektedir.
   - Önerilerinizi ss@gazi.edu.tr veya udemirezen@gmail.com adresine iletebilirsiniz.
6. Yapay Zekâ Ve Büyük Veri Kitap Serisi projesi hakkında bazı ön bilgilendirmeler aşağıda verilmiştir.
   - Proje kapsamında 18 başlık altında kitaplar yazılması planlanmaktadır.
   - Ülkemizde bu alanda yayınlanacak olan en kapsamlı kaynak olacaktır.
   - Kitaplar, akademik bir dille yazılacaktır.
   - Kitaplar, etik kurallara uygun olarak yazılacaktır.
   - Kitaplar için (her bir kitap cildi için) ISBN numarası alınacaktır.
   - Kitap bölümlerinde telifler “açık erişim” kapsamında olacaktır.
   - Kitaplar dijital olarak hazırlanacak ve kamuoyu ile paylaşılacaktır.
   - Kitaplar, Öğretim Elemanları Teşvik Programı kapsamına girecek şekilde basılacaktır.
   - Kitapların sponsorlukla basılması planlanmakta olup, yazarlara belirli sayıda basılı kopya hediye edilecektir.
   - Dağıtımlar, açık erişim çerçevesinde kamuoyuna dağıtılacaktır.
   - Bu kitap serisinin hazırlanması ve sunulması için bir web sayfası oluşturulacaktır. Bu sayfa daha sonra duyurulacaktır.

7. Yapay Zekâ Ve Büyük Veri Kitap Serisi projesine katkı sağlamak isteyen uzmanlarımızı bu seriye destek vermeye davet ediyoruz.

Bu kitap serisi oluşturulurken sunulan konu başlıkları, ülkemizde bu alanda çalışan akademisyenler, uzmanlar ve çalışanlar ile paylaşılmış ve bu kitap serisine katkı sağlamaları istenilmiştir. Zamanı uygun olan, katkı vermek isteyen uzman veya akademisyenler belirlenen bir konuda bölüm yazarı olmaları için davet edilmişlerdir. Belirlenen süre içerisinde bölümlerini tamamlayan yazarlarımızın eserleri ise uygun olan ciltlerde basılmaktadır. Bundan sonraki süreçte, belirlenen diğer konular belirli sürelerde tamamlanıp takip eden ciltlerde yayımlanacaktır. Yapay zekâ ve Büyük Veri konularını kapsamlı bir bakış açısıyla sunmayı amaçlayan ve farklı başlıkları bir araya toplayan bu eserin, ülkemizde yapay zekâ ve büyük veri konusunda yapılan çalışmalara katkı sağlaması beklenmektedir.

Bu kitap serimizin dördüncü cildinde, 9 farklı bölüm sunulmuştur. Büyük veri ve yapay zekânın kapsamlı ve farklı açılardan irdelendiği bu ciltte; yorumlanabilir, açıklanabilir yapay zekâ yaklaşımları, graf sinir ağları, kuantum sinir ağları, gerçek saha problemlerine büyük veri yaklaşımı, kod karıştırmada yapay zekâ kullanımı gibi konular kapsamlı olarak ele alınmış, teorik ve pratik çalışmalar verilmiş, karşılaşılabilecek yeni riskler ve problemler ve sonuçta alınması gereken önlemler ve yapılması gerekenler farklı alanlar için sunulmuş ve elde edilen sonuçlar değerlendirilmiştir. Her bir bölüm; bu alana katkı sağlayan, alanında eğitim almış, tez hazırlamış, çalışmalar yapmış değerli akademisyen, kamu çalışanı ve üst düzey yöneticiler tarafından hazırlanmıştır. Bölümler mümkün olduğunca birbirini takip eden sırada hazırlanmış, konu bütünlüğü ve devamlılığına dikkat edilerek sunulmuştur. Her bölüm tarafımızdan değerlendirilmiş, yazarlara konu içeriği ve başlıklarla ilgili olarak bazı önerilerde bulunulmuş, düzeltmeler yapılması istenilmiş ve sonuçta yapılan değişiklikler dikkate alınarak bu kitap hazırlanmıştır.

Bu kitabın, yapay zekâ ve büyük veri konularında yapılacak çalışmalara ışık tutması, yeni çalışmaların yapılmasına katkı sağlaması, bu konuda yapılacak olan işbirliklerini geliştirmesi ve en önemlisi ise ülkemizde bu alanda duyulan ihtiyacı bir nebze de olsa karşılaması, açık kaynak olarak sunulması ile de kaynaklara erişimi kolaylaştırıcı **bir başvuru kitabı serisi** olması beklenmektedir. Bu eser serisi **açık kaynak** olarak, gerek bu proje için hazırlanmış olan internet sayfasında gerekse üniversite ve ilgili kurumların internet sayfalarında yayımlanmaktadır.

Bu kitapta yazarlarımız; alan uzmanlıklarına göre kitap bölümlerini hazırlamışlar, literatürden faydalanmışlar, kişisel, kurumsal ve ulusal bilgi birikimlerini hazırladıkları bölümlerde sunmuşlar, hazırladıkları bölümlerin açık kaynak olarak yayımlanmasını kabul etmişler ve bu kitabın basımı ve dağıtımı ile ilgili olarak herhangi bir telif hakkı talep etmemişlerdir. Yazarlarımıza, bu yöndeki katkıları için kitap serisinin editörleri olarak çok özel teşekkürlerimizi ve şükranlarımızı sunarız.

Açık Kaynak, Açık Kod, Açık Bilimin gelişmesi dileğiyle..

Prof. Dr. Şeref SAĞIROĞLU (ss@gazi.edu.tr)

- Gazi Üniversitesi Yapay Zekâ ve Büyük Veri Analitiği Güvenliği Uygulama ve Araştırma Merkezi (Gazi AI Center) Müdürü
- Gazi Üniversitesi MF Bilgisayar Mühendisliği Bölümü Öğretim Üyesi
- Elektrik Mühendisleri Odası Ankara Şube Başkanı
- GÜ FBE Büyük Veri Analitiği Güvenliği ve Mahremiyeti ABD Kurucu Başkanı
- GÜ FBE Bilgi Güvenliği Mühendisliği ABD Kurucu Başkanı
- Gazi Üniversitesi Büyük Veri ve Siber Güvenlik Laboratuvarı (BIDISEC) Birim Sorumlusu
- IPv6 Forum Turkey Başkanı
- Bilgi Güvenliği Derneği Ulusal Bilim Kurulu Başkanı
- ISCTurkey Eş Başkanı
- UBMK Eş Başkanı
- Ülkem İçin Yapay Zekâ Derneği Kurucu Üyesi
- Akademik Bilişim Vakfı Kurucu Üyesi
- Bilgi Güvenliği Derneği Üyesi
- Türkiye Bilişim Derneği Üyesi
- Türk Bilim Araştırma Vakfı Üyesi
- Geleceği Önemseyenler Derneği Kurucu Üyesi

Dr. M. Umut DEMİREZEN (umut@demirezen.tech)

Ağustos 2022 tarihinden itibaren Huawei Turkey R&D, Intelligent Applications DC Departmanında, Principal Engineer olarak, Innovation and Research Manager olarak görev yapmaktadır. Daha önce Roketsan A.Ş.'de Yönetici Baş Mühendis olarak Yapay Zekâ alanında araştırma ve geliştirme faaliyetlerini 2018-2022 yılları arasında yürütmüştür. Daha da önce Savunma Teknolojileri ve Mühendislik Anonim Şirketi (STM A.Ş.) de Yapay Zekâ Teknolojileri Grup yöneticisi olarak görev yapmış olan Dr. M. Umut DEMİREZEN, yine STM A.Ş.'de, Siber Güvenlik ve Büyük Veri ArGe Grup Yöneticiliği görevini yürütmüştür. Lisans, Yüksek Lisans ve Doktora çalışmalarını sırasıyla 2000, 2003 ve 2015 yıllarında tamamlamıştır. 2000 yılında Türk Deniz Kuvvetleri'ne Mühendis Teğmen olarak katılmış ve daha sonra sırasıyla Genelkurmay MEBS Başkanlığı, Genelkurmay İstihbarat Başkanlığı ve Başbakanlık'ta görev yaparak, 2013 yılında STM A.Ş.'de SSB adına Kıdemli Danışman olarak çalışmalar yürütmüştür. Siber Güvenlik, Büyük Veri, Veri Bilimi ve Derin Öğrenme alanlarında birçok projede araştırmacı ve yönetici olarak görev alan Demirezen, alandan bağımsız olarak milli büyük veri ve siber güvenlik teknolojilerinin geliştirilmesi göreviyle birçok projede aktif olarak görev almış ve STM Siber Füzyon Merkezinin kurulumu ve faaliyete geçirilmesi çalışmalarında bulunmuştur. Ayrıca yapay zekâ teknolojilerinin askeri ve sivil alanlarda kullanımı üzerine aktif olarak araştırmalarını sürdürmekte olup, veri bilimi teknikleri ile uygulama ve kullanım alanlarından bağımsız olarak veri analitiği ve akıllı optimizasyon yöntemleri konularında çalışmalarına devam etmektedir.

2018 yılından itibaren ROKETSAN A.Ş.'de Yapay Zekâ teknolojilerinin farklı alanlarda geliştirilmesinden sorumlu olarak görev yapmış olup, Sürü Zekâsı (Swarm Intelligence) ve Otonom Sistemler için Yapay Zekâ alanında araştırma ve geliştirme çalışmalarını yürütmektedir. Sürü Zekâsı alanında gerçekleştirilen AR-GE çalışmalarını teknik olarak yönetmiş ve Sürü Zekâsı Tabanlı Otonom Drone Teknolojilerinin ülkemize kazandırılmasına öncülük etmiştir. Halen İstanbul Teknik Üniversitesi Aerospace Research Center (ARC)'de Aerospace alanında yapay zekâ teknolojilerinin uygulanması konularında araştırmalar yürütmekte olup, İTÜ Havacılık ve Uzay Mühendisliği,

Gazi Üniversitesi Bilgisayar Mühendisliği bölümlerinde Yapay Zekâ, Veri Bilimi ve Büyük Veri Konularında Lisans ve Lisansüstü dersler vermektedir. Aynı zamanda T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, Yapay Zekâ ve Büyük Veri Teknolojileri Birim Müdürlüğü görevini de eş zamanlı olarak yürütmüştür. 2021-2022 yılları arasında Yapay Zekâ Politikaları Derneği (AIPA) Danışma Kurulu Üyeliği görevinde bulunmuş olup, Eylül 2022 yılından itibaren Yapay Zekâ Politikaları Derneği (AIPA) Başkan Yardımcılığı görevini yürütmektedir.

İÇİNDEKİLER

Bölüm Yazarları iii
Editörlerden v

BÖLÜM 1
YORUMLANABİLİR MAKİNE ÖĞRENMESİ 1
Feyza Yıldırım Okay, Mustafa Yıldırım

1.1. GİRİŞ 2
1.2. YORUMLANABİLİR MAKİNE ÖĞRENMESİ (INTERPRETABLE MACHINE LEARNING-IML) 5
1.3. IML YÖNTEMLERİN SINIFLANDIRILMASI 9
1.4. IML'DE KULLANILAN YÖNTEMLER 11
2.4.1. Görsel tabanlı IML Yöntemler 12
2.4.2. Pertürbasyon tabanlı IML Yöntemler 13
1.5. IML'İN HEDEF KİTLESİ VE KULLANIM ALANLARI 15
2.5.1. Sağlık 16
2.5.2. Finans 18
2.5.3. Otonom Sistemler 19
1.6. ÖRNEK UYGULAMA: SHAP İLE SAĞLIK VERİSİNİN YORUMLANMASI 24
2.6.1. SHAP ile Yerel Yorumlanabilirlik 27
2.6.2. SHAP ile Global Yorumlanabilirlik 31
1.7. IML YÖNTEMLERİN KALİTESİNİ ÖLÇME VE DEĞERLENDİRİLME KRİTERLERİ 32
1.8. IML UYGULAMALARINDA KARŞILAŞILAN ZORLUKLAR 35
1.9. SONUÇ VE DEĞERLENDİRMELER 36
KAYNAKLAR 38

BÖLÜM 2
AÇIKLANABİLİR YAPAY ZEKÂ METODOLOJİLERİ VE UYGULAMALARI 43
Utku Köse

2.1. GİRİŞ 44
2.2. BİR KARA KUTU OLARAK YAPAY ZEKÂ 45
2.2.1. Yapay Zekâ ve Yorumlanabilirlik 47
2.2.2. Açıklanabilir Yapay Zekâ 48
2.2.3. Açıklanabilir Yapay Zekânın Uygulama Alanları 51
2.3. SAYDAM TASARIM (YORUMLANABİLİRLİK) 53
2.3.1. Benzetilebilirlik 54
2.3.2. Ayrıştırılabilirlik 54
2.3.3. Algoritmik Saydamlık 55
2.3.4. Yorumlanabilir Makine Öğrenmesi Modelleri 56

2.4. KARA KUTUNUN AÇIKLANMASI 60
2.4.1. Katmanlı Sinir Ağlarında ve Karma Modellerde Açıklanabilirlik 61
2.4.2. LIME Yöntemi ile Açıklanabilirlik 65
2.4.3. SHAP Yöntemi ile Açıklanabilirlik 67
2.4.4. Derin Öğrenme'de Açıklanabilirlik 70
2.5. SONUÇ VE DEĞERLENDİRMELER 73
KAYNAKLAR 76

Bölüm 3
GRAF ALGORİTMALARI VE YAPAY ZEKÂYA KATKILARI 81
Elif Eryaşar, Şerife Büyükköse

3.1 GİRİŞ 82
3.2. EN KISA YOL BULMA ALGORİTMALARI 85
3.1.1. Dijkstra Algoritması 85
3.2.2. Bellman-Ford Algoritması 89
3.2.3. Floyd-Warshall Algoritması 91
3.3. EN KÜÇÜK YOL AĞACI BULMA ALGORİTMALARI 93
3.3.1. Prim Algoritması 94
3.3.2. Kruskal Algoritması 96
3.3.3. Sollin Algoritması 99
3.4. GRAF DOLAŞMA ALGORİTMALARI 100
3.4.1. DFS Algoritması 100
3.4.2. BFS Algoritması 101
3.5. GRAF RENKLENDİRME ALGORİTMASI 103
3.6. SONUÇ VE DEĞERLENDİRMELER 105
KAYNAKLAR 106

BÖLÜM 4
KUANTUM DERİN ÖĞRENME 109
Hande Çavşi Zaim, Metin Yılmaz, Esra Nergis Yolaçan

4.1. GİRİŞ 110
4.2. KUANTUM HESAPLAMANIN TEMELLERİ 111
4.3. KUANTUM DERİN ÖĞRENME YAPILARI 115
4.3.1. Kuantum Sinir Ağı (QNN) 116
4.3.2. Kuantum Evrişimli Sinir Ağı (QCNN) 121
4.3.3. Kuantum Tekrarlayan Sinir Ağı (QRNN) 122
4.3.4. Kuantum Çekişmeli Üretici Ağlar (QGAN) 123
4.4. KUANTUM DERİN ÖĞRENME PLATFORMLARI 126
4.5. SONUÇ VE DEĞERLENDİRMELER 127
KAYNAKLAR 128

BÖLÜM 5
ROBOT İŞLETİM SİSTEMİNDE (RİS) YAPAY ZEKÂ TEMELLİ ANORMAL DURUM TESPİT YAKLAŞIMLARI 133
Hande Çavşi Zaim, Esra Nergis Yolaçan

5.1. GİRİŞ 134
5.2. ROBOT İŞLETİM SİSTEM (RİS) MİMARİSİ 135
5.3. ROBOT İŞLETİM SİSTEMİNDE (RİS) GÖRÜLEN ANORMAL DURUMLAR 136
5.4. YAPAY ZEKÂ TABANLI ANORMALLİK TESPİTİ YAKLAŞIMLARI 138
5.5. ROBOT İŞLETİM SİSTEMİNDE UYGULANAN ANORMAL DURUM TESPİT YAKLAŞIMLARI 140
5.6. SONUÇ VE DEĞERLENDİRMELER 144
KAYNAKLAR 147

BÖLÜM 6
EĞİTİMDE YAPAY ZEKÂ, SANAL GERÇEKLİK VE SANAL EVREN (METAVERSE) 149
Betül Ersöz, Halil İbrahim Bülbül

6.1. GİRİŞ 150
6.2. EĞİTİMDE YAPAY ZEKÂ 153
6.3. EĞİTİMDE SANAL VE ARTIRILMIŞ GERÇEKLİK 156
6.4. EĞİTİMDE SANAL EVREN (METAVERSE) 160
6.4.1. Web 3.0 ve Blok Zinciri 168
6.4.2.Akıllı Sözleşmeler 169
6.4.3. Nitelikli Fikri Tapu (Non-Fungible Token- NFT) 171
6.4.4. Kripto Varlıklar 171
6.5. SANAL EVREN (METAVERSE), SİBER GÜVENLİK VE OLASI TEHDİTLER 172
6.7. SONUÇ VE DEĞERLENDİRMELER 176
KAYNAKLAR 180

BÖLÜM 7
BÜYÜK VERİ VE YAPAY ZEKÂ TABANLI KAPALI ALAN KONUM TESPİTİ: İNAVİ 185
Murat Akın, Emrah Yılmaz, Tuncay Küçükpehlivan, Şeref Sağıroğlu

7.1. NAVİGASYON VE NAVİGASYONUN ÖNEMİ 186
7.2. İÇ MEKAN KONUMLAMA ÇALIŞMALARI 187
7.2.1. Veri Toplama 188
7.2.2. Harita oluşturulması 188
7.2.3. Konum Tespiti 189
7.2.4. Rota Oluşturma ve Yönlendirme 190
7.3. GELİŞTİRİLEN SİSTEM MİMARİSİ 192
7.3.1. Sistemde Kullanılacak Veri Türleri 193
7.3.2. Alt Sistem ve Yapılar 193
7.3.3. Veri Depolama ve Gereklilikler 195
7.4. DENEYSEL SONUÇLAR 195
7.4.1. Haritanın Elde Edilmesi 196

7.4.2. Veri Toplanması 197
7.4.3. Konum Tespiti 202
7.5. SONUÇ VE DEĞERLENDİRMELER 206
KAYNAKLAR 207

BÖLÜM 8
GÜVENLİ KODLAMADA YAPAY ZEKÂ KULLANIMI: KOD KARIŞTIRMA (CODE OBFUSCATION) YAKLAŞIMLARI 211
Çağrı Şahin

8.1. GİRİŞ 212
8.2. GÜVENLİ KODLAMA 213
8.3. KOD KARIŞTIRMA NEDEN ÖNEMLİ 214
8.4. TERSİNE MÜHENDİSLİK 215
8.4.1. Programlama Dillerine Göre Tersine Mühendislik 215
8.4.2. Makine Öğrenmesi ve Tersine Mühendislik 216
8.5. KODLARA YAPILAN SALDIRILAR 217
8.6. KOD KARIŞTIRMA TEKNİKLERİ 218
8.7. KOD KARIŞTIRMA VE YAPAY ZEKÂ 227
8.7.1. Kod Karıştırma Teknikleri Tespiti 227
8.7.2. Kod Benzerliği Analizi 229
8.8. GÜVENLİ KODLAMA VE YAPAY ZEKÂ 230
8.8.1. Kod Oluşturma 230
8.8.2. Kod Güvenliği 233
8.8.3. Sistem Güvenliği 234
8.9. SONUÇ VE DEĞERLENDİRMELER 236
KAYNAKLAR 238

BÖLÜM 9
YAPAY ZEKÂ İLE ÇOCUK ÇİZİMLERİ ANALİZİ
S. Mine Okumuş, Pelin Canbay

9.1. GİRİŞ 243
9.2. ÇOCUK ÇİZİMLERİ ÜZERİNE YAPILAN ANALİZLER VE YORUMLAMALAR 246
9.2.1. Bir Ağaç Çiz Testi 247
9.2.2. Goodenough-Harris Bir Adam Çiz Testi 248
9.2.3. Bir Aile Çiz Testi 250
9.2.4. Duygu ve Renk Çemberi Testi 251
9.3. ÇOCUK ÇİZİMLERİ ANALİZİNDE KULLANILAN YAPAY ZEKÂ YÖNTEMLERİ 253
9.3.1. Bir Kişi Çiz Testi için Yapay Zekâ Destekli Çözüm Örneği 254
9.3.2. Bir Ağaç Çiz Testi için Yapay Zekâ Destekli Çözüm Örneği 255
9.4. SONUÇ VE DEĞERLENDİRMELER 257
KAYNAKLAR 258

DİZİN 261

Bölüm 1

YORUMLANABİLİR MAKİNE ÖĞRENMESİ

Feyza Yıldırım Okay - Mustafa Yıldırım

Yorumlanabilir makine öğrenmesi (Interpretable Machine Learning-IML), makine öğrenmesi algoritmalarının kara kutu gibi davranmasına bir çözüm olarak geliştirilen ve işleyişi hakkında kullanıcılarına ek bilgiler sunan yenilikçi bir yaklaşımdır. Bu kitap bölümünde, IML ve diğer benzer tanımlamalar, IML'nin önemi, popüler IML yöntemleri ile bu yöntemlerin kullanım alanları, ölçme değerlendirme metrikleri ve IML kullanımının getirdiği zorluklar üzerinde durulmaktadır. Konunun daha iyi anlaşılması için popüler IML yöntemlerinden biri olan SHAP ile göğüs kanseri verisinin yorumlanmasına yönelik örnek bir uygulama ile ilgili detay bilgiler verilmiştir. Son olarak, IML tekniklerinin gelecekteki yeri, önemi ve gelecek çalışmalara katkıları değerlendirilmiş, bu konuda yeni çalışmak isteyenlere de önerilerde bulunulmuştur.

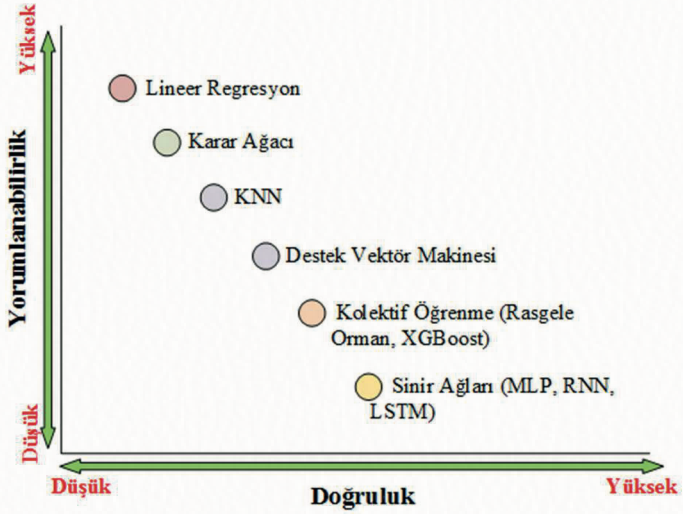
1.1. GİRİŞ

Yapay zekâ (YZ) konusundaki güncel gelişmeler ile birlikte Makine Öğrenmesi (Machine Learning-ML) teknikleri daha güçlü öngörücüler haline gelmekte ve buna bağlı olarak hastalık teşhisi, film tavsiyesi, gözetleme sistemi veya saldırı tespiti gibi farklı uygulama senaryolarında ML teknikleri yaygın olarak kullanılmaktadır. ML, tutarlı ve güvenilir tahminler yapmak için karmaşık örüntüleri öğrenmede oldukça başarılı bir tahmin yöntemidir. ML teknikleri, Şekil 1.1'deki gibi gözetimli, gözetimsiz ve takviyeli öğrenme olmak üzere 3 alt gruba ayrılır. Her bir ML yöntemi, geçmişten günümüze farklı alanlarda farklı sorunlara büyük çözümler ve fırsatlar sunmaktadır. Özellikle, kolektif (ensemble) ve derin öğrenme modellerinin güçlendirilmesiyle birlikte film tavsiyesi, makine çevirisi, konuşma tanıma [1] gibi birçok gerçek dünya uygulamalarında büyük şirketler tarafından tercih edilir hale gelmiştir. ML modelleri öğrenme sürecinde hız, tekrarlanabilirlik ve ölçekleme açısından insanlara göre üstünlüğe sahiptir. Bu modeller uygulandıktan sonra bir görevi daha hızlı tamamlayabilir, tutarlı sonuçlar üretebilir ve başka bir makineye kolayca ve ucuza aktarılabilirler. Öte yandan, bir insanın bir görevi tamamlaması için eğitilmesi hem uzun zaman alır hem de kişi genç ise ek maliyet gerektirir. Ayrıca, kişi bilgisi kopyalanarak veya aktarılarak çoğaltılamaz [2].

Şekil 1.1. ML algoritmalarının sınıflandırılması [3]

ML metotlarında doğruluk ve yorumlanabilirlik arasında ters bir ilişki vardır. Bu metotlar; yüksek doğrulukta tahminlemeler sağlamasına rağmen

açıklanabilirlikten, şeffaflıktan veya anlaşılabilirlikten uzaktır. Diğer bir ifadeyle, modeller bir kara kutu gibi davranarak, belirli kararlara nasıl ulaşılacağı konusunda kullanıcılara yeterli bilgi vermemektedir. Şekil 1.2'de görüldüğü gibi model tahmin doğruluğu arttıkça yorumlanabilirlik azalmaktadır. Diğer durumda ise model tahmin doğruluğu azaldıkça yorumlanabilirlik artmaktadır. Günümüzde tahmin başarısı yüksek algoritmalar daha fazla tercih edilmektedir. Ancak, özellikle sağlık, finans, otonom sistemler gibi kritik senaryolarda model tahmin başarısının yanı sıra kararı etkileyen sebeplerin de bilinmesi son derece önemli hale gelmektedir. Bu durum örneğin bir film tavsiye sistemi için çok önemli bir problem teşkil etmezken, bir hastalık teşhis sistemi veya kredi skoru belirleme sisteminde kritik öneme sahiptir.



Şekil 1.2. ML algoritmalarında yorumlanabilirlik ve doğruluk arasındaki ters ilişki

Yorumlanabilir Makine Öğrenmesi (IML), karmaşık ML tekniklerinin anlaşılmazlık/şeffaf olmayış sorununa umut verici bir çözüm sunan yeni bir teknolojidir [2]. Özellikle, yüksek başarıya sahip ancak yorumlanabilirliği az karmaşık ML modellerine uygulanarak bu modellerin yorumlanabilirliğinin artması hedeflenmektedir. IML sayesinde, ML iç işleyişinin belirli kararlara nasıl yol açtığı konusunda şeffaflık sağlar ve kullanıcıların karar verme sürecinden haberdar olmalarını sağlar. Böylece, model doğruluk başarısı sabit kalırken, her bir model için daha yüksek yorumlanabilirlik sağlanmış olur.

Bu kitap bölümü 9 bölümden oluşmaktadır. Kitabın 2. Bölümünde, IML yöntemlerinin tanımı, önemi ve avantajları sunulmuştur. 3. bölümde, IML yöntemlerinin hangi şekilde sınıflandırıldıkları detaylı olarak açıklanmaktadır. Sonrasında, 4. bölümde literatürde yaygın olarak kullanılan IML yöntemleri tanıtılmaktadır. 5. Bölümde, IML yöntemlerinin hedef kitlesi ve yaygın olarak kullanıldıkları alanlar sıralanmıştır. 6. bölümde ise popüler IML yöntemlerinden biri ile örnek bir uygulama gerçekleştirilmiştir. 7. bölümde, IML'in kalitesini ölçmede ve değerlendirmede kullanılan kriterler açıklanmıştır. 8. Bölümde, IML'in zorluklarından bahsedilmiş ve elde edilen sonuçlar ile IML konseptinin gelecekteki potansiyeli konusunda genel değerlendirilmeler sunulmuştur.

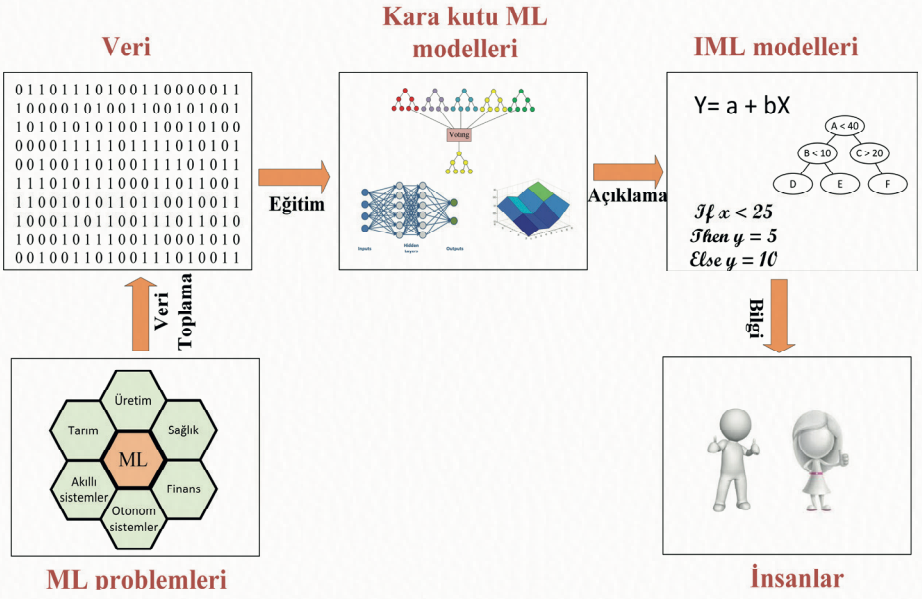
Konunun daha iyi takip edilmesi için ise kitap bölümünün organizasyonu Tablo 1.1'de özet olarak verilmektedir.

Tablo 1.1. Bölüm Organizasyonu

Bölüm Adı	İçerik
Bölüm 1.1. Giriş	Makine öğrenmesi Yorumlanabilir makine öğrenmesi (IML) Motivasyon Organizasyon
Bölüm 1.2. IML	Tanımı Önemi Avantajları Farkındalık
Bölüm 1.3. IML yöntemlerin sınıflandırılması	Model-öncesi, model-içi, model-sonrası Global, lokal Modele-özgü, modelden-bağımsız
Bölüm 1.4. IML'de kullanılan yöntemler	Görsel tabanlı yöntemler; Pertürbasyon tabanlı yöntemler
Bölüm 1.5. IML'in hedef kitlesi ve kullanım alanları	Sağlık Finans; Otonom sistemler
Bölüm 1.6. Örnek uygulama	SHAP ile sağlık verisinin yorumlanması
Bölüm 1.7. IML yöntemlerin kalitesini ölçme ve değerlendirme kriterleri	Kalitesini ölçmede kullanılan özellikler; Değerlendirme kriterleri
Bölüm1.8. IML uygulamalarında karşılaşılan zorluklar	Kısıtlar ve zorluklar
Bölüm 1.9. Sonuç ve değerlendirmeler	Sonuç ve Değerlendirmeler

1.2. YORUMLANABİLİR MAKİNE ÖĞRENMESİ

Yorumlanabilir Makine Öğrenmesi (IML) terimi yeni bir kavram olmasına rağmen, herhangi bir uzman sistemin sonucunu açıklama problemi 1970'lerden beri çözülmeye çalışılan oldukça eski bir problemdir. Son zamanlarda ise özellikle 2016'dan sonra literatürde yeniden ilgi görmeye başlamıştır. IML terimi, literatürdeki uygulayıcılar ve akademisyenler tarafından Açıklanabilir Yapay Zekâ (XAI) ile birbirinin yerine kullanılabilmektedir. Çoğu çalışma, ML modellerinin açıklanabilirliğine odaklanır ve bu modeller genellikle IML olarak adlandırılır. Bu iki kavram birbirinin yerine kullanılabilir olsa da, IML, XAI'nın ML algoritmalarındaki mantığını açıklamaya odaklanan önemli bir parçasıdır. Bilim dünyasında 'yorumlanabilir' terimi 'açıklanabilir' teriminden daha çok tercih edilmektedir. Öte yandan, genel kullanım olarak bakıldığında açıklanabilir teriminin daha çok kullanıldığı da görülmektedir [4].



Şekil 1.3. IML modellerin genel akışı [2]

Bu iki kavram dışında benzer anlamlara karşılık gelen literatürde farklı tanımlamalar mevcuttur. Tablo 1.2'de IML ve bu kavrama benzer olarak kullanılan diğer kavramlar listelenmektedir.

Tablo 1.2. IML ve benzer kavramlar [4]

Kavram	Tanım
Yorumlanabilir ML (IML)	Kullanıcıların hem görüp hem de girdi çıktı ilişkilerini nasıl eşleştirebildiğini açıklayan sistemlerdir. Birçok çalışmada ise XAI ile aynı anlamlara gelmektedir.
Açıklanabilir YZ (XAI)	Yapay zekânın şeffaflık ve güven endişelerine yönelik yapılan hareket ve girişimlerdir. DARPA [5]'ya göre yüksek performans sürdürülürken yorumlanabilirliğin sağlanmasıdır. Bu kavram, IML kavramına oldukça yakın bir anlam içermektedir.
Sorumlu YZ	Toplumsal değerleri, ahlak ve etik değerleri içine alan bir yapay zekâ kavramıdır. Temel noktaları, hesap verilebilirlik, sorumluluk ve şeffaflık ilkelerine dayalıdır.
Doğru YZ	Modelin tüm tahminleri arasından doğru yaptığı tahminleri dikkate alan bir performans metriğidir.
Üçüncü Nesil YZ	Modelin yeni görevlerle karşılaştıkça öğrendiği ve akıl yürüttüğü ve insanlarla iletişim içerisinde olduğu, kısacası sürekli öğrenen bir açıklayıcı model oluşturmayı amaçlar
Yapay Genel Zekâ	İnsanın yapabildiği zekâ gerektiren görevleri başarıyla gerçekleştirebilen makinelerin tasarlanmasını amaçlamaktadır.

Rastgele ormanlar, derin öğrenme ve gradyan artırma yöntemleri gibi en güçlü ancak karmaşık ML modelleri, lineer regresyon, karar ağacı gibi basit ML modellerine kıyasla, çeşitli gerçek dünya sınıflandırma, regresyon ve tahmin problemlerinde doğruluk açısından daha iyi performansa sahiptir. Ancak, aynı zamanda kullanıcılarına nasıl karar verecekleri ve tahmin yaparken hangi belirli kararların verildiği konusunda da muğlaktırlar. Bu, yalnızca sonucun değil, aynı zamanda sonucu etkileyen kararların nasıl verildiğinin de önemli olduğu bazı senaryolarda zorlu bir konudur.

ML modellerinin yorumlanabilirliğinin olmaması, örneğin sağlık alanında ML modellerinin yaygın olarak benimsenmesini sınırlar. Model belirli kararlar üzerinde tasarlanıp değerlendirildiğinden ve genellikle sınırlı istatistiksel ve ML bilgisine dayandığından, klinisyenler ML modellerine pek güvenmezler [6]. Örneğin bir hastaya kanser teşhisi konulabilir. Buna göre tedavi yöntemine karar verilmesi istenebilir. Ancak hasta veya doktor, kanser tanısında hangi sendromların daha etkili veya etkisiz olduğuna dair bilgi-

lere ulaşamaz. Tedavinin özellikle yanlış teşhis edildiğinde hastaya daha fazla zarar verme olasılığı da vardır. Sonucun insan yaşamı gibi son derece kritik olduğu durumlarda, modelin daha açık ve yorumlanabilir olması gerekir. Finansta da kredi bazlı puan modellerinde bu durumla karşılaşılmaktadır. Finansal kurumlar, bireylere mi yoksa şirketlere mi borç verecekleri konusunda yasaların öngördüğü adalete sahip olmalıdır. Buna dayanarak, borçluların kredilerinin neden reddedildiğine dair daha ayrıntılı açıklamalar - neden kodları - gereklidir. Otonom sistemler için IML modellerinin önemi düşünüldüğünde ise, örneğin sürücüsüz kontrol sistemiyle yönetilen bir arabanın yoldan geçen bir yayaya çarpması durumunda tüm kusur otonom sistemde gibi gözükür. Ancak, otonom aracın neden o şekilde davrandığı ile ilgili bir gerekçeye varılamaz. IML ile otonom aracın davranış şekline sebep olan ihtimaller daha iyi bir şekilde analiz edilip, hatanın neden kaynaklandığını bulmak çok daha kolay ve anlaşılır olmaktadır.

Algoritmaların yorumlanabilir olmasının önemli diğer bir kritik gerekçesi ise algoritmanın eğitim setinde beklenmedik sapmalar (bias) yapma ihtimalidir. Örneğin, bir kredi başvuru kabul uygulamasında algoritmanın bir bölge, grup veya ırktan olan başvuruları doğrudan reddetmesi gibi yanlılık problemleri doğurabilir. Eğitim veri setinde; bölge, grup, cinsiyet veya ırkla ilgili değişken olmasa bile algoritma eğitimi sırasında bir gruba olumlu veya olumsuz yanlılık yapma durumu olabilir; ML algoritmalarının yorumlanabilir olabilmesi ile yeni bilgiler de açığa çıkabilir; algoritma yorumlanabilirse daha önce fark edilmemiş önemli ek bilgi ve örüntüleri keşfetme imkânı sağlayabilir.

IML modellerine duyulan ihtiyaç, problemin türüne veya önemine göre tartışmalı bir konudur. Bunt ve ark. [7] çalışmalarında "Açıklamalar Her Zaman Önemli midir?" sorusunu gündeme getirmişlerdir. Buna göre, kullanıcının modellerin davranış kalıplarını anlamasının önemi değerlendirilirken, bu analizi gerçekleştirmenin ekstra maliyeti de dikkate alınmalıdır. Örneğin, bir film tavsiye sistemi için yorumlanabilirliğin maliyeti, faydalarından daha ağır basabilirken, diğer yandan, afet yönetimi veya teşhis sistemleri gibi kritik uygulamalar için kararın arkasındaki nedenleri bulmak ise kritik önem taşımaktadır. Doshi-Velez ve Kim [8] raporunda; eğer hatalı sonuçların uygulama alanına etkisi çok şiddetli olmadığında ya da problem yeterince iyi çalışılıp gerçek uygulamalar üzerinde doğrulandıysa sistem hatalı olsa bile sisteme güvenilebilir. Bu tip durumlarda ise ML algoritmasının yorumlanabilirliği kritik olmamakta ve sadece yüksek performanslı sonuçlar yeterli olmaktadır.

IML'yi kullanımının üstünlükleri Şekil 1.4'te verildiği gibi dört temel başlıkla gruplandırılabilir [4].

(i) ML modellerinin iç işleyişinin bir tanımını vermek yerine sonuçları doğrulamayı amaçlar. Beklenmeyen bir karar alındığında, sonuçlar daha adil ve güvenilir kararlar olması için daha fazla gerekçe ile desteklenir.

(ii) ML modellerinin davranış kalıplarını inceler ve kontrol eder. Ardından, öngörülemeyen hataları ve kusurları hızlı bir şekilde belirleyebilir ve sorunsuz bir karar verme süreci için bunları düzeltebilir.

(iii) ML modellerinin insanlar tarafından daha anlaşılır hale getirilmesini sağlayarak iyileştirilmesini sağlar. İnsanlar tarafından daha net bir akıl yürütme zinciriyle, daha akıllı kararlar almak daha kolay hale gelir.

(iv) ML sorunlarına yeni bakış açıları kazandırmaya ve gizli gerçekleri ve bilgileri keşfetmeye yardımcı olur.

Şekil 1.4. IML'in üstünlükleri [4]

IML konusunun ML dünyasındaki farkındalığı hem kamusal hem de endüstriyel boyuttadır. Amerika başta olmak üzere dünyanın birçok ülkesinde yapay zekânın açıklanabirlirliği konusunda çalışmalar başlatılmıştır. Bunlardan en bilineni Defense Advanced Research Projects Agency (DARPA) [5], akademik ve askeri araştırılmalarda fayda sağlayacak açıklanabilir yapay zekâ programlarını finanse etmiştir. Ayrıca, yapay zekâların açık, şeffaf olması ve insanlar tarafından anlaşılabilir olmasına yönelik çeşitli raporlar sunulmuştur [9] .

Bunun dışında, Almanya Manifestosu [10] ve Fransa Başkanlığı [11] tarafından yapılan açıklamalarda YZ modellerinin sadece yüksek doğrulukla değil, verilen kararlara nasıl ulaştığı ve hedef kitlelere karşı şeffaflık sağlamaları konusunda çeşitli açıklamalar yapmışlardır. İngiltere Bilim akademisi [12], benzer şekilde yayınladığı bir raporda makine öğrenimi uygulamalarıyla ilişkili sosyal konular olarak yorumlanabilirlik ve şeffaflığın yanı sıra sorumluluk ve hesap verebilirliğin önemini kabul etmektedir.

Yorumlanabilirlik konusundaki endüstriyel çalışmalar ise başta Google [13] olmak üzere, H2O[14], DataRobot [15] gibi popüler firmalar yorumlanabilirlik konusunu ticarileşmiş ML çözümleri ve ürünlerindeki odak noktalardan biri haline getirmiştir. IBM benzer şekilde IBM Watson AIOpenScale [16] isimli YZ platformunun güven, şeffaf ve açıklanabilir özelliklerini vurgulamıştır.

1.3. IML YÖNTEMLERİN SINIFLANDIRILMASI

Literatürdeki IML yöntemleri, tekniğin ne zaman uygulanabilir olduğuna (model öncesi, model içi, model sonrası), örneklerin kapsamına (yerel, global) göre farklı şekillerde sınıflandırılabilir [2, 17] ya da belirli bir modele özgü olup olmadığına göre sınıflandırılabilmektedir (Şekil 1.5).

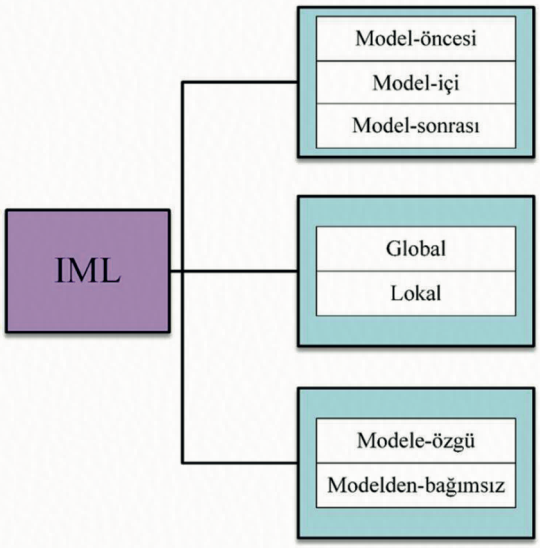
Yorumlanabilirlik modelinin oluşturulmasından önce, sırasında veya sonrasında uygulanabilirliğine bağlı olarak IML teknikleri üçe gruba ayrılır:

(i) Model-öncesi (pre-model) teknikler: Veriler ML modeline eğitilmeden önce uygulanır. Öğrenme aşamasının ilk adımı olarak sapmaları ortadan kaldırır, verilerin keşfedilmesini ve anlaşılmasını sağlar. Modeli eğitmek için kullanılan anlamlı sezgisel özellikleri çıkararak yorumlanabilirlik ve seyreklik sağlar. Model öncesi tek-

nikler, öznitelik seçimi, kümeleme ve bilgi kazancı, k-means, PCA gibi bazı istatistiksel yöntemleri içerir.

(ii) Model-içi (in-model) teknikler: Şeffaflık olarak adlandırılan bu teknikler modelin nasıl çalıştığı sorusunu yanıtlar. Bu teknikler özünde yorumlanabilir olan (kısıtlamalar yoluyla veya değil) ML modelleriyle ilgilidir. Seyreklik, monotonluk, nedensellik ya da alan bilgisinden gelen fiziksel engeller gibi kısıtlamalar yoluyla modelin karmaşıklığına içsel yorumlanabilirlik getirir.

(iii) Model-sonrası (post-model) teknikler: Agnostik olarak da adlandırılan model-sonrası teknikler, eğitim sürecinden sonra modelin belirli yöntemlerle analiz edilmesini ve yorumlanmasını sağlar. Lipton'un çalışmasına [18] göre bu teknikler model bize başka ne anlatıyor sorusunu cevaplar.



Şekil 1.5. IML Tekniklerin Sınıflandırılması

Bir diğer popüler kategorizasyon ise yorumlanabilirliğin kapsamına, yani global ve yerel yorumlanabilirliğe dayanmaktadır.

(i) Yerel teknikler: Tek bir örnek üzerinde tahmin yapmaya odaklanır ve modelin bu örnek üzerindeki tahmine nasıl ulaştığını anlamaya

çalışır. Genelleyecek olursak, belirli tahminlerin neden ve nasıl yapılabileceğine ilişkin yerel bir anlayış sağlar. Bu vekil model, optimal bir çözüm sağlamamakla birlikte, yorumlanabilirliği korurken oldukça iyi bir yakınsama sağlar. Bunun arkasındaki mantık ise, yerel olarak, tahminin bazı özelliklere karmaşık bir bağımlılıktan ziyade yalnızca doğrusal veya monoton olarak bağlı olabileceğidir.

(ii) Global teknikler: Modeli bir bütün olarak tanımlamayı amaçlar. Daha spesifik olarak, bir modelin, veri özelliklerinin ve öğrenilen bileşenlerin her birinin, örneğin ağırlıkların, parametrelerin bütünsel bir görünümüne dayalı olarak nasıl kararlar aldığını kavramayı gerektirir. Özellikle hangi özelliklerin daha önemli olduğu ve bunlar arasında ne tür etkileşimlerin gerçekleştiği sorusuna cevap aramaktadır.

Son olarak, belirli bir model sınıfı için uygulamaya bağlı olarak modele-özgü ve modelden-bağımsız olarak iki gruba ayrılabilir. **Modele-özgü teknikler**; belirli model sınıflarıyla sınırlıdır. Doğrusal bir modelde regresyon ağırlıklarının yorumlanması modele-özgü bir yorumdur, çünkü özünde yorumlanabilir modellerin yorumu her zaman modele-özgüdür. Yalnızca belirli bir modelin (örn. sinir ağları modele) yorumlanması da benzer şekilde modele-özgüdür. Bu teknikler yaygın olarak, oldukça kapalı bir yapıya sahip olmasına rağmen üstün bir tahmin performansına sahip olan Derin Sinir Ağları için tasarlanmaktadır [19, 20]. **Modelden-bağımsız teknikler**; model sınıflarını sınırlamadan herhangi bir ML modeli için kullanılabilir. Bu yöntemlerin en büyük avantajı yöntemi herhangi bir modele bağımlı olmadan farklı senaryolar için tekrar kullanılabilmesidir. Model-spesifik yöntemlere göre, modelin ağırlıkları ya da yapısal bilgileri gibi iç çalışmalarına erişim sağlayamaz. Daha çok girdi çıktı analizlerini yapar [2].

1.4. IML'DE KULLANILAN YÖNTEMLER

IML kavramının yapay zekâ dünyasında popülerlik kazanmasıyla birlikte, farklı IML yöntemleri geliştirilmeye başlanmıştır. Bunlar arasında literatürde en yaygın olarak kullanılan iki IML yöntemi, SHAP ve LIME'dır. Bunla-

ra ilave olarak; Anchors, LORE, ILIME, DLIME, MAPLE, CIU, FI, DT, MATS, Doctor XAI, PDP, ICE, ALE, GIRP, GINN, SHAP+GIRP, Proto-Dash, AlphaPortfolio ve Lasso vb. yöntemler de vardır (Bakınız Şekil 1.3). Bu bölümde literatürde en çok tercih edilen post-model tekniklerini metodolojilerine göre ayırarak inceleyeceğiz. Buna göre, IML yöntemler, görsel tabanlı ve pertürbasyon tabanlı olmak üzere ikiye ayrılmaktadır.

1.4.1. Görsel tabanlı IML Yöntemler

Görsel tabanlı IML teknikler, modelin yorumlanması görevini kolaylaştırmak için öznitelik ilişki tekniklerinin çıktısını görselleştirmek için ilginç yollar önerir. Bu yöntemlerin tasarımı, içyapısı göz ardı edilerek herhangi bir ML modeline sorunsuz bir şekilde uygulanabilmelerini sağlamak zorunda olduğundan, açık olmayan bir modelden yalnızca girdilerden ve çıktılardan görselleştirmeler oluşturmak karmaşık bir iştir. Bu nedenle bu teknikler daha az tercih edilmektedir [21].

i) Kısmi Bağımlılık Grafiği (PDP)

Kısmi bağımlılık grafiği (PDP) [22], hedef fonksiyon ile birkaç hedef özellik arasındaki bağımlılığı ortaya çıkarır. Eşitlik 1.1'den görülebileceği gibi, kısmi bağımlılık $f_{xs}(xs)$ Monte Carlo yöntemi olarak da bilinen eğitim verilerindeki ortalamaların hesaplanmasıyla tahmin edilir. Hedef özellikler için bir grid kurulduktan sonra, eğitim setindeki tüm hedef özellikler grid noktalarının değeri olacak şekilde ayarlanır. Sonrasında tahminler yapılır ve her bir grid için ortalamalar hesaplanır. PDP'nin dezavantajı, bir hedef özelliğinin 2B grafikler oluşturması ve ikisinin de 3B grafikler oluşturmasıdır. Öte yandan, bir insanın daha yüksek boyutlarda grafikleri anlaması oldukça zor olabilir.

$$f_{xs}(xs) = \frac{1}{n}\sum_{1}^{n} f(x_s, x_c^i) \tag{1.1}$$

Burada, s ve c indeksleri tahmin edicinin birer altkümesidir.

ii) Bireysel Koşullu Beklenti (ICE)

Bireysel Koşullu Beklenti (ICE) [23], PDP'ye benzer daha yeni bir yöntemdir. Aradaki fark, PDP'nin marjinal dağılım üzerinden ortalamayı hesap-

larken, ICE hepsini tutmaktadır. ICE grafiğindeki her satır, her bir birey için tahminleri temsil eder. ICE, tüm örneklerin ortalamasını almadan, heterojen ilişkileri ortaya çıkarır. Ancak sadece tek bir hedef özellik ile sınırlıdır. Bunun sebebi, iki özellik insan gözüyle tanımlanamayan üst üste binen yüzeylerle sonuçlanmaktadır.

iii) Birikmiş Yerel Etkiler (ALE)

Birikmiş Yerel Etkiler (ALE) [24], tahminlerdeki değişikliklerin ortalamasını alır ve bunları yerel gridler üzerinde biriktirir. PDP ile farkı, ALE eğrisinin her noktasındaki değerin, tüm gridden ziyade küçük bir pencerede hesaplanan ortalama tahminin farkı olmasıdır. Böylece ALE, bağlantılı ilişkilerin özelliklerinin etkisini ortadan kaldırır.

1.4.2. Pertürbasyon tabanlı IML Yöntemler

Pertürbasyona dayalı algoritmalar, girdilerin farklı varyasyonları ile eğitimli bir makine öğrenimi modelinin yinelemeli olarak soruşturulmasıyla oluşturulur. Bu bozulmalar, açıklama, bulanıklaştırma, kaydırma veya maskeleme vb. için belirli özellikleri sıfır veya rastgele karşı olgu örnekleriyle değiştirerek veya bir grup piksel (süper piksel) seçerek özellik düzeyinde olabilir [25].

i) LIME

LIME [26]; ML algoritmalarının veri seti içindeki tekil örneklerinin yorumlanmasında kullanılan bir model olup, tekil örneklerin açıklanmasında yorumlanabilir vekil bir model kullanır. Bu vekil model lineer regresyon, karar ağacı gibi yorumlanabilir bir modeldir. Çalışması sezgiseldir. ML algoritmalarını tamamen bir kara kutu olarak kabul eder. ML algoritmalarının girdilerini ve tahminlerini alır. Amacı, ML'nin bir örneklem için belirli bir tahmini neden yaptığını anlamaktır. Bu amaç doğrultusunda seçilen örnekleme yakın olan yeni örneklemler oluşturulur. Yakınlık derecesine göre ağırlıklandırılır. LIME, ağırlıklandırılmış yeni veri setinde yorumlanabilir vekil modelle eğitilir. Vekil modelin değişkenlere verdiği ağırlıklarla yorumlanabilirlik sağlanmış olur. LIME, matematiksel olarak aşağıda verilmiştir [27]:

$$explanation(x) = \arg\min_{g \in G} L(f, g, \pi_x) + \Omega(g) \quad (1.2)$$

x: açıklanmak istenen örnek,

g: lineer regresyon veya karar ağacı gibi vekil model,

f: rastgele orman gibi tahminleri yapan asıl model,

π_x ise x örneğinin yakınlık ölçeği,

L: en küçük ortalama kare hatası gibi kayıp fonksiyonunu minimum yapan fonksiyon

$\Omega(g)$: vekil modelin öznitelik sayısını düşük tutmak için kullanılan fonksiyondur. Ancak genellikle pratikte kullanıcı öznitelik sayısını uzman kendi belirler.

ii) **SH**apley **A**dditive Ex**P**lanations (SHAP)

SHAP, literatürdeki karmaşık ML modellerini açıklamak için kullanılan popüler bir yaklaşımdır. Lundberg ve Lee [27] tarafından Shapley değerlerine dayalı bir oyun-teorik yaklaşım olarak önerilmiştir. SHAP bir örneğin niteliklerini çıktıya açıklamayı amaçlar. Bu amaçla, Shapley değerleri, takım oyunlarında bir oyuncunun takım puanına marjinal katkısını ölçmeye çalışır. Shapley değerleri, takımdaki her bir oyuncunun takımdan birer birer çıkarıldığında toplam puanın nasıl değiştiğini kontrol eder. Böylece her oyuncunun skora olan marjinal katkısı ölçülebilir. Ölçme tek tek çıkarılarak gerçekleştirildiği için, oyuncu sayısındaki artışla birlikte hesaplama karmaşıklığı katlanarak artmaktadır. Benzer şekilde, SHAP her bir özniteliği bir oyuncu olarak kabul eder ve her bir özniteliğin tahmin sonucu üzerindeki etkilerini ölçerek açıklar. SHAP, LIME ve Shapley değerlerini bütünleştirir ve aşağıdaki gibi biçimlendirilebilir [27].

$$g(z') = \varphi_0 + \sum_{j=1}^{M} \varphi_j z'_j \quad (1.3)$$

g: açıklanabilir model olmak üzere M maksimum öznitelik sayısı,

$z' \in [0,1]$ M basitleştirilmiş yeni veri setidir.

$\varphi_j \in R$ olmak üzere j özniteliğinin tahmine olan katkısı, yani Shapley değeridir. Shapley değeri φ_i aşağıdaki formülle hesaplanmaktadır.

$$\varphi_i = \sum_{S \subseteq M \setminus i} \frac{|S|!(|M|-|S|-1)!}{|M|!}\left[f(S \cup i) - f(S)\right] \qquad (1.4)$$

φ_i: öznitelik i kullanılarak yapılan tahmin ile i olmaksızın yapılan tahmin arasındaki farkı göstermektedir. Böylelikle i özniteliğinin marjinal katkısı belirlenmiş olmaktadır.

S: i özniteliği olmaksızın oluşturulan bir alt öznitelik kümesidir. S i öznitelik altkümesine i'nin eklendiği durumu,

M: öznitelikleri ve

$S \subseteq M \setminus i$ ise i özniteliği hariç olası alt kümeleri göstermektedir.

1.5. IML'İN HEDEF KİTLESİ VE KULLANIM ALANLARI

IML yöntemleri son zamanlarda popüler hale gelen kavramlardan biridir. ML yöntemlerinin kara kutu gibi davranmasına yönelik eleştirilere bir çözüm sunan yaklaşımlardan biridir. Arrieta ve ark. çalışmalarında [21] yorumlanabilirliği hedef kitleleri de dikkate alarak yeniden yorumlamışlardır. Buna göre, IML kavramını, "Bir hedef kitle göz önüne alındığında, işleyişini net veya anlaşılması kolay hale getirmek için ayrıntılar veya nedenler üreten bir zekâdır." olarak tanımlamışlardır. Çalışmalarında IML'in hedef kitlelerini beş ana gruba ayırtmaktadırlar. Bunlar;

(i) alan uzmanları / model kullanıcıları,

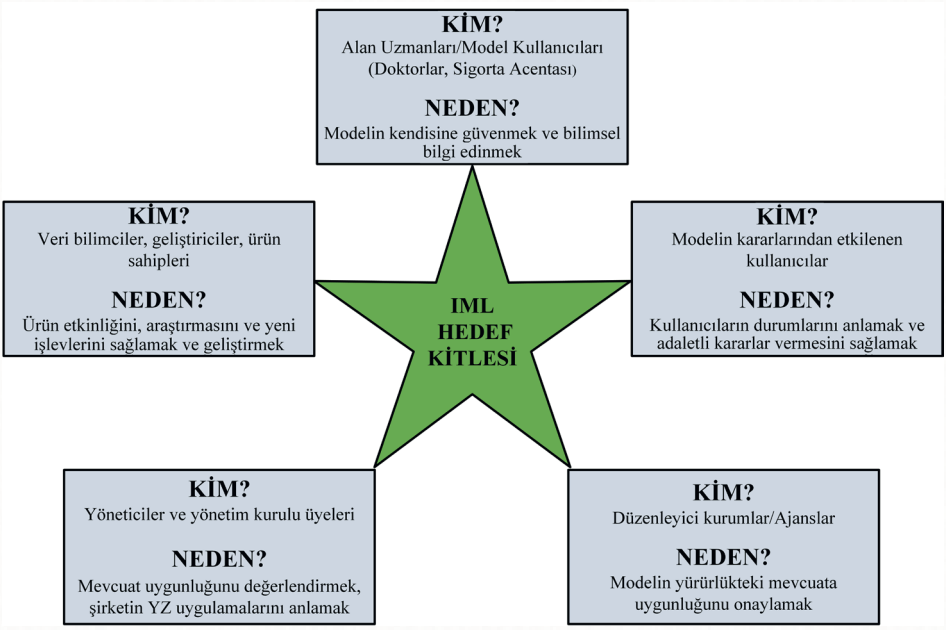
(ii) modelin kararlarından etkilenen kullanıcılar,

(iii) düzenleyici kurumlar / ajanslar,

(iv) yöneticiler / yönetim kurulu üyeleri, ve

(v) veri bilimciler, geliştiriciler ve veri sahipleridir.

Her bir hedef kitle için Şekil 1.6'dan görülebileceği gibi "Neden model yorumlamaya ihtiyaç duyarız?" kısmını açıklamaktadır.



Şekil 1.6. IML'in hedef kitlesi [21]

IML kavramı literatürde oldukça yeni bir kavram olduğundan çalışma alanları sınırlıdır. Ancak, sağlık finans gibi sadece sonucun değil sonuca giden adımların da önemli olduğu kritik uygulamalarda son zamanlarda yaygın bir şekilde kullanılmaktadır.

1.5.1. Sağlık

Geleneksel makine öğrenimi tekniklerinin sağlık alanlarına yayılmasıyla birlikte, bu modellerin yorumlanması da uygulayıcılar açısından büyük önem kazanmıştır. Sağlık uygulayıcıları, IML kullanarak, erken teşhiste en önemli göstergeleri derhal ortaya çıkarmak için önceki tıbbi bilgileriyle modellerden gelen öngörüleri birleştirebilir. ElShawi ve ark. [28] LIME, Anchors, SHAP, LORE, ILIME ve MAPLE'den oluşan altı popüler yerel post-hoc IML modelini, farklı veri setlerinde kimliklendirme, kararlılık, ayrılabilirlik, benzerlik, zaman, önyargı tespiti ve güven tespiti metriklerine göre değerlendirmişlerdir. Gerçek sağlık veri setleri üzerinde farklı yerel IML modelleri karşılaştırılarak kapsamlı bir analiz gerçekleştirilmiş-

tir. Sonuçlara göre, MAPLE değerlendirme ölçütlerinin çoğunda üstünlüğe sahiptir. Panigutti ve ark. [29] sağlık alanında çok etiketli sıralı ve ontoloji bağlantılı verileri (MIMIC-III) ele alan ilk post-hoc IML modeli olan Doctor XAI'yi önermişlerdir. Açıklama sonuçları, sadakat, isabet ve açıklama karmaşıklığı açısından değerlendirilmektedir. Sonuçlar, ontolojik bilgiden yararlanmanın açıklama performansını arttırdığını göstermektedir. Zafar ve ark. [30], sağlık durumunun belirlenmesinde LIME'den daha kararlı açıklamalar elde etmeyi amaçlayan deterministik LIME (DLIME) önermektedir. Çalışmalarında, eğitim verilerini farklı kümelere bölmek için hiyerarşik kümelemeden yararlanmışlardır. Sonuçlar LIME ile karşılaştırıldığında DLIME ile daha tutarlı açıklamalar elde edilmektedir. Knapic ve ark. [31] SHAP, LIME ve Contextual Value and Utility (CIU) gibi farklı post-hoc IML modellerini karşılaştırarak video kapsül endoskopisinden toplanan gastral görüntüler üzerinden açıklama sürelerini ve desteklerini değerlendirmişlerdir. Sonuçlar, CIU'nun kararı daha hızlı yorumlamada SHAP ve LIME'den daha iyi olduğunu açıkça göstermektedir. Ayrıca, karar vermede daha fazla desteğe sahiptir. Hakkoum ve ark. [32] meme kanseri teşhisi için LIME, Özellik Önemi (FI) ve Kısmi Bağımlılık Grafiğini (PDP) dürüstlük açısından ölçmüşlerdir. Radyal Temel Fonksiyon Ağının (RBFN) doğruluk sonuçları Çok Katmanlı Algılayıcıdan (MLP) daha yüksek olmasına rağmen, MLP'nin yorumlanabilirliği DT sınıflandırıcı ile daha tutarlıdır. Ayrıca, anlama düzeyini artırmak için global ve yerel yorumlanabilirliği birleştirmeyi önerirler. Wu ve ark. [33] çalışmalarında Covid-19 hastalığının teşhisi konulmuş 92 kişilik bir veritabanından yararlanarak Covid-19 ile biyo belirteçleri belirlemek için enfeksiyon şiddet tahminini yapmışlardır. Buna göre, karar ağacı, rastgele orman, Gradient Boosting ve sinir ağları olmak üzere 4 farklı ML algoritması üzerinde çeşitli IML yöntemlerini uygulamışlardır. Sonuçlar, bu veriler üzerindeki güncel tıbbi araştırmalar ve modellerle tutarlı olduğu görülmüştür. Ayrıca, IML modeller daha büyük ve açık başka veri setlerinde de denenerek sonuçları doğrulanmıştır. Sonuç olarak, N-Terminal pro-Brain Natriuretic Peptide (NTproBNP), C-Reaction Protein (CRP), lactic dehydrogenase (LDH), ve lymphocyte (LYM)'nin COVID-19 hastalarının şiddet seviyesini gösteren en önemli dört biyo-belirteç olduğu gösterilmiştir.

1.5.2. Finans

Yapay zekâ alanındaki son yenilikler, uzmanların anlayışlı ve daha doğru kararlar vermelerini sağladığından, kredi puanlaması yüksek popülerlik kazanmıştır. IML yöntemleri ile hem doğru hem de yorumlanabilir kararlar alınabilir. Ancak literatürde IML ile finansal konuları ele alan sınırlı sayıda çalışma bulunmaktadır. Demajo ve ark. [34] kredi puanlamasını tahmin etmek için XGBoost algoritması üzerinde üç farklı açıklamaya (global, yerel özellik tabanlı ve yerel örnek tabanlı) olanak tanıyan 360 derecelik bir açıklama çerçevesi önermektedir. Global açıklamalar için SHAP+GIRP, yerel özellik tabanlı açıklamalar için Anchors ve yerel örnek tabanlı açıklamalar için ProtoDash yöntemlerini kullanmıştır. Önerilen modeller, Home Equity Line of Credit (HELOC) ve Lending Club (LC) veri kümeleri üzerinde uygulanmıştır. Sonuçlar, modeli global olarak yorumlamak için Boole kuralları üreten BRCG yöntemi ile karşılaştırılır. Ayrıca önerilen modeli tutarlılık, basitlik, doğruluk, etkinlik, kolay anlaşılırlık, detay yeterliliği ve güvenilirlik açısından analiz etmek için işlevsel temelli, uygulama temelli ve insan temelli olmak üzere üç farklı değerlendirme yaklaşımı benimsenmiştir. Bussmann ve ark. [35] paydaşlar arası kredilendirmede kredi riskini tahmin etmeye çalışmışlardır. Bu sorunu çözmek için, sınırlı yorumlanabilirlik ile yüksek tahmin başarısına sahip XGBoost'u önermişlerdir. Çalışmalarında, Shapley değeri, her bir şirketin kredibilitesi açısından modele yorumlanabilirlik sağlamak için kullanılmıştır. Önerilen model, Avrupa Dış Kredi Değerlendirme Kurumu'ndan (ECAI) alınan veriler üzerinde eğitilmiş ve ROC eğrileri değerlendirilerek optimal lojistik regresyon ile karşılaştırılmıştır. Paydaşlar arası kredi talebinde bulunan 15.000 küçük ve orta ölçekli şirketin deneysel analizlerine bakıldığında, hem riskli hem de riskli olmayan borçluların, kredi puanlarını açıklamak ve anlamak için kullanılabilecek bir dizi benzer finansal özelliğe göre gruplandırılabildiği görülmektedir. Bu sayede, kişilerin gelecekteki finansal davranışlarını da tahmin etmek mümkündür. Chen ve ark. [36] diğer sinir ağlarına benzer doğruluk sağlayan, 'iki katmanlı toplamalı risk modeli' olarak bilinen global olarak yorumlanabilir bir model önermişlerdir. Model, özelliklerin alt gruplarından daha küçük modellerin oluşturulduğu ve nihayetinde nihai varsayılan olasılığı üretmek için birleştirildiği alt ölçeklere ayrıştırılabilir yapıdadır. Modelin ayrıştırılabilir do-

ğası, global yorumlanabilirlik için kural tabanlı özet açıklamalar ve yerel yorumlama için yerel vaka tabanlı açıklamalar sağlayarak, modelin tahminlerine en çok katkıda bulunan faktörlerin listesini tanımlayan anlamlı bileşenler üretmesine izin vermektedir. Ito ve ark. [37], finansal belgeleri görselleştirmek için Gradyan Sinir Ağı (GINN) modeli önermişlerdir. Böylece uzman olmayanların bile piyasa duyarlılığını kolayca anlayabilmelerini amaçlamışlardır. Yahoo Finance Board'daki gönderilerden ve Japon finans haber makalelerinden iki farklı veri seti toplanmıştır. Önerileri, temel MLP ve artı MLP ile karşılaştırıldığında F-skoru açısından üstünlük göstermektedir. Ayrıca, GINN tarafından elde edilen duygu puanları ile kişilerin algıları arasındaki uyumu ölçmek için insan yorumlanabilirlik testleri yapılmıştır. Cong ve ark. [38] performansı örneklem dışı bir şekilde artırmak ve sağlamlık sağlamak için AlphaPortfolio adlı takviye tabanlı bir portföy yönetim modeli tanıtmışlardır. Ayrıca yazarlar, ML modellerinin daha şeffaf ve yorumlanabilir olmasını sağlayan polinom duyarlılık analizi yaparak modeli doğrusal bir modele yansıtmışlardır. Önerilen model, CRSP, CRSP CompustatMerged ve Financial Ratio Firm Level olmak üzere üç farklı WRDS veritabanı üzerinde yürütülmüştür. Deneysel sonuçlara göre, önerilen modelin, özellikle bazı finansal kısıtlamalar tarafından kısıtlandığında, mevcut ML modellerine göre daha iyi performans göstermektedir. Ayrıca, karşı-olgusal açıklama gibi bazı IML yöntemleri, özelliklerin sayısı arttığında zorlayıcı olabildiği gözlemlenmiştir. Grath ve ark. [39], önemli ve en yakın komşu temelli stratejiler olan pozitif karşı-olgusal ve iki ağırlıklı karşı-olgusal öne rerek bu sorunun üstesinden gelmeyi amaçlamışlardır. HELOC kredi başvuru veri seti üzerinde gerçekleştirilen deneyler, ağırlıklı karşı-olgu oluşturma stratejisinin daha yorumlanabilir kararlar alınmasını sağlarken daha küçük karşı-olgular önererek temel karşı-olgu stratejilerine göre daha iyi bir performans sağladığını göstermektedir.

1.5.3. Otonom Sistemler

Franco ve ark. [40] anlamlı tahminler ve açıklamalar sağlamak üzere karar ağaçlarından yararlanarak otonom sistemler için yorumlanabilir bir gerçek zamanlı gözetleme sistemi oluşturmuşlardır. Bu gözetleme sistemi

sayesinde, sistemin durumunu sürekli olarak kontrol eder ve güvenli düzeltme eylemi sağlamak için karar ağacı tarafından tahmin edilen mantıksal açıklamadan yararlanır. Önerilen modelin başarısı hem simülasyon hem de otonom sistem operasyonları için fizibilite ve uygulanabilirliğini gösteren deneylerde test edilmiştir. Schmidt ve ark. [41] ise otonom araçlarda güvenli ve etkin davranış politikası oluşturmak için yeni bir işlem hattı sağlamışlardır. Buna göre, bu işlem hattı mevcut ödül yapılarını ve eğitim tekniklerini değiştirerek güvenli davranış için yorumlanamayan bir Güçlendirmeli Öğrenme (RL) ajanını eğitmektedir. Ardından, bir karar ağacı şeklinde politikaya yakınsayan bir dizi kural çıkarmaktadır. Sonuç olarak, çıkarılan bu politikanın derin sinir ağı (DNN) ajanı kadar performanslı ve güvenli olduğu ve aynı zamanda kolayca yorumlanabilir olduğu gösterilmiştir.

Otonom sistemlerde davranış tahmininin ana kullanımlarından biri, robot hareket planlaması ve kontrolü hakkında bilgi vermektir. Ancak, planlama ve kontrol algoritmalarının çoğu, entegrasyonlarını engelleyebilecek olan yörünge tahmin yöntemleri tarafından yaygın olarak çıkarılan ajan izciklerinden (yani sıralı yol noktaları kümelerinden) ziyade sistem dinamikleri hakkında akıl yürütmektedir. Bu amaçla, Ivanovic ve ark. [42] akış aşağı planlama ve kontrol kullanımı için daha uygun olan yörünge tahmini için bir çıktı temsili olarak Afin Zamanla Değişen Sistemlerin Karışımlarını (MATS) önermiştir. Sonuçta, MATS ile lineer olmayan gerçek dünya sistemleri için bile geçerli bir yörünge tahmin gösterimi sunulmaktadır. Ayrıca, böyle bir tahmin gösteriminin, eklenen yapı nedeniyle ham tahmin performansında küçük bir gerileme pahasına, çok modlu tahminler hesaba katıldığında bile, aşağı akış planlama ve kontrol karmaşıklığında önemli azalmalar sağladığı gösterilmiştir.

IML yöntemlerinin genel kullanım alanlarına bakıldığında sağlık ve finans alanında sıklıkla kullanıldığı görülmesine rağmen, son zamanlarda otonom sistemler başta olmak üzere diğer alanlara da yayıldığı görülmektedir. IML yöntemlerin kullanıldıkları alanlara göre özetlenmiş hali Tablo 1.3'deki gibidir.

Tablo 1.3. Kullanım Alanlarına göre IML Yöntemler

Uygulama Alanı	Çalışma	Amaç	Veri seti	IML yöntemleri	Makine öğrenmesi modelleri	Kapsamı	Değerlendirme Metrikleri	Önemli Bulgular
Sağlık	ElShawi ve ark. [28]	Ölüm ve diyabet teşhisi	Ölüm ve diyabet veri setleri ile uyuşturucu incelemesi ile yan etkileri veri setleri	LIME, Anchors, SHAP, LORE, ILIME, MAPLE	RF	Yerel	Kimliklendirme, kararlılık, ayrılabilirlik, benzerlik, zaman, önyargı tespiti, güven	MAPLE yöntemi çoğu metrik için diğerlerinden üstün performans göstermiştir.
	Panigutti ve ark. [29]	Hastaların sonraki ziyaret tahmini	MIMIC-III veri seti	Doctor XAI	RNN with GRU	Yerel	Sadakat, isabet, açıklama karmaşıklığı	Ontolojik bilgiyi kullanmak açıklama performansını artırır.
	Zafar ve ark. [30]	Göğüs kanseri teşhisi	Meme kanseri, karaciğer hastaları ve hepatit hastaları	DLIME, LIME	RF, NN	Yerel	İstikrar	DLIME kararlı açıklamalar üretirken, LIME kararsız nesiller üretir.
	Knapic ve ark. [31]	Karar destek sistemi	Video kapsül endoskopisinden toplanan gastral görüntüler	SHAP, LIME, CIU	CNN	Yerel	Zaman, destek	CIU, kararı daha hızlı yorumlamada SHAP ve LIME'den daha iyidir. Ayrıca, karar vermede üstün desteğe sahiptir.
	Hakkoum ve ark. [32]	Göğüs kanseri teşhisi	Wisconsin orjinal veri seti	LIME, FI, PDP	MLP,RBFN	Yerel, global	Dürüstlük	RBFN'nin doğruluk sonuçları MLP'den daha yüksektir, MLP'nin yorumlanabilirliği DT sınıflandırıcı ile daha tutarlıdır. Global ve yerel yorumlanabilirliği birleştirmeyi önerirler, böylece anlama düzeyi artar.
	Wu ve ark. [33]	Enfeksiyon şiddet tahmini	Çin'de toplanan bir özel veri seti ve İsrail ve Sao Paulo'dan toplanan açık veri setleri	PDP, ICE, ALE, SHAP, LIME	DT, RF, GB, NN	Global	Monotonluk, sadakat	N-Terminal pro-Brain Natriuretic Peptide (NTproBNP), C-Reaction Protein (CRP), lactic ve dehydrogenase (LDH), lymphocyte (LYM)'nin COVID-19 hastalarının şiddet seviyesini gösteren en önemli dört biyobelirteç olduğu gösterilmiştir.

Finans	Demajo ve ark. [34]	Kredi skorlama	Home Equity Line of Credit (HELOC) and Lending Club (LC) veri setleri	SHAP+GIRP, Anchors, ProtoDash	XGBoost	Yerel özellik tabanlı, yerel örneklem tabanlı, global	Benzersiz kural sayısı, ortalama kural koşulu sayısı, kuralların tutarlılığı, tamlık oranı	İşlevsel temelli, uygulama temelli ve insan temelli analizlerin kullanımı yoluyla yapılan değerlendirme, sağlanan açıklamaların basit ve tutarlı olduğunu ve ayrıca doğruluk, etkinlik, kolay anlaşılırlık, ayrıntılı yeterlilik ve güvenilirlik için test edilen önceden belirlenmiş altı hipotezi karşıladığını göstermektedir.
	Buss-mann ve ark. [35]	Kredi skorlama	Avrupa Dış Kredi Değerlendirme Kurumundan (ECAI) gelen veriler	Shapley değerleri	XGBoost	Yerel	-	Hem riskli hem de riskli olmayan borçluların, kredi puanlarını açıklamak ve anlamak için kullanılabilecek bir dizi benzer finansal özelliğe göre gruplandırılabilir. Böylece gelecekteki finansal davranışlarını tahmin etmek mümkündür.
	Chen ve ark. [36]	Kredi skorlama	FICO veri seti	İki katmanlı toplamalı risk modeli	İki katmanlı toplamalı risk modeli	Global	Benzerlik, Tutarlılık, Seyreklik, Destek	Model, ayrıştırılabilir yapısı sayesinde global yorumlanabilirlik için kural tabanlı özet açıklamalar ve yerel yorumlama için yerel vaka tabanlı açıklamalar sağlayarak, modelin tahminlerine en çok katkıda bulunan faktörlerin listesini tanımlayan anlamlı bileşenler üretmesine izin vermektedir.
	Ito ve ark. [37]	Finans metinlerini görselleş-tirme	Yahoo Finance Board'daki gönderiler, haber makaleleri	GINN	NN	Yerel	F-skoru, benzerlik	GINN sayesinde çeşitli cümle bağlamlarında verilen önemli kavramlar görselleştirilebilir. Bu tür bir görselleştirme, uzman olmayan kişilerin finansal belgeleri kolayca anlamasına yardımcı olur.
	Cong ve ark. [38]	Portföy yönetimi	WRDS veritabanından toplanan üç veri seti: CRSP, CRSP Compustat Merged, and Financial Ratio Firm Level	AlphaPortfolio, Lasso	LSTM, RNN, Transformer	Global	Örnek dışı Sharpe oranı, dayanıklılık	AlphaPortfolio çeşitli ekonomik ve ticari kısıtlamalar getirildikten sonra bile mükemmel örnek dışı performanslar sağlar. Önerilen model ticaret ve yatırım danışmanlığı için uygulayıcılar tarafından hemen kullanılabilir.

	Grath ve ark. [39]	Kredi risk değerlendirmesi	HELOC veri seti	Karşı Olgusal Açıklamalar için iki ağırlıklı strateji	LR, GradBoost, SVC ve MLP	Global	Tahmin gücü, üretilen karşı-olgusal açıklamaların ortalama boyutu	Özellik öneminden oluşturulan ağırlıkların daha kısa ve etkili karşı-olgular sağlar ve bu nedenle son kullanıcılar için daha etkili ve anlaşılır açıklamalar sunar.
Otonom Sistemler	Franco ve ark. [40]	Otonom sistemlerin çalışma zamanının izlemesi	Güvenli/güvensiz yörüngelerden oluşan bir veri seti	DT tabanlı sistem, karşı-olgusal kurallar	-	Yerel	Fizibilite, uygulanabilirlik	Önerilen yöntem, otonom araçlarda güvenli veya güvensiz bir durum olup olmadığını tahmin etmek için yorumlanabilir çalışma zamanı izleme yeteneğine sahiptir. Yorumlanabilirlik sayesinde, sonuca hangi sebeplerin neden olduğu hakkında daha fazla bilgiye sahip olunur ve düzeltici önlemleri almak daha kolay hale gelir.
	Schmidt ve ark. [41]	Güvenli ve etkin davranış politikaları	-	DT	RL	-	Güvenilirlik, doğrulanabilirlik	Çıkarılan bu politikanın derin sinir ağı (DNN) ajanı kadar performanslı ve güvenli olduğu ve aynı zamanda kolayca yorumlanabilir olduğu gösterilmektedir.
	Ivanovic ve ark. [42]	Yörünge Tahmini	nuScenes veri seti	MATS	-	Global	Son yer değiştirme hatası	MATS ile lineer olmayan gerçek dünya sistemleri için bile geçerli bir yörünge tahmin gösterimi sunulmaktadır. Ayrıca, aşağı akış planlama ve kontrol karmaşıklığında önemli azalmalar sağlanmıştır.

1.6. ÖRNEK UYGULAMA: SHAP İLE SAĞLIK VERİSİNİN YORUMLANMASI

Bu bölümde IML modellerinin en önemlilerinden biri olan SHAP algoritması ile örnek bir uygulama gerçekleştirmiştir. Bu uygulama ile IML modellerin kullanımının kolaylığı vurgulanmakta ve sonuçların nasıl yorumlandığı gösterilmektedir. Uygulama Python programlama dili ile gerçekleştirilmiştir. Editör olarak Jupyter Notebook kullanılmıştır.

SHAP, post-model IML tekniklerinden biri olduğu için öncelikle ML modelleriyle eğitilmiş ve daha sonra SHAP uygulanarak yorumlanabilirlik katılmıştır. Çalışmanın okuyucular tarafından kolayca tekrarlanabilmesi için olabildiğince ayrıntılı olarak tüm adımlar açıklanmıştır.

Uygulamada öncelikle veri manipülasyonu ve temel matematiksel işlemler için ihtiyaç duyulan Numpy ve Pandas kütüphaneleri eklenmiştir. Numpy, geniş bir matematiksel işlem, rasgele sayı üretimi ve doğrusal cebir fonksiyonları sunan hızlı ve çok amaçlı kütüphanedir. Pandas ise temelde Numpy kütüphanesini kullanan hızlı, güçlü, esnek ve kolay kullanıma sahip veri analizi ve manipüslasyonu kütüphanesidir. Aşağıdaki kod satırı ile bu kütüphaneler çalışmaya eklenmiş ve Numpy ve Pandas için yaygın olarak kullanılan np ve pd kısaltmaları atanmıştır.

```
import numpy as np
import pandas as pd
```

Uygulamada Wisconsin Üniversitesi tarafından yayınlanan göğüs kanseri veri seti [43] kullanılmıştır. Veri seti, UCI ML veri seti deposundan elde edilmiştir. Toplam 699 adet hastadan alının tümör hücresi görüntüsü üzerinden elde edilen 9 farklı değişken ve bir adet sınıf değişkeni içermektedir. Veri seti UCI web sitesinden 'csv' olarak indirilmiş ve her bir sütun isimlendirilmiştir. Aşağıdaki kod ile veri Python tarafından okunmuştur.

```
df = pd.read_csv("breast-cancer-wisconsin.csv")
```

Aşağıda veri ML ile analiz edilebilmesi için ön işlemden geçirilmiştir. Her bir hastayı tanımlayan "id" sütunu ve kayıp veri içeren satırlar silinmiş-

tir. Ayrıca anlaşılması kolay olması için sınıf değişkeninde ’2’ yazan sütunlar ‘iyi huylu’, kalanlar ‘kötü huylu’ olarak değiştirilmiştir.

```
df = df.drop("id", axis = 1)
df = df[ (df['Yoğunluk'] != '?' ) & ( df['Yoğunluk'].notnull() ) ]
df["Sınıf"] = np.where(df[Sınıf]== 2, "iyi huylu", "kötü huylu" )
```

Şekil 1.7’de verideki ‘iyi huylu’ ve ‘kötü huylu’ hasta sayısı ve verinin ilk 10 satırı gösterilmektedir.

```
print(df.Sınıf.value_counts())
df.head(10)
```

```
iyi huylu     444
kötü huylu    239
Name: Sınıf, dtype: int64
```

	Yarı Çap	Doku	Çevre Uzunluğu	Alan	Pürüzsüzlük	Yoğunluk	İç Bükeylik	İç Bükey Nokta Sayısı	Simetri	Sınıf
0	5	1	1	1	2	1	3	1	1	iyi huylu
1	5	4	4	5	7	10	3	2	1	iyi huylu
2	3	1	1	1	2	2	3	1	1	iyi huylu
3	6	8	8	1	3	4	3	7	1	iyi huylu
4	4	1	1	3	2	1	3	1	1	iyi huylu
5	8	10	10	8	7	10	9	7	1	kötü huylu
6	1	1	1	1	2	10	3	1	1	iyi huylu
7	2	1	2	1	2	1	3	1	1	iyi huylu
8	2	1	1	1	2	1	1	1	5	iyi huylu
9	4	2	1	1	2	1	2	1	1	iyi huylu

Şekil 1.7. Model eğitiminde kullanılan sağlık veri seti örneği

ML algoritması için veri seti sklearn kütüphanesinin train_test_split fonksiyonu kullanılarak eğitim ve test verisi olarak ayrılmıştır. Eğitim ve test verisinin de ‘iyi huylu’ ve ‘kötü huylu’ oranının korunması için stratify parametresine sınıf değişkeni atanmıştır. Veri seti %70-%30 olarak eğitim ve test olarak ayılmıştır.

```
from sklearn.model_selection import train_test_split
X = df.loc[:, ~df.columns.isin(['Sınıf'])]
y = df.loc[:, 'Sınıf']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, stratify = df['Sınıf'], random_state = 19)
print("Eğitim veri seti sayısı:", X_train.shape[0])
print("Test veri seti sayısı:", X_test.shape[0])
```

```
Eğitim veri seti sayısı: 478
Test veri seti sayısı: 205
```

Şekil 1.8. Eğitim ve test veri seti sayısı

ML algoritması olarak ağaç tabanlı güçlü Gradient Boosting algoritmasının optimize edilmiş hali olan XGBoost [44] seçilmiştir. Bilindiği üzere XGBoost algoritması boosting kullanan ağaç tabanlı karmaşık ML algoritmalarındandır. Çok yüksek başarı sağlamasına rağmen sonuçları yorumlanamazdır. SHAP ile bu eksiklik giderilmeye çalışılmıştır. Şekil 1.9'da verilen algoritma eğitim verisi ile eğitilmiş ve test verisi üzerinden tahmin yapılmıştır. ML algoritmasının sınıflandırma başarısı Şekil 1.9'da görülmektedir.

```
import xgboost
model = xgboost.XGBClassifier().fit(X_train, y_train)
y_pred = model.predict(X_test)
```

```
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred)
```

	precision	recall	f1-score	support
iyi huylu	0.98	0.96	0.97	133
kötü huylu	0.93	0.96	0.95	72
accuracy			0.96	205
macro avg	0.95	0.96	0.96	205
weighted avg	0.96	0.96	0.96	205

Şekil 1.9. XGBoost model sınıflandırma sonucu

1.6.1. SHAP ile Yerel Yorumlanabilirlik

SHAP yöntemi hem yerel hem de global yorumlanabilirliği destekleyen post-model bir tekniktir. Bu nedenle, örnek olarak sunulan uygulamada SHAP ile önce yerel sonra da global yorumlanabilirliğin nasıl uygulandığı ve yorumlandığı açıklanmıştır.

Yerel yorumlanabilirlik, tüm verileri temsil etmek yerine tek bir örneğin açıklamasına odaklanır. Örnek veri setimiz için düşündüğümüzde her bir örnek bir hasta olacaktır. Elimizde XGBoost gibi karmaşık bir ML algoritması ile eğitilmiş modelimize SHAP algoritmasıyla yorumlanabilirlik eklemek için öncelikle SHAP kütüphanesinin indirilmesi gerekmektedir. İndirme "pip install shap" komutu ile yapılmaktadır. SHAP algoritması ile SHAP değerleri (shap_values) için eğitilen XGBoost modeli ve kanser eğitim verisi kullanılmıştır (Bkz. Şekil 1.10).

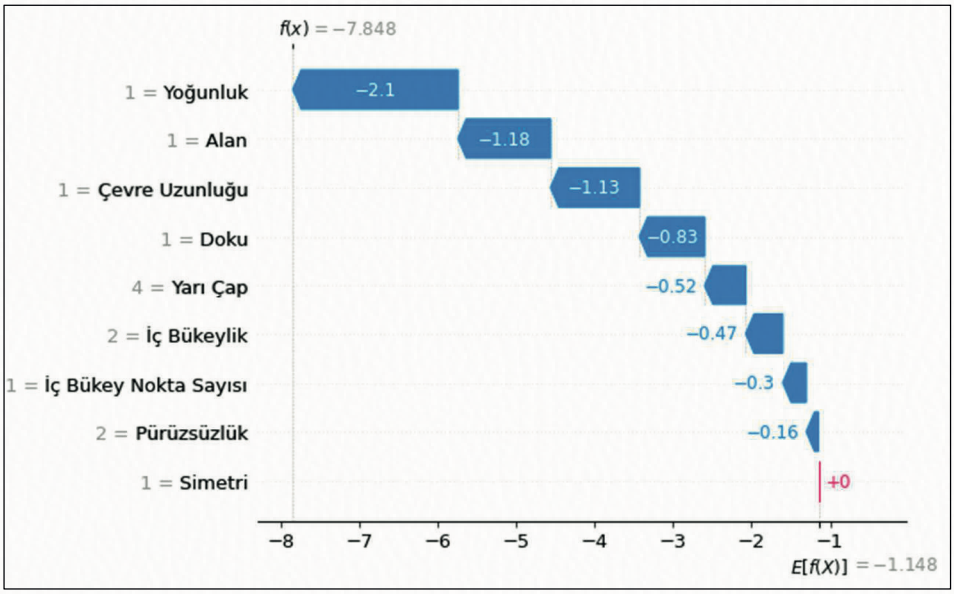
```
import shap
explainer = shap.Explainer(model)
shap_values = explainer(X_train)
```

Şekil 1.10. XGBoost model sınıflandırma sonucu

SHAP algoritması ortalama değerden değişkenlerin ayrı ayrı marjinal katkılarını bularak hesaplanmaktadır. Aşağıdaki grafikte ilk hastanın SHAP değerinin nasıl ortaya çıktığı görselleştirmiştir. Bu aynı zamanda ilk hastanın

yerel yorumlanabilirliğinden elde edilen sonuçlardır. Tüm veri seti kullanılarak ortalama SHAP değeri $E(f(x))$ ile gösterilen -1.148 olarak hesaplanmıştır. İlk hastanın SHAP değeri ise $f(x)$ ile gösterilen -7.848'dir. Bu değere en çok katkı sağlayan değişken aşağıdaki grafikte (Şekil 1.11) en üstte gösterilmektedir. Bu değişkenlerin isimleri ve değerleri y ekseninde gösterilmektedir. Barın rengi SHAP değerine olan marjinal katkının pozitif veya negatif olduğunu, barın üstündeki değer ise marjinal katkısının büyüklüğünü göstermektedir. Eğer barın rengi mavi ve büyüklüğü negatifse o hastanın hücresinin 'iyi huylu' olma ihtimali artmaktadır. Aksi şekilde, barın rengi kırmızı ve büyüklüğü pozitif ise, hücrenin 'kötü huylu' olma ihtimali artmaktadır. Buna göre, bu örnek hasta için tüm değişkenler SHAP değerini düşürmüş ve en yüksek katkıyı -2.1 ile yoğunluk değişkeni sağlamıştır. Sonuç olarak, hastaya ait $f(x)$ değeri, $E\big(f(x)\big)$ değerinden düşük olması hastanın hücresinin 'iyi huylu' olabileceğini göstermektedir. $f(x)$ değeri, $E\big(f(x)\big)$ değerinden negatif yönde uzaklaştıkça bu ihtimal daha da artmaktadır.

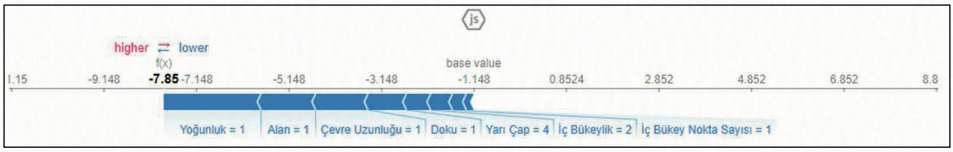
```
shap.plots.waterfall(shap_values[0])
```



Şekil 1.11. İlk hasta için SHAP değeri için şelale grafiği

Şekil 1.12 ise, lokal yorumlanabilirliği gösteren bir diğer grafiktir. Bu grafikte şelale grafiğine benzer şekilde bir hastanın SHAP değerinin belirlenmesinde en çok katkı sağlayan değişkenleri ve bunların marjinal katkılarını göstermektedir. Barların rengi katkının pozitif veya negatif olduğunu, büyüklüğü ise marjinal katkının büyüklüğünü göstermektedir.

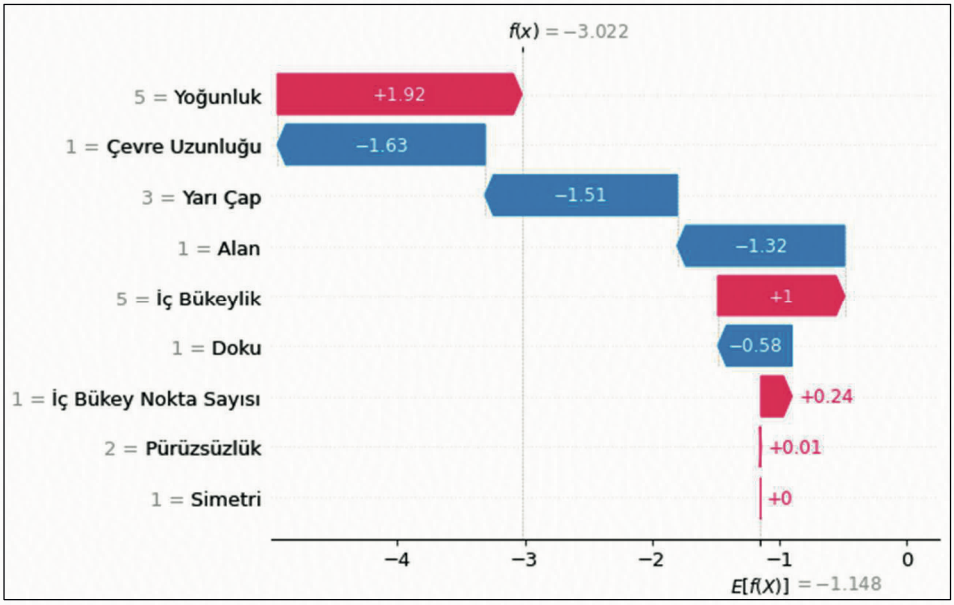
```
shap.initjs()

shap.force_plot(explainer.expected_value, shap_values.values[0,:],
X_train.iloc[0,:])
```



Şekil 1.12. İlk hasta için SHAP kodu ile yerel yorumlanabilirlik grafiği

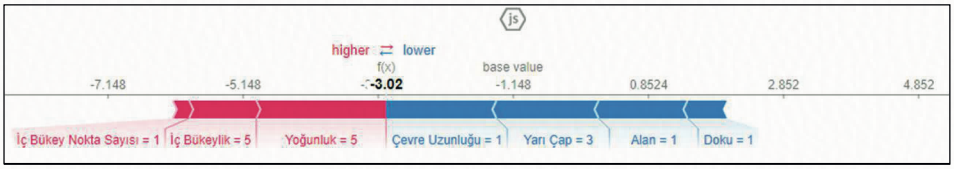
Şekil 1.13 ve Şekil 1.14'de bir başka hastaya ait lokal yorumlanabilirlik sonuçları görülmektedir. Görüldüğü gibi yoğunluk, iç bükeylik ve iç bükeylik nokta sayısı kırmızı renkte ve pozitif büyüklüktedir. Bu durum hastanın hücresinin 'kötü huylu' olma ihtimalini artırmaktadır. Ancak, tüm değişkenler dikkate alındığında $f(x)$ değeri, $E(f(x))$ değerinden negatif yönde uzaklaştığı için hastanın hücresinin 'iyi huylu' olma ihtimali yüksektir olarak yorumlanır. İlk hasta ile kıyaslandığında ise ilk hastanın $f(x)$ değeri $E(f(x))$ değerinden negatif yönde daha uzak olduğu için, ilk hastanın hücresinin 'iyi huylu' olma ihtimali daha yüksektir denilebilir.

```
shap.plots.waterfall(shap_values[200])
```



Şekil 1.13. Örnek hasta için SHAP değeri şelale grafiği

```
shap.initjs()
shap.force_plot(explainer.expected_value,
shap_values.values[200,:], X_train.iloc[200,:])
```

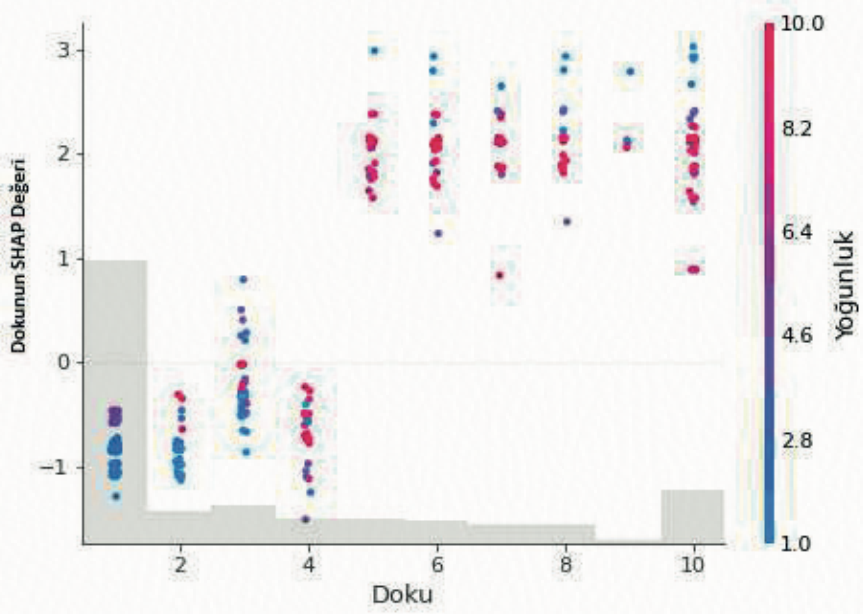


Şekil 1.14. Örnek hasta için SHAP kodu ile yerel yorumlanabilirlik grafiği

Yorumlanabilirlik için önemli bir diğer SHAP grafiği bağımlılık (dependence) grafiğidir. Şekil 1.15'te doku değişkeninin bağımlılık grafiği görülmektedir. Grafikte x ekseni doku değişkeninin aldığı değerleri, y ekseni ise doku değişkeninin SHAP değerini göstermektedir. Görüldüğü gibi doku değeri arttıkça SHAP değeri de artmaktadır. Yani hücrenin 'kötü huylu'

olma ihtimali artmaktadır. Bağımlılık grafiği çizilen değişkenin en çok ilişki içinde olduğu diğer değişken otomatik olarak seçilmektedir. Bu grafiğe göre doku değişkeninin en çok ilişki içinde olduğu değişken yoğunluk değişkenidir. Renkler, yoğunluk değişkeninin değerine göre değişmektedir. Yoğunluğu az olanlar mavi, fazla olanlar ise kırmızı ile gösterilmektedir.

```
shap.plots.scatter(shap_values[:, "Doku"], color=shap_values)
```



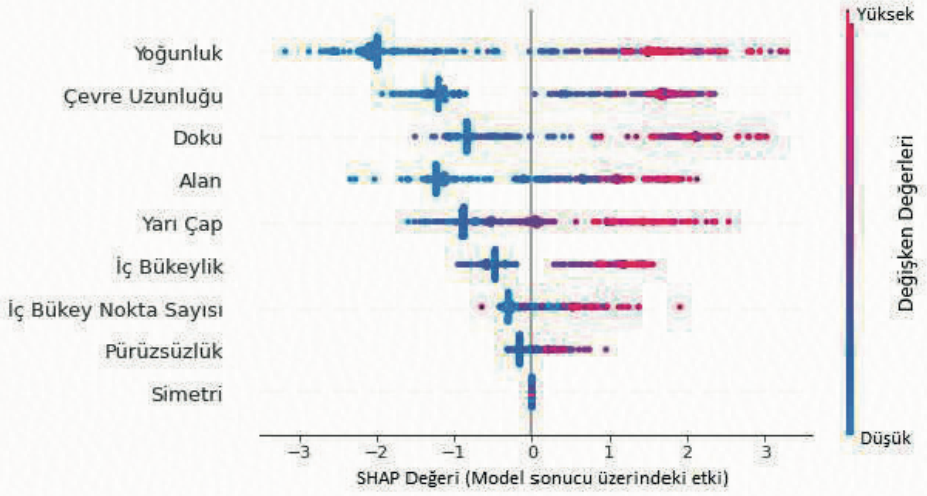
Şekil 1.15. Doku değişkeni için kullanılan kod ve elde edilen bağımlılık grafiği

1.5.2. SHAP ile Global Yorumlanabilirlik

Global yorumlanabilirlik kısaca modelde kullanılan tüm veri seti üzerinden değişkenlerin modele katkısının ölçülmesidir. Şekil 1.16'da XGBoost ile eğitilen modelin global yorumlanabilirlik grafiği görülmektedir. Grafikte en üstteki değişken modele en çok katkı sağlayan değişkendir. Her bir nokta bir hastayı göstermektedir. Renkler değişkenlerin değerlerini, x ekseni ise SHAP değerini göstermektedir. Buna göre bizim veri setimizde yoğunluk değişkeni tahmine en çok marjinal katkıyı sağlayan değişkendir. Bu değiş-

kenin değeri büyüdükçe SHAP değeri de büyümekte ve bu durum hastanın 'kötü huylu' hücreye sahip olma ihtimalinin arttığını göstermektedir. Grafiğe genel olarak bakıldığında hemen hemen tüm değişkenlerin sonuç ile pozitif korelasyona sahip olduğu görülmektedir.

```
shap.summary_plot(shap_values, X_train)
```



Şekil 1.16. SHAP ile global yorumlanabilirlik grafiği

1.7. IML YÖNTEMLERİN KALİTESİNİ ÖLÇME VE DEĞERLENDİRİLME KRİTERLERİ

Yorumlanabilirliğin kalitesini ölçmek çeşitli özellikler kullanılabilir. Bu özelliklere göre yorumlanabilirliğin kalitesi değerlendirilebilir [45].

- Doğruluk: Mevcut kararların açıklamalarının görünmeyen durumları kapsayacak şekilde genelleştirilmesini ifade eder. Açıklamalar ile tahmin sonuçları arasındaki benzerlikle ölçülür.
- Sadakat: Bir açıklamanın ne kadar iyi olduğunu açıklar. ML modelinin davranışını ifade eder.
- Tutarlılık: Aynı görevi yerine getirmek için farklı modellerin eğitilmesi durumunda açıklamalar arasındaki farkın derecesini ölçer.

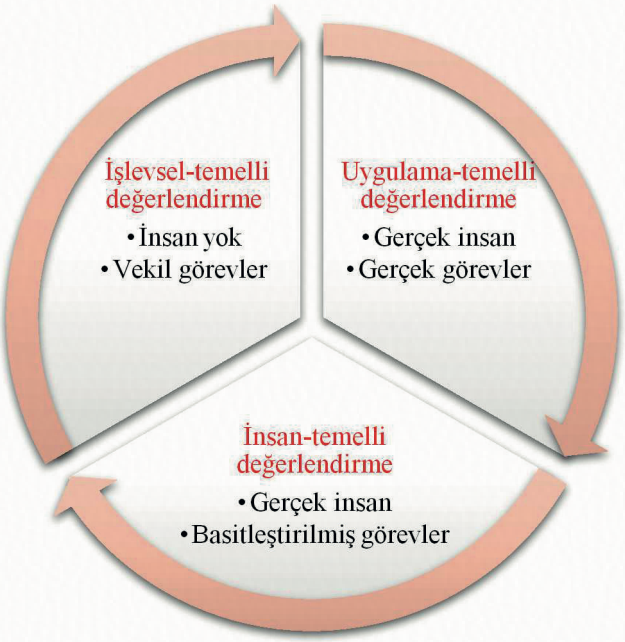
- Kararlılık: Benzer örneklemler için bir uygulamanın sonuçları arasındaki farkın derecesidir. Tutarlılık dışında, bu özellik aynı model tarafından elde edilen açıklamalarla ilgilenir.
- Anlaşılabilirlik: Açıklamaların okunabilirliği ve boyutu ile ilgilidir. Diğer bir deyişle, açıklamaların bir insan tarafından ne kadar iyi anlaşıldığını ölçmeye çalışır.
- Kesinlik: ML modelinin kesinliği ile ilgili açıklamaların yansımasını değerlendirir.
- Önem Derecesi: Açıklamaların, özelliklerin önemini ne kadar iyi sunduğunu ölçer.
- Yenilik: Açıklanan bir örneğin eğitim setine dahil olup olmadığını dikkate alarak kesinliğini değerlendirir.
- Temsilcilik: Modelin tamamının mı yoksa bir kısmının mı davranışını kapsadığını dikkate alarak bir modelin temsil edilebilirliğini değerlendirir.

Ayrıca, Doshi-Velez ve Kim [8] yorumlanabilirliğin değerlendirilmesi için üç ana deney seviyesi önermektedir. Bu seviyeler, aşağıda uygulama maliyetine ve sonuçların geçerliliğine göre azalan düzende sıralanmaktadır:

- **Uygulama-temelli değerlendirme (gerçek insanlar, gerçek görevler):** Gerçek bir uygulama içinde son kullanıcı deneylerinin yürütülmesini gerektirir. Bu deney, gerçek dünyadaki bir uygulamadaki açıklama kullanılarak ve aynı zamanda bir alan uzmanı olan son kullanıcı tarafından test edilip değerlendirilerek gerçekleştirilir. Örneğin; finansal olarak bir şirketin batacağını tahmin eden bir yazılımın bir finans uzmanı tarafından değerlendirilmesidir. Bunun en temel değerlendirmesi, bir insanın aynı kararı açıklamakta ne kadar iyi olacağıdır [2].
- **İnsan-temelli değerlendirme (gerçek insanlar, basitleştirilmiş görevler):** Hedef uygulamanın özünü koruyan daha basit insan-denek deneyleri yürütmeyi ifade eder. Aradaki fark, bu deneylerin alan uzmanlarıyla değil, meslekten olmayan kişilerle yapılmasıdır. Alan uzmanı gerekmediği için deneyler daha ucuzdur ve daha fazla insan üzerinde denenebilir. Örneğin, bir oyun/tahmin modelinde insanlardan modelin gerçek çıktısının ne olduğunu doğru anlayıp an-

lamadığını belirlemek için modelin gerçek çıktısı ile bozuk çıktılar arasındaki farkın bulunmasını istemektir [2].

- **İşlevsel-temelli değerlendirme (insan yok, vekil görevler):** İnsan deneyi gerektirmez. Bu tür bir değerlendirmede, yorumlanabilirliğin kalitesini ölçmede bir vekil olarak yorumlanabilirliğin bazı resmi tanımlamaları kullanılır. Bu tip deneyler, metodun tam olgunluğa ulaşmadığı durumlarda ya da insan deneylerinin etik olmadığı durumlarda tercih edilmektedir. İnsanlar dahil olmadığı için zaman ve maliyet açısından daha etkindir. Örneğin, bir karar ağacı modeli kullanıcılar tarafından zaten bilinen bir algoritmadır. Bu durumda, yorumlanabilirliğin kalitesini ölçmede vekil olarak ağacın derinliği dikkate alınabilir. Kısa ağaçlarda daha iyi yorumlanabilirlik skoru elde edilir. Ayrıca, büyük ağaçlarla karşılaştırıldığında tahmin performansının iyi kaldığı kısıtı eklenebilir [2]. Bu değerlendirme en iyi kullanılan modelin sınıfı insan düzeyinde bir değerlendirmede başka biri tarafından zaten değerlendirilmişse işe yarar [17].



Şekil 1.7. IML Değerlendirmede Kullanılan Kriterler [8]

Uygulama temelli değerlendirme, nihai hedefteki yorumlanabilirliği son kullanıcılarla değerlendirdiği için kesinlikle en uygun değerlendirmedir. Buna rağmen çok maliyetlidir ve sonuçları farklı alanlarda karşılaştırmak zordur. İşlevsel olarak temellendirilmiş değerlendirme, hiçbir insan denek gerektirmediğinden ve bu değerlendirme için tanımlanan vekiller genellikle farklı alanlarda karşılaştırılabilir olduğundan, spektrumun diğer ucunda görünür. Bununla birlikte, tanımlanabilecek vekiller gerçek yorumlanabilirlik ölçüleri olmadığından ve insan geri bildirimi olmadığından, işlevsel olarak temellendirilmiş değerlendirmeden gelen sonuçların geçerliliği düşüktür. İnsana dayalı değerlendirme, uygulamaya dayalı değerlendirmeden daha düşük maliyetli, ancak işlevsel olarak temellendirilmiş değerlendirmeden daha yüksek geçerliliğe sahip bir ara çözüm olarak kullanılmaktadır. Sonuçlar insan geri bildiriminden gelir, ancak değerlendirilen uygulama alanını göz ardı eder [17].

1.8. IML UYGULAMALARINDA KARŞILAŞILAN ZORLUKLAR

IML'in ML dünyasına getirdiği üstünlüklerin ve fırsatların dışında bazı zorlukları da bulunmaktadır [21, 25, 46]. Bunlar listelendiğinde;

- Başta, IML kapsamındaki bazı önemli tanımlamaların tutarsız ya da belirsiz olmasıdır. Henüz, yeni bir kavram olmasından kaynaklanan standartlaşmamış bir terminolojiye sahiptir. Örneğin, herkes algoritmik açıklamaların sadık olması konusunda hemfikirdir. Ancak sadakatin hedef modele mi, veri üretim sürecine mi olması gerektiği belirsizdir. Bu anlam karmaşıkları kafa karışıklığına ve gereksiz tartışmalara yol açmaktadır [46].
- IML yöntemlerinin büyük çoğunluğu, beklenen hata oranlarını ölçmemektedir. Bu da, herhangi bir bilimsel hipotezin gerektirdiği gibi, algoritmik açıklamaları bazı basitlik (severity) testlerine tabi tutmayı zor hale getirmektedir [46].
- Yorumlanabilir sistemlerin eksiksizliğinin ve doğruluğunun nicel bir ölçüsünün bulunmamasıdır. Bir modelin yorumlanabilirliğinin kalitesi ölçülürken gözlemcinin alan bilgisine göre değişkenlik göstermemesi gerekir. Buna en uygun nesnel ölçüm metriklerinin belirlenmesi gerekmektedir [25].

- ML sonuçların belirli bir problem için nasıl davrandığı daha açık bir şekilde ifade edildiğinde, kötücül kişiler bu bilgileri kendi çıkarları için kullanabilmektedir. Örneğin, saldırgan bir kişi yorumlanabilirlik sayesinde bir ML algoritmasının belirli bir çıktıya ulaşmak için hangi bilgileri kullandığını öğrenebilir. Bu bilgiyle ML algoritmasını manipüle etmeye çalışır ve farklı bir çıktı sonucuna neden olmak için girdilerdeki minimum değişiklikleri keşfetmeye başlar [21].

1.9. SONUÇ VE DEĞERLENDİRMELER

Bu bölümde, özellikle son yıllarda popülerlik kazanan ve ML konusundaki temel eleştirilere bir çözüm sunan IML kavramı tanıtılmış, kompleks ve yorumlanabilirliği düşük olan ML algoritmaları yorumlanabilir hale gelmiş, ML yöntemlerinin performansı sürdürülürken yorumlanabilir hale getirilerek, bu yaklaşım olan doğruluk ve yorumlanabilirlik açıklanmıştır. IML yöntemlerinin kullanılmasıyla birlikte ML yöntemlerinin yüksek performansları sürdürülürken farklı üstünlükleri içeren yeni çözümlerin geliştirileceği muhakkaktır. Bu üstünlükler ise aşağıda özetlenmiştir. Bunlar;

- IML yöntemleri sadece sonucun değil sonuca giden sebeplerin de önemli olduğu kritik ML problemlerinde tahmin sonucunun neden ve nasıl elde edildiğiyle alakalı bilgi verir.
- Ayrıca, ML yöntemlerinin eğitim setindeki olası sapmalardan kaçınılmış olur.
- Yerel yorumlanabilirlik özelliği sayesinde, ML yöntemlerinin global sonuçlarından farklı olarak belirli bir örneğin sonucunu nasıl elde ettiğiyle ilgili yerel kararlar alabilir.
- Daha önce keşfedilmemiş yeni bilgiler ortaya çıkarılabilir.
- Yukarıdaki özellikleri içeren IML'nin;
- ML dünyasına kattığı faydalar sayesinde kara kutu gibi davranan ML modellerine şeffaflık özelliği kazandırdığı, kritik kararların kullanıcıların da anlayabileceği bir şekilde saydam bir şekilde ifade edilmesine katkılar sağlamaktadır.

- En büyük dezavantajlardan biri ise henüz standartlaşmamış olmasıdır. IML denilince ilk akla gelen post-model teknikler olmasına karşın, lineer regresyon, karar ağacı gibi yöntemleri de IML yöntemleri arasında ele almak mümkündür. Ayrıca, terimlerin ve problemlerin matematiksel ifadelerinde eksiklik ve belirsizlikler mevcuttur.

Yorumlanabilir makine öğrenmesi yaklaşımlarının geliştirilmesine yönelik olarak önerilerimiz;

- Ülkemizde IML yaklaşımlarının kullanımının henüz yeterli seviyede olmadığı görülmektedir. Bu konuya daha çok önem verilmesi faydalı olacaktır.
- Ülkemizdeki IML ve benzer kavramları dikkate alan tez çıktısı sayısı çok kısıtlıdır [47, 48]. Terzi [47] doktora tezinde Telekom ve kredi kartı dolandırıcılığı için üç farklı tespit yöntemi geliştirmiş ve bu yöntemlerin birinin yorumlanabilmesini sağlamak amacıyla yeni bir ısı haritası tabanlı açıklanabilir model geliştirmişlerdir. Yıldırım [48] ise doktora tezinde büyük veri teknolojileri kullanılarak ML algoritmalarıyla temerrüt tahminini yapmış ve sonuçların yorumlanmasında IML yaklaşımı kullanmıştır. Bu çalışmaların akademik çevrede yaygınlaştırılması adına, lisans ve lisansüstü öğrencilerin IML konusuna yönelmesi teşvik edilmelidir.
- IML yöntemlerinin sonuçlarının doğruluğunun incelenmesi bir diğer önemli araştırma konusudur. Bu konuda ülkemizde görebildiğimiz kadarıyla henüz bir çalışma yapılmamıştır. Bu konudaki çalışmaların artırılması IML sonuçlarına olan güveni artıracak, IML yöntemlerinin daha yaygın bir şekilde gerçek dünya problemlerinde uygulanmasına olanak sağlayacaktır.
- Cumhurbaşkanlığı Dijital Dönüşüm Ofisi ile Gazi Üniversitesi işbirliğinde yapılan Türk Beyin Projesi kapsamında açıklanabilir yapay zekâ yaklaşımının kullanıldığı raporlanmıştır [45]. Ülkemizde sağlık alanında yapılan ve kamuoyunda paylaşılan bu gibi projelerde açıklanabilir yapay zekâ yaklaşımı kullanılmış olması sevindirici olup buna benzer projelerin sayısının kamuda ve özel sektörde arttırılması faydalı olacaktır.

KAYNAKLAR

[1] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Communications of the ACM,* vol. 63, no. 1, pp. 68-77, 2019.

[2] C. Molnar, *Interpretable machine learning: A Guide for Making Black Box Models Explainable*. 2018.

[3] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.

[4] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," *IEEE access,* vol. 6, pp. 52138-52160, 2018.

[5] D. Gunning, "Defense advanced research projects agency (darpa) program information: Explainable artificial intelligence (xai)(2018)," ed, 2018.

[6] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, and L. Cilar, "Interpretability of machine learning-based prediction models in healthcare," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 10, no. 5, 2020.

[7] A. Bunt, M. Lount, and C. Lauzon, "Are explanations always important? A study of deployed, low-cost intelligent interactive systems," in *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, 2012, pp. 169-178.

[8] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608,* 2017.

[9] OSTP, "Preparing for the Future of Artificial Intelligence;," CreateSpace Independent Publishing Platform: Scotts Valley, CA, USA, 2016.

[10] I. S. AI, "Dutch Artificial Intelligence Manifesto," 2018. [Online]. Available: http://ii.tudelft.nl/bnvki/wpcontent/

[11] C. Villani. "AI for Humanity—French National Strategy for Artificial intelligence." https://www.aiforhumanity.fr/en/ (accessed 20.11.2021)

[12] R. Society. "Machine Learning: The Power and Promise of Computers that Learn by Example." https://royalsociety.org/topics-policy/projects/machine-learning (accessed 20.11.2021)

[13] Google. "Responsible AI Practices—Interpretability." https://ai.google/education/responsible-ai-practices?category=interpretability (accessed 20.11.2021)

[14] H2O. "H2O Driverless AI." https://www.h2o.ai/products/h2o-driverless-ai/ (accessed 20.11.2021)

[15] DataRobot. "Model Interpretability." https://www.datarobot.com/wiki/ interpretability/ (accessed 20.11.2021)

[16] IBM. "Trust and Transparency in AI." https://www.ibm.com/watson/trust-transparency (accessed 20.11.2021)

[17] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics,* vol. 8, no. 8, p. 832, 2019.

[18] Z. C. Lipton, "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery," *Queue,* vol. 16, no. 3, pp. 31-57, 2018.

[19] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, "Explaining nonlinear classification decisions with deep taylor decomposition," *Pattern Recognition,* vol. 65, pp. 211-222, 2017.

[20] P.-J. Kindermans *et al.*, "Learning how to explain neural networks: Patternnet and patternattribution," *arXiv preprint arXiv:1705.05598,* 2017.

[21] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion,* vol. 58, pp. 82-115, 2020.

[22] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics,* pp. 1189-1232, 2001.

[23] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *journal of Computational and Graphical Statistics,* vol. 24, no. 1, pp. 44-65, 2015.

[24] D. W. Apley and J. Zhu, "Visualizing the effects of predictor variables in black box supervised learning models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 82, no. 4, pp. 1059-1086, 2020.

[25] A. Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (xai): A survey," *arXiv preprint arXiv:2006.11371,* 2020.

[26] M. T. Ribeiro, S. Singh, and C. Guestrin, "" Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135-1144.

[27] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st international conference on neural information processing systems*, 2017, pp. 4768-4777.

[28] R. ElShawi, Y. Sherif, M. Al-Mallah, and S. Sakr, "Interpretability in healthcare: A comparative study of local machine learning interpretability techniques," *Computational Intelligence,* 2020.

[29] C. Panigutti, A. Perotti, and D. Pedreschi, "Doctor XAI: an ontology-based approach to black-box sequential data classification explanations," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 2020, pp. 629-639.

[30] M. R. Zafar and N. M. Khan, "DLIME: A deterministic local interpretable model-agnostic explanations approach for computer-aided diagnosis systems," *arXiv preprint arXiv:1906.10263,* 2019.

[31] S. Knapič, A. Malhi, R. Salujaa, and K. Främling, "Explainable Artificial Intelligence for Human Decision-Support System in Medical Domain," *arXiv preprint arXiv:2105.02357,* 2021.

[32] H. Hakkoum, A. Idri, and I. Abnane, "Assessing and comparing interpretability techniques for artificial neural networks breast cancer classication," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* pp. 1-13, 2021.

[33] H. Wu *et al.*, "Interpretable machine learning for covid-19: an empirical study on severity prediction task," *IEEE Transactions on Artificial Intelligence,* 2021.

[34] L. M. Demajo, V. Vella, and A. Dingli, "Explainable ai for interpretable credit scoring," *arXiv preprint arXiv:2012.03749,* 2020.

[35] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable AI in fintech risk management," *Frontiers in Artificial Intelligence,* vol. 3, p. 26, 2020.

[36] C. Chen, K. Lin, C. Rudin, Y. Shaposhnik, S. Wang, and T. Wang, "An interpretable model with globally consistent explanations for credit risk," *arXiv preprint arXiv:1811.12615,* 2018.

[37] T. Ito, H. Sakaji, K. Izumi, K. Tsubouchi, and T. Yamashita, "Ginn: gradient interpretable neural networks for visualizing financial texts," *International Journal of Data Science and Analytics,* vol. 9, no. 4, pp. 431-445, 2020.

[38] L. W. Cong, K. Tang, J. Wang, and Y. Zhang, "AlphaPortfolio for investment and economically interpretable AI," *SSRN,* 2020.

[39] R. M. Grath *et al.*, "Interpretable credit application predictions with counterfactual explanations," *arXiv preprint arXiv:1811.05245,* 2018.

[40] C. Di Franco and N. Bezzo, "Interpretable Run-Time Monitoring and Replanning for Safe Autonomous Systems Operations," *IEEE Robotics and Automation Letters,* vol. 5, no. 2, pp. 2427-2434, 2020.

[41] L. M. Schmidt, G. Kontes, A. Plinge, and C. Mutschler, "Can You Trust Your Autonomous Car? Interpretable and Verifiably Safe Reinforcement Learning," in *IEEE Intelligent Vehicles Symposium (IV)(2021)*, 2021, pp. 171-178.

[42] B. Ivanovic, A. Elhafsi, G. Rosman, A. Gaidon, and M. Pavone, "Mats: An interpretable trajectory forecasting representation for planning and control," *arXiv preprint arXiv:2009.07517,* 2020.

[43] W. N. S. William H. Wolberg, Olvi L. Mangasarian. "Breast Cancer Wisconsin (Diagnostic) Data Set." https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29 (accessed 21.11.2021)

[44] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.

[45] M. Robnik-Šikonja and M. Bohanec, "Perturbation-based explanations of prediction models," in *Human and machine learning*: Springer, 2018, pp. 159-175.

[46] D. Watson, "Conceptual Challenges for Interpret-able Machine Learning," *Available at SSRN 3668444,* 2020.

[47] D. S. Terzi, "Sınıf dengesiz büyük veride dolandırıcılık tespiti ve açıklanabilirlik," Doktora, Bilgisayar Mühendisliği, Gazi Üniversitesi, 2020.

[48] M. Yıldırım, "Büyük Veride Çizge Teorisiyle Temerrüt Tahmini ve Makine Öğrenmesi Modellerinin Yorumlanması," Doktora, Bilgisayar Mühendisliği, Gazi Üniversitesi, 2021.

Bölüm 2

AÇIKLANABİLİR YAPAY ZEKÂ METODOLOJİLERİ VE UYGULAMALARI

Utku Köse

Gelişen Yapay Zekâ (YZ) çözümler, her geçen zaman çok daha zorlu problemler için başarılı sonuçlar ortaya koymaktadır. Bu çözümlerin gelişi-mi, bilgi ve iletişim teknolojilerindeki etkin değişimler ve yine YZ literatüründeki yenilikçi çalışmalarla bağlantılıdır. Ancak YZ'nin farklı alanlara nüfuz etmesi ve gelişen-değişen çözümlerin çok daha fazla parametreyle baş edebilen gelişmiş zeki sistemleri (örneğin Derin Öğrenme tabanlı sistemler) ortaya çıkarması farklı tartışmaları da ortaya çıkarmaktadır. Zeki sistemlerin insan hayatında çok daha fazla yer edinmesi güvenlik ve insan-makine arası rol paylaşımı konularında tartışmaları da alevlendirmektedir. Bu tartışmalar içerisinde en dikkat çeken konulardan biri de kullandığı parametre ve veri kümesi artan zeki sistemlerin, elde ettikleri çözümlere nasıl ulaştıkları ve bu bağlamda ne kadar güvenilir oldukları hususudur. Bu yöndeki endişeler, geliştirilen YZ çözümlerinin yorumlanabilirliği kapsamında araştırmaları tetiklemiş, ilgili araştırmalar Açıklanabilir Yapay Zekâ (XAI) adı altında incelenen bir ilgi alanının ortaya çıkmasına sebep olmuştur. Günümüzde insan bakış açısıyla anlaşılırlığı oldukça düşen; yani kara kutuya dönüşen YZ çözümlerinin sağlıklı değerlendirilmesi için çeşitli çözümler sunan bu ilgi alanı, YZ'nin geleceği açısından da kritik öneme sahip durumdadır. Açıklamalardan hareketle bu bölüm altında XAI'nın temelleri açıklanmış ve genel olarak Makine Öğrenmesi özelinde kullanılan bazı yöntemler ele alınmıştır.

2.1. GİRİŞ

Tanımlanan bir problem için en iyi sonuçları üreten zeki sistem kurgusunun nasıl yapılacağı, bu sistemin hangi veriler ve parametreler eşliğinde tasarlanacağı ve yine hangi teknikler üzerine inşa edileceği yönündeki sorunlar, Yapay Zekâ (YZ) geliştiricisi birçok kişiyi yakından ilgilendirmektedir. YZ dünyası her ne kadar farklı karakteristikte birçok tekniği içerse de, literatürün en kritik aktörleri arasında öğrenen teknikler; yani Makine Öğrenmesi teknikleri yer almaktadır. Makine Öğrenmesi kapsamında geliştirilecek bir sistemin önce hedef problem için eğitilmesi, ardından test sürecinden geçirilerek değerlendirilmesi gerekmektedir. Yapılan test süreçleri kimi zaman istenilen başarımda bulgular ortaya koymamakta, dolayısıyla YZ sisteminin hedef problemi çözememe sorunu ile karşı karşıya kalınmaktadır. Sorun ister veriden kaynaklı, ister kurulan Makine Öğrenmesi modelinden kaynaklı olsun, istendik düzeyde eğitilmiş bir sistemi elde etmek için sorunun kaynağını bulmak gerekmektedir. Bu sorun makine öğrenmesi modelinin parametrelerinden, modelin kendisine, hatta kullanılan veriye kadar birçok farklı faktörle ilişkili olabilmektedir. Sorun bu faktörlerden herhangi birine tekabül edebileceği gibi, birden fazlasıyla da başa çıkmak gerekebilmektedir. Söz konusu sorunlar önemli bir başka problemi de YZ'nin yorumlanması problemidir.

YZ'nin yorumlanması, YZ tabanlı oluşturmuş olduğumuz bir zeki sistemin; modeliyle, uğraştığı verilerle ve bize sunduğu sonuçlarla topyekun değerlendirebilmek olarak tanımlanabilmektedir [1, 2]. Bir zeki sistemin dayandığı teknik, kullanıldığı alan…vs. gibi konular (kendi içlerinde önemli görünseler de) sistemi sadece yüzeysel olarak algılamaya sebep olmaktadır. Eğer zeki sistem yorumlanabilirse;

i. sistemin başarısının ardındaki sırları anlaşılabilmekte,
ii. başka problemlere uyarlanıp uyarlanamayacağın tespit edilebilmekte, ve
iii. geliştirmelere ihtiyacı olup olmadığı

değerlendirilebilmektedir. Neticede YZ'yi yorumlamak, bir insanoğlu olarak şu sorulara daha sağlıklı cevaplar verilebilmesini mümkün hale getirmektedir: İlgili zeki sistemin başarısının bir sınırı var mıdır? Başarısız olduğunda insanlar bunu anlayabilir mi? Başarısız olduğunda bu sorunu çözebilmek olası mıdır? Özellikle karmaşık zeki sistemlerin çözümlere nasıl ulaştığını insanlar anlayabilir mi? İnsanlar bir zeki sisteme çözümleri konusunda güvenebilir mi?

Özellikle son soru, YZ'nin bir teknoloji olarak insanlığa olan katkısını, avantajlarını ve dezavantajlarını tartışmaya açabilecek önemli bir sorudur. YZ'yi olan güven konusu özellikle 21. Yüzyıl'ın önemli tartışma konuları arasında yer almaktadır [3-13]. Nihayetinde, tıptan ekonomiye, eğitimden hukuk alanına, hatta iyiden iyiye siber güvenlik gibi kritik alanlarda yaygın kullanım alanı bulan zeki sistemlerin, insan hayatında kritik rol oynayan bu alanlardaki etkisini değerlendirmek; teknik bilimler olduğu kadar sosyal bilimleri de ilgilendiren önemli bir araştırma konusu haline gelmiş durumdadır. Ancak önceleri yorumlanabilir doğaya sahip olan YZ çözümleri, artan karmaşıklık düzeyleri nedeniyle yorumlanması zor ya da mümkün olmayan yapılara (kara kutulara) dönüşmeye başlamıştır. Günümüz koşulları altında bakıldığında, mevcut zeki sistemler başarılı sonuçlar üretebilmek adına yorumlanabilme düzeylerinden feragat etmek zorunda kalmaktadır.

Bu bölüm altında, insanlığı zeki sistemleri yorumlama problemini çözme çabaları içerisine yönlendiren ve son birkaç yıldır önemi gittikçe artan Açıklanabilir Yapay Zekâ (XAI) kavramı ele alınmıştır. Buna göre, öncelikli olarak YZ'nin kara kutu olarak bilinmesine sebep olan hususlar, gelişmeler ışığında ifade edilmiş, ardından farklı seviyelerde yorumlanabilirliği ve açıklanabilirliği sağlayan yöntemlerin Makine Öğrenmesi ve bu alt-alanın günümüz yansıması olan Derin Öğrenme altında incelenmesi sağlanmıştır.

2.2. BİR KARA KUTU OLARAK YAPAY ZEKÂ

Her ne kadar temelindeki bilimsel gelişmelerle birlikte daha uzak tarihsel noktalara tekabül ediyor olsa da, YZ'nin kavramsal bağlamdaki ilk ortaya çıkışı 1940'lı yıllara uzanmaktadır. Bu süre baz alındığında, neredeyse asırlık bir geçmişe doğru giden bu bilim ve araştırma alanı, kapsamı içerisinde zekice tasarlanmış algoritmaları ve birtakım matematiksel / mantıksal paradigmaları barındırmaktadır. Hayatın özünde yer alan Matematik ve Mantık gibi alanlardan güç alan YZ tabanlı algoritmalar, zeki sistemlerin tasarımı noktasında birer Lego parçası gibi bir araya getirilerek; gerçek hayattaki problemlere esnek ve güçlü çözümler üretebilmektedir. Tarihsel sürece bakıldığında, YZ tekniklerinin modern hayatın farklı alanlarına nüfuz etmesi çok gecikmemiştir. YZ'nin farklı alanlara etkin ve kolay bir şekilde uygulanabilmesinin sebepleri arasında

i. esnek algoritma kalıplarına sahip olması,
ii. Matematik ve Mantık'ın gücünden destek alması,

iii. verilere duyarlı olması,
iv. destekleyici teknolojiler (elektronik, bilgisayar, iletişim teknolojileri…vs.) sayesinde daima zinde olması / gelişmesini sürdürmesi

yer almaktadır.

YZ kapsamında ilk ortaya çıkan teknikler / algoritmalar, içerdikleri kural tabanlı (rule based) algoritma adımları ve dış dünyaya nispeten daha açık yapıda olan parametreleri sayesinde kolaylıkla yorumlanabilmiştir. Yani bu teknikler ile oluşturulan zeki sistemlerin davranışları, işletilen kurallar, kullanılan parametreler, değişen değerler, vs. neticesinde insanoğlu tarafından anlaşılabilmiştir. Ancak yıllar ilerledikçe şu gelişmeler de YZ çözümlerinin gidişatını derinden etkilemiştir:

- Sayısal teknolojilerinin kullanımının artmasıyla birlikte daha karmaşık bilgisayar tabanlı sistemlere olan ihtiyaç artmıştır.
- İnsanlık, teknolojik unsurların sağladığı pratiklik karşısında kayıtsız kalmamış; günlük hayattaki problemlerini çözebilmek adına sayısal sistemlere daha fazla ihtiyaç duymuştur.
- Var olan YZ teknikleri gerçek hayattaki daha karmaşık problemlerle karşı karşıya kaldığında başarısız sonuçlar elde edilmiştir.
- Karmaşık problemlere karşı performansı daha iyi çözümlere ihtiyaç duyuluyordu; temel yapıdaki çözümler miladını teker teker doldurmuştur.
- YZ'nin farklı alanlardaki karar süreçlerine destek sağlayan çözümleri olarak ifade edilebilen Karar Destek Sistemleri'nin [14-16] popülaritesi ve etkinliği katlanarak artmıştır.

Söz konusu gelişmeler karşısında YZ cephesinin çözümü daha karmaşık tekniklerin geliştirilmesi olmuştur. Ancak daha karmaşık tekniklerin geliştirilmesi, daha fazla parametreyle uğraşılması ve yorumlanabilme olanaklarının azalması durumunu da beraberinde getirmiştir. Basit bir örnekle; yapay sinir hücresi ya da başka bir ifadeyle algılayıcı (perceptron) adı verilen ve ağırlıklandırılmış girişleri fonksiyonlardan geçirerek anlamlı hale getirmeye çalışan bilgi işleyicisi [17], XOR problemini çözememesi nedeniyle Çok Katmanlı Algılayıcı (Multi-Layer Perceptron) olarak ifade edilen Yapay Sinir Ağları tekniğinin tasarlanmasına sebep olmuştur (akt. [18]). Sonuç olarak, gerçek dünya problemlerine daha başarılı sonuçlar sunabilen

ancak yorumlanabilme düzeyi zorlaşmış / azalmış teknikler YZ literatürüne kazandırılmaya başlanmıştır. Bu durum özellikle Makine Öğrenmesi ile bağlantılı olmakla beraber; 21. Yüzyıl'da Derin Öğrenme'nin hız kazanmasıyla daha kötü bir hal almıştır. Bu durum, katmanlı sinir ağlarında ve birden fazla Makine Öğrenmesi tekniklerinin bir araya getirildiği karma (hibrit) çözümlerde de gözlenebilmektedir. Açıklamalardan hareketle günümüz çoğu sinir ağları modellerini, karma Makine Öğrenmesi çözümlerini ve Derin Öğrenme tabanlı YZ tekniklerini birer kara kutu (black box) olarak ifade etmek mümkündür. Buradaki 'kara kutu' ifadesi tekniğin ya da daha geniş ölçekte bir zeki sistemin çalışma mekanizmasına insan olarak hakim olamamak (ya da bunun oldukça zor olması) anlamına gelmektedir [1, 19].

2.2.1. Yapay Zekâ ve Yorumlanabilirlik

YZ açısından yorumlanabilirlik (interpretability) kavramı, zeki sistemdeki parametrelerin elde edilen çözümü anlatabilmesiyle ilişkilendirilmektedir. Özellikle Makine Öğrenmesi çözümlerine atfedilen yorumlanabilirlik, geleneksel teknikler altında nispeten kolaylıkla değerlendirilebilirken, YZ'nin amiral gemisi sinir ağları modelleri, karma modeller ve Derin Öğrenme ile birlikte birtakım önemli yenilikler gerektiren bir unsur haline gelmiştir. Katmanlı sinir ağlarını, karma Makine Öğrenmesi modellerini ve Derin Öğrenme'yi içeren çözümler tıp alanında oldukça başarılı sonuçlar ortaya koymakta, yine otonom veya sürücüsüz otomobillerin geliştirilmesi, finans alanı kapsamında öngörü tabanlı sistemlerin oluşturulması, toplumsal güvenlik adına suçlu tespiti, hukuksal karar verme yönelimli çözümlerin elde edilmesi gibi; hayatın her noktasında, özellikle kritik alanlarda YZ egemenliği üst düzeyde görülmektedir. Ancak ortaya çıkan kara kutu olma problemi, böylesine kritik alanlarda bu tür çözümlerin uygulanması, uygulansa da çözümlere güvenilmesi bağlamında insanlarda kararsızlıklara yol açmaktadır. Zeki sistemlerin kanıtlanamayan, gerekçeleri açıklanamayan kararlar oluşturması ve hatta eylemler içerisine girmesi an geldiğinde son derece tehlikeli sonuçlara yol açabilecektir. Unutulmamalıdır ki, çözümlere nasıl ulaştığı bilinmeyen bir kara kutu sistemin hata yapma olasılığının tespit edilmesi de oldukça zor; belki de imkansız olacaktır. Sayısal teknolojilerin yaygınlığı insanlarda etik ve güvenlik ihlallerine karşı endişeleri de artırmaktadır. Karar ve çözüm süreçlerinin teslim edileceği YZ'de, yorumlanamaz ise etik ve güvenlik endişeleri denizinde boğulmuş olacaktır. Zeki sistemlerin yorumlanabilirliği işte bu nedenlerden dolayı gerek günümüz ge-

rekse gelecek açısından biz insanlar için oldukça önemli bir faktördür. Buraya kadar tespit edilen problemlerden yola çıkarak yorumlanabilirlik daha net tanımlanabilmekte: bir zeki sistemin çalışma ve çözüm üretme mekanizmasının bütün şeffaflığıyla anlaşılabilmesi olarak ifade edilebilmektedir.

Günümüz zeki sistemleri ile yorumlanabilirlik faktörünü bir arada düşündüğümüzde, tasarım / planlama aşamasında yorumlanabilirlik kriterlerinin de dikkate alınması elzem bir hale gelmiştir. Bunun başlıca sebepleri, Arrieta vd. (2020) tarafından üç temel sebebe dayandırılmaktadır [20]:

- Tasarlanan zeki sistem tarafsız olabilecek, dolayısıyla eğitilmesi sürecinde veri setinde ortaya çıkabilecek önyargı (bias) faktörleri temizlenebilecektir.
- Yorumlanabilirlik, zeki sistemlerin karar mekanizmalarını manipüle edecek; muhtemel düşmancıl örnekleri [8, 21] ortaya çıkartarak, daha güvenilir bir sistem oluşturulmasını kolaylaştırmaktadır.
- Yorumlanabilirlik sayesinde elde edilen çıktıların gerçekten mantığa / nedenselliğe uygun olduğu bilinmekte; böylece sisteme güven sağlama alınmış olmaktadır.

Anlaşılacağı üzere, yorumlanabilirlik YZ literatürü ve insanlık için önemi gittikçe artan bir kavram haline gelmiştir. Dünyanın dört bir yanında araştırmacılar, Makine Öğrenmesi tabanlı sistemlere eklenecek yorumlanabilme unsurları ile sebep-sonuç ilişkisi kurulmuş sistemler elde etme yönünde alternatif çabalar içerisine girmişlerdir. Önceleri Yorumlanabilir YZ adı altında toplanan bu çabalar geleneksel Makine Öğrenmesi için kullanılırken, katman yapıları detaylanan sinir ağlarını, karma modelleri ve Derin Öğrenme çözümlerini karşılama noktasında zamanla XAI kavramına dönüşmüştür.

2.2.2. Açıklanabilir Yapay Zekâ

Anlaşabilme ve performans açısından insan ve makine (Yapay Zekâ-YZ) arasındaki boşluk her geçen zaman daha fazla artmaktadır. Açıklanabilir Yapay Zekâ (Explainable Artificial Intelligence - XAI), bu boşluğu tasarımsal açıdan telafi edebilme noktasında önemli araştırma alanlarından biri haline gelmiştir. Öyle ki, artık tasarlanan bir zeki sistemin sunmuş olduğu çözümlerin sebeplerini tartabilecek seviyeye insanı taşımış olacaktır.

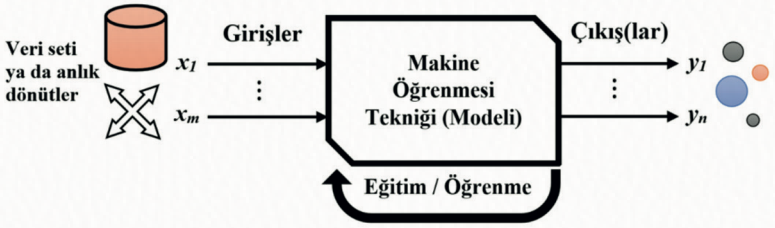
XAI kavramı genel olarak: kapsamı ve karmaşıklığı nedeniyle kapalı kutu niteliğindeki bir YZ / Makine Öğrenmesi sisteminin (zeki sistem) çeşit-

li unsurlar yardımıyla sebep-sonuç ilişkisini açıklayabilecek hale getirilmiş versiyonu olarak tanımlamak mümkündür [19, 20]. Bu tanımdaki kapsam, hedef problem ya da veri bütününün detayını işaret ederken, karmaşıklık ise sistemin kendi içerisindeki tekniklerin / algoritmaların / parametrelerin zorluğunu ifade etmektedir. İngilizce ifade ediliş şeklinden yola çıkılarak kısaca XAI olarak da ifade edilen Açıklanabilir YZ, esasında yorumlanabilir ve –dolayısıyla– güvenilir bir zeki sistem tasarlama sanatı görevi görmektedir. Literatürü incelediğimizde ilk işaretleri Yorumlanabilir YZ kavramı eşliğinde 2012 yılına kadar tekabül eden (akt. [20]) XAI, genel olarak şu amaçlar doğrultusunda çözüm üretmektedir [22]:

- Yüksek düzey öğrenme performansını sağlarken, mümkün olduğunca en iyi açıklanabilir modelleri üretmek.
- İnsanların zeki sistemleri anlayabilmesini, onlara güvenebilmesini ve onları etkin bir şekilde yönetilmesini sağlamak.

XAI modellerinin ifade edilen amaçları sağlayabilmesinin en etkin yolu tabi ki kullanılan Makine Öğrenmesi tekniklerinin doğasını anlamaktır. Diğer yandan mesele insan ve makine (zeki sistem) arasındaki farkı kapamak olduğu için insan faktörü de dikkate alınmalıdır. Neticede insanın neyin açıklamasını beklediği ve neyi nasıl anladığı sorunsalları, açıklanabilme özelliği / işlevleri içerecek bir zeki sistem için önemli konular olacaktır. Dolayısıyla XAI, Psikoloji, Sosyoloji gibi Sosyal Bilimler taraflı konulardan da destek almaktadır [23].

XAI esasında kara kutu sistemlerin bir şekilde insan anlaşılırlık seviyesine çekmek; yani sistemleri 'saydam' hale getirmek anlamına gelmektedir. Açıklanabilir yapıda modellerin oluşturulması, Makine Öğrenmesi sürecinin de bir sistem olarak ele alınmasını gerektirmektedir. Buna göre bir Makine Öğrenmesi modeli girişler olarak ele aldığı değerlerle çıkışta ulaşılması istenilen değerler arasındaki bağlantıyı kuran sistem olarak ifade edilebilecektir (Şekil 2.1).



Şekil 2.1. Bir sistem olarak Makine Öğrenmesi.

Makine Öğrenmesi'ni bu şekilde bir sistem olarak ele aldığımızda XAI taraflı müdahale edebilmek adına birkaç yol vardır:

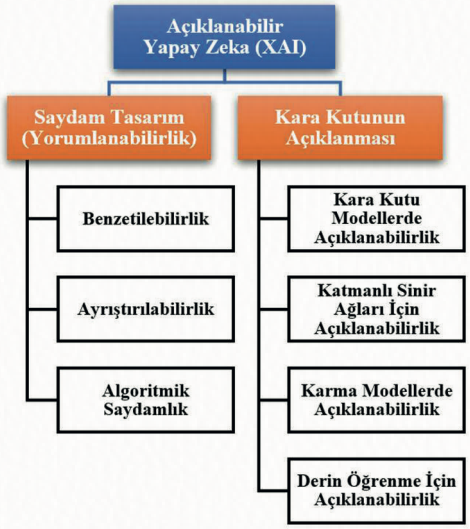
- Elde edilen çıkışlar ile girişler arası köprü kuran iç parametrelerden anlam çıkarmak.
- Model içerisinden çıkışları şekillendiren mekanizmaları tespit edip dış ortamda gözlemlenir hale getirmek.
- Matematiksel ve mantıksal müdahalelerle model içi mekanizmaları açıklanabilir hale dönüştürmek.
- Model içi mekanizmaları anlamlandıracak bir başka teknik / algoritma kullanmak.

Müdahale noktasında bu yollardan birden fazlasını kullanmak da mümkün olmaktadır. Nihayetinde bir Makine Öğrenmesi sistemini açıklanabilir hale getirmek; yani XAI bir tür tersine mühendislik (reverse engineering) anlamına gelmektedir. Bu müdahaleler literatürde iki temel sınıf altında anlam kazanmaktadır [24]: (1) saydam tasarım ve (2) kara kutunun açıklanması. Arrieta vd. (2020) bu sınıflandırmayı (1) saydam modelleme (transparent modelling) ve (2) sonradan açıklanabilirlik (post-hoc explainability) olarak adlandırmayı daha uygun görmüştür [18]. Esasında saydam modelleme geleneksel Makine Öğrenmesi teknikleri açısından yorumlanabilirlik anlamına gelirken, sonradan açıklanabilirlik tam anlamıyla XAI'yı işaret etmektedir. Ancak mevcut gelişmeler ışığında söz konusu çözümlerin tümünü tek bir XAI çatısı altında ifade etmek daha isabetli bir sınıflandırma sağlayabilmektedir. Bu bağlamda, her iki sınıflandırma genel olarak şöyle tanımlanabilmektedir (Şekil 2.2):

- **Saydam Tasarım (Yorumlanabilirlik):** Saydam modeller, kendi içlerinde yorumlanabilirlik için gereken birtakım unsurları içerdikleri için artık açıklanabilirlik olarak kabul edilen hakimiyet aşamasına ulaşmak oldukça kolaydır. Bu noktada özellikle temel Makine Öğrenmesi teknikleri kapsam içerisine girmekle beraber, saydam modellerin karakteristiklerine göre yorumlanabilirlik üç özelliğe dayandırılmaktadır [20]:

  i. Benzetilebilirlik (Simulatability),
  ii. Ayrıştırılabilirlik (Decomposability) ve
  iii. Algoritmik Saydamlık (Algorithmic Transparency).

- **Kara Kutunun Açıklanması:** Kara kutu modellere sonradan müdahalelerle açıklanabilirlik eklenmektedir. Bu durum parametre sayıları ve çalışma mekanizmaları karar kutu olmalarına sebep olan birtakım tekniklerde, katmanlı sinir ağlarında ve karma Makine Öğrenmesi modellerde parametrelere hakimiyeti artıracak müdahalelerle sağlanırken, Derin Öğrenme'de daha spesifik yöntemleri gerekli kılmıştır.



Şekil 2.2. XAI kapsamına dair genel bir sınıflandırma.

2.2.3. Açıklanabilir Yapay Zekânın Uygulama Alanları

XAI özellikle kara kutu modellerin farklı alanlardaki uygulamalarında birçok kazanımı da beraberinde getirme potansiyeline sahiptir. Buna göre, XAI'nın kullanımı sayesinde sağlık alanında güvenilir teşhis ve planlamalar mümkün olabilecek, hatta bilimsel ilerlemeler kolay gerçekleştirilecektir. Yine eğitim alanında, idarecilerin ya da öğreticilerin daha etkin planlamalar konusunda güvenilir destek alması, sebepleri tutarlı öğrenci başarı tahminleri sayesinde yeni öğrencileri başarılı sonuçlara götürecek unsurlarla tanıştırmaları daha kolay olacaktır. Finans alanında ekonomi verilerinin gelecek akış süreçlerini yorumlanabilir bir şekilde öğrenmek, önlemlerin ve güvenilir kararların alınmasını hızlandıracaktır. Bu konuda genel bir bakış açısı Tablo 2.1'de sunulmuştur.

Tablo 2.1. XAI ve örneklerle uygulama alanları.

Uygulama Alanı	Örnekler
Sağlık	Güvenilir teşhis uygulamaları Daha verimli tedavi planlamaları Daha güvenilir ilaç keşifleri Bilgi birikimi ve bilimsel ilerleme
Fizik, Kimya, Biyoloji	Sonuçları daha güvenilir deneyler Daha verimli çözümlemeler Bilgi birikimi ve bilimsel ilerleme
Eğitim	Daha etkin eğitim planlamaları Sebep-sonuç ilişkisi daha net başarı tahminleri Bilgi birikimi ve bilimsel ilerleme
Sosyoloji	Daha etkin kitlesel tahminler Daha güvenilir kitlesel benzetimler Bilgi birikimi ve bilimsel ilerleme
Finans	Sebep-sonuç ilişkisi daha net finansal tahminler Erken önlemler için daha etkin tahminler Daha güvenilir finansal stratejilerin belirlenmesi Bilgi birikimi
Savunma Sanayi	Daha güvenilir tasarım / modellemeler Daha güvenli savunma teknolojileri Daha güvenli otonom sistemler Bilgi birikimi
Felsefe	İnsan-bilgisayar arası daha etkin iletişim Olguların daha tutarlı mantıksal açıklanması Bilgi birikimi
Hukuk	Güvenilir sebep-sonuç ilişkisi tespitleri Adil karar desteği uygulamaları Bilgi birikimi
Elektronik, Mekatronik	Daha güvenilir robotik sistemler Daha güvenilir erken arıza teşhis sistemleri Daha güvenilir karmaşık kontrol sistemleri Bilgi birikimi ve bilimsel ilerleme
Ağır Sanayi	Daha güvenilir ve hızlı üretim planlamaları Üretim bileşenlerinin daha tutarlı kullanımı Bilgi birikimi
İş Dünyası	Daha güvenilir yönetimsel planlamalar / kararlar Data etkin maliyet / zarar optimizasyonları Data etkin ticari stratejilerin belirlenmesi

Anlaşılacağı üzere zeki sistemlerin çıktılarını anlayabilmek, çözümlerin 'daha iyi' olmasına yol açmaktadır. En azından olaya insan-teknoloji ilişkisi doğrultusunda bakıldığı zaman bu doğru bir tespit haline gelmektedir.

XAI'ya dair uygulama örneklerini de baz alarak, genel hedef kitlesini ve her bir farklı hedef kitlesinin neden XAI'ya ihtiyaç duyacağına açıklama getirmek de oldukça kolaydır. Neticede bir doktorun XAI kullanım amacı ile bir firma yöneticisinin kullanım amacı arasında çeşitli farklılıklar söz konusu olmaktadır. Bunların sebebi uygulama alanlarındaki dinamikler ile birlikte zeki sistemlerle elde edilen çıktıların kullanımı amaçları doğrultusunda şekillenmektedir. XAI'nın hedef kitlesi esasen herkestir. Ancak detaylardan arınmak suretiyle ve Arrieta vd. (2020) tarafından ortaya konulan tespitlerden de [20] yola çıkarak, Tablo 2.2'de olduğu gibi bir dağılım kurulabilmektedir.

Tablo 2.2. XAI'da hedef kitle ve önde gelen kullanım amaçları.

Hedef Kitle	Kullanım Amaçları
Doktorlar, sağlık çalışanları, eğitimciler	Güvenilir bilgi ve karar desteği elde etmek Bilgi keşfi sağlamak
Veri bilimcileri, mühendisler, yazılım-donanım geliştiricileri	Erişilebilir, şeffaf bir sistem geliştirmek Bilgi dönüşümü gerçekleştirmek Etkileşim sağlamak
Hukukçular, yöneticiler, karar vericiler / politika yapıcılar	Güvenilir sebep-sonuç ilişkisi kurmak Adil değerlendirme ve karar desteği elde etmek
Fiziksel çalışmalarda bulunanlar	Güvenilir eylem ve karar desteği elde etmek Etkileşim sağlamak
Alan uzmanları, spesifik meslek grupları	Alana / mesleğe özgü güvenilir bilgi ve karar desteği elde etmek
Sıradan kullanıcılar (*diğer kitleler dışında kalanlar*)	Etkin ve verimli bilgi edinmek Güvenilir sebep-sonuç ilişkisi kurmak Etkileşim sağlamak

2.3. SAYDAM TASARIM (YORUMLANABİLİRLİK)

Kara kutu haline gelinceye kadar Makine Öğrenmesi modellerinin (zeki sistemlerin) açıklanabilmesini sağlayan yorumlanabilme; yani saydam tasarım faktörü, modelin iç dinamiklerine hakim olmayı gerektirmektedir. Sadece çeşitli parametrelerin değerlerini baz alarak kurulan bir zeki sistemin çıktılarını yorumlamak, bu çıktıları girişlerle ya da iç dinamiklerle ilişkilendir-

meyi içermektedir. Bu açıdan saydam tasarım üç temel yaklaşıma tekabül etmektedir: Benzetilebilirlik (Simulatability), Ayrıştırılabilirlik (Decomposability) ve Algoritmik Saydamlık (Algorithmic Transparency).

2.3.1. Benzetilebilirlik

Benzetilebilirlik kavramı, saydam bir modelin insan tarafından mantıksal / matematiksel bağlamda yorumlanabilecek düzeyde açık olduğunu ifade etmektedir. Bu durumu dikkate aldığımızda, seyrek lineer / doğrusal (sparse linear) yapıda olan modellerin benzetilebilirlik özelliği taşıdığını düşünmek yanlış olmayacaktır [20, 25].

Benzetilebilirlik özelliğine örnek olarak algılayıcı (perceptron) adı verilen ve Yapay Sinir Ağları mimarilerinin temel yapıtaşı olan işlem biriminin dinamikleri verilebilmektedir. Tipik bir algılayıcının çıkışlarından elde edilen değerler, giriş değerlerinin işleme konulduğu fonksiyonların ilişkisel durumlarından yola çıkılarak açıklanabilmektedir. Bu durum, modelin fonksiyonel işleyişi insan tarafından görülüp anlaşılabiliyor anlamına gelmektedir. Neticede saydam, yani yorumlanabilir bir YZ modeline, Benzetilebilirlik özelliğini dikkate alarak açıklanabilirlik eklemek için modelde kullanılan fonksiyonları ve mantıksal aşamaları anlamlandırmak, bunları kullanıcıya uygun arayüzler üzerinden aktarmak gerekmektedir.

2.3.2. Ayrıştırılabilirlik

Ayrıştırılabilirlik bir saydam modelin, girişleri, çıkışları, parametreleri ve kullandığı algoritmik mekanizmalarla anlaşılması anlamına gelmektedir. Kısacası ayrıştırılabilirlik özelliği belirgin bir saydam modelin, ekstra çabalara ihtiyaç duymadan; çeşitli bileşenlerinin dikkate alınması suretiyle insanlar tarafından yorumlanması mümkün olmaktadır [20, 26].

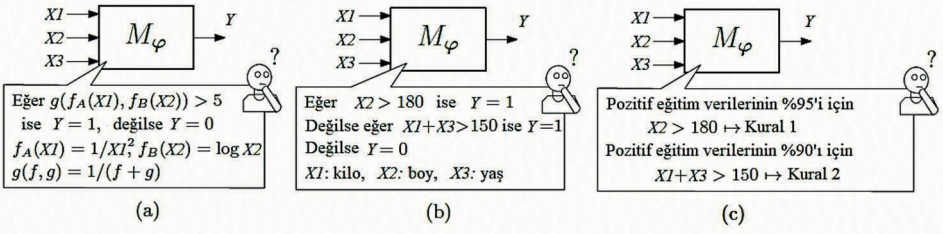
Ayrıştırılabilirlik yaklaşımı ile ilgili olarak; örneğin bir Makine Öğrenmesi modelinin belirli bir girişinin aldığı değerler sayesinde çıkışın nasıl değişebileceği anlaşılıyorsa, bu modelin çıkarımları da açıklanabilir demektir. Bu senaryo her zeki sistemde her zaman mümkün olmasa da nispeten basit öğrenme-çözme mekanizmalarına sahip; ama yeterince etkili tekniklerde görülebilmektedir. Ayrıştırılabilirlik özelliğini kullanarak girişler ve çıkışlar arası ilişkileri gösteren ya da bunlar arasındaki köprüleri, algoritmik mekanizmalar üzerinden oluşturan bir açıklama altyapısı inşa edilebilmektedir.

2.3.3. Algoritmik Saydamlık

Saydam tasarım kapsamındaki üçüncü yaklaşım Algoritmik Saydamlık olarak isimlendirilmektedir. Algoritmik Saydamlık genel olarak saydam bir modelin ürettiği çıkış değerlerini bu değerlerin sebebi olan girişlere nasıl dayandırdığının, insan tarafından anlaşılmasını ifade etmektedir [20]. Yani saydam modelin algoritmik yapısı, dışarıdan bir gözlemciye olan bitenler konusunda çeşitli mantıksal ipuçlarını vermekte, böylelikle açıklanabilirlik adına gerekli bağlantıları kurma fırsatını doğurmaktadır.

Seyrek lineer modeller, değişen girişlere karşı çıkışlarda ne şekilde eğilimler olduğu konusunda kolay yorumlarda bulunmamıza imkan vermekte, dolayısıyla Algoritmik Saydamlık özelliğine de göstermektedir. Algoritmik Saydamlık özelliğine en güzel örnekler kural tabanlı (rule based) modeller üzerinden verilebilmektedir. Örneğin, bir Bulanık Mantık (Fuzzy Logic) tabanlı sistemde girişlerin durumuna göre tetiklenen kurallar söz konusu olduğu için elde edilen çıkışları kurallara, kuralları da giriş değerlerine bağlamak; böylelikle çıkartımlardan giriş değerlerine sebep-sonuç ilişkisi kurmak mümkün hale gelmektedir.

Şekil 2.1'de ifade edilen Makine Öğrenmesi'nin bir sistem olarak gösteriminden yola çıkılarak, [20] tarafından açıklanan saydam tasarım yaklaşımları Şekil 2.3'te olduğu gibi tanımlanabilmektedir. İlgili şekilden de anlaşılacağı üzere, Benzetilebilirlik'te (a) modeldeki parametrelerden yorumlama daha net iken, Ayrıştırılabilirlik (b) model için parametrelerle insan arasında algoritmik mekanizmalar üzerinden bağlantı oluşturulmasını gerekli kılmakta, Algoritmik Saydamlık (c) ise daha fazla insan yorumlaması ile modelin giriş-çıkış arası çıkartımlarının anlamlandırılmasını içermektedir.



Şekil 2.3. Saydam Tasarım (yorumlanabilirlik) kapsamında üç farklı yaklaşım: (a) Benzetilebilirlik, (b) Ayrıştırılabilirlik, (c) Algoritmik Saydamlık [20].

Saydam tasarım için, yani diğer bir deyişle Makine Öğrenmesi kapsamında yorumlanabilirlik sağlamak için, ilgili yaklaşımların farklı teknikler / modeller içerisinde değişken seviyelerde olduğunu görmek mümkündür. Bu noktada önemli olan husus, açıklanabilirlik imkanlarının Benzetilebilirlik özelliğinden Algoritmik Saydamlık özelliğine doğru soyutlaşmasıdır. Yani Benzetilebilirlik özelliği daha baskın olan bir modeli doğrudan gözlemleyerek açıklamak daha kolay olurken, Algoritmik Saydamlık özelliği ön planda olan bir model, daha mantıksal bir yapıyla anlaşılabilmektedir.

2.3.4. Yorumlanabilir Makine Öğrenmesi Modelleri

Saydam tasarım ya da saydam model ifadeleri yorumlanabilir Makine Öğrenmesi olarak daha pratik ifade edilebilmektedir. Buna göre, Makine Öğrenmesi kapsamında literatüre kazandırılan temel tekniklerin esasında model bağımlı yapıda yorumlanabilir olduklarını söylemek mümkündür. Yani, bu tür teknikler yorumlanabilme özelliklerini kendi yapılarından almakta, dolayısıyla bir modelin yorumlanabilmesi sadece kendine has parametrelerle ilgili olmaktadır. Bu bakımdan model bağımlı yorumlanabilirlik altında da incelenebilecek olan birtakım modellerin yorumlanabilme yaklaşımları genel olarak şöyle açıklanabilmektedir:

i. Algılayıcı Modelinde Yorumlanabilirlik

Algılayıcı (perceptron) modelleri, katman tabanlı sinir ağları tasarlanmadan önce literatüre kazandırılmış, basit çıkartım araçları olarak bilinmektedir. Saydam tasarım / yorumlanabilirlik yaklaşımları bakımından Benzetilebilirlik özellikleri gösteren bu modellerde varsayılan çıkartım mekanizması şu denkleme dayanmaktadır:

$$Y = w1.X1 + w2.X2 + w3.X3 + \cdots + wn.Xn \qquad (2.1)$$

Denklemde *Y* algılayıcının ürettiği çıkışı, *X1, X2, X3, ..., Xn* giriş özelliklerinin değerlerini ve *w1, w2, w3, ..., wn* ilgili giriş özelliklerinin ağırlıklarını tanımlamaktadır. Eğitim süreciyle birlikte bir algılayıcıda ağırlık (*w*) değerlerinin her bir giriş özelliğinin önemini gösterecek düzeyde değişmesi söz konusu olmaktadır. Dolayısıyla ağırlık (*w*) değeri yüksek giriş özellikleri çıkış üzerinde daha etkili olurken, düşük olanlar daha az etkili ya da önemsiz

olarak yorumlanmaktadır. Elbette algılayıcı mimarisi daha sonraları farklı gelişmelere sahne olmuş; sırasıyla ağırlıklı toplanan giriş özellik değerlerinin aktivasyon fonksiyonlarından geçirilmesi, bu durumdaki çok sayıda algılayıcının kolektif kullanılması ve nihayetinde katmanlı ağlara geçiş sağlanmıştır. Bu durumda aktivasyon fonksiyonlarının çıkışa olan etkilerinin yorumlanabilmesi problemi ve akabinde artan parametre sayıları nedeniyle yorumlanabilmenin nasıl sağlanabileceği gibi sorular XAI çözümlerini tetiklemiştir. Başlangıçta gayet yorumlanabilir olan algılayıcı, tarihsel süreçte güçlü ancak gittikçe kara kutu haline gelen sinir ağları mimarilerinin temelinde yer almaya devam etmektedir.

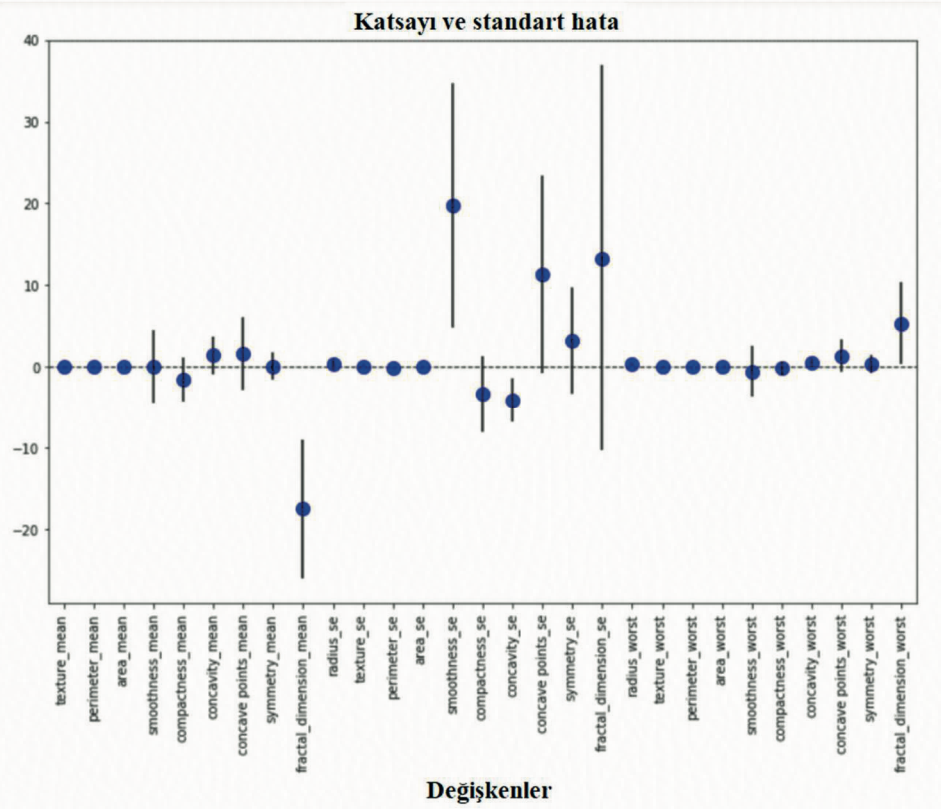
ii. Lineer Regresyon Modelinde Yorumlanabilirlik

Lineer Regresyon, Benzetilebilirlik yaklaşımı içeren basit bir tekniktir. Buna göre, tipik bir Lineer Regresyon modelinin matematiksel ifade ediliş şekli şöyledir:

$$Yi = S + w1.X1i + w2.X2i + w3.X3i + \cdots + wn.Xni \qquad (2.2)$$

Denklem 2.2'de *Yi*, i. örneğin çıkış değeri anlamına gelmekte, yine ilgili örnek için *X1, X2, X3, ..., Xn.* değerleri giriş özelliklerine tekabül etmektedir. Yine denklemde *S* gibi birtakım sabitler yer alabileceği gibi kurulan bir modelde çıkışı oluşturacak giriş değerlerinin ağırlıkları da *w1, w2, w3*, …, *wn* ile belirlenmektedir. Burada Lineer Regresyon'daki amaç uygun ağırlık (*w*) değerleri bulmaktır.

Lineer Regresyon ile kurulan bir modelde bir veri seti ile eğitim gerçekleştirildikten sonra elde edilen ağırlık (w) değerleri, her bir girişin çıkış üzerindeki etkisini yorumlamaya imkan vermektedir. Daha ileri düzey yorumlanabilirlik ise t-testi ile sağlanabilmekte, buna göre yüksek değerde ve varyansı az olan ağırlığa sahip bir özellik diğerlerine göre daha önemli (çıkış değeri üzerinde daha etkin) olarak yorumlanmaktadır. Buradan hareketle, eğitimi gerçekleştirilmiş bir modelde her bir giriş özelliğinin elde edilen ağırlık değerleri ve standart hata değerlerinden yola çıkılarak, girişlerin güvenli bir çıkartım sağlayıp sağlamadığını gösteren bir grafik elde etmek mümkün olmaktadır. Şekil 2.4 bu bağlamda örnek bir dağılımı göstermektedir [27].



Şekil 2.4. Lineer Regresyon tabanlı bir modelde giriş özelliklerinin yorumlanabilirliği [27].

iii. Naive Bayes Modelinde Yorumlanabilirlik

Bayes Teoremi'nden güç alan ve farklı sınıfların olasılıksal tespiti üzerinden karar oluşturan Naive Bayes tekniğinde, hangi sınıfın ağır bastığının bulunması noktasında farklı giriş özelliklerinin olasılıksal dağılımları dikkate alınmaktadır:

$$P(C_k|x) = \frac{1}{Z} P(C_k) \prod_{i=1}^{n} P(x_i|C_k) \tag{2.3}$$

Denklem 2.3'de Z ile bütün sınıf olasılıklarının toplamda 1'e eşit olması sağlanmakta; buradan hareketle, bir sınıfın olasılığı da sınıf olasılığı kere sınıfa verilen her bir özelliğin Z ile normalleştirilmiş olasılık değeri olarak anlaşılmaktadır [2]. Bu çıkartım farklı giriş özelliklerinin hedef sınıflar üze-

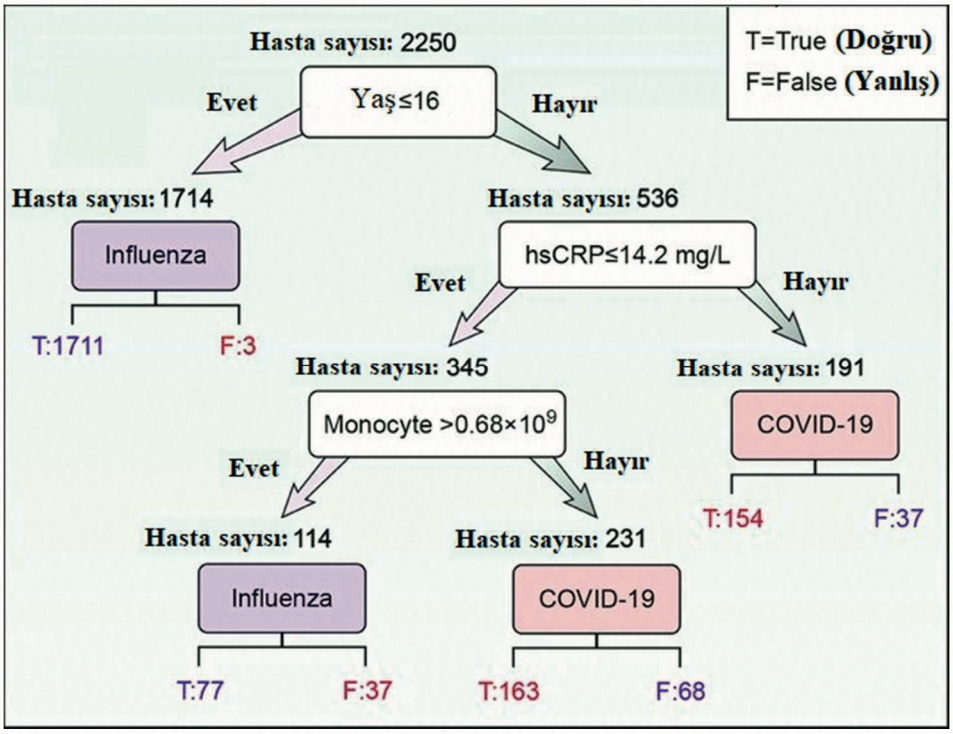
rindeki etkilerinin yorumlanmasına olanak sağlayan bir Benzetilebilirlik yaklaşımını işaret etmektedir.

iv. kNN Modelinde Yorumlanabilirlik

Sınıflandırma problemleri kapsamında kullanılan kNN tekniği, sınıfı tespit edilmeye çalışılan özellik değerlerine en yakın konumda yer alan k adet eğitim verisinin sınıflarına göre karar oluşturma yeteneğine sahiptir. Bu bağlamda, en yakın konumdaki k adet verinin (giriş verilerinin) hedef sınıflarından hangisi çoğunluktaysa karar sınıf o olarak tespit edilmektedir. Yorumlanabilirlik açısından bakıldığında, kNN modellerinde benzerlik fonksiyonlarından k değerlerine kadar birçok farklı unsur söz konusu olmakta ve en önemlisi, bu modelde eğitim sürecinin yansımalarını gösteren model içi parametreler var olmamaktadır. Bu durumda kNN modellerinin farklı k değerleri ya da benzerlik fonksiyonları üzerinden çalıştırılarak, Ayrıştırılabilirlik yaklaşımı sayesinde değişen bu değerlerin çıkış sınıfındaki etkileri yorumlanabilir hale getirilebilmektedir. Ancak kuşkusuz ki bu durum doğrudan parametre hakimiyetinden ziyade değişen koşullardan hareketle model gidişatının değerlendirilmesi anlamına gelmektedir [2, 20].

v. Karar Ağaçları Modelinde Yorumlanabilirlik

Karar Ağaçları tekniği, eğitim veri setinden hareketle görsel bir ağaç akışı elde edilmesini mümkün kıldığı için yorumlanabilirliği yüksek modellerin oluşturulmasını sağlayan bir teknik olarak bilinmektedir. Elde edilen ağaç yapılarındaki akışlar ve değerler süreci doğrudan yorumlayabilmeye fırsat tanıdığı için Benzetilebilirlik, farklı model değerleri ile ağaç akışlarının nasıl değişebildiğini görmek de söz konusu olduğu için Ayrıştırılabilirlik yaklaşımları Karar Ağaçları modellerindeki saydam tasarım yaklaşımları arasında yer almaktadır. Örneğin, Zhou vd. [28] tarafından ortaya konulan oldukça güncel bir çalışma, Influenza ve COVID-19 arasındaki ayrımı güvenli ve yorumlanabilir bir ağaç görseli üzerinden yapılmasını sağlayan Karar Ağaçları tekniği üzerine kurgulanmıştır (Şekil 2.5). Elde edilen görselden hareketle kurulan modelin klinik anlamda yanlış bir karar süreci mi izlediği yoksa yanlış değerler üzerinden mi kararlara ulaştığı gibi yorumlamalar yapmak oldukça kolay olmaktadır.



Şekil 2.5. Influenza ve COVID-19 ayrımında yorumlanabilir Karar Ağaçları kullanımı [28].

Buraya kadar ifade edilen yorumlanabilir modeller, kendilerine has özellikler nedeniyle farklı açılardan yorumlanmaya ihtiyaç duydukları için XAI literatürü açısından modele bağımlı tasarım süreçlerini bir süre gerekli kılmış, ancak Makine Öğrenmesi çözümlerinin kara kutu olma düzeylerinin artışı, modele bağımlılığın ötesinde ihtiyaçları tetiklemiştir. Neticede kara kutunun açıklanması olarak da adlandırılabilecek söz konusu çabalar, modelden bağımsız (yani tasarlanan modele entegre edilebilen) açıklanabilirlik çözümlerinin tanımlanmasına yol açmış, bu süreç günümüz Derin Öğrenme modellerine kadar, artan bir ivmeyle devam etmiştir.

2.4. KARA KUTUNUN AÇIKLANMASI

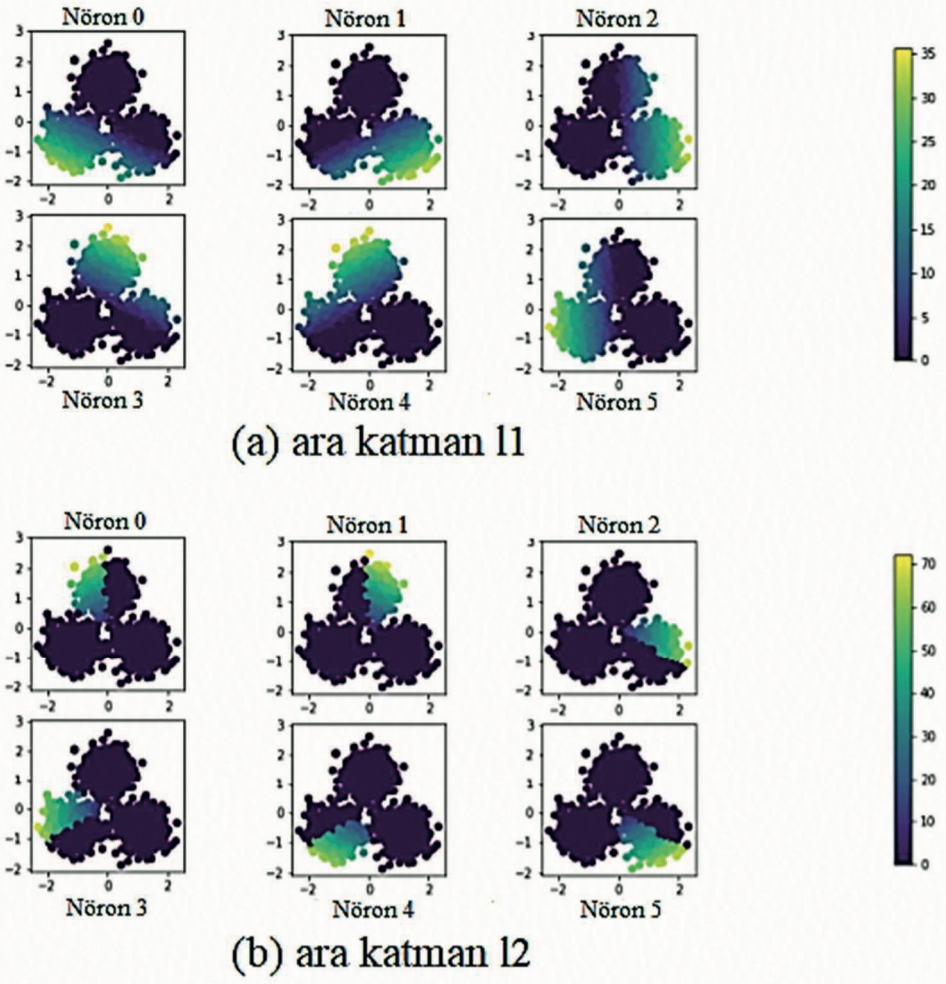
Kara kutunun açıklanması problemi, günümüz XAI araştırmalarının merkezinde yer almaktadır. Saydam tasarımdan farklı olarak artık modele bağımlı çözümlerden ziyade farklı Makine Öğrenmesi modellerine uygulanabilen

(entegre edilebilen) yöntemler, kara kutu düzeyindeki YZ destekli çözümlerin açıklanabilmesine olanak sağlamaktadır. Kara kutunun açıklanması bağlamındaki önde gelen yöntemler genel olarak şöyledir:

2.4.1. Katmanlı Sinir Ağlarında ve Karma Modellerde Açıklanabilirlik

Katmanlı sinir ağları genel olarak çok katmanlı algılayıcı (MLP: Multi-Layer Perceptron) olarak bilinen ve farklı katmanlar arasında organize edilmiş algılayıcıların bir arada olduğu mimariler için kullanılan bir ifade olarak bilinmektedir. Bununla birlikte, literatürde yer alan farklı amaçta sinir ağları mimarileri de (örneğin, Elman, Kohonen, SOM) bu kapsam içerisine girmekte, hatta günümüz Derin Öğrenme odaklı sinir ağları teknikleri de aynı yönde kabul edilebilmektedir. Özünde ilerleyen paragraflar altında da ifade edileceği üzere; model bağımsız yöntemlerle açıklanabilir hale getirilebilen katmanlı sinir ağları için birtakım çalışmalar kapsamında tasarlanan açıklanabilirlik çözümleri de mevcuttur.

Lin vd. (2020) [29] tarafından tasarlanan bir çözüm, Doğrusal Diskriminant Analizi (LDA - Linear Discriminant Analysis) olarak bilinen ve Gaussian (normal dağılımlı) sınıflandırma öncesinde daha makul sayıda özellik sayısına ulaşarak süreci daha pratik hale getiren; bunu da özelliklerin doğrusal birleşimini bularak yapan bir teknikten yola çıkmıştır. Kısaca FF-MLP olarak adlandırılan çözüm, iki sınıflı bir LDA'dan yola çıkarak, giriş ve çıkış katmanları dışında l1 ve l2 olarak adlandırılan iki ara katmanın yer aldığı bir MLP mimarisi tasarımına dayanmaktadır. Bu noktada, sırasıyla giriş katmanından l1 katmanına çoklu paralel LDA'lar ile yarım-uzay bölümlenmesi (giriş değerleri uzayının bölümlenmesi), l1 katmanından l2 katmanına her bir Gaussian modalitesinin bir algılayıcı tarafından temsil edildiği alt-uzay izolasyonu (odaklanılan ilgi alanlarının girişten gelen diğer alanlardan ayrılması) ve l2 katmanından çıkış katmanına da ilgili Gaussian modelitelerin farklı sınıflara tekabül ettiği sınıf bazlı alt-uzay birleşimi (odaklanılan ilgi alanlarının birleştirilmesi) gerçekleştirilerek açıklanabilirlik altyapısı elde edilmektedir [29]. Esasında geleneksel geri yayım algoritması tabanlı MLP'ye alternatif bir mimari oluşmasına sebep olan bu çözüm, Şekil 2.6'da gösterildiği gibi farklı sınıflara ilgi durumlarının görselleştirilebildiği bir açıklanabilirlik seviyesi yakalanmasını mümkün kılmaktadır.

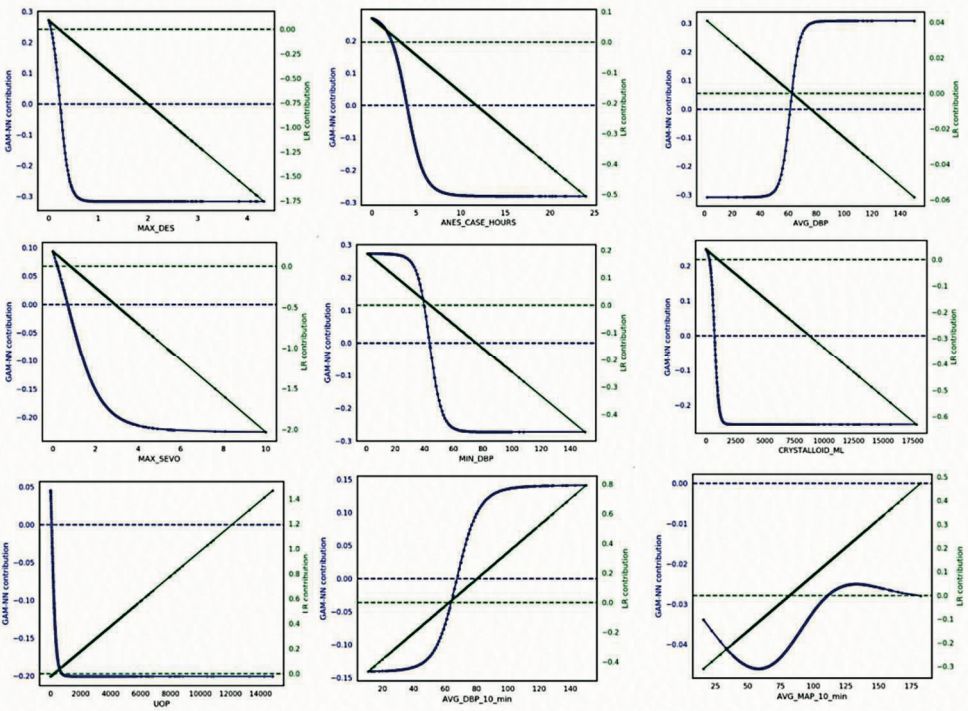


Şekil 2.6. FF-MLP ile sınıf ilgilerinin ara katmanlardaki algılayıcılarda yansımaları [29].

Katmanlı sinir ağları, matematiksel çözüm mekanizmaları olan parçaların bir arada olduğu mimari yapılar sunabildiklerinden dolayı zamanla farklı tekniklerle iş birliğine giderek karma modellenmeler içerisine de girmiştir. Örneğin, ANFIS tekniği katmanlı sinir ağları ve Bulanık Mantık kurallarının bir arada kullanıldığı bir karma çözüm olarak, farklı problemlere başarılı sonuçlar üretebilmiştir. Bu aşamada, ANFIS tekniğinin içerisinde taşıdığı kural tabanlı karakteristikler, bu tekniğin açıklanabilirliğini sağlama nokta-

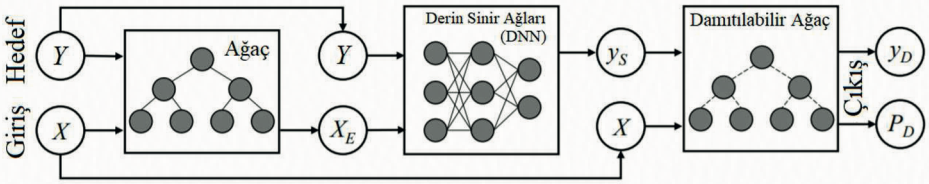
sında önemli fikirler verebilmektedir. Nitekim, Aghamohammadi vd. [30] tarafından kalp krizi riskini tahmin etme amaçlı geliştirilen ANFIS modelinde açıklanabilirlik faktörü söz konusu model dinamikleri dikkate alarak tasarlanmıştır. Çalışmadaki karma çözüm yaklaşımı, ANFIS özelinde yer alan üyelik fonksiyon parametrelerinin Genetik Algoritma yardımıyla optimize edilerek, model başarımını değerlendiren RMSE (Root Mean Square Error) değerinin minimize edilmesi yönünde bir ANFIS-GA sisteminin oluşturulmasını içermiştir. ANFIS-GA sisteminde kalp krizi tahmini sürecinde etkisi olan giriş özelliklerinin önem derecelerini belirlemek ve böylelikle açıklanabilirlik sağlamak adına, ilgili test veri seti üzerinde Önem Değerlendirme Fonksiyonu (IEF - Importance Evaluation Function) kullanılmış ve ilgili fonksiyon bulgularına göre her bir test veri örneği üzerinde giriş özelliklerinin etkileri açıklanabilmiştir [30].

Lee vd. (2021) [31] tarafından gerçekleştirilen bir diğer çalışma, lineer (doğrusal) olmayan değişkenlerin hesaplanabilmesini sağlayan Genelleştirilmiş Katkı Modelleri (GAMs - Generalized Additive Models) tekniğinin katmanlı sinir ağlarına uygulanarak, özelliklere açıklama getirilebilmesi yönünde bir çözüm ortaya koymuştur. Her ne kadar gelecek paragraflar altında ayrıca ele alınan Derin Öğrenmeye yönelik açıklanabilirlik kapsamına tekabül ediyor olsa da söz konusu çalışmada Derin Sinir Ağları (DNN) modeline uygulanan GAM yardımıyla hastane içi ölüm-riski sınıflandırma probleminde sınıflandırmayı en çok hangi giriş özelliklerinin etkilediği belirlenebilmiştir. Kısaca GAM-NN olarak ifade edilen modelde, giriş özellikleri ilk olarak farklı sayılarda ReLu ya da Tanh taşıyan algılayıcılarla karşılandığı gizli katmanlara geçmekte, ardından Tanh fonksiyonlu bir adet algılayıcı taşıyan yeni ara katmanlar tarafından etki değerlerinin -1 ile 1 aralığına taşınması sağlanmaktadır. Nihayetinde Tanh fonksiyonlu katmanlardan elde edilen değerler birleştirilip tek bir logistik katmana giriş olarak verilerek, logistik ağırlıkların özellik giriş değerleriyle çarpımı üzerinden her bir özelliğin çıkıştaki etkileri hesaplanabilmektedir [31]. Böylelikle Şekil 2.7'de gösterildiği gibi her giriş özelliğinin x ekseni boyunca farklı örnekler (hastalar) için değişken etkileri y eksenindeki GAM-NN (mavi renk) ve Lineer Regresyon (yeşil renk) kapsamında görselleştirilebilmektedir [31].



Şekil 2.7. GAM-NN sistem kapsamında farklı giriş özelliklerinin veri örnekleri (çıkışlar) üzerindeki etkileri [31].

Katmanlı sinir ağlarında açıklanabilirliği sağlamak için karar ağaçları tabanlı modellerin kullanımı literatürde zaman zaman başvurulan bir çözüm olarak göze çarpmaktadır. Örneğin, [32] tarafından ortaya konulan çalışma karar ağaçları yardımıyla DNN modellerine açıklanabilirlik ekleme üzerine kurgulanmıştır. Kısaca TNT (Tree-Network-Tree / Ağaç-Ağ-Ağaç) olarak adlandırılan çözümde, ilk olarak hedef veri seti için karar kurallarını çıkaracak bir ağaç modeli kullanılmakta, böylelikle DNN sistemi için nispeten anlamlandırılmış girişlerin kullanılması sağlanmaktadır. Bu girişlerden yeni sınıflar üreten DNN çıktı verileri tekrar bir başka ağaç modeli tarafından, bilinen giriş özelliklerine dair yorumlanabilir düzeyde bir karar süreci gösterimi elde etmek için kullanılmaktadır (Şekil 2.8).



Şekil 2.8. TNT yaklaşımı ile katmanlı sinir ağlarına yönelik karma açıklanabilirlik çözümü [32].

Anlaşılacağı üzere, taşıdıkları karmaşıklık düzeyi arttıkça kara kutu düzeyine ulaşan katmanlı sinir ağlarında gerek karma modellemelerin doğasından yola çıkarak gerekse özel birtakım düzenlemelerle açıklanabilirlik sağlanabilmektedir. Söz konusu özel düzenlemeler günümüzde Derin Öğrenme şemsiyesi altında incelenen teknikler için özellikle görsel veriler üzerinden daha anlamlı çıktılar elde edilmesine olanak sağlamaktadır. Bu husus, ilerleyen paragraflar altında ayrıca ele alınmıştır.

2.4.2. LIME Yöntemi ile Açıklanabilirlik

Local **I**nterpretable **M**odel-agnostic **E**xplanations tanımının kısaltmasıyla ifade edilen LIME yöntemi, kara kutu formundaki Makine Öğrenmesi modelinin, hedef veri setinden üretilen çeşitli yapay örneklere karşı tahminlerine bakarak, orijinal verinin varlığı ve yokluğu durumlarına karşı olan değişiklikler üzerinden kara kutuya yakınsanmış yorumlanabilir bir çözüm üretmektedir [33]. LIME yöntemi hem metin hem de görüntü tabanlı verilerle çalışabilmesiyle model bağımsız olma konusunda büyük avantajlar sağlamaktadır. Netice olarak veriye odaklanan bir yöntem olması nedeniyle kara kutu olmasalar bile daha karmaşık problemlerin çözümlendiği regresyon tabanlı teknikler ve Karar Ağaçları ile kullanılabilirken, yine kara kutu formundaki katmanlı sinir ağları mimarileri ya da Destek Vektör Makineleri gibi bazı spesifik tekniklerde de çalışmaktadır [33, 34].

LIME yönteminin temel mantığı aşağıdaki denklemde ifade edilen ve x özelliğine karşı Ω karmaşıklık değerindeki modelin orijinal tahmin f ve muhtemel açıklama değerleri (g) kümesi G dikkate alınmak suretiyle, en iyi açıklamayı verecek L hata değerini minimize etmeye çalışan bir optimizasyon sürecine tekabül etmektedir [2, 33]:

$$açıklama(x) = \arg\min_{g \in G} L(f, g, \pi_x) + \Omega(g) \quad (2.4)$$

Şekil 2.9, bir gıdanın yenilebilir mi yoksa zehirli mi olduğunu tanımlayan (tahmin edilmesini sağlayan) özellik değerlerinin LIME yöntemi sayesinde elde edilen açıklanabilir dağılımlarını göstermektedir [34].

Şekil 2.9. LIME yöntemi ile özellik değerlerinin tahmin üzerindeki etkileri [34].

LIME yöntemi görsel veriler üzerinde işlemler yapan kara kutu tekniklerde kullanıldığında ise Şekil 2.10'da gösterilenlere benzer bir açıklama yapısı elde edilmektedir. Şekil 2.10 (a)'daki görsel, fotoğraf üzerinden kedi tespit eden bir modele ait olup, kedi üzerindeki yeşil görselleştirmeler tahmini doğrulayan noktaları açıklamakta, kırmızı görselleştirmeler (köpek üzerinde yer alanlar) ise kedi tahmini aleyhindeki noktaları açıklamaktadır. Benzer şekilde Şekil 2.10 (b) ise fotoğrafta ekmek ve çilek tespiti yapan bir modelde ekmek tespitine hassas açıklamalar betimlenmektedir [2, 34].

Şekil 2.10. LIME yöntemi ile görsel veriler üzerinde açıklama görselleştirmeleri [2, 34].

2.4.3. SHAP Yöntemi ile Açıklanabilirlik

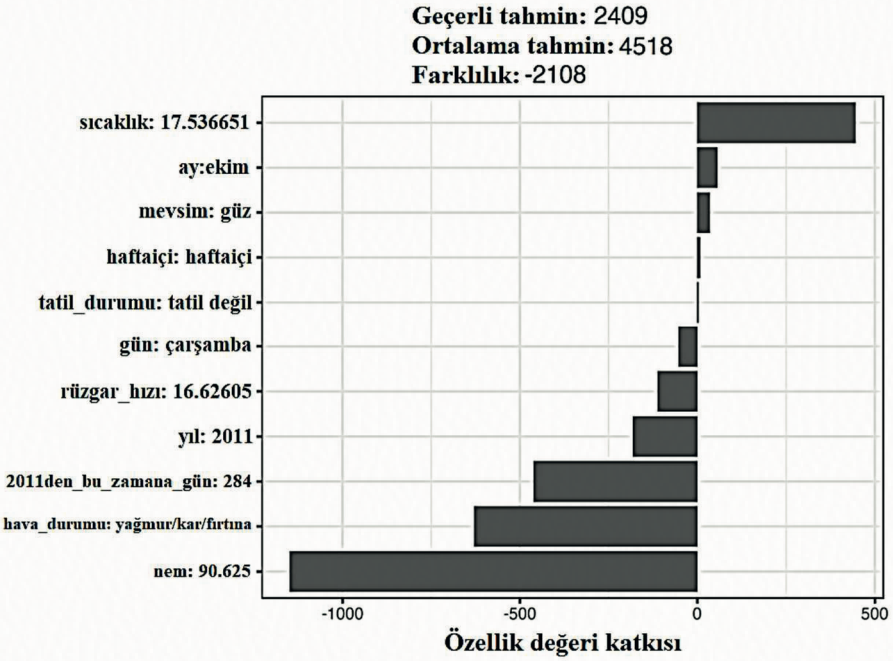
Lundberg ve Lee (2017) [35] tarafından literatüre kazandırılmış olan ve **SH**apley **A**dditive ex**P**lanations ifadesinin kısaltılmasıyla ifade edilen SHAP yöntemi, tıpkı LIME gibi model bağımsız bir açıklanabilirlik çözümüdür. Yöntemin temeli Shapley değerleri adı verilen ve Koalisyon Tabanlı Oyun Teorisi'nden (Coalitional Game Theory) esinlenilen; ilgili veri setindeki her bir giriş özelliğinin birer oyuncu olduğu ve modelin gerçekleştirdiği tahminde her birinin ne kadar payı olduğunu açıklayan değerlere dayanmaktadır. Buna göre, modeldeki belli bir *X* örneği ile elde edilen *Y* çıktı değerinin, model ortalama tahmin değerinden olan farkına her bir giriş özelliğinin ne kadar katkıda bulunduğu belirlenmek suretiyle giriş özelliklerinin çıktıdaki rolü açıklanabilmektedir.

Aşağıdaki denklemden de anlaşılacağı üzere Shapley değerleri, ortalama tahmin değerinde yer almayan her bir özellik için çoklu integral hesabı yapılarak elde edilmektedir. İlgili denklem, *X1*, *X2*, *X3* ve *X4* olmak üzere dört giriş özelliği olan bir modelde *X1* ve *X3* giriş özelliği değerleri olan S birlikteliğinin (koalisyonunun) tahmin değerlendirmesini yapmaktadır [2]:

$$val_x(S) = val_x(\{1,3\}) = \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}\,(X1, X2, X3, X4) d\mathbb{P}x2, x4 - E_x(\hat{f}(X)) \qquad (2.5)$$

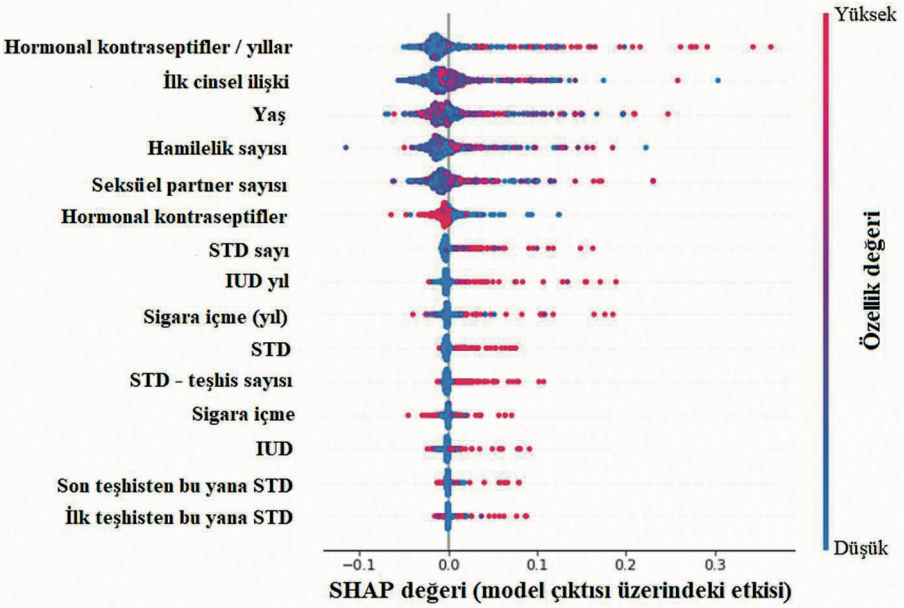
SHAP yöntemi ile elde edilen açıklanabilir değerler, farklı görselleştirme süreçleriyle insanlar için çok daha anlaşılır hale getirilebilmektedir. Şekil 2.11, hava koşullarına göre kaç adet bisiklet kiralanabileceğini tahminleyen bisiklet kiralama veriseti (bike rental dataset) kapsamında farklı giriş özelliklerinin çıkış üzerindeki pozitif ve negatif yönde etkilerini gösteren bir çubuk grafik yaklaşımını göstermektedir [2]. Grafikten anlaşılacağı üzere ilgili tahmin senaryosu kapsamında sıcaklık özelliği tahminde pozitif yönde bir etki gösterirken, havanın yağmurlu, karlı ya da fırtınalı olması ile birlikte nem durumu negatif yönde bir etki göstermektedir.

SHAP yönteminden yola çıkılarak tasarlanabilen bir başka görsel türü de özet plot grafiği olarak bilinmektedir. Bu grafik türünde her bir giriş özelliğinin farklı giriş veri örnekleri için çıkış üzerinde ne kadar etkileri olduğu tek bir grafik yardımıyla okunabilmektedir.

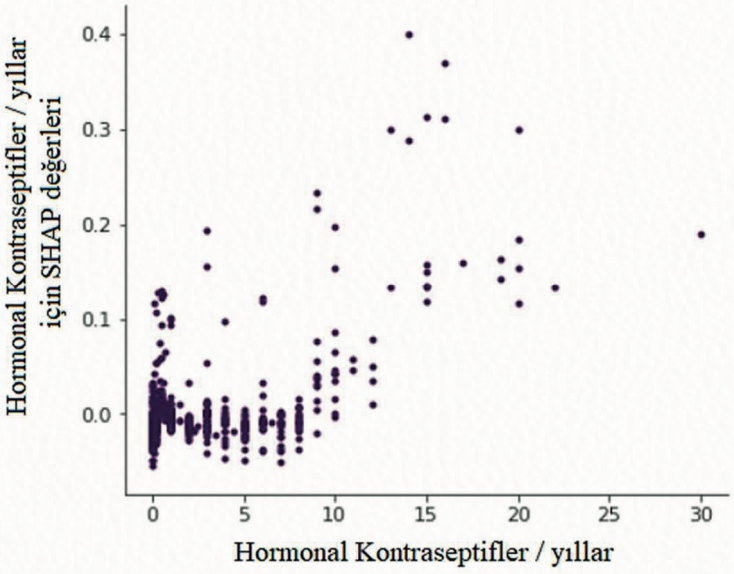


Şekil 2.11. SHAP yöntemi ile farklı giriş özelliklerinin çıkış üzerindeki pozitif ve negatif yönde etkileri [2].

Şekil 2.12'de gösterilen grafiğe göre, kadınlarda rahim ağzı kanseri risk veri setinde hormonal kontraseptifler özelliğinde düşük yıl değerlerinin olması kanser risk değerini (tahmin değerini) azaltırken, tersi durumda artmasına sebep olmaktadır. Grafikte her bir nokta ilgili giriş özelliği ve ilgili giriş verisi için Shapley değerlerini ifade etmekte, yine her özelliğin etkisi yüksek ve düşük aralıkta renklerle görselleştirilmektedir [2]. Bu noktada, her bir giriş özelliğindeki farklı değerlerin tahmin üzerindeki olasılık tabanlı etkisini de tekil ölçekte görselleştirmek mümkündür. Bu şekildeki grafikler SHAP bağımlılık grafiği olarak adlandırılmakla beraber, Şekil 2.12'de görülen hormonal kontraseptifler özelliğine has bağımlılık değer değişimleri Şekil 2.13'te olduğu gibi elde edilebilmektedir [2].



Şekil 2.12. SHAP yöntemi kapsamında özet plot grafiği örneği [2].



Şekil 2.13. Hormonal kontraseptifler özelliği için SHAP bağımlılık grafiği [2].

SHAP yöntemi kapsamında elde edilen değerler için farklı gruplandırmalarla giriş özellikleri arası etkileşim grafikleri ya da ilgili giriş özelliklerinin etki benzerliklerinin tek bir görselde açıklanabildiği kümeleme grafikleri de üretilebilmektedir [2]. Bütün bu görselleştirme süreçleri farklı açıklanabilirlik avantajları sağlamakla beraber, özünde Shapley değerleri bu açıklama süreçlerinde her zaman anahtar rol oynamaktadır.

2.4.4. Derin Öğrenme'de Açıklanabilirlik

Makine Öğrenmesinin günümüz şartlarına göre uyarlanarak, daha gelişmiş modellerle desteklendiği formatı olan Derin Öğrenme, çoğunlukla karmaşıklığı daha yüksek sinir ağları teknikleriyle bilinmektedir. Bu teknikler arasında yer alan Evrişimsel Sinir Ağları (CNN) özellikle görüntü verileri üzerindeki başarılarıyla, Uzun-Kısa Dönem Hafıza (LSTM) modeli birbirleriyle bağlantılı veri yığınlarına karşı dayanıklılığıyla ve Derin Sinir Ağları (DNN), Derin İnanç Ağları (DBN), Çekişmeli Üretici Ağlar (GAN) gibi diğer alternatif teknikler de geleneksel Makine Öğrenmesi ile çözülemeyecek zorluktaki problemleri çözebilmeleriyle Derin Öğrenme literatürüne pozitif yönde katkıda bulunan aktörler arasında yer almaktadır. Ne var ki bütün bu teknikler ve geliştirilen her türlü alternatif Derin Öğrenme çözümleri, baş etmeye çalıştıkları veri yığınları ve yüksek başarım ile birlikte optimum performans beklentileri neticesinde kara kutu düzeyinde tasarlanmak durumunda kalmaktadır. Dolayısıyla XAI yönündeki çalışmalar, bu tekniklerin de bir şekilde açıklanabilirlik seviyesine gelmesi adına çabaları barındırmaktadır.

i. Katman Bazlı Uygunluk Yayılımı

Literatürde Katman Bazlı Uygunluk Yayılımı (LRP-Layer-Wise Relevance Propagation) olarak bilinen Derin Öğrenme modelinde çıkışta elde edilen sonuçları üzerinde giriş verilerinin katkı düzeylerinin açıklanmasına dayanan bir yöntemdir [36]. LRP ile aşağıda verilen temel denklemden yola çıkılarak, elde edilen çıktının ağ üzerinden giriş katmanına kadar ağ ağırlıkları ve aktivasyon değerlerini kullanarak yayılması ve böylelikle girişlerle çıkış arasında anlaşılabilir köprüler kurulması sağlanmaktadır. Denklem kısaca, çıktıdan gelen R (relevance) ile ifade edilen etki değerinin çıktı katmanından önceki katmanlarına doğru iteratif bir şekilde; a aktivasyon değerindeki j. algılayıcı ile ardışık katmanda yer alan j. algılayıcı arasındaki etki değerlerinin hesaplanmasını sağlayarak ilgili bağlantıyı kurmayı amaçlamaktadır [37]:

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{0,j} a_j w_{jk}} R_k \quad (2.6)$$

LRP yardımıyla elde edilen giriş-çıkış arası açıklanabilirlik değerleri, giriş verisi üzerinde hangi bölümlerin dikkate alındığı hakkında yerinde bilgi vermekte, böylelikle bir Derin Öğrenme modelinin çıktılara ulaşırken girişteki veride gerçekten doğru noktalara odaklanıp odaklanmadığı anlaşılabilmektedir. Bu bağlamda, özellikle görüntü verileri açısından bir örnek olarak Şekil 2.14'deki örnek bulgular, fotoğraflar üzerinde cinsiyet ve yaş tahmininde bulunan GoogleNet Derin Öğrenme modelinin odaklandığı piksel alanlarını göstermesi bakımından LRP modelinin çalışma mantığını tanımlamaktadır [38].

Şekil 2.14. LRP ile fotoğraf giriş verileri üzerinde odaklanılan alanların açıklanması [38].

Derin Öğrenme modellerinin en etkin kullanıldığı problem çözümlemeleri görüntü verileriyle ilişkili olduğu için LRP yöntemi görüntü üzerinde odaklanılan bölgeleri göstererek, -tıpkı bir insan gibi- çıkartıma önderlik eden odak noktalarını işaret edebilmektedir.

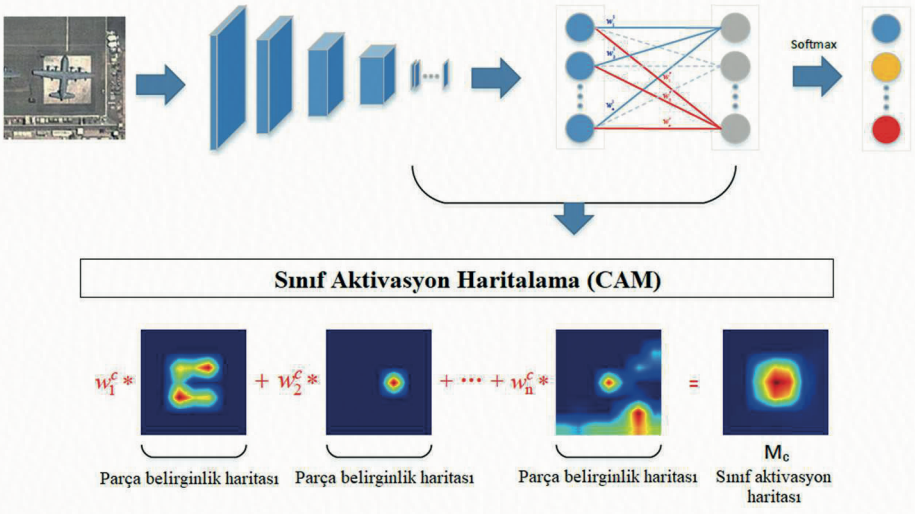
ii. Sınıf Aktivasyon Haritalama

Kısaca Sınıf Aktivasyon Haritalama (CAM - Class Activation Mapping) olarak da bilinen yöntemi, CNN başta olmak üzere görüntü üzerinde

çalışan Derin Öğrenme modellerinin açıklanabilirliği için yaygın bir biçimde kullanılan bir çözüm olarak kabul edilmektedir. Tıpkı LRP gibi giriş görüntü verisi üzerinde odaklanılan alanları işaret eden CAM yöntemi, çözümü ısı haritaları oluşturarak tasarlamaktadır. Böylelikle giriş görüntü verisi üzerinde odaklanılan alanların önem seviyeleri ısı haritası renk dağılımlarına göre ayırt edilebilmektedir. Söz konusu yöntem matematiksel olarak, ağırlıklı giriş özellikleri harita değerlerinin bir araya toplanarak ortak aktivasyon haritaları oluşturulmasını sağlaması ve bu harita değerlerinden de giriş görüntüleri üzerinde ısı haritaları oluşturulması yönünde bir çözüm ortaya koymaktadır. Aşağıdaki denklemde ve Şekil 2.15'te gösterildiği üzere, CAM yöntemi oldukça basit bir mantığa dayanmakta; buna göre, farklı algılayıcılardan gelen ağırlıklı özelliklerin toplanarak sınıf aktivasyon haritasının oluşturulması yoluna gitmektedir [39]:

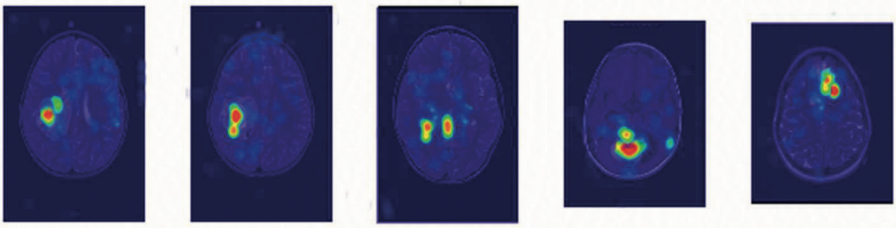
$$w1 * f1 + w2 * f2 + \cdots + wn * fn \tag{2.7}$$

CAM tabanlı açıklanabilirlik çözümleri özellikle sınıflandırma problemlerinde etkin hale gelmektedir. Yine literatür incelendiğinde, CAM yönteminin çözüm kapsamını ve verimliliğini geliştiren alternatif varyasyonlar: Grad-CAM ve Axiom-based Grad-CAM'de alternatif CAM yöntemleri olarak kullanılmaktadır [40, 41].



Şekil 2.15. CAM yöntemi ile açıklanabilirlik mekanizması [39].

CAM yöntemi görüntünün önemli olduğu alanlarda pratik açıklanabilirlik sağlamak için oldukça etkilidir. Bu konuda örneğin YZ'nin oldukça etkin olduğu sağlık alanında, medikal görüntüler üzerinden elde edilen teşhislerin açıklanabilmesi çok daha etkili bir hale gelmektedir. Şekil 2.16'da [42] tarafından gerçekleştirilen çalışmada, özel bir CNN modelinde beyin tümörü teşhisinde CAM yöntemiyle elde edilen görsel bulguların etkinliği görülmektedir. Söz konusu çalışma elde edilen bulguları doktorlar ile birlikte de değerlendirmiş ve geliştirilen sistem insan odağındaki değerlendirmelerden de pozitif yönde dönütler almıştır [42].



Şekil 2.16. CAM yöntemi ile beyin tümörü görüntüleri üzerinde açıklanabilirlik [42].

Derin Öğrenme'de açıklanabilirlik, bu alanın 2020'li yıllardaki etkin YZ alanı olması nedeniyle literatürde oldukça ilgi çekmektedir. Alternatif Derin Öğrenme tabanlı zeki sistem çözümleri için XAI çözümleri halen geliştirilmekte olup, literatürün bu yönden hızlı ilerlemesi güvenilir ve güçlü zeki sistemlerin tasarlanması aşamasında umut veren bir gelişme olarak görülebilmektedir. Bu noktada, Derin Öğrenme'de artan surette devam eden parametre optimizasyonu problemi, yenilikçi açıklanabilirlik çözümlerine olan ihtiyacı da her zaman canlı tutacaktır.

2.5. SONUÇ VE DEĞERLENDİRMELER

Bu bölüm altında, XAI kavramı, konunun arka-planında yer alan sebepler ve saydam tasarım (yorumlanabilirlik) ve kara kutu sistemlerde açıklanabilirlik boyutlarında ele alınmış; farklı açıklanabilirlik yöntemleri genel hatlarıyla vurgulanmıştır. Anlaşılacağı üzere, XAI çözümlerinin çok boyutlu etkileri bulunmakta, bu etkiler günümüzün ve geleceğin güvenli, denetlenebilir ve geliştirilebilir zeki sistemlerinin elde edilmesinde anahtar rol oynamaktadır. Detaylı irdelendiğinde, XAI ile bağlantılı olarak şu değerlendirmelere ulaşılmaktadır;

- Günümüz dijital araçları, gelişen iletişim teknolojileri ve etkinliği hızla artan hesaplama teknolojileri sayesinde insan hayatında sıradanlaşan unsurlar haline gelmektedir. Ancak bu araçların sıradanlaşması ve hayatın her alanına nüfuz etmesi, onların daha basit arayüzler üzerinden geliştirilmesini gerekli kılmaktadır. Bu geliştirme ise güven ve kontrol gibi insani gereklilikleri sağlamak adına XAI gibi unsurları ön plana çıkarmaktadır.
- XAI özellikte günümüzde insan nezdinde anlaşılırlığı gittikçe bulanıklaşan zeki sistemlerin davranışlarının anlaşılması noktasında etkin bir düzeyde kullanılmaktadır. Her ne kadar çözümleri yorumlanabilir seviyede olan geçmiş modeller nedeniyle daha derin bir geçmişe sahip olsa da XAI'nın Derin Öğrenme ve karma Makine Öğrenmesi çözümleri nedeniyle YZ'nin geleceğini, güvenliğini ve hatta etiğini ilgilendiren kritik bir araştırma alanı haline geldiğini ifade etmek yanlış olmayacaktır.
- Son kullanıcı tarafında basit arayüzler pratiklik sağlasa da içerisinde YZ tabanlı mekanizmalar içeren araçlarda zeki sistem mekanizmalarının güven düzeyi ve hata potansiyelleri ancak XAI ile değerlendirilebilmektedir. Bu noktada yorumlanabilirlik faktörünü XAI düzeyine taşıyan Derin Öğrenme'de CAM yöntemi görüntü verileri üzerinde takip edilirliği sağladığı gibi, LIME yöntemi de hem görüntü hem de metin tabanlı veriler üzerinde çözümler sağlayabilmektedir. Yine SHAP yöntemi de dayandığı Koalisyon Tabanlı Oyun Teorisi itibariyle güçlü bir bilimsel yaklaşım ihtiva etmektedir. XAI ailesi daha birçok alternatif çözüme ev sahipliği yapmakla birlikte, sadece Derin Öğrenme tabanlı zeki sistemler değil, geleneksel teknikler içerebilen karma modeller de XAI sayesinde insan nezdinde seviye atlamaktadır. XAI bu açıdan uluslararası YZ politikalarını ve standartlarını da tetikleyen en önemli unsurlardan biri haline gelmektedir.
- XAI ülkemizde ve dünyada görece yeni bir kavram olsa da oldukça hızlı gelişmektedir. Halihazırda YZ'nin uygulandığı farklı alanlarda XAI sayesinde insan-makine arası daha etkin etkileşimler oluşturulabilmektedir. Bu konuda XAI yöntemlerinin genel değerlendirmeleri farklı yöntemlerin etkinliği konusunda fikir vermektedir [43-46]. Bununla birlikte XAI'nın disiplinlerarası potansiyelinin analizi [47] ve hatta XAI yöntemlerinin kullanıcı deneyimi çerçevesinde yorumlanması [42, 48-50] yönünde güncel çalışmalar da dikkat çekicidir.

- Bridle'ın (2018) [51] işaret ettiği gibi, gittikçe insan algısı karşısında bulanıklaşan teknolojilerin anlaşılırlığını ve güvenliğini sağlamak gelecek süreçte de XAI kapsamındaki etkin gelişmeler sayesinde mümkün olacaktır. Bu gelişmeler hiç kuşkusuz ki insan-makine etkileşimi nedeniyle Bilgisayar Bilimleri ve Mühendisliği, Yazılım Mühendisliği, Elektrik-Elektronik Mühendisliği, Kullanılabilirlik Mühendisliği, Psikoloji ve Sosyoloji gibi daha nice teknoloji ve insan odaklı disiplinlerin iş birliğini artan düzeylerde gerekli kılacaktır.
- YZ'nin günümüzde çok daha fazla veri üzerinde başarılı bir şekilde çalışabilmesi ve nihayetinde daha fazla enerji tüketimine ihtiyaç duyması, çeşitli düzenlemeleri de beraberinde getirmektedir. Özellikle kişisel mahremiyeti (veri gizliliğini) dikkate alan, çevreye zararlı olmayan, şeffaf, adil ve yorumlanabilir zeki sistem tasarımları yönündeki ihtiyaçlar, Sorumlu YZ (Responsible AI) kavramının da oluşmasına neden olmuştur [52]. Sorumlu YZ, mahremiyetin önemli olduğu uygulamalarda [53-55], sağlık uygulamalarında [56-58], toplumsal gözetim uygulamalarında [59, 60] ve hatta iklim krizi [61] gibi önemli toplumsal sorunlarda ön plana çıkmaktadır. Bununla birlikte yorumlanabilirlik ve XAI faktörleri Sorumlu YZ'nin en önemli bileşenleri arasında bulunmaktadır. Bu açıdan özellikle XAI'nın geleceğin Sorumlu YZ sistemleri açısından kritik bir araştırma konusu olduğu anlaşılmaktadır.
- Ülkemiz açısından dikkate alındığında, XAI kullanımının milli teknolojilerimizde güvenliği ve denetimi sağlama noktasında gerekli olacağı aşikardır. Bununla birlikte, Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Başkanlığı ile Sanayi ve Teknoloji Bakanlığı iş birliğinde olmak üzere, çeşitli paydaşların da katılımıyla hazırlanmış olan Ulusal Yapay Zekâ Stratejisi 2021-2025'in vizyonu, "müreffeh bir Türkiye için çevik ve sürdürülebilir yapay zekâ ekosistemiyle küresel ölçekte değer üretmek" olarak ifade edilmektedir [62]. Bu vizyon ve ilgili stratejik öncelikler yönünde tasarlanacak çevik, sürdürülebilir YZ ekosisteminde XAI kullanımının sağlanması, ülkemiz stratejik öncelikleri ve değerleriyle uyumlu zeki sistemlerin elde edilmesi bakımından oldukça önemlidir. Bu yöndeki milli zeki sistemler, ülkemizi uluslararası arenada lider konuma yükseltme görevini de üstlenmiş olacaklardır.

KAYNAKLAR

[1]. Kaçar, T. (2021). Açıklanabilir Yapay Zekâ. *Yapay Zekâ Etiği* (Ed. U. Köse). Nobel Yayıncılık.

[2]. Molnar, C. (2018). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Çevimiçi: https://www.firmai.org/bit/ (Erişim 21 Mayıs 2022).

[3]. Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

[4]. Keskinbora, K. H. (2019). Medical ethics considerations on artificial intelligence. *Journal of Clinical Neuroscience*, *64*, 277-282.

[5]. Köse, U., & Vasant, P. (2017). Fading intelligence theory: A theory on keeping artificial intelligence safety for the future. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)* (1-5). IEEE.

[6]. Köse, U. (2018a). Are we safe enough in the future of artificial intelligence? A discussion on machine ethics and artificial intelligence safety. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *9*(2), 184-197.

[7]. Köse, U. (2018b). Yapay zekâ: Geleceğin biliminde paradokslar. *Popüler Bilim Dergisi*, *25*(261), 12-21.

[8]. Köse, U. (2018c). Yapay zekâ ve gelecek: Endişelenmeli miyiz?. *Bilim ve Ütopya*, *24*(284), 39-44.

[9]. Macrae, C. (2019). Governing the safety of artificial intelligence in healthcare. *BMJ Quality & Safety*, *28*(6), 495-498.

[10]. Shaw, J. (2019). Artificial intelligence and Ethics. *Harvard Magazine*, *30*.

[11]. Turchin, A., & Denkenberger, D. (2020). Classification of global catastrophic risks connected with artificial intelligence. *AI & Socıety*, *35*(1), 147-163.

[12]. Yampolskiy, R. V. (2013). Artificial intelligence safety engineering: Why machine ethics is a wrong approach. In *Philosophy and Theory of Artificial Intelligence* (389-396). Springer, Berlin, Heidelberg.

[13]. Yampolskiy, R. V. (Ed.). (2018). *Artificial Intelligence Safety and Security*. CRC Press.

[14]. Bonczek, R. H., Holsapple, C. W., & Whinston, A. B. (2014). *Foundations of Decision Support Systems*. Academic Press.

[15]. Power, D. J. (2002). *Decision Support Systems: Concepts and Resources for Managers*. Greenwood Publishing Group.

[16]. Sebetci, Ö. (2020). *Karar Destek Sistemleri*. Seçkin Yayıncılık.

[17]. Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

[18]. Arı, A., & Berberler, M. E. (2017). Yapay sinir ağları ile tahmin ve sınıflandırma problemlerinin çözümü için arayüz tasarımı. *Acta Infologica*, *1*(2), 55-73.

[19]. Orman, A., Köse, U., & Yiğit, T. (2021). Açıklanabilir Evrişimsel Sinir Ağları ile Beyin Tümörü Tespiti. *El-Cezeri*, *8*(3), 1323-1337.

[20]. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Chatila, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115.

[21]. Yuan, X., He, P., Zhu, Q., & Li, X. (2019). Adversarial examples: Attacks and defenses for deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(9), 2805-2824.

[22]. Gunning, D. (2017). Explainable artificial intelligence (XAI). *Defense Advanced Research Projects Agency (DARPA).* Çevrimiçi: https://www.darpa.mil/ attachments/XAIProgramUpdate.pdf (Erişim 13 Haziran 2022).

[23]. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1-38.

[24]. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), 1-42.

[25]. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

[26]. Lou, Y., Caruana, R., & Gehrke, J. (2012). Intelligible models for classification and regression. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (150-158).

[27]. Fabien, M. (2019). Interpretability and explainability (Part 1). *Explorium.AI.* Çevrimiçi: https://www.explorium.ai/blog/interpretability-and-explainability-part-1/ (Erişim 23 Temmuz 2022).

[28]. Zhou, X., Wang, Z., Li, S., Liu, T., Wang, X., Xia, J., & Zhao, Y. (2021). Machine Learning-Based Decision Model to Distinguish Between COVID-19 and Influenza: A Retrospective, Two-Centered, Diagnostic Study. *Risk Management and Healthcare Policy*, *14*, 595.

[29] Lin, R., Zhou, Z., You, S., Rao, R., & Kuo, C. C. J. (2020). From Two-Class Linear Discriminant Analysis to Interpretable Multilayer Perceptron Design. *arXiv e-prints*, arXiv-2009.

[30]. Aghamohammadi, M., Madan, M., Hong, J. K., & Watson, I. (2019). Predicting Heart Attack through Explainable Artificial Intelligence. In *International Conference on Computational Science* (pp. 633-645). Springer, Cham.

[31]. Lee, C. K., Samad, M., Hofer, I., Cannesson, M., & Baldi, P. (2021). Development and validation of an interpretable neural network for prediction of postoperative in-hospital mortality. *NPJ Digital Medicine*, *4*(1), 1-9.

[32]. Li, J., Li, Y., Xiang, X., Xia, S. T., Dong, S., & Cai, Y. (2020). TNT: An Interpretable Tree-Network-Tree Learning Framework using Knowledge Distillation. *Entropy*, *22*(11), 1203.

[33]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*.

[34]. Ribeiro, M. T. C. (2020). *Lime: Explaining the predictions of any machine learning classifier*. Çevimiçi: https://github.com/marcotcr/lime (Erişim 26 Temmuz 2022).

[35]. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768-4777).

[36]. Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K. R. (2019). Layer-wise relevance propagation: an overview. *Explainable AI: interpreting, explaining and visualizing deep learning*, 193-209.

[37]. Lindwurm, E. (2019). InDepth: Layer-Wise Relevance Propagation: A deep dive into LRP for explaining neural network predictions. *Towards Data Science.* Çevrimiçi: https://towardsdatascience.com/indepth-layer-wise-relevance-propagation-340f95deb1ea (Erişim 18 Ağustos 2022).

[38]. Lapuschkin, S., Binder, A., Muller, K. R., & Samek, W. (2017). Understanding and comparing deep neural networks for age and gender classification. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 1629-1638).

[39]. Fu, K., Dai, W., Zhang, Y., Wang, Z., Yan, M., & Sun, X. (2019). Multicam: Multiple class activation mapping for aircraft recognition in remote sensing images. *Remote Sensing*, *11*(5), 544.

[40]. Fu, R., Hu, Q., Dong, X., Guo, Y., Gao, Y., & Li, B. (2020). Axiom-based grad-cam: Towards accurate visualization and explanation of cnns. *arXiv preprint arXiv:2008.02312*.

[41]. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

[42]. Kumar, A., Manikandan, R., Köse, U., Gupta, D., & Satapathy, S. C. (2021). Doctor's dilemma: An explainable subtractive spatial lightweight CNN for brain tumor diagnosis. *ACM Transactions on Multimedia Computing Communications and Applications*, *17*(3s), Article No.: 105, 1-26.

[43]. Van der Waa, J., Nieuwburg, E., Cremers, A., & Neerincx, M. (2021). Evaluating XAI: A comparison of rule-based and example-based explanations. *Artificial Intelligence*, *291*, 103404.

[44]. Speith, T. (2022). A Review of Taxonomies of Explainable Artificial Intelligence (XAI) Methods. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2239-2250).

[45]. Sahakyan, M., Aung, Z., & Rahwan, T. (2021). Explainable artificial intelligence for tabular data: A survey. *IEEE Access*, *9*, 135392-135422.

[46]. van der Velden, B. H., Kuijf, H. J., Gilhuijs, K. G., & Viergever, M. A. (2022). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical Image Analysis*, 102470.

[47]. Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., ... & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)?–A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence*, *296*, 103473.

[48]. Liao, Q. V., & Varshney, K. R. (2021). Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790*.

[49]. Deperlioğlu, Ö., Köse, U., Gupta, D., Khanna, A., Giampaolo, F., & Fortino, G. (2022). Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: analysis with doctor evaluation. *Future Generation Computer Systems*, *129*, 152-169.

[50]. Chromik, M., & Butz, A. (2021). Human-XAI interaction: a review and design principles for explanation user interfaces. In *IFIP Conference on Human-Computer Interaction* (pp. 619-640). Springer, Cham.

[51]. Bridle, J. (2018). *New dark age: Technology and the end of the future*. Verso Books.

[52]. Ghallab, M. (2019). Responsible AI: requirements and challenges. *AI Perspectives*, *1*(1), 1-7.

[53]. Yang, Q. (2021). Toward responsible ai: An overview of federated learning for user-centered privacy-preserving computing. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *11*(3-4), 1-22.

[54]. Krausova, A. (2022). Socially Responsible Virtual Assistant for Privacy Protection: Implementing Trustworthy AI. *Proceedings of 4th International Workshop on eXplainable and Responsible AI and Law co-located with 18th International Conference on Artificial Intelligence and Law (ICAIL 2021),* Sao Paolo, Brazil.

[55]. Chu, W. (2022). A decentralized approach towards responsible ai in social ecosystems. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 16, pp. 79-89).

[56]. Fosso Wamba, S., & Queiroz, M. M. (2021). Responsible artificial intelligence as a secret ingredient for digital health: Bibliometric analysis, insights, and research directions. *Information Systems Frontiers*, 1-16.

[57]. Trocin, C., Mikalef, P., Papamitsiou, Z., & Conboy, K. (2021). Responsible AI for digital health: a synthesis and a research agenda. *Information Systems Frontiers*, 1-19.

[58]. Smith, H. (2021). Clinical AI: opacity, accountability, responsibility and liability. *AI & Society*, *36*(2), 535-545.

[59]. Sugianto, N., Tjondronegoro, D., Stockdale, R., & Yuwono, E. I. (2021). Privacy-preserving AI-enabled video surveillance for social distancing: Responsible design and deployment for public spaces. *Information Technology & People*. DOI: 10.1108/ITP-07-2020-0534.

[60]. Clarke, R. (2021). Responsible application of artificial intelligence to surveillance: What prospects?. *Information Polity*, (Preprint), 1-17.

[61]. Köse, U. (2021). Responsible AI for Climate Crisis. *UN Climate Change Conference UK 2021 (COP26).* Glasgow, Scotland. Çevrimiçi: https://www.youtube.com/watch?v=cu8_7btnarM&t=49m55s (Erişim 10 Ağustos 2022).

[62]. Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Başkanlığı. (2021). *Ulusal Yapay Zekâ Stratejisi 2021-2025.* Çevrimiçi: https://cbddo.gov.tr/SharedFolderServer/Genel/File/TR-UlusalYZStratejisi2021-2025.pdf (Erişim 14 Ağustos 2022).

Bölüm 3

GRAF ALGORİTMALARI VE YAPAY ZEKÂYA KATKILARI

Elif Eryaşar - Şerife Büyükköse

Bu kitap serisinin ilkinde Yapay Zekâ ve Büyük Veri Analizleri için Graf Teorisi yaklaşımının bir devamı şeklindedir. Bu çalışmaların önemi ise yapay zekâ alanında karşılaşılan gelişmelerin altlığını matematik oluşturmaktadır. Algoritma bir problemi çözmek için izlenmesi gereken yolun, bir başlangıç noktasından başlayarak bitiş noktasına doğru adım adım ilerleyerek ifade edilmesidir. Algoritmalar genel olarak matematik ve mühendislik problemlerinin çözümünden bilgisayar ve iletişim teknolojilerinin geliştirilmesine, yapay zekâ algoritmalarından simülasyon çalışmalarına, protokol tasarımından yazılım geliştirme süreçlerinde pek çok yerde kullanılırlar. Bu bölümde bazı makine öğrenme ve derin öğrenme yaklaşımlarının temelini oluşturan graf algoritmaları tanıtılmış ve bunların kullanımı ile ilgili açıklamalar ile bunların yapay zekâya katkıları sunulmuştur.

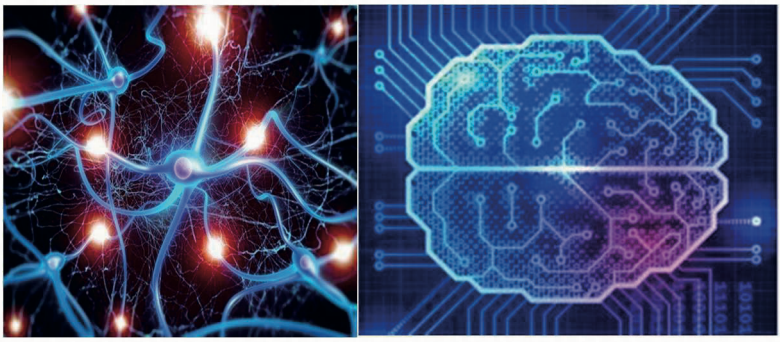
3.1. GİRİŞ

Graf teorisi (*İngilizce: graph theory*), **çizge teorisi** veya **çizit teorisi** olarak isimlendirilen matematik dalıdır. Bu kitap serisinde Graf tabiri kullanımı tercih edilmiştir. Graf, düğümler (köşeler) ve bu düğümleri birbirine bağlayan kenarlardan (bağıntılardan) oluşan bir tür ağ yapısıdır.

1805 de Fransız matematikçi Adrien-Marie Legendre'nin, astronomik gözlemlerden güneşin etrafındaki yörüngeleri belirlemek için kullandığı regresyondaki **en küçük kareler yöntemi** günümüzün makine öğrenme modellerinin çoğuna temel oluşturdu. 1958 yılına gelindiğinde Amerikalı psikolog ve bilgisayar bilimcisi Frank Rosenblatt kendi başına öğrenebilecek ilk algoritmik model olarak bilinen ve bir yapay sinir ağı (YSA) olan **algılayıcı algoritmasını** geliştirdi. 1965 de Ukraynalı matematikçi Alexey Grigorevich Ivakhnenko, mantığı günümüzün derin öğrenme mantığına çok benzeyen birkaç yapay sinir ağının üst üste istiflendiği ve bir yapay sinir ağı katmanının faaliyetinin beslendiğini denetleyen çok katmanlı yapay sinir ağları için ilk genel **öğrenme algoritmasını** önerdi. 1998 de Brin ve Page web sayfalarını, diğer web sayfalarına bağlantı verdikçe daha üst sıralarda tutan **Page-Rank algoritmasını** yayınlayarak Google'ın arama motorunun ilk prototipini oluşturdu. Sergey Brin ve Larry Page'in geliştirdiği bu algoritma, World Wide Web'de daha fazla içerik ve veri oluşturulmasına, ve bu verilerin kullanılmasına yol açarak internet aramalarında devrim yarattı.

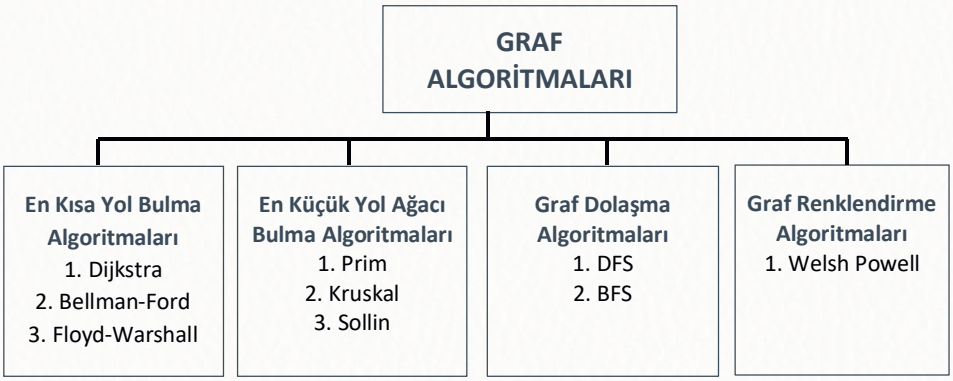
Almanların kırılamaz dediği "enigma" haberleşme ağını kırmakla görevlendirilen Alan Turing "**Makineler düşünebilir mi?**" sorusu ile bilgisayarların insan davranışları sergileyebileceğini ve ilerleyen yıllarda insan-makine ayrımı yapmanın imkansız hale geleceğini öne sürmüştü. Bilgi işleme gücünü kullanarak çok büyük miktarda verileri eş zamanlı olarak işleyebilen **yapay zekâ teknolojisi**, bu sayede bilgisayarlara insan gibi düşünme, analiz ve hareket etme gibi yetenekler katmıştır. Yapay zekânın işleyişini sağlayan algoritma ise bir bilgisayarın bu davranışları sergileyebilmesi için geliştirilen programlar bütünüdür. Bu algoritmalar sayesinde bilgisayarlar kendisinden istenen herhangi bir işlemi rahatlıkla yerine getirebilirler. **Yapay zekâ** temelli algoritmaları, **danışmanlı**, **danışmansız** ve takviyeli veya kuvvetlendirmeli öğrenme algoritmaları olarak üç farklı grupta sınıflandırabiliriz. Bu algoritmalar, istenilen işlem türüne göre bilgisayarların öğrenebilmesine imkan tanımaktadır. **Danışmansız öğrenme algoritmaları**, veri-

len bilgilerden öğrenerek bir model oluşturması beklenir. Bu algoritmalara hedef çıktılar verilmez. **Danışmanlı algoritmalarda** ise birbirinden farklı ve çeşitli veri havuzlarından beslenen sistemler, alınan sonuçları hedeflenen çıktılar ile kıyaslar ve nihai sonucu almaya çalışır. **Takviyeli veya kuvvetlendirmeli algoritmalar** ise hedefler yerine bilgilerin doğruluğunu baz alınır. **Yapay zekâ algoritmaları** ile öğrenme, genelleme yapabilme, ayrıştırabilme ve analiz edebilme yetisine sahip olan bilgisayarlar ve sistemler, mevcut verileri sezgisel, duyusal ve görsel olarak işleyerek elde edilen verileri en iyi performansı elde etmek için kullanır. Sonuç olarak, Şekil 3.1'den de görülebileceği gibi hücresel yapıların aktivasyonu ve elektronik devrelere aktarılabilmiş olup beynin elektronik olarak devre şeması sembolik olarak gösterilmiştir.



Şekil 3.1. Hücresel yapıların aktivasyon ve devre şeması

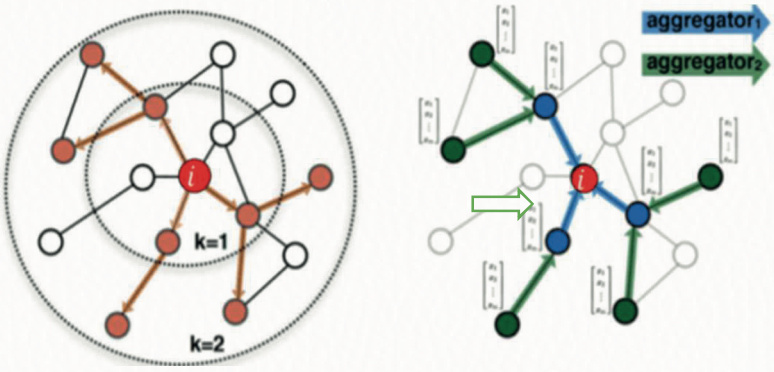
Temeli 1730 yıllarına dayanan graf teori, pek çok bilim dalında farklı problemlerin modellenmesinde kullanılmaktadır. Graf, hem belirli problemlerin doğrudan çözümünü sağlamakta hem de birçok problemin çözümünde ara işlem olarak yardımcı olmaktadır. Fizik, kimya gibi temel bilimlerde, mühendislik uygulamalarında, sosyal bilimlerde ve tıp biliminde birçok problemlerin çözümü ve modellenmesi graf teori kullanılarak yapılmaktadır. Dolayısıyla graf üzerine geliştirilen algoritmalar, bilim insanlarının ve araştırmacıların yoğun ilgisini çekmektedir. Graf teori kullanılarak geliştirilen algoritmalar kullandıkları graf özellikleri cinsinden Şekil 3.2'de verildiği gibi; en kısa yol bulma algoritmaları, en küçük yol ağacı bulma algoritmaları, Graf dolaşma algoritmaları, Graf renklendirme algoritmaları olarak dört ana grupta incelenebilir.



Şekil 3.2. Graf algoritmalarının sınıflandırması

Nesne algılama, makine çevirisi ve konuşma tanıma gibi makine öğrenimi görevlerine CNN, RNN veya otomatik kodlayıcılar gibi uçtan uca derin öğrenme kavramları ile geliştirilen Derin Öğrenme, Öklid verilerinin resimler, metinler, videolar gibi gizli kalıplarını yakalamada iyidir ancak verilerin nesneler arasında karmaşık ilişkiler ve karşılıklı bağımlılıklara sahip graflar olarak temsil edilen Öklid olmayan alanlardan üretildiği uygulamalarda yeterli olamayabilir. Geleneksel Makine Öğrenimi ve Derin Öğrenme araçları basit veri türlerinde uzmanlaşmıştır. Bundan dolayı yapay zekâ alanındaki gelişmeler kullanılan algoritmaların da geliştirilmesine sebep olmuştur.

Graf Sinir Ağları (GNN), graflara doğrudan uygulanabilen sinir ağlarıdır ve düğüm düzeyinde, kenar düzeyinde ve graf düzeyinde tahmin görevlerini gerçekleştirmenin kolay bir yolunu sunan yani graflarla açıklanan veriler üzerinde çıkarım yapmak için tasarlanmış bir derin öğrenme yöntemleri sınıfıdır. GNN'ler, makinelerin bir şeyleri görselleştirmesini sağlamak ve görüntü sınıflandırma, görüntü tanıma veya nesne algılama gibi görevleri gerçekleştirmek için kullanılan Evrişimsel Sinir Ağlarının (CNN) yapamadıklarını yapabilir. Graf teorisinde, Düğüm Gömme kavramı uygulanır. Bu, düğümleri d boyutlu bir gömme alanına grafın gerçek boyutundan ziyade düşük boyutlu alana eşlemek anlamına gelen bir kavramdır. Bu işlemde temel basamak, graftaki benzer düğümleri birbirine yakın gömmektir. Amacımız ise düğümleri, gömme alanındaki benzerliğin ağdaki benzerliğe yaklaşması için eşlemektir. İşlem yapılırken bir düğümün komşularına ve bu komşuların komşularına nasıl bağlandığını izleyerek devam edilir. Yani bir düğüme bağlı noktalarla bir yol oluşturulur. Her düğümün bir özellik vektörü vardır. Şekil 3.3'de hücresel bağlantıların grafsal gösterimi verilmiştir.



Şekil 3.3. Hücresel bağlantıların grafiksel gösterimi

Algoritmalar, graftaki varlıklar arasındaki ilişkiyi anlama ve iki varlık arasında bir bağlantı olup olmadığını tahmin etmeye çalışırlar. Sosyal ağlarda, sosyal etkileşimleri anlamak veya kullanıcılara olası arkadaşlar önermek çok önemlidir. Bu anlamda kullanılan bu yöntem aynı zamanda tavsiye sistemi problemlerinde ve suçla ilişkili derneklerinin öngörülmesinde de kullanılmıştır. Ayrıca bu yöntemle **köşe kümelemesi** yani grafın düğümlerini kenar ağırlıklarına veya kenar mesafelerine bağlı olarak yoğun olarak bağlı bölge grupları halinde kümelemekte de kullanılır.

3.2. EN KISA YOL BULMA ALGORİTMALARI

En Kısa Yol (Path of Minimum Length) Problemi: Temelde graf modellerinde kullanılan iki düğüm arasındaki en kısa, yani en düşük maliyetli yolu bulma problemidir. Bir düğümden diğer tüm düğümlere veya her bir düğümden diğer tüm düğümlere olan en kısa yol ya da en düşük maliyetli yolu bulmak için çeşitli algoritmalar geliştirilmiştir. Bu algoritmalar aşağıda özetlenmiştir.

3.2.1. Dijkstra Algoritması

Graf yapısındaki her bir nokta, yan bilgi veya her düğümle ilişkili nitelikler ve özellikleri gösterir. Yani her düğümün kendisiyle ilişkilendirilmiş bir görüntüsü olabilir. Bu durumda, bir graf hakkında karar vermeye çalışan bir algoritma, bu görüntü düğümleri için, içine gömülü bir CNN alt yorda-

mına sahip olabilir. Böylece bir algoritma olan özyinelemeli bir sinir ağı, çalıştırdığımız yaprakları kelimeler, kenarları da düğümleri yani kelimeleri birleştiren ifadeler şeklinde temsil eden bir ağaç olarak nitelendirilebilir. Bu durumda da sinir ağları graf ile modellenirken sinir ağını temsil eden grafın geren ağaçlarını bulma önem taşır. Grafın geren ağaçları (spanning tree) ya da bir düğümden bir başka düğüme giden en uygun yolu bulan algoritmalar en kısa yol bulma algoritmaları olarak adlandırılır.

Dijkstra algoritması, belirli bir başlangıç düğümünden en kısa yolu belirlemeyi amaçlar. Bu algoritmada en kısa yolu bulmak için bir başlangıç düğümü seçilir ve daha sonra bu başlangıç düğümünden diğer tüm düğümlere olan en kısa yol hesaplanır. Algoritma ağırlıklı ve yönlü graflar için geliştirilmiş olup graf üzerindeki kenarların ağırlıkları sıfır veya sıfırdan büyük sayılar olmalıdır. Dijkstra algoritması negatif ağırlıklar için kullanılmaz. Ayrıca bu algoritma en kısa yolu belirlerken Greedy yaklaşımını kullanır.

Greedy yaklaşımı; graf üzerinde dolaşırken bir düğümden herhangi bir başka düğüme gitmek için en uygun seçimi belirleyen bir yöntem olup mevcut şartlar altında en iyi olan seçimin yapılması gerektiğini savunan yaklaşımdır. Genelde kısa vadede çözüm bulma arayışındadır. Her zaman en düşük maliyetli ya da en karlı çözümü vermeyebilir. Ancak çoğunlukla istenilen sonuca ulaştırır.

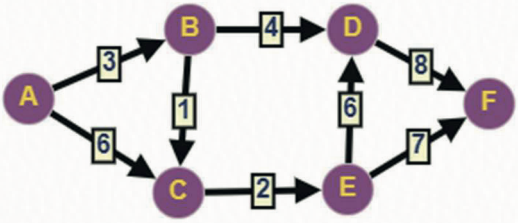
Dijkstra algoritması belirttiğimiz üzere bir kaynak yani başlangıç düğümüne ihtiyaç duyar ve etiketleme yöntemi ile çalışır. Burada iki çeşit etiketleme tipi vardır. Birincisi geçici ve diğeri ise kalıcı etiketlemedir. Geçici etiketleme yaparken; iki düğüm arası maliyet hesaplanır ve yazılır. Geçici etiketleme yapılan aynı yol üzerinde daha kısa bir yola rastlanmaz ise geçici etiket kalıcı etikete dönüşür. Yani iki düğüm için daha iyi bir yolun bulunamayacağı anlaşıldığında geçici etiket kalıcı etikete dönüşür.

Algoritma uygulanırken öncelikle başlangıç düğümüne sıfır değeri (maliyeti) atanır. Geriye kalan tüm düğümler ise sonsuz maliyetli olarak yazılır. Burada başlangıç düğümüne atanan değer kalıcı, diğer atanan değerler ise geçicidir. İkinci adımda, düğümlerin maliyet değerleri için güncelleme yapılır. Üzerinde bulunulan düğümden gidilecek diğer geçici etikete sahip düğüm bulunarak bu düğümün maliyet değerleri güncellenir. Bir başka bir düğüm kümesine giderek en küçük maliyet değeri bulunur. Gelinen en kü-

çük maliyetli düğümün etiketi kalıcı yapılır. Üçüncü ve son adımda ise yeni ulaşılan düğümünden gidilebilecek diğer düğümlerin tamamı kalıcıysa algoritma son bulur. Ancak kalıcı değilse ve geçici düğümler hala devam ediyorsa ikinci adıma tekrar dönülerek aynı işlemler tekrar edilir. Maliyet değerleri güncellenerek hiç geçici düğüm kalmayana kadar yani her düğüm kalıcı etiketli olana devam edilir.

Google haritalarda, coğrafi haritalarda, telefon şebekesi, IP yönlendirme, yazılım dünyasında, navigasyon cihazlarında iki nokta arasındaki en kısa yolu belirlemede vs. kullanılır.

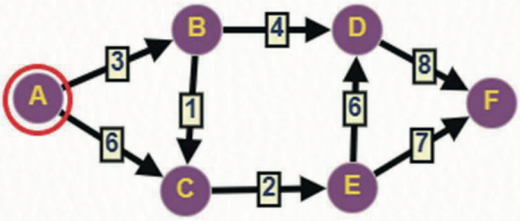
Örnek 3.1:



Yan tarafta verilen grafa Dijkstra algoritması uygulanması aşağıdaki gibi olacaktır.

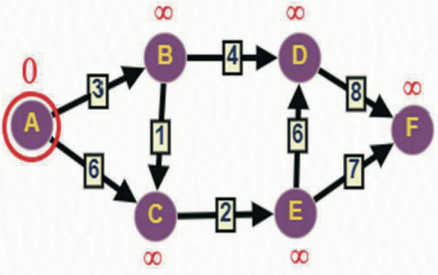
1. Adım:

Öncelikle bir başlangıç düğümü seçilir. Verilen graf için başlangıç düğümü A olarak seçilmiştir.



2. Adım:

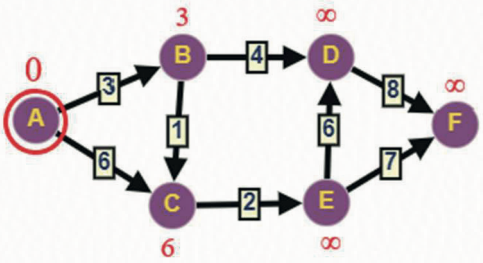
Başlangıç düğümüne 0,geriye kalan tüm düğümlere ise ∞ değeri atanır. Algoritma çalışırken; $(d[u] + c(u,v) < d[v]) \Rightarrow \{d[v] = d[v] + c(u,v)\}$ önermesini kullanarak ilerleyecektir. Burada $d[u]$ bulunan düğümün maliyeti, $c(u,v)$ bulunan düğüm ile ulaşılmak istenen düğüm arasındaki kenarın maliyeti ve son olarak $d[v]$ ulaşılmak istenilen yolun maliyetidir. Eğer $d[u] + c(u,v) < d[v]$



koşulu sağlanırsa ulaşılmak istenilen yolun maliyeti yani d[v]; $d[v] = d[v] + c(u, v)$ olarak güncellenir.

3. Adım:

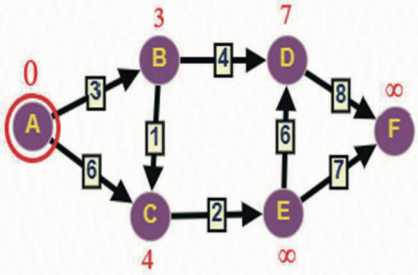
A düğümünden B ve C düğümlerine yol olup, $0 + 3 < \infty$ olduğundan dolayısıyla $\infty = 0 + 3$ elde edilir. O halde B düğümünün maliyeti 3 olarak güncellenir.



Benzer şekilde $0 + 6 < \infty$ olup $\infty = 0 + 6$ olacaktır. C düğümünün maliyeti de 3 olarak güncellenir.

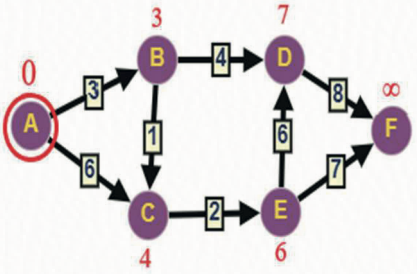
4. Adım:

B düğümünün maliyeti 3 ve C düğümünün maliyeti 6 olmak üzere, B düğümünün maliyeti daha az olduğu için sıradaki düğüm seçimi B olacaktır. B düğümü içinde aynı önerme kullanılacaktır. Daha sonra C ve D düğümlerinin maliyeti güncellenecektir. Yani $3 + 4 = 7 < \infty$ olduğundan D düğümünün maliyeti 7 ve $3 + 1 = 4 < 6$ olduğundan C düğümünün maliyeti 4 olarak güncellenecektir.



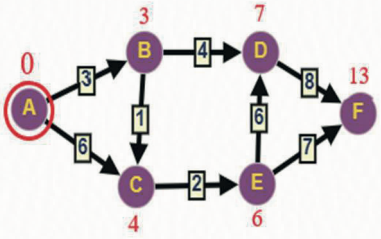
5. Adım:

Bir sonraki maliyeti en küçük düğüm C düğümü olup, C düğümünden E düğümüne gidilecektir. $4 + 2 = 6 < \infty$ olup E düğümünün maliyeti 6 olarak güncellenir.



6. Adım:

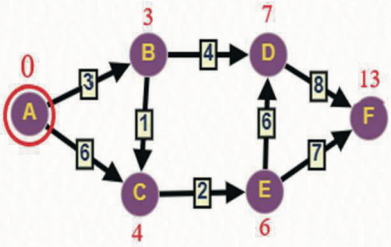
E düğümünden D ve F düğümlerine yol vardır. E düğümünden D düğümüne gidilirken $6 + 6 = 12 < 7$ koşulu sağlanmadığı için D düğümünde maliyet güncellemesi yapılmaz.



F düğümü için ise $6 + 7 = 13 < \infty$ olup F düğümünün maliyeti 13 olarak güncellenir.

7. Adım:

Son olarak D düğümünden de F düğümüne bir yol vardır. Ancak $7 + 8 = 15 > 13$ olduğundan F düğümünde herhangi bir maliyet güncellemesi yapılmaz. İşlem burada sona erer.



3.2.2. Bellman-Ford Algoritması

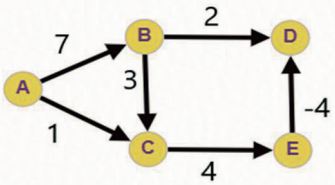
Sosyal ağlarda, genellikle bir kişiye baktığınızda bu kişinin ne tür arkadaşları ve etkileşimleri olduğuna karar vermeye çalışılır. Yani düğümün kendisi veya kenarları (komşu noktaları) hakkında tahminlerde bulunulmaya çalışılır. Bunun içinde düğümün yakın çevresi, düğümden hareketle her yöne doğru k adım atarak, en yakın çevrede k=1 alınarak, muhtemelen önemsenen ya da aranan bilgilerin çoğu yakalanabilir.

Sinir ağlarındaki her bir düğüm için gösterimleri öğrenmek için bir sinir ağı modelini zorlayabilirsiniz; bu, daha sonra bunlar hakkında aşağı doğru önemsediğiniz şeylere ya da ulaşmaya çalıştığınız tahminlerde bulunmada yardımcı olabilir. Amaç bu düğüm, önemsediğimiz diğer şeylere ne kadar yakın sorusuna cevap almaktır. Yapay zekâ çalışmalarında graf modellemesi yapılırken bu amaçlar için bir başlangıç noktasından hedef noktalara giden aşağıdaki algoritmalar kullanılır.

Bellman-Ford algoritmasının amacı da Dijkstra algoritmasının amacı gibi bir başlangıç düğümünden hedef düğümlere olan en kısa yolu bulmaktır. Bu algoritma Dijkstra algoritmasının aksine kenar ağırlıkları negatif olan graflar içinde çalışır. Ancak bu algoritma Dijkstra algoritmasına göre daha yavaş çalışmaktadır. Burada dikkat edilmesi gereken nokta eksi maliyetli döngüler olmamasıdır. Algoritma Dijkstra algoritmasında olduğu gibi en küçük değere sahip olan kenardan gitmek yerine graf üzerindeki tüm kenarları test eder. Bu sayede Greedy yaklaşımının handikabına düşmez ve her düğüme sadece bir kere bakarak en kısa yolu bulmuş olur.

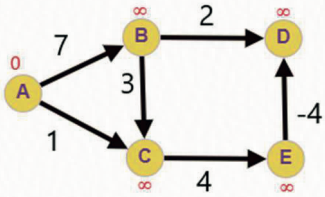
Yine ilk olarak başlangıç noktasına 0, diğer düğümlere sonsuz değeri atanır. Daha sonra her düğüm ile komşusu arasındaki maliyetler bulunur. Daha sonra bu maliyetler karşılaştırılır ve en kısa yol ya da maliyet bulunur. Dijkstra algoritması gibi Google haritalar, coğrafi haritalar ve kablosuz algılayıcı ağlar üzerinde paketlerin anahtarlanması gibi alanlarda kullanılır.

Örnek 3.2:

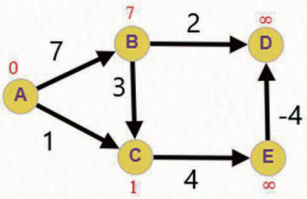


Yan tarafta verilen graf için Bellman-Ford algoritma uygulanması aşağıdaki gibidir.

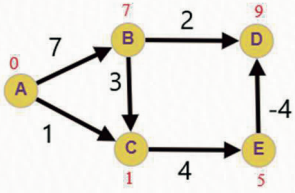
1



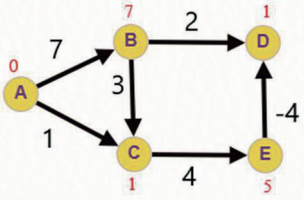
2



3



4



3.2.3. Floyd-Warshall Algoritması

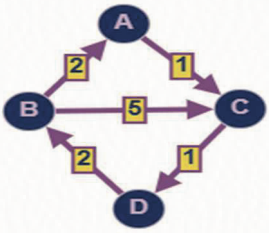
Bu algoritma, pozitif veya negatif kenar ağırlıklarına sahip ağırlıklı bir grafta tüm düğümler arasındaki en kısa yolları bulmak için kullanılan bir algoritmadır. Ancak bu algoritma kullanılırken de negatif döngülerin bulunmaması gerekmektedir.

Başlangıçta komşuluk listesine göre bir matris inşa edilir. Birbiri ile ilişkisi olmayanlara sonsuz diğer durumda ise doğrudan ağırlıklar yazılır. Sonrasında bu matris adım adım güncellenerek bir düğümden diğer tüm düğümlere giden en kısa yollar çıkartılır.

Algoritma en kısa yol bulmak için kullanılan programlarda, düzenli bir ifadede herhangi bir tekrarlı olayın tespiti, ağ programlanmasında vs. kullanılır.

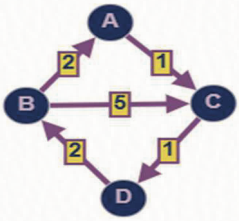
Örnek 3.3:

Yan tarafta verilen grafa Floyd-Warshall algoritması uygulanması aşağıdaki gibidir.



1. Adım:

Öncelikle verilen grafın derece matrisi oluşturulur. Bu derece matrisinin köşegenleri her zaman 0'dır. Ve birbiriyle komşu olmayan düğümler için ise ∞ değeri kullanılacaktır. Diğer durumlarda kenar ağırlıkları aynen yazılacaktır.



Burada kullanılan önerme, $d(u,v) > d(u,k) + d(k,v)$ ise $d(u,v) = d(u,k) + d(k,v)$ dir. Burada d(u,v), u. satırın v. sütun değerini temsil eder. Benzer şekilde $d(u,k)$ u. satırın k. sütununu ve $d(k,v)$ de k. satırın v. sütun değerini belirtecektir. Algoritmanın kullandığı uzaklık matrisinde düğüm sayısı kadar satır ve sütun sayısı olduğu aşikardır. Algoritma çalışırken belli adımlarda bazı satır ve sütunları sabit tutacaktır. Önermedeki k ise hangi satır ve

	A	B	C	D
A	0	∞	1	∞
B	2	0	5	∞
C	∞	∞	0	1
D	∞	2	∞	0

sütunu sabit tutacağımızı temsil edecektir. Örneğin k=1 için önerme d(u,v) > d(u,1)+d(1,v) halini alacaktır.

2. Adım:

$k = 1$ için;

$d(2,3) = 5, d(2,1) = 2$ ve $d(1,3) = 1$ olup $d(2,1) + d(1,3) = 3$ olduğundan $5 > 3$ olup istenilen şart sağlanır. Dolayısıyla matrisin 2. satır 3. Sütunu 5 olarak güncellenir.

	A	B	C	D
A	0	∞	1	∞
B	2	0	5	∞
C	∞	∞	0	1
D	∞	2	∞	0

Benzer şekilde diğer satır ve sütun işlemleri de yapılırsa grafın yeni matrisi yandaki gibi olacaktır.

3. Adım:

$k = 2$ için benzer işlemler yapıldığında uzaklık matrisi yan taraftaki gibidir.

	A	B	C	D
A	0	∞	3	4
B	2	0	5	6
C	∞	∞	0	1
D	4	2	7	0

4. Adım:

$k = 3$ için benzer işlemler yapıldığında uzaklık matrisi yan taraftaki gibidir.

	A	B	C	D
A	0	∞	3	∞
B	2	0	5	∞
C	∞	∞	0	1
D	4	2	7	0

5. Adım:

Son olarak $k = 4$ için benzer işlemler yapıldığında uzaklık matrisi yan taraftaki gibidir. İşlem sona erer.

	A	B	C	D
A	0	6	3	4
B	2	0	5	6
C	5	3	0	1
D	4	2	7	0

3.3. EN KÜÇÜK YOL AĞACI BULMA ALGORİTMALARI

Derin Yürüyüş (DeepWalk): Bu algoritma, denetimsiz bir şekilde öğrenilen düğümlerin gömülmesini öneren ilk algoritmadır. Eğitim süreci açısından kelime gömmeye çok benzemektedir. Algoritma iki adım içerir:

1. Düğüm dizileri oluşturmak için graftaki düğümlerde rastgele yürüyüşler gerçekleştirilir.
2. Adımda oluşturulan düğüm sıralarına göre her düğümün gömülmesini öğrenmek için skip-gram'ı çalıştırılır.

Rastgele yürüyüşün her adımında, bir sonraki düğüm önceki düğümün komşusundan eşit olarak örneklenir. Her sıra daha sonra 2|w| + 1 uzunluğundaki alt dizilere kesilir; burada w, atlama gram'ındaki pencere boyutunu gösterir.

Bu algoritma temelinde ikili bir ağaç kullanır. Bu ikili ağaçta, tüm yapraklar graftaki köşelerdir. İç düğümlerin her birinde, hangi yolun seçileceğine karar vermek için ikili bir sınıflandırıcı vardır. Belirli bir köşenin olasılığını hesaplamak için, kök düğümden çıkılan k uzunluğundaki yol boyunca alt yolların her birinin olasılığını hesaplamak yeterlidir. Her düğümün olasılığı hesaplanırken alt toplamları 1 olduğundan, tüm köşelerin olasılığının toplamının 1'e eşit olması özelliği kullanılır. n yaprak sayılı bir ikili ağaç için en uzun yol O(log(n)) fonksiyonu ile sınırlandırıldığından, bir öğenin hesaplama süresi artık O(log|V|) değerine dönüşür.

Derin öğrenme ve yapay zekâ çalışmalarında derin yürüyüş önemli bir yer tutar. Derin yürüyüşün temelinde ise en küçük yol ağacını bulan graf algoritmaları yatmaktadır.

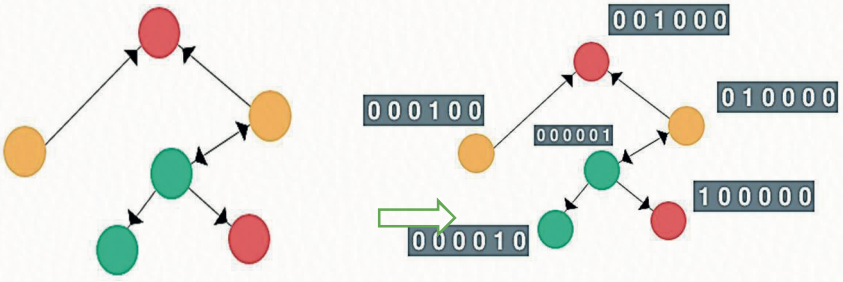
En Küçük Yol Ağacı (Minimum Spanning Tree) Problemi: Ağaç graf, içinde döngü içermeyen özel bir graf olup yol ağacı ise graf üzerindeki tüm düğümleri kapsayan ağaç şeklinde bir yoldur. En küçük yol ağacı ise grafdaki kenarları toplamda en düşük maliyeti oluşturacak ve tüm düğümleri içerecek şekilde üretilen yani en küçük maliyetli bir ağaçtır.

En küçük yol ağacını belirlemek için belirli algoritmalar kullanılır. Geliştirilen ve şu an yaygın olarak kullanılan bazı algoritmalar Prim, Kruskal ve Sollin algoritmalarıdır. Bu algoritmalar aşağıda kısaca tanıtılmıştır.

3.3.1. Prim Algoritması

Yapay sinir ağlarında kullanılan bir grafın düğümleri belirli özellikler (yaş, cinsiyet, boy ağırlık vb)yükleyerek göstermek istersek modeldeki grafı ağırlıklı graf olarak belirleyebiliriz. Buradaki amaç düğümler kodlanarak sinir ağlarını karışıma getirerek grafı değiştirilebilmektir. Tüm düğümler tekrarlayan birimlere veya herhangi bir sinir ağına dönüştürülür ve tüm kenarlar basit ileri beslemeli sinir ağlarına yardımcı olur. Düğümlerin ve kenarların dönüştürülmesi tamamlandıktan sonra, graf düğümleri arasında iletilerin geçişi gerçekleştirilir. Belirli bir referans düğümün etrafındaki çevreleyen düğümlerden yönlendirilmiş kenarlardan mesajların itilmesini içerir. Bu işlem yapılırken farklı kenar türleri için biri tek yönlü diğeri çift yönlü olmak üzere farklı bir sinir ağı kullanılabilinir. Bu şekilde, düğümler arasındaki uzamsal ilişkileri yakalamaya devam edebilirsiniz.

GNN'ler açısından, tek bir referans düğümü için komşu düğümler mesajlarını (katıştırmalarını) kenar sinir ağları üzerinden referans düğümündeki tekrarlayan birime iletir. Şekil 3.4'den de görülebileceği gibi GNN'lerin şematik gösterimi yapılan ve referans tekrarlayan ünitenin yeni gömülmesi, söz konusu tekrarlayan fonksiyonun mevcut gömmeye uygulanması ve komşu düğüm gömmelerinin kenar sinir ağı çıktılarının bir toplamı ile güncellenir.



Şekil 3.4. GNN'lerin şematik gösterimi

Prim algoritması ağırlıklı bir grafın kenarlarından yararlanarak grafın tüm düğümlerini kapsayan en küçük ağacı bulmak için tasarlanmıştır. Algoritma maliyeti en az olan kenardan başlayarak adım adım diğer kenarların

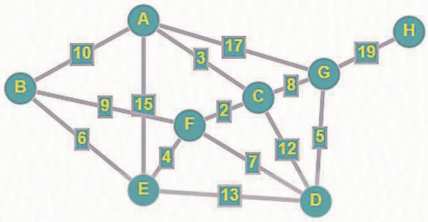
belirlenmesine dayanır. Yol ağacı belirlenmesi sırasında birden fazla ağaç yoktur. Başlangıçta bir kenardan oluşan bir yol ağacı vardır ve sonraki adımlarda bu yol ağacına eklenen kenarlar bu yol ağacı üzerindeki düğümlerden birine komşu olmalıdır. Burada bir döngü oluşturulmamasına dikkat edilmelidir çünkü döngü oluşturulursa ağaç yapısı bozulacaktır.

Öncelikle bir başlangıç düğümü seçilir. Daha sonra bu başlangıç düğümünün komşuları belirlenir, başlangıç düğümüne olan uzaklıklar tespit edilip en az maliyetli olan yol seçilir. Burada dikkat edilecek nokta, ulaşılan düğümden devam edilmeyecek oluşudur. Burada artık hem başlangıç düğümünden hem de varılan düğümden gidilebilecek maliyeti en az olan yol tekrar edilecektir. İşlem, graf üzerindeki tüm düğümleri dolaşana kadar devam eder.

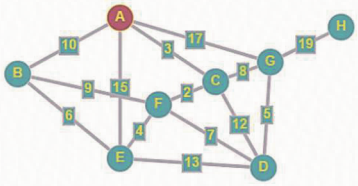
Bu algoritma ağ tasarımı, gezgin satıcı problemi, Steiner Ağaç Problemi, kümeleme analizi gibi problemleri çözmek için kullanılır.

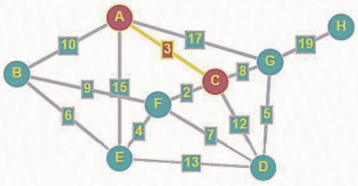
ÖRNEK 3.4:

Şekilde verilen grafa Prim algoritması uygulanması aşağıdaki gibidir.

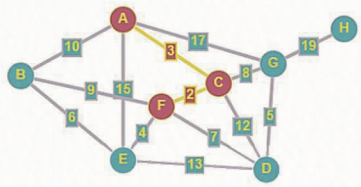


1 **2**

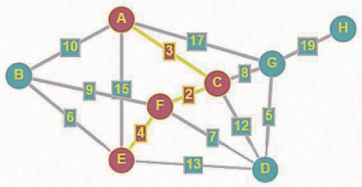




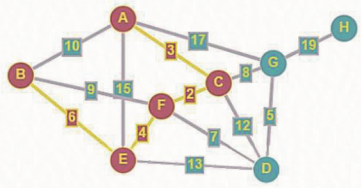
3



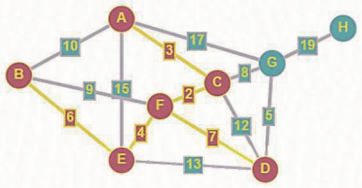
4



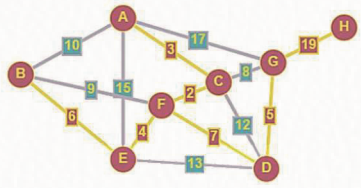
5



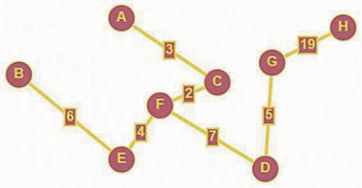
6



7



8



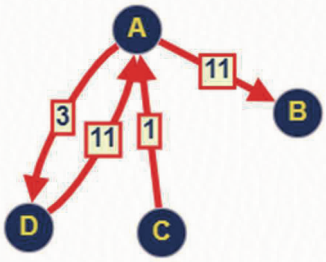
3.3.2. Kruskal Algoritması

Kruskal algoritması, bir grafın tüm düğümlerini değerlendirerek en kısa olacak şekilde dolaşılmasını amaçlar. En küçük yol ağacı problemini bulmak için kullanılan bir algoritmadır.

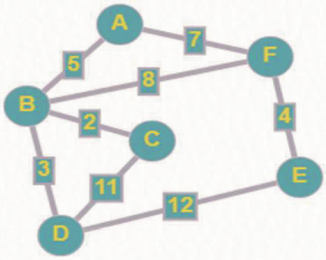
Temelde her seferinde en iyi kenarın seçilmesi vardır. n kenarlı bir graf için en az maliyetli bir kenar yapısı ile başlanır ve kalan kenarlar arasından en düşük maliyetli seçilerek ilerlenir. Elde etmek istenilen bir ağaç yapısı olduğuna göre döngü oluşturmayacak şekilde $(n - 1)$ kenar eklenene kadar devam edilir. Aynı değerli kenarlarda döngü oluşturulmaması şartı ile seçim keyfi yapılabilir.

Örneğin bir yere yaptığınız gezi sırasında zamanınız kısıtlı ise Kruskal algoritmasını kullanmak sizin için oldukça faydalı olacaktır.

Burada belirtmeliyiz ki yönlü graflarda Kruskal algoritması kullanılırken herhangi iki düğüm arasında birden fazla yol var ise bu her iki yolun ağırlığı aynı olmalıdır. Örneğin aşağıdaki grafta A ve D düğümleri arası 3 ve 11 değerlikli olduğundan Kruskal algoritması kullanılmaz. Çift yönlü farklı uzunluklu yollar vardır. Bu algoritma ağ yapılarında, grafik çizimlerinde, istatistik biliminde vs kullanılmaktadır.

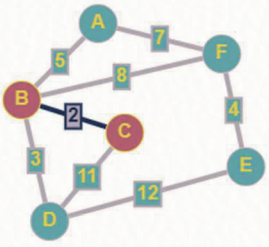


ÖRNEK 3.5: Aşağıda verilen grafta Kruskal algoritması davranışı incelenmektedir.

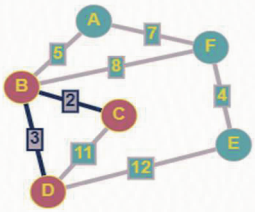


1. Adım:

Grafta öncelikle en kısa kenar belirlenir. Şayet aynı uzunlukta başka kenarlar mevcutsa bu durumda graf üzerinde rastgele seçim yapılabilir. Verilen grafta en kısa kenar B ve C düğümlerinin oluşturduğu kenar olup bu kenar koyu mavi renk ile belirtilir.



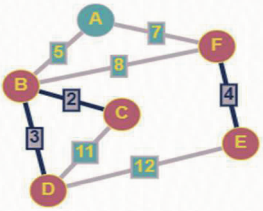
2. Adım:



BC kenarı seçildikten sonra kalan kenarlar bir döngü oluşturmayacak şekilde seçilmelidir. O halde bir sonraki seçim en kısa kenar olan BD kenarı olacaktır.

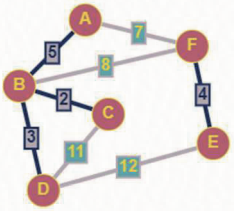
Burada dikkat edilmesi gereken bir husus ise artık DC kenarının kullanılamayacak olmasıdır. Aksi halde BCD döngüsü oluşacaktır.

3. Adım:



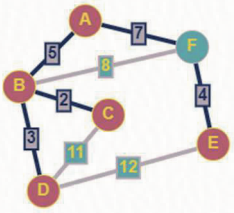
Kalan kenarlar içerisinde en kısa olan kenar EF kenarı olup herhangi bir döngü de oluşturmadığı için seçilen bir sonraki kenar EF kenarı olacaktır.

4. Adım:



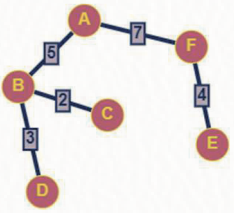
Kalan kenarlar içerisinde bir sonraki en kısa olan kenar AB kenarıdır. Ayrıca herhangi bir döngü de oluşturmayacağı için bir sonraki seçim AB kenarı olacaktır.

5. Adım:



Seçilecek bir sonraki kısa kenar ise herhangi bir döngü oluşturmayacak olan AF kenarıdır.

6. Adım:



Geriye kalan kenarlar seçildiğinde döngüler oluşacağından (Örneğin BF kenarı seçilirse ABF döngüsü oluşur) seçilecek başka bir kenar kalmamıştır. İşlem sona erer.

3.3.3. Sollin Algoritması

Bu algoritma, doğrudan paralel programlamaya yatkın bir algoritmadır. Sollin algoritması aynı anda birden çok ağaçla başlar ve sonraki adımlarda ise elde edilen ağaçların birleşerek tek bir yol ağacına dönüştüğü algoritmadır. Kruskal ve Prim'in algoritmalarına göre sonuca daha az adımda ulaşabilen Sollin algoritmasının ara işlemleri belirtilen diğer iki algoritmaya göre daha fazla olabilir.

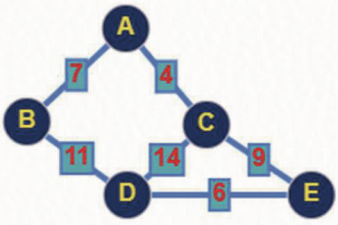
Bu algoritmada, birden fazla kenar seçimi yapılır ve yol ağacı oluşacak şekilde alt ağaçlar elde edilir. En küçük yol ağacı ara işlemlerde alt ağaç olarak bulunan önceki ağaçların birbirlerine eklenmek suretiyle bulunur.

Bu algoritma belirli programlama alanlarında, ağ tasarımı vs kullanılabilir.

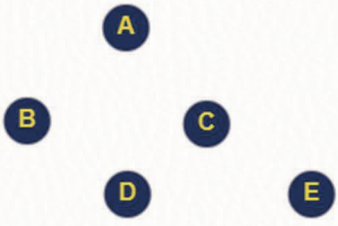
ÖRNEK 3.6:

Sollin algoritması uygulanması aşağıdaki gibidir.

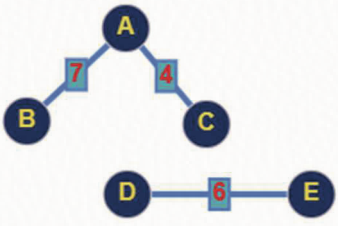
1



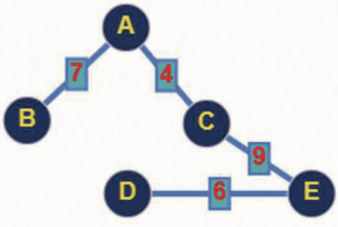
2



3



4



3.4. GRAF DOLAŞMA ALGORİTMALARI

Graf üzerinde dolaşma, bir problemi çözecek bir biçimde grafın düğümleri arasındaki ilişkiye dayanarak bu düğümlerin oluşturduğu kenarlar üzerinde hareket etmektir. Bilinen en önemli iki dolaşma algoritması kısaca DFS (Depth First Search-Önce Derinlik Araması) ve BFS (Breadth First Search-Önce Genişlik Araması) algoritmalarıdır.

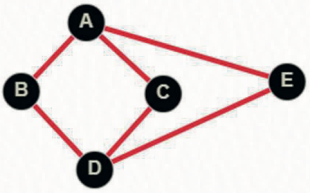
3.4.1. DFS Algoritması

Seçilen bir başlangıç düğümünün bir kenarından başlayarak o kenar üzerinden gidilebilecek en uzak (derin) düğüme kadar gidilmesi esasına dayanan graf üzerinde dolaşma yöntemlerinden birisidir. Özel olarak ağaçlarda sıklıkla kullanılır.

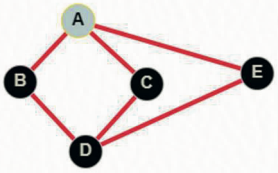
Öncelikle bir başlangıç düğümü seçilerek bu başlangıç düğümünden gidilebilecek en uzak yere kadar gidilir. Daha sonra gidilecek herhangi bir komşu kalmayınca tekrar geri dönülür ve benzer şekilde gidilebilecek en uzak düğüme gidilir. İşlem bu şekilde gidilecek herhangi bir düğüm kalmayana kadar devam eder.

Bu algoritma sınav çizelgesi hazırlama, haritalar üzerinde bölge renklendirme, belirli ağ sistemlerinde vs. kullanılır.

ÖRNEK 3.7:



Yukarıda verilen grafa DFS algoritması uygulanması aşağıdaki gibi gerçekleşir.

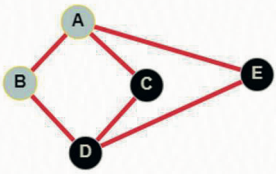


1. Adım:

Başlangıç noktası keyfi olarak A olarak belirlenir. A noktasına uğranılmış olduğu belirtilmek üzere işaretlenir.

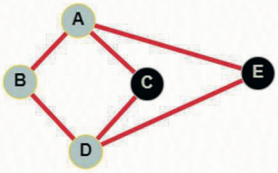
2. Adım:

Daha sonra A düğümünden gidilebilecek komşuluklara bakılır. Bunlar B,C ve E düğümleri olmak üzere alfabetik sıra göz önünde bulundurulduğunda uğradığımız nokta B olacaktır. (Belirlenen başka bir özelliğe göre de değişebilir.) B düğümü de işaretlenir.



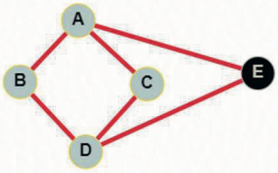
3. Adım:

Artık gezinmeye B düğümünden başlıyoruz. B düğümünün komşusu D düğümü olmak üzere benzer şekilde D düğümü de işaretlenir.



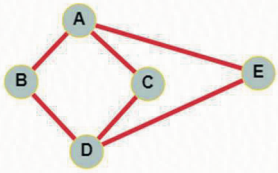
4. Adım:

Benzer şekilde D düğümünden devam ediyoruz. D düğümünün komşuları C ve E düğümleri olmak üzere alfabetik sıra göz önünde bulundurulduğunda bir sonraki uğrayacağımız düğüm C düğümü olacaktır.



5. Adım:

Son olarak C düğümün komşuları olan A ve D düğümleri ziyaret edildiği için algoritma D düğümüne geri döner. D düğümünün komşusu olan E düğümü ziyaret edilir. Ziyaret edilmemiş bir düğüm kalmamıştır. Dolaşma algoritması sona erer.



3.4.2. BFS Algoritması

Genişlik araması olarak geçen bu algoritma; bir grafın düğümlerini başlangıç düğümüne daha yakın olanlara yani komşuluklara öncelik vererek arayan bir algoritmadır. Dolayısıyla DFS'den farkı başlangıç düğümünün bir kenarı üzerinden en uzak düğüme gitmez, bu başlangıç düğümünün tüm

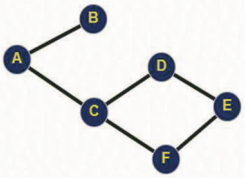
komşu düğümlerine gidilmesi ile başlanır. Burada ara düğümlere de başlangıç düğümüne olduğu gibi yaklaşılır. Bir ara düğüme gelindiğinde bu ara düğümün ziyaret edilmemiş tüm komşu düğümlerine gidilir.

Seçilen başlangıç düğümünün tüm komşu düğümleri ziyaret edilir. Ziyaret edilmemiş düğümler varsa en kısa maliyet ilkesi göz önüne alınarak bir başka ara düğüm seçilir.

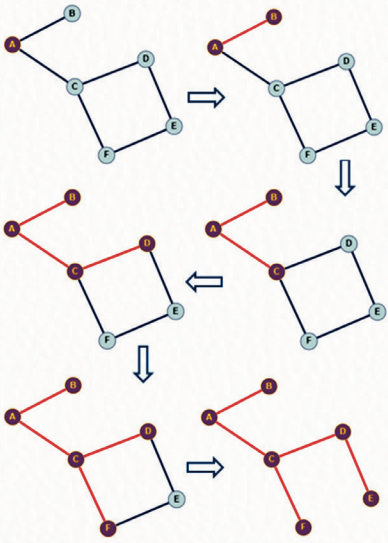
Benzer şekilde bu ara düğümünde ziyaret edilmemiş tüm komşu düğümleri ziyaret edilir. İşlem bu şekilde ziyaret edilmemiş düğüm kalmayana kadar devam eder.

Bu algoritma bulunduğumuz yere yakın olan yerleri bulmak amacıyla GPS sistemlerinde, sosyal ağ sitelerinde, iki düğüm arası en kısa rotayı belirleme gibi problemlerin çözümünde kullanılır.

ÖRNEK 3.8:



Yan tarafta verilen grafa BFS algoritması uygulanması sonucunda aşağıdaki sonuç elde edilir.



DFS VE BFS algoritmaları kıyaslandığında sonuç bulma noktasında BFS algoritmasının daha avantajlı olduğunu söylenilebilir. Ancak çözümün derinlerde olduğu durumlarda DFS algoritması kullanılmalıdır.

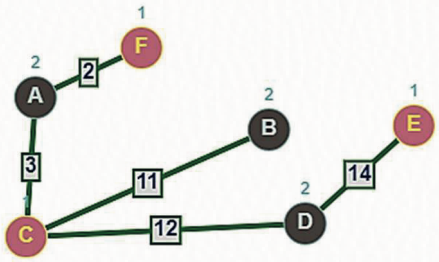
3.5. GRAF RENKLENDİRME ALGORİTMASI

Graf renklendirme, bir haritanın renklendirilmesinde, birbirine komşu iki bölgenin farklı renklerde olması koşuluyla en az kaç renk kullanılması gerektiği ve yine yeterli renk sayısı için, en karmaşık bir haritada dahi renklendirme yaparken kullanılan renk sayısının bir alt sınırın olup olmadığı sorusunun cevabının aranması sırasında ortaya çıkmıştır.

Graf renklendirme, graf üzerindeki tüm düğümlere renk atama işlemidir. Burada birbirine komşu olan her düğüme farklı renklendirme yapılır. Amaç minimum sayıda renk kullanarak tüm düğümlere kendilerine komşu olan düğümlerden farklı bir renk vermektir.

Burada aynı renge sahip birden fazla düğüm olabilir. Renklendirme yapılırken kullanılan renk sayısı kromatik (chromatic) sayı olarak adlandırılır.

Aşağıda verilen grafın kromatik sayısının iki olduğu kolayca görülebilir.



Uygulama Alanları

- Harita renklendirme,
- İşlemcilerin işlem sırasını belirleme,
- Ders ve sınav programı ayarlama,
- Hava alanlarında iniş ve kalkış sırasını belirleme,
- Kümeleme işlemlerinde,
- Çeşitli bilişim alanları

vs. graf renklendirme yoğun bir şekilde kullanılır. Bunlara verilebilecek en iyi örneklerden birisi Welsh-Powell Algoritmasıdır. Bu algoritma aşağıda kısaca özetlenmiştir.

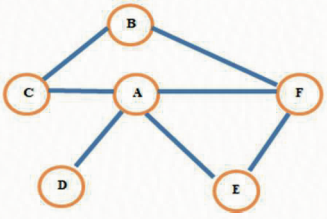
Welsh-Powell Algoritması, graf renklendirmede yaygın olarak kullanılan algoritmalardan birisidir. Bu algoritma düğümlerin derecelerine bağlı olarak çalışır. Öncelikle graftaki düğümlerin dereceleri belirlenir. Daha sonra bu düğümlerin dereceleri büyükten küçüğe sıralanır. İlk renk birinci sıradaki düğüme verilir. Kullanılan ilk renk bu düğüme komşu olmayan diğer düğümlere verilir. Bu işlemden sonra diğer renge geçilerek aynı işlem bütün düğümler renklendirilinceye kadar devam eder.

ÖRNEK 3.9:

Aşağıda verilen grafa Welsh-Powell algoritması uygulaması verilmiştir.

1. Adım:

Grafın her bir düğümünün derecesi belirlenir.



DÜĞÜM	DERECE
A	4
B	2
C	2
D	1
E	2
F	3

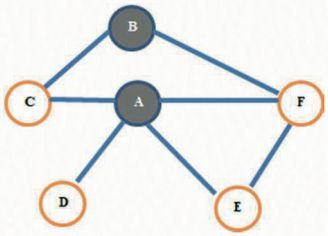
2. Adım:

Daha sonra dereceleri büyükten küçüğe sıralanacaktır. Aynı dereceye sahip düğümleri alfabetik özelliğine göre sıralayabiliriz. (Belirlenen başka bir özelliğe göre de sıralanabilir.)

DÜĞÜM	DERECE
A	4
F	3
B	2
C	2
E	2
D	1

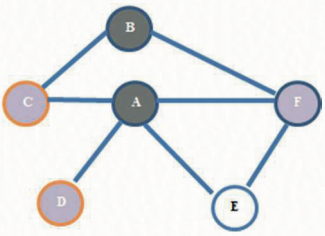
3. Adım:

Renklendirmeye derecesi en yüksek olan düğümle başlanır. Daha sonra bu düğüme komşu olmayan tüm düğümlere aynı renk verilir.



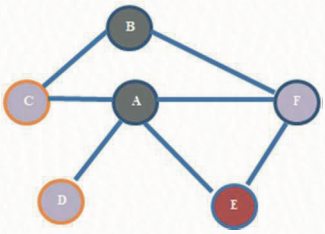
4. Adım:

Bir sonraki renklendirme işlemi komşu olmayan düğümlerden derecesi en yüksek olan düğümle yapılır. Yani temelde olan işlem, boyanmamış düğümler arasından en yüksek dereceli düğüme bir renk verip ardından bu düğüme komşu olmayan tüm düğümlere aynı renk verilmesidir.



5. Adım:

Renklendirilmeyen son düğüme de farklı bir renk verilir. İşlem sona erer.



Bu problemde ise kromatik sayı üçtür.

3.6. SONUÇ VE DEĞERLENDİRMELER

Yapay zekâ alanında kullanılan günümüz yazılım ve donanım seçeneklerinin giderek insan anlayışına daha da yakınlaşmaktadır. Yapay zekâ algoritmaları, mimarileri ve teknolojileri geliştikçe kullanılan algoritmalar da gelişmektedir.

Bu çalışmada, yapay zekâ ve makine öğrenmede kullanılan graf algoritmaları ele alınmıştır. Makine öğrenimi algoritmaları, daha fazla veriyle karşılaştıkça kendilerini daha iyi performans gösterecek hale getirir. Yapay zekâ ile ilgili birçok problemde durum uzayının veya çözüm ağacının gösterilmesinde graf yapıları kullanılmaktadır. **Durum uzayı** oluşabilecek tüm ihtimalleri kapsayan kümedir. **Çözüm ağacı** ise bu ihtimallerden bizi amacımıza ulaştıran kısımlarının graflarla ifade edilmesidir.

Ayrıca birçok önemli fonksiyonel beyin ağı son yıllarda graf teori kullanılarak modellenmekte, karakterize edilmekte, yapay zekâ algoritmalarına uyarlanmakta ve başarılı sonuçlar elde edilmektedir. Başlangıçta bir dolaşma problemini çözmek için ifade edilen graf teori bugün beyni modellemek, karakterize etmek ve bu sayede önemli problemleri çözmek için kullanılmaktadır. HUB'lar, bir graftaki birçok komşuya sahip düğümlerdir. Beyindeki yer hücrelerinin bağlanabilirliği, beyin tarafından çevrede gezinmek için kullanılır. İki konum arasındaki en kısa yol, yer hücreleri arasındaki bağlantıların izlenmesi ile hesaplanabilir. Bu, düğümlerin anatomik bağlantılar ile tanımlandığı yapısal beyin ağının bir örneğidir. Bu örneklerin sayılarının artırılması gereklidir.

Yapay Zekâ ve Büyük Veri Kitap Serisinin ilkinde de belirttiğimiz gibi, matematiksel yaklaşımlar güncel algoritmaların, mimarilerin ve teknolojilerin geliştirilmesi için en önemli altyapıyı oluşturmaktadır. Bu yapıların tekrar gözden geçirilmesi ve bu yaklaşımların YZ metotlarına adapte edilmesi ve yeni çözümlerin geliştirilmesi ile de bu gelişimlerin daha da hızlanacağı değerlendirilmektedir.

TEŞEKKÜR

Yazarlar, 2022/1, 2210/A nolu proje kapsamında verdikleri destek için Türkiye Bilimsel ve Teknolojik Araştırma Kurumuna (TÜBİTAK) teşekkür ederler.

KAYNAKLAR

[1]. Anand R., An Illustrated Guide to Graph Neural Networks, https://medium.com/dair-ai/an-illustrated-guide-to-graph-neural-networks-d5564a551783 (erişim tarihi: 29.09.2021)

[2]. Bondy, J. A.; Murty, U. S. R. (2008). Graph Theory, Springer, ISBN 978-1-84628-969-9.

[3]. Büyükköse, Ş., Graf Teori, Nobel Akademik Yayıncılık, 2021.

[4]. Cansu, S., Graflar Üzerinde En Kısa Yol Algoritmalarının Karşılaştırılmasına İlişkin Bir Uygulama, Muğla Sıtkı Kocaman Üniversitesi Fen Bilimleri Enstitüsü İstatistik Ana Bilim Dalı. Yüksek Lisans Tezi, 2017, Muğla Sıtkı Kocaman Üniversitesi.

[5]. Çölkesen, R., Bilişim Matematiği, Papatya Yayıncılık, 2015.

[6]. Çölkesen, R., Graf Algoritması Papatya Yayıncılık Eğitim ve TBV ISBN: 975-6797-38-X, sayfa: 426-432, 2006.

[7]. Dasgupta S., Papadimitriou C. And Vazirani U., Algorithms, McGraw Hill, 2006

[8]. Demirkol, A., Demirkol, Ö. E., Dijkstra ve Bellman-Ford En Kısa Yol Algoritmalarının Karşılaştırılması, Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 2003, 7(3): p. 55-62.

[9]. Domingos P., Master Algortima Yapay Öğrenme Hayatımızı Nasıl Değiştirecek (Çeviren: Tufan Göbekçin) Paloma Yayınevi, 2019

[10]. Duman, S., Graf Boyama Üzerine, Yaşar Üniversitesi Fen Bilimleri Enstitüsü. Yüksek Lisans Tezi, 2014, Yaşar Üniversitesi:İzmir.

[11]. Frohlich J., Mapping Brain Connectivity Using Graph Theory https://knowingneurons.com/2015/10/21/mapping-brain-connectivity-using-graph-theory/ (erişim tarihi: 25.08.2021)

[12]. Huang K., A Gentle Introduction to Graph Neural Networks, https://towardsdatascience.com/a-gentle-introduction-to-graph-neural-network-basics-deepwalk-and-graphsage-db5d540d50b3 (erişim tarihi: 29.09.2021)

[13]. Kirişçi, M., Öncel, D., Büyüyen Bir Şirkette Bilgisayar Ağları için Minimum Spanning Trees İçeren Yeni Bir Algoritma, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 2017, 16(32): p. 1-14.

[14]. Karcı, A., Çizge Algoritmaları ve Çizge Bölmeleme, Fırat Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı. Yüksek Lisans Tezi, 1998, Fırat Üniversitesi:Elazığ.

[15]. Karaboğa D., Yapay Zekâ Optimizasyon Algoritmaları, Nobel Yayınevi, 2020

[16]. Kyzy, G. K., Akıllı Ulaşım Sistemine Yönelik En Kısa Yol Bulma Algoritmalarının Performans Analizi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar ve Bilişim Mühendisliği Ana Bilim Dalı. Yüksek Lisans Tezi, 2019, Sakarya Üniversitesi:Sakarya.

[17]. Menzli A., GNN and Some of GNN Applications, https://neptune.ai/blog/graph-neural-network-and-some-of-gnn-applications (erişim tarihi: 29.09.2021)

[18]. Motwoni R. And Raghavan P., Randomized Algorithms, Cambridge University Press, 1995

[19]. Needham, M.,Hodler, A. E.,Graph Algorithms, O'Reilly Media Yayıncılık, 2019.

[20]. Nicholson C., A Beginner's Guide to Graph Analytics and Deep Learnings 29.09.2021 https://wiki.pathmind.com/graph-analysis (erişim tarihi: 29.09.2021)

[21]. Sağlam, A., Minimum Yayılan Ağaç Tabanlı Sıralı Görüntü Bölütleme, Selçuk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı. Yüksek Lisans Tezi, 2016, Selçuk Üniversitesi: Konya.

[22]. Şahin, F., Şehir İçi Ulaşımda Zaman Bağımlı Rota Belirleme, Gazi Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı. Yüksek Lisans Tezi, 2009, Gazi Üniversitesi, Ankara.

[23]. Tekindal, B., Benzer, R., Benzer, A. İ., Yapay Zekâ Uygulamalarında Kullanılan Arama Algoritmalarının Kıyaslanması: Savunma Bilimleri Dergisi, 2007, 6(2): p. 88-103.

Bölüm 4

KUANTUM DERİN ÖĞRENME

Hande Çavşi Zaim - Metin Yılmaz - Esra Nergis Yolaçan

Bu kitap bölümünde kuantum hesaplamanın temel prensipleri, kuantum derin öğrenme algoritmaları, kuantum derin öğrenme platformları ve kuantum derin öğrenmenin günümüzde hangi alanlarda kullanıldığı incelenmiştir. Kuantum derin öğrenme algoritmalarının klasik modellerden nasıl esinlendiği gösterilmiştir. Kitap bölümünün kuantum derin öğrenme alanında çalışma yapmak isteyen araştırmacılara rehber niteliğinde olması amaçlanmıştır.

4.1. GİRİŞ

1959'da Arthur Samuel'in "bilgisayarların açıkça programlanmadan öğrenme yeteneği" olarak tanımladığı makine öğrenimi, bilgisayar biliminin bir alt alanıdır [1]. Yapay zekâdaki örüntü tanıma ve hesaplamalı öğrenme teorisi çalışmasından geliştirilen makine öğrenimi, verilerden öğrenebilen ve veriler üzerinde tahminler yapabilen algoritmaların incelenmesini ve oluşturulmasını araştırır [2,3]. Makine öğrenimi, iyi performansa sahip açık algoritmalar tasarlamanın ve programlamanın zor veya mümkün olmadığı alanlarda bir dizi hesaplama görevinde kullanılır. Bu alanlara e-posta filtreleme, ağa izinsiz girişlerin tespiti, optik karakter tanıma ve bilgisayarlı görü alanları örnek olarak verilebilir [1,4]. Bir tür makine öğrenimi olan derin öğrenme, insan beyninin işleyişini taklit etmek amacıyla birden fazla gizli katman içeren yapay sinir ağları olarak tanımlanabilir [1]. Bu derin ağlar, özellik çıkarma ve dönüştürme için birçok doğrusal olmayan işlem birimi katmanından oluşur ve her ardışık katman bir önceki katmanın çıktısını girdi olarak kullanır. Makine öğreniminde olduğu gibi derin öğrenmede de algoritmalar danışmanlı (denetimli) veya danışmansız (denetimsiz) olabilir. Özetle derin öğrenme, verilerin öğrenim temsillerine dayalı daha geniş bir makine öğrenimi yöntemleri ailesinin parçasıdır. Derin sinir ağları, evrişimli derin sinir ağları, derin inanç ağları ve tekrarlayan sinir ağları gibi çeşitli derin öğrenme mimarileri, bilgisayarlı görme, otomatik konuşma tanıma, doğal dil işleme, ses tanıma ve biyoinformatik gibi alanlara uygulanmıştır. Derin öğrenme, uygulandığı problemler değerlendirildiğinde, makine öğrenimi algoritmalarından daha başarılı sonuçlar göstermiştir [5,6].

Kuantum bilgisayar, yalnızca klasik fizik yasalarına dayanan herhangi bir makinenin yapamayacağı şeyleri yapmak için kuantum mekaniğini kullanmak üzere tasarlanmış bir makinedir [7]. Kuantum hesaplamanın nihai uygulamaları, kriptografik sistemlerin kırılmasından yeni ilaçların tasarımına kadar uzanan geniş bir alana sahiptir. Bu uygulamalar, herhangi bir klasik makine öğrenimi veya derin öğrenme algoritmalarında hızlanma veya verimlilik artışı sağlamayı hedeflemektedir. Büyük ölçekli genel amaçlı kuantum bilgisayarlar henüz mevcut olmasa da kuantum algoritmaları teorisi 20 yılı aşkın bir süredir aktif bir çalışma alanı olmuştur.

Bu bölümde kuantum hesaplamanın temel prensipleri, kuantum derin öğrenme algoritmaları ve günümüzde hangi alanlarda kullanıldığı ayrıntılı olarak

ele alınmıştır. Kuantum hesaplama ve kuantum algoritmaları günümüzde halen gelişmekte olan popüler konulardır. Bu çalışmanın, kuantum derin öğrenme algoritmalarının avantajları, klasik modellerinden nasıl esinlenildiği ve hangi alanlarda çalışmalar yapıldığı gösterilerek bu alanda çalışma yapmak isteyen araştırmacılara bir rehber niteliğinde olması amaçlanmıştır.

4.2. KUANTUM HESAPLAMANIN TEMELLERİ

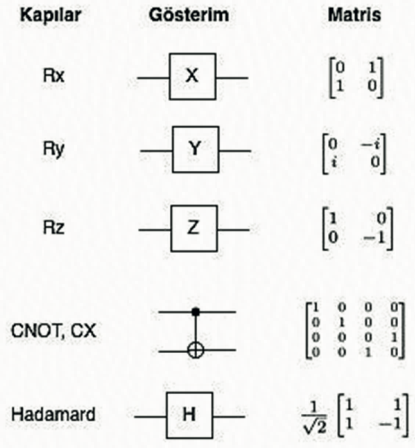
Kuantum hesaplama, bilgisayar bilimi ile kuantum mekaniğini birleştiren hızla büyüyen bir araştırma alanıdır [8]. İlk olarak 1982'de Feynman, bir kuantum sistemini simüle etmek için bilgisayarın kuantum mekaniksel olarak çalışması gerektiğine veya bir kuantum bilgisayara ihtiyaç duyulduğuna dikkat çekmiştir [9]. Kuantum hesaplamanın pratik uygulaması için ilk teklif ise 1993'te sunulmuştur. Klasik bilgisayarlardaki temel bilgi birimi bit yerine, kuantum bilgisayarlar kübitleri kullanmaktadır. Kuantum hesaplamanın klasik hesaplamaya kıyasla avantajı, kübitler tarafından sergilenen süperpozisyon ve dolanıklık özellikleridir. 0 veya 1 gibi klasik bitlerin aksine, süperpozisyon özelliği iki durumun bir kombinasyonunda bir kübitin var olmasına izin verir. Kübitler klasik bitlere göre içerisinde daha fazla bilgi depolayabildikleri için çok yüksek boyutlu verilerde kübitlerle işlem yapmak işlem süresini kısaltır. Bir kübit genel olarak $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ şeklinde ifade edilir. Burada, $|0\rangle$ ve $|1\rangle$ kuantum durumlarını, α ve β her birine karşılık gelen genlikleri ifade eder. $\alpha^2 + \beta^2 = 1$'dir. Bir kübit gözlemlendiğinde, kübitin temel durumlardan birinin değerini alması söz konusudur. Gözlemlenen her durumun alınan değer olma olasılığı o kübitin katsayı genliğinin karesiyle orantılıdır. Yani $|0\rangle$ ve $|1\rangle$'in olasılıkları sırasıyla α^2 ve β^2 ile orantılıdır. Süperpozisyon özelliği, kuantum hesaplama sistemlerinin olasılıksal durum kombinasyonları üzerindeki paralel hesaplamalar nedeniyle, klasik hesaplamaya kıyasla üstel hızlanmalar sağlamaktadır. Dolanıklık özelliği ise; kübitlerin birbirleriyle korelasyonunu sağlamaktadır. Kübitler arasındaki bu dolanıklık durumundan kaynaklı olarak bir kübit durumunu tensör ürünü olarak ifade etmek mümkün değildir. Bu durumda bir kuantum durumu Eşitlik 4.1'de görüldüğü gibi yazılabilir.

$$|\phi+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \tag{4.1}$$

Kübitlerin kuantum durumları çeşitli kuantum kapıları tarafından manipüle edilebilir. Kuantum kapıları üniter matrislerle (U) ifade edilir. Örnek olarak başlangıç durumu $|\psi_0\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle$ olan bir kuantum sistemi için, üniter bir U dönüşümü gerçekleştiren bir U kapısı Eşitlik 4.2'deki gibi ifade edilebilir.

$$U|\psi_0\rangle = U \sum_{i=0}^{2^n-1} \alpha_i |i\rangle = \sum_{i=0}^{2^n-1} \beta_i |i\rangle \quad (4.2)$$

Herhangi bir üniter dönüşüm, bir dizi temel kuantum kapısı içerir ve sonlu bir kapı dizisi olarak tanımlanır [10]. Kuantum kapıları tersine çevrilebilir olmak için sınırlandırılmıştır ve aynı boyutta bir çıktı sağlamak için giriş durumunda çalışır. Genel olarak, kuantum kapıları, tersi karmaşık eşleniği olan kare matrisler olan üniter matrislerle temsil edilir. N kübitlik bir kuantum sisteminin 2N temel kuantum durumunun süper pozisyonu mevcuttur. Kuantum sistemlerde en yaygın olarak Hadamard ve CNOT kapıları kullanılır. Hadamard kapıları bir kübit üzerinde hareket eder ve $|0\rangle$ ve $|1\rangle$ kuantum durumlarını $\frac{|0\rangle + |1\rangle}{\sqrt{2}}$ ve $\frac{|0\rangle - |1\rangle}{\sqrt{2}}$ durumlarına eşler. CNOT kapısı 2 kübit üzerinde çalışır ve $|a, b\rangle$ kuantum durumunu $|a, a \oplus b\rangle$ durumuna eşler. Yani, ilk bit kopyalanır ve ilk bit 1 ise ikinci bit çevrilir. R_x, R_y, R_z kapıları ise tek kübit üzerinde çalışır ve sırasıyla kübitleri x, y, z eksenleri üzerinde belirli bir açıyla döndürme işlemi gerçekleştirir. Genel kuantum kapıları Şekil 4.1'de gösterildiği gibidir.



Şekil 4.1. Kuantum kapıları [8].

Kuantum makine öğrenimi, kuantum hesaplama ve makine öğrenimi alanlarının birlikte kullanılması sonucu oluşmuş bir çalışma alanıdır. Lamontagne ve ark. [11], makine öğrenimi ve kuantum hesaplamayı birleştirmek amacıyla dört yaklaşım belirlemişlerdir. Bu yaklaşıma göre, veriler bir kuantum (Q) veya klasik (C) sistem tarafından üretilebilir. İşleme cihazı olarak bir kuantum veya klasik sistem kullanılabilir. Yaklaşımlar, veri ve işleme yöntemine göre dört grupta tanımlanmıştır. İlk yöntem klasik verilerin klasik olarak işlenmesi (CC) durumudur. Burada kuantum makine öğrenimi bağlamında, kuantum araştırmalarından ödünç alınan yöntemlere dayalı makine öğrenimi ifade edilir. İkinci yöntem kuantum verilerin klasik olarak işlenmesi (QC), makine öğreniminin kuantum hesaplamaya nasıl yardımcı olabileceğini araştırır. Birçok makine öğrenimi modeli, bir kuantum bilgisayarın iç durumunu mümkün olduğunca az ölçümle belirlemek için faydalı olabilir. Üçüncü yöntem klasik verilerin kuantum olarak işlenmesi (CQ), metin, görüntü, zaman serisi gibi klasik sistemlerden alınan klasik veri kümelerini işlemek için kuantum hesaplamayı kullanır. Dördüncü yöntem, kuantum bilgisayarlar tarafından işlenen kuantum verilerini (QQ) araştırır. Veriler fiziksel bir deneyde bir kuantum sisteminin ölçülmesiyle ve elde edilen değerin başka bir kuantum cihazına beslenmesiyle elde edilir veya bir kuantum sisteminin dinamiklerini simüle etmek için bir kuantum bilgisayar kullanılabilir. Kuantum makine öğreniminin klasik makine öğrenimine kıyasla avantajları çalışma zamanı karmaşıklığı, örnek karmaşıklığı, gürültüye karşı dayanıklılık ve model karmaşıklığı açısından ele alınabilir [12]. Çalışma zamanı karmaşıklığı hesaplama karmaşıklığı olarak da bilinir. Kuantum makine öğrenimi algoritmaları, klasik yöntemlere kıyasla daha hızlı hesaplamalar sağlar. Örnek karmaşıklığı, verilerden genelleme yapmak için gereken örnek sayısını ifade eder. Klasik ve kuantum örnek karmaşıklığının polinomsal olarak eşdeğer olduğu gözlemlenmiştir. Makine öğrenimindeki gürültü terimi, bozuk örneklerin varlığını ifade eder. Kuantum makine öğrenimi, klasik yöntemlere kıyasla bozuk verilere sahip olsa bile daha fazla veri kullanarak etkili şekilde öğrenme gerçekleştirebilir. Model karmaşıklığı, makine öğrenmesi modellerinin ifade kapasitesini temsil eder. Karmaşık modellerde aşırı uyum gösterme (overfitting) söz konusudur. Kuantum modellerinin, belirli veri kümelerindeki kalıpları yakalamada yararlı bir yöntem olduğu kanıtlanmıştır. ML modelleri, tanıdıkları kalıpları yeniden üretebilir ve tersine, üretebilecekleri kalıpları tanıyabilirler. Yakın zamanda yapılan

araştırmalar klasik olarak örneklenmesi çok daha zor olan olasılık dağılımlarından kolayca örneklemeler gerçekleştirebileceğini göstermiştir [13,14]. Bu dağılımlar gerçek dünya problemi dağılımlarıyla örtüşecek olursa, klasik modellere göre önemli bir kuantum avantajı sağlamak mümkün olacaktır.

Kuantum makine öğrenimi algoritmalarından en yaygın olarak kullanılanları, kuantum destek vektör, kuantum k-ortalama kümeleme ve kuantum temel bileşenler analizi algoritmalarıdır. K-ortalama kümeleme algoritmasında kümeleme işlemini gerçekleştirmek için üç farklı kuantum alt rutini kullanılır. Bunlar SwapTest, DistCalc ve Grovers optimizasyonudur. SwapTest alt programından gelen örtüşme, DistCalc alt programında mesafeyi hesaplamak için kullanılır. a ve b vektörü arasındaki öklid uzaklığı |a − b| üç adımda hesaplanır [15]. Bunlardan ilki durum hazırlığıdır. Eşitlik 4.3'teki gibi iki kuantum durumu hazırlanır.

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0,a\rangle + |1,b\rangle), |\emptyset\rangle = \frac{1}{\sqrt{Z}}(|a||0\rangle, |a||1\rangle) \quad (4.3)$$

Burada $Z = |a|^2+|b|^2$'dir. İkinci adım örtüşmeyi bulmaktır. Burada SwapTest yöntemi kullanılarak $\langle\psi|\emptyset\rangle$ örtüşmesi hesaplanır. Üçüncü adım öklid mesafesini hesaplamaktır. Öklid mesafesi Eşitlik 4.4'teki gibi hesaplanır.

$$Mesafe = 2Z|\langle\psi|\emptyset\rangle|^2 \quad (4.4)$$

Mesafe hesaplamaları gerçekleştirildikten sonra, en yakın ağırlık merkezi Grovers optimizasyon alt programı kullanılarak bulunur. Kuantum k-ortalama kümeleme algoritması kuantum paralelliğini kullanarak üstel bir hızlanma sağlar. Kuantum ve klasik destek vektör makineleri arasındaki temel fark, klasik durumda açık bir çekirdek fonksiyon denkleminin genellikle bilinmesidir. $K(x_i, x_j) = \exp\left(\frac{-|x_i-x_j|^2}{2\sigma^2}\right)$ olarak tanımlanan Radyal Temelli Fonksiyon (RBF) çekirdeği buna örnek olarak verilebilir. Burada, σ optimize edilebilen serbest bir parametredir. Kuantum durumunda, çekirdek matrisi bunun yerine bir kuantum devresi kullanılarak hesaplanır. Kuantum destek vektör makinesinin ilk adımı, klasik x_i verilerini $|\psi(x_i)\rangle$ kuantum durumuna kodlamaktır. Bu kodlama devresinin U(x) amacı, her bir parçacı-

ğın momentum koordinatlarından oluşan klasik bir n boyutlu dizi olan her olayı 2n boyutlu bir kuantum durumuna dönüştürmektir. Girdi verileri, bir olay için düşünülen parçacıkların tüm momentum değişkenlerini içeren bir $x \in \mathbb{R}$ n vektörü olarak temsil edilebilir. Bu vektör, Eşitlik 4.5'te n kübit kullanılarak $|\psi(x)\rangle$ kuantum durumuna eşlenir [16].

$$|\psi(x)\rangle = U(x)|0\rangle^{\otimes n} = \left(\prod_{i=0}^{L} U(x)H^{\otimes n}\right) \tag{4.5}$$

Kuantum temel bileşenler analizi, yoğunluk matrislerinin öz vektörlerini ve öz değerlerini ortaya çıkaran yeni bir durum ve süreç tomografisi ilkesidir. Kuantum temel bileşenler analizinin avantajlarını ve dezavantajlarını kuantum sıkıştırılmış algılama yöntemiyle kıyaslayarak net bir şekilde göstermek mümkündür. İki yöntem arasındaki birincil fark; Kuantum temel bileşenler analizinin özvektörleri oluşturması ve bunları O(R log d) zamanında karşılık gelen öz değerlerle ilişkilendirmesidir. Sıkıştırılmış algılama yöntemi ise, tam yoğunluk matrisinin O(Rd log d) zamanında klasik tanımını yeniden yapılandıran bir durum ve süreç tomografi yöntemidir. Yalnızca tek kübitlik hazırlıklar ve ölçümler kullanılır. Kuantum temel bileşenler analizi, öz vektörler üzerinde durum tomografisi yapmak için de kullanılabilir ve bileşenlerini O(Rd log d) zamanında ortaya çıkarır. Öz değerlerin ve öz vektörlerin bu klasik tanımı, daha sonra tam yoğunluk matrisini O(Rd log d) zamanında yeniden üretmek için kullanılabilir [17].

4.3. KUANTUM DERİN ÖĞRENME YAPILARI

Günümüzde, derin öğrenme artan veri ve hesaplama kaynakları göz önüne alındığında bilgisayar bilimlerindeki en aktif araştırma alanlarından biri haline gelmiştir [18]. Derin öğrenme bilgisayarla görme, doğal dil işleme, pekiştirmeli öğrenme, ses tanıma gibi çeşitli alanlarda çalışmalara katkı sağlamıştır [19-24]. Bu çalışmalara paralel olarak, hesaplama açısından daha az maliyetli teknikler kullanılmasından dolayı zorlu problemleri çözmeye odaklanan kuantum hesaplama alanında da önemli ilerlemeler kaydedilmiştir. Shor [25], RSA gibi güvenlik protokollerinin zafiyetlerini açığa çıkaran polinomsal zamanda asal çarpanlara ayrılan sayılarla çalışan yeni bir algoritma önererek kuantum alanındaki en büyük atılımı başlatmıştır. Kuantum ve derin öğrenme teknolojilerindeki bu gelişim kuantum derin öğrenme alanını

son yıllarda popüler bir araştırma alanı haline getirmiştir. Günümüzde kuantum derin öğrenme aynı kuantum makine öğrenimi gibi keşif aşamasındadır. Ancak, son zamanlarda kuantum sinir ağlarının ses tanıma, görüntü tanıma ve finans alanlarında gerçek dünya problemleri üzerindeki bazı uygulamaları ortaya çıkmaya başlamıştır [26-28]. Tüm bu çalışmalara rağmen, kuantum derin öğrenme için algılayıcının, mimarinin, optimizasyon algoritmalarının ve kayıp fonksiyonlarının doğru genelleştirilmesi önemli bir zorluktur. Kuantum derin öğrenme alanındaki yapılan çalışmalar, eğitim verimliliği ve model kalitesi açısından geleneksel derin öğrenme algoritmalarından daha iyi performans gösteren kuantum derin öğrenme algoritmalarını araştırmaktadır. Genel olarak, derin ağlar yüksek düzeyde bağlantılı karmaşık bir grafik ağa kodlanmış birçok soyutlama düzeyi içermektedir. Kuantum hesaplamalardaki gelişme klasik modellerin çözemediği zorlu problemleri daha iyi bir performansla çözmek için umut verici çözümler sunmaktadır.

4.3.1. Kuantum Sinir Ağı (QNN)

Klasik sinir ağları, ileri beslemeli sinir ağlarıdır. Bu mimarilerinden dolayı girdi olarak verilen çıktının tek yönlü hesaplaması söz konusudur. $(X_1, X_2, \dots, X_n) \rightarrow y$ ifadesinde X_n giriş verisi, y çıktı olarak ifade edilir. Klasik sinir ağlarının aksine kuantum mekaniği doğası gereği tersine çevrilebilir dönüşümlere bağlıdır. Kuantum hesaplamada sinir ağı girdilerinin çıktılara dönüştürülmesi için girdiye yardımcı bir bit eklenir. Bu durum matematiksel olarak; $(X_1, X_2, \dots, X_n, 0) \rightarrow (X'_1, X'_2, \dots, X'_n, y)$ şeklinde ifade edilebilir. Girdi temsillerini üniter yapmak için $(X_1, X_2, \dots, X_n)$ girdi temsili $|\psi\rangle_{1,\dots,n}$ kuantum durumu aracılığıyla gösterilir. Burada girdi çıkışı y'ye karşılık gelen yardımcı bir kübit eklenir. Böylece kuantum ayarında üniter hale getirilmiş bir tersinir dönüşüm elde edilmiş olur. Kuantum sinir ağları girdi olarak salt kuantum verilerini veya klasik verilerin kuantum durumlarına dönüştürülmüş halini alabilir. Kuantum verileri temsil edilirken, $|\psi\rangle_{1,\dots,n}$ ifadesi n boyutlu $H^{2^n} = H^2 \otimes \dots \otimes H^2$ Hilbert uzayında, 2n hesaplama tabanının bir süper pozisyonu olarak temsil edilir. Burada H^2 iki boyutlu Hilbert uzayını $\{|0\rangle, |1\rangle\}$ temsil eder. H^{2^n} ise; $\{|0,0,\dots,0\rangle,\dots,|1,1,\dots,1\rangle\}$ uzayını temsil eder. Böylece $|\psi\rangle_{1,\dots,n}$; Eşitlik 4.6'daki gibi hesaplanabilir.

$$|\psi\rangle_{1,\dots,n} = \sum_{i=0}^{2^n} \alpha_i \, |Z_i\rangle \tag{4.6}$$

Burada, $\alpha_i, i \in \{1, \dots, 2^n\}$ durumlara atanan karmaşık genlikleri temsil etmektedir ve $|z_i\rangle \in H^{2^n}$'dir.

Gerçek kuantum verilerini elde etmek her durumda mümkün olmadığından literatürde yapılan çalışmalar, klasik verileri kuantum verilere çevirme yöntemleri üzerine yoğunlaşmıştır. Bunlardan biri Fahri ve ark.'nın kullandığı X_i, i ∈ {1,….,n} ise X=$(X_1, X_2, \dots, X_n)$ girişini bir eşik değerinden geçirdiği yöntemdir [29]. Burada eşik değerini geçen her bir bit ikili gösterimi sonucu elde edilmiş X_n boyutuna karşılık gelen |0⟩ veya |1⟩ kübitleriyle temsil edilir ve her bir x, H^{2^n} Hilbert uzayında yer alır. Ancak bu yaklaşım verilerde yüksek bilgi kaybına yol açabilmektedir. Buna çözüm olarak Allcock ve ark., *X*'lerin süperpozisyonu olarak daha detaylı bir temsil gerçekleştirmeyi önermişlerdir [30]. Burada |i⟩ her i ∈ {1,….,n} boyutu için i. Konumunda kübit 1'li |0…..,1,……0⟩ kuantum durumunun hesaplama temelini göstersin. Bu durumda *X* verisi Eşitlik 4.7'deki kuantum durumu olarak hesaplanabilir.

$$|\psi\rangle_{1,\dots,n} = \sum_{i=1}^{n} b_i \, |i\rangle, b_i = \frac{X_i}{||X||} \tag{4.7}$$

Ham kuantum verileri veya klasik verilerden kuantum verilerine dönüştürülmüş veriler ikinci adımda öğrenilebilir kuantum devleri kullanılarak modellenmiş sinir ağı modeline girdi olarak beslenir. Bu model, $|\psi\rangle_{1,\dots,n}$ |0⟩ kuantum durumunu $|\psi'\rangle_{1,\dots,n}$ |y⟩ kuantum çıktılarına dönüştürmek için permütasyon matrislerini kullanır. Matematiksel olarak bir permütasyon kare matrisi P, girdileri 0 veya 1 olacak şekilde $P.P^T$=I ifadesini verir. Bu bir kapalı kuantum devresini temsil eder. Ancak bu devrelerle oluşturulan modellerin giriş verileri üzerinde yakalayabileceği kalıp zenginliği kısıtlıdır. Bu nedenle öğrenilebilir serbest parametreleri ifade eden üniter matrislerin kullanılması daha derin kalıpların yakalanması işlemini arttırır. Herhangi bir üniter matris U, $U=e^{iH}$ olarak ifade edilir. Burada H bir Hermit matrisi ifade eder. Her bir Hermit matrisi, Pauli matrislerinin ({σ_x , σ_y , σ_z}) ve birim matrisin tensör ürünlerinin kombinasyonları olarak yazılabilir. Bu durumda K bitleri üzerindeki U üniter matrisinin hesaplaması Eşitlik 4.8'deki gibidir.

$$U = \exp[i(\sum_{j_1=0,\dots,j_{k=0}}^{3,\dots.3} \sigma_{j_1} \otimes \dots . \otimes \sigma_{j_k})] \tag{4.8}$$

Burada σ_i, $\{i = 0,1,2,3\}$ için sırasıyla $\{I_{2x2}, \sigma_x \sigma_y \sigma_z\}$ anlamına gelir ve $\alpha_{j_1, j_2, \ldots j_K}$ eğitilebilir serbest parametrelerdir.

$$U^{n+1}(\Theta)|\psi\rangle_{1,\ldots,n}|0\rangle \tag{4.9}$$

Kuantum sinir ağı, giriş verilerindeki ayrıntılı desenleri yakalamak için klasik derin sinir ağının yapısına benzer şekilde birkaç varyasyonel devrenin birleşmesinden oluşur. N ardışık üniter matrisin gösterimi Eşitlik 4.10'daki gibidir.

$$U(\Theta) = U_N(\Theta_N)U_{N-1}(\Theta_{N-1}) \ldots U_1(\Theta_1) \tag{4.10}$$

Burada U_i (Θ_i), i. katmana karşılık gelen üniter matrisi gösterir ve $\Theta = \{\Theta_1, = \{\Theta_N\}$ tüm parametrelerin kümesidir.

$|\psi\rangle_{1,\ldots,n}$ kuantum giriş verilerinin çıktısı olan $|y\rangle$ durumu hesaplama temelli veya karışık bir kuantum durumu olabilir. Bu durumda elde edilen çıktı durumunun giriş verisinin üniter işlemlerden sonra elde edilen son kuantum durumundaki son kübite karşılık gelen $p^{çıkış}$ ile gösterilmesi gerekir. Gözlemlenen gerçek çıktı ve kuantum durumu arasında popüler bir yakınlık ölçüsü olan eğitim verilerinin ortalaması yöntemi kullanılarak elde edilen çıktının gerçek duruma yakınlığının maliyet fonksiyonu (C) Eşitlik 4.11'deki gibi hesaplanabilir.

$$C = \frac{1}{N}\sum_{x=1}^{N} \langle y_x | p_x^{çıkış} | y_x \rangle \tag{4.11}$$

Fahri ve Neven ise; QNN çıktısını ölçmek için σ_y olarak adlandırılan bir Pauli operatörünü ölçmeyi önermişlerdir [29]. Bu ölçüm farklı olası sonuçlarda farklı olasılık değerleri verir. $|\psi\rangle_{1,\ldots,n}|0\rangle$ girişi için oluşan çıktıların birden çok olasılık değerinin ortalaması Eşitlik 4.12'deki gibi ölçülür.

$$y^{çıkış} = \langle \psi_{1,\ldots,n} 0 | U(\Theta)\dagger \quad YU(\Theta) \quad | \psi_{1,\ldots,n} 0 \quad \rangle \tag{4.12}$$

Bu durumda C kaybı, ortalama karesel hata veya 0/1 kayıp olarak tanımlanabilir. Eşitlik 4.13'te görüldüğü gibidir.

$$C_{MSE} = \frac{1}{N}\sum_{x=1}^{N} |y_x - y_x^{çıkış}|^2 \text{ veya } C_{0/1} = \frac{1}{N}\sum_{x=1}^{N} 1[y_x = y_x^{çıkış}] \tag{4.13}$$

Burada y_x, X girişi için orijinal çıktıya, $y_x^{çıkış}$ ortalama QNN çıktısına karşılık gelir. Kuantum derin öğrenmede de klasik derin öğrenmeye benzer şekilde kuantum sinir ağları (Θ) parametre kümesine sahip üniter matrisleri üzerindeki kayıp fonksiyonu en aza indirgemek için birinci dereceden optimizasyon tekniklerini kullanarak öğrenme gerçekleştirir. Bu kural Eşitlik 4.14'te görüldüğü gibidir.

$$\Theta \leftarrow \Theta - \eta \frac{\partial C(\Theta)}{\partial \Theta} \tag{4.14}$$

Burada Θ öğrenilen parametreleri, C kayıp fonksiyonunu, η adım boyutunu ifade etmektedir.

Dabson ve ark., kuantum makine öğrenimi modellerinin potansiyel yeteneklerini ölçmek ve değerlendirmek için Modified NIST (National Institute of Standards and Technology) veri tabanlarında çalışabilen bir kuantum sinir ağı önermişlerdir [31]. Bu kuantum sinir ağı modeli bir kuantum simülatörü üzerinde MNIST veri seti kullanılarak gerçekleştirilmiştir. Gelecekte gerçek bir kuantum donanımı üzerinde çalıştırılması ve simüle edilmiş kuantum platformlarıyla kıyaslanması hedeflenmiştir. Çalışmanın sonucunda kuantum sinir ağı mimarisinin makine öğrenimi modelinin eğitim süresini önemli ölçüde azalttığı gözlemlenmiştir. Ayrıca kuantum makine öğreniminin teknik hazırlık düzeyini oluşturduğu düşünülmektedir.

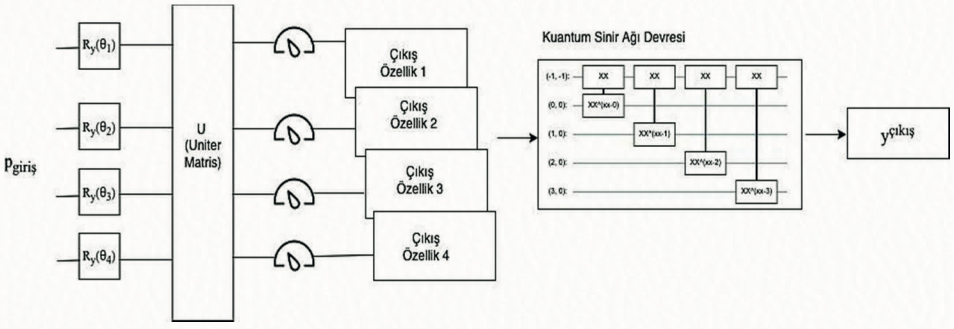
Chen ve ark., parametre öğrenme sürecinde geleneksel geri yayılım algoritmasını geliştirmek için genetik algoritmayı benimseyen ve kuantum hesaplamaya dayalı yeni bir kuantum ileri beslemeli sinir ağı modeli önermişlerdir [32]. Çalışma, Matlab simülasyonları üzerinde gerçekleştirilmiştir. Deneysel sonuçlar hızlı kuantum sinirsel hesaplamayı en iyi şekilde kullanan ağ modelinin, fonksiyon yaklaşımında ve eğitici kısa vidyonun görüntünün (video) yayılma kapasitesinin tahmininde geleneksel yöntemlere kıyasla daha verimli sonuçlar verdiğini göstermektedir. Uygulanan modelin hava tahmini, el yazısı tanıma, ses tanıma ve diğer alanlarda yaygın olarak kullanılabileceği düşünülmektedir.

Safari ve ark., kuantum teknolojisi ve akıllı şebekelerin hava tahmini olarak bilinen bir uygulama alanını analiz etmişlerdir [33]. Akıllı şebekeler, akıllı bir şehrin ve onun tedarik kaynaklarının önemli temelleri arasındadır. Bu ne-

denle hava tahmini veya talep yanıtı gibi faktörler bu alanda çok önemli bir role sahiptir. Uzmanların hava durumu tahmini gibi bu faktörleri tahmin etmek için yüksek doğrulukta modeller kullanmaları gerekir. Toplanan büyük verinin çok büyük miktarda olduğu geleceğin akıllı sürdürülebilir şehirlerinin (SSC'ler) akıllı şebekelerinde, klasik yapay zekâ ve yapay sinir ağı modelleri kullanmak verimsiz olacaktır. Bu nedenle bu çalışmada akıllı şebekelerin hava tahmini için kuantum sinir ağı modeli kullanılmıştır. Kuantum sinir ağı modeli diğer klasik yöntemlerle karşılaştırılmıştır. Kuantum teknolojisinin ve kuantum sinir ağlarının oldukça hızlı işleme ve yürütme hızı kullanılarak diğer klasik modellerle birleştirilebileceği sonucuna varılmıştır.

Sengupta ve ark., Hindistan'daki COVID-19 hastalarının bilgisayarlı tomografi (BT) görüntü verileri üzerinde yürütülen simülasyon teknikleri aracılığıyla yeni bir görüntü sınıflandırma ve segmentasyonu yöntemi ortaya koymuşlardır [34]. Sınıflandırma için kuantum sinir ağları modelinden yararlanılmıştır. Önerilen model, belirli sınıflandırma görevi için geleneksel derin öğrenme modellerinden daha iyi performans göstermiştir. Kuantum simülasyonu kullanılarak gerçekleştirilen kuantum sinir ağı modeli sınıflandırma işleminin eğitim süresini kısaltmıştır. GPU üzerinde gerçekleştirilen eğitim süresi 1 sa. 30 dk. Olmasına rağmen kuantum simülatörü üzerinde bu eğitim süresi 52 dakika olarak bulunmuştur. Ancak kuantum sinir ağı modelinin tıbbi cihazlara entegre edilerek uygulama senaryolarının değerlendirilmesi gelecekteki hedefler arasındadır.

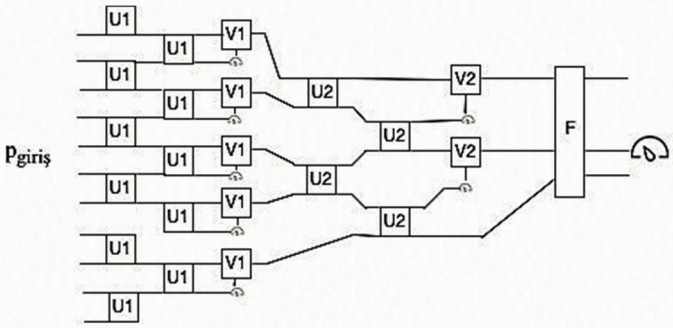
Literatür çalışmaları sonucunda kuantum sinir ağı modelini Şekil 4.2'deki gibi göstermek mümkündür.



Şekil 4.2. Kuantum sinir ağı modeli.

4.3.2. Kuantum Evrişimli Sinir Ağı (QCNN)

Kuantum evrişimli sinir ağı klasik evrişimli sinir ağının evrişim ve havuzlama katmanlarından esinlenerek bir kuantum devre modeli aracılığıyla modellenmiştir [35]. Ancak iki boyutlu ve üç boyutlu evrişim katmanlarının aksine bir boyutlu evrişim katmanı uygulanması açısından farklılıklar gösterir. Burada evrişim katmanı $p_{giriş}$ durum yoğunluğu üzerinde yarı yerel bir üniter işlem olarak modellenmiştir. Bir üniter matrisi U_i ile gösterilir ve önceden tanımlanmış bir derinliğe kadar birkaç ardışık girdi kübitine uygulanır. Havuzlama katmanı bazı kübitler üzerinde ölçümler yapar ve yakındaki kübitlere üniter dönüşler uygular. V_i ile temsil edilir. Döndürme işlemi kübitler üzerindeki ölçümlerle belirlenir. Özetle havuzlama katmanı boyut indirgeme ve doğrusal olmama fonksiyonlarını birleştiren bir katmandır. Evrişim ve havuzlama üniterlerinin katmanlarından sonra tam bağlı (F) katmanı uygulanır. F çıkışındaki son bir ölçüm kuantum evrişimli sinir ağının çıkışını vermektedir. Kuantum evrişimli sinir ağı mimarisi Şekil 4.3'teki gibidir.



Şekil 4.3. Kuantum evrişimli sinir ağı kuantum devre modeli [35].

Kuantum sinir ağı modelinde üniterlerin parametreleri öğrenilerek öğrenme işlemi gerçekleşir. Öğrenilen parametrelerin amacı optimizasyon yöntemleriyle kayıp fonksiyonunu en aza indirgemektir.

Zhang ve ark., polimetrik sentetik açıklıklı radar (PolSAR) görüntü sınıflandırmasında benek gürültüsünün olumsuz etkilerinden dolayı piksellerin yerel uzamsal ilişkisinin modellenememesi sorunundan yola çıkmışlardır ve renk özelliklerinin iyi bir görsel ifade sağlayabileceği, görüntü yorumlama için iyi performans gösterebileceği göz önüne alındığında öğrenilmiş

süper pikselleri ve kuantum evrişimli sinir ağı modelini birleştiren denetimli bir görüntü sınıflandırma yaklaşımı önermişlerdir [36].

Veri veya modelin boyutu çok büyük miktarda olursa, geleneksel evrişimli yöntemler verimli bir şekilde öğrenmede zorlanırlar. Buna çözüm olarak, kuantum evrişimli sinir ağı bir kuantum hesaplama ortamı kullanarak mevcut yöntemlerin öğrenme performansını iyileştirmeyi sağlar. Oh ve ark., geleneksel evrişimli sinir ağı yapısını kuantum hesaplama ortamına uygulayarak yeni bir model önermişlerdir [37]. Önerilen model Multi-scale Entanglement Renormalization Ansatz (MERA) kullanarak, O(log(n)) derinliği ile hesaplanabilen bir modeldir. Bu çalışmada ikinci olarak, geleneksel evrişimli sinir ağı modeline kuantum hesaplamaları kullanılarak bir kuantum evrişim katmanı eklenmiştir. Bu yöntem model performansını iyileştirmek için bir çözüm sunar ve küçük kuantum bilgisayarlarda da kullanılabilir. Çalışmada sunulan bir diğer uygulama da TensorFlow-quantum platformu aracılığıyla kuantum evrişimli sinir ağı modelinin ve geleneksel evrişimli sinir ağı modelinin sonuçlarının karşılaştırılmasıdır. Bunun için MNIST veri seti kullanılmıştır.

Rajesh ve ark., kuantum evrişimli sinir ağının bilgisayarlı görme, sinyal ve görüntü işleme, eczacılık, kriptografi ve diğer çeşitli alanlardaki uygulamalarını incelemişlerdir [38]. Aynı zamanda, kuantum bilgisayarlar, kuantum hesaplama algoritmaları, yazılım ve donanım desteği geliştirmede önemli noktaları göstererek, gelecekteki çalışmalar hakkında tartışmışlardır.

4.3.3. Kuantum Tekrarlayan Sinir Ağı (QRNN)

Tekrarlayan sinir ağları, zaman serileri söz konusu olduğunda denetimli problemlerin üstesinden gelmek için güçlü bir çözümdür. İleri beslemeli sinir ağlarının bir modifikasyonudur ancak önceki çıktıların (t-1) mevcut durum (t) için kullanılabileceği yerlerde işlevseldir. Bu yöntem bir tür belleğe sahiptir ve istenen çıktıyı tahmin ederken geçmiş veriyi kullanır. Cao ve ark., kuantum nöronun değiştirilmiş bir versiyonuna dayanarak kuantum tekrarlayan sinir ağı modeli önermişlerdir [39]. Tekrarlayan sinir ağlarının kuantum versiyonlarını geliştirmek için Hibat-Allah ve ark., bir kuantum yaklaşık temel durumunu öğrenmek amacıyla değişken dalga fonksiyonlarını kullanan bir kuantum tekrarlayan sinir ağı modeli önermişlerdir [40]. Burada doğrusal olmayan aktivasyon fonksiyonlarını simüle etmek amacıyla

bazı döndürme hareketleri uygulanmaktadır. Roth, kafes (lattice) vektörlerinin çevirilerini, tekrarlayan sinir ağı zaman indeksine eşleyerek toplu kuantum sistemlerini simüle etmek amacıyla tekrarlayan bir eğitim yaklaşımı önermiştir [41].

Bauch, kuantum tekrarlayan sinir ağı modelini kullanarak bazı deneysel sonuçlar elde etmiştir [42]. Bu çalışmada MNIST veri setini kullanmıştır. Kuantum devrelerini klasik bir bilgisayarda simüle etmiştir ve verimi arttırmak için görüntülerin boyutlarını düşürmüştür. Ayrıca birkaç boyutsallık azaltma teknikleri uygulayarak da sonuçları test etmiştir. Kuantum tekrarlayan sinir ağı kullanarak yaptığı deneyler sonucunda %95 doğruluk elde etmiştir.

Taha ve ark., iki elektroensefalografi (EEG) sinyallerinin sınıflandırılması için otomatik gerilemeli (AR) bir model ve bir kuantum tekrarlayan sinir ağı modelini içeren hibrit bir yaklaşım önermişlerdir [43]. Bu hibrit modelin EEG sinyallerindeki doğal belirsizliği yakalayıp ölçebildiği gözlemlenmiştir. Kuantum tekrarlayan sinir ağının eğitim performansı kuantum sinir ağı ve kuantum dalgacık sinir ağı modelleri ile karşılaştırılmıştır. Kuantum tekrarlayan sinir ağı modelinin ortalama doğruluğu 6 saniyede %88.28452 olarak bulunmuştur. Elde edilen 14.714086 doğruluk/zaman oranı önerilen modelin üstünlüğünü göstermektedir. Özetle, AR ve kuantum tekrarlayan sinir ağı modelinden oluşan hibrit yaklaşım, diğer modellere kıyasla yüksek doğruluk, kısa işlem süresi ve daha güvenilir bir yapıya sahiptir.

4.3.4. Kuantum Çekişmeli Üretici Ağlar (QGAN)

Çekişmeli üretici ağlar, Goodfellow ve ark. tarafından geliştirilen bir derin ağ sınıfı ağıdır [44]. Çekişmeli üretici ağlar bir üretici G ve bir ayrımcı D modelini içeren teorik oyun senaryosuna dayanmaktadır. Çekişmeli üretici ağlar, verilerin bazı bilinmeyen R işlemi tarafından üretilen bir pR(x) dağıtımından geldiğini varsayar. θ_G ile parametrelendirilmiş üretici ağı, girdi olarak yapılandırılmamış rastgele değişken bir Z alır ve doğrudan $X=G(\theta_G, Z)$ örneklerini üretir. Böylece üreteç dağılımı pG(x) oluşmuş olur. θ_D ile parametrelendirilmiş rakip ayrımcı ağı, eğitim verilerinden (yani gerçek verilerden) alınan örnekler ile üretici ağdan alınan örnekler arasında bir ayrım yapmaya çalışır. Bu işlemler sonucunda ayırıcı ağ, $D(\theta_D, X)$ tarafından verilen ve X'in modelden alınan sahte örnek (pG(x)) yerine gerçek bir eğitim örneği (pR(x)) olma

olasılığını gösteren bir olasılık değeri verir. Klasik çekişmeli üretici ağların kuantum alanına genişletildiği çeşitli uzantıları önerilmiştir. Bunlardan biri Dallaire-Demers ve Killoran tarafından önerilen kuantum çekişmeli üretici ağıdır [45]. Klasik versiyonuna benzer şekilde $|\lambda\rangle$ etiketi verilen bir R veri kaynağının yoğunluk matrisinin $R(|\lambda\rangle) = \rho\lambda R$ olduğu varsayılmıştır. Burada çekişmeli üretici ağın amacı gerçek veri kaynağı R'yi taklit eden G üretici ağını öğrenmektir. Kuantum durumunda G ağı, üniter matrisleri θ_G ile parametrelendirilmiş bir kuantum devresi olarak tanımlanır. Böyle bir üreteç, girdi olarak $|\lambda\rangle$ etiketini ve ek bir $|z\rangle$ durumunu alır. Bunun sonucunda $G(\theta_G, |\lambda\rangle, z\rangle) = \rho\lambda G(\theta_G, z)$ kuantum durumunu üretir. Burada ek girdi durumu $|z\rangle$'nin rolü iki yönlüdür. İlki üretilen verilerin dağılımı içinde entropi sağlayan yapılandırılmamış bir gürültü kaynağı olarak görülmesidir. İkincisi üretici için bir kontrol işlevi olarak görülmesidir. Ayırıcı ağ ise; θ_D ile parametrelendirilmiş üniter kapılar içeren bir kuantum devresinden oluşur. G bu ayırıcı ağ D ile çekişmeli bir biçimde eğitilir. Eğitim sırasında D, belirli bir girdi durumunun gerçek veri kaynağından mı geldiğini yoksa G tarafından mı üretildiğini belirlemeyi amaçlarken, G, D'nin bunları gerçek olarak sınıflandırması için daha iyi girdiler üretmeyi amaçlar. Belirli bir girdi gerçek R kaynağından geliyorsa D ağı çıkış olarak $|gerçek\rangle$ çıktısını, G ağından gelen veri girdi olarak geliyorsa D ağı çıkış olarak $|sahte\rangle$ çıktısını vermelidir. Kuantum çekişmeli üretici ağlarının optimizasyon hedefi çekişmeli bir görev olarak Eşitlik 4.15'te görüldüğü gibi formüle edilir.

$$\min_{\theta_G} \max_{\theta_D} \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} \Pr((D(\theta_D, |\lambda\rangle,\ R|\lambda\rangle)) = |gerçek\rangle) \cap (D(\theta_D, |\lambda\rangle,\ G(\theta_{G,}|\lambda, z\rangle)) = |sahte\rangle)) \quad (4.15)$$

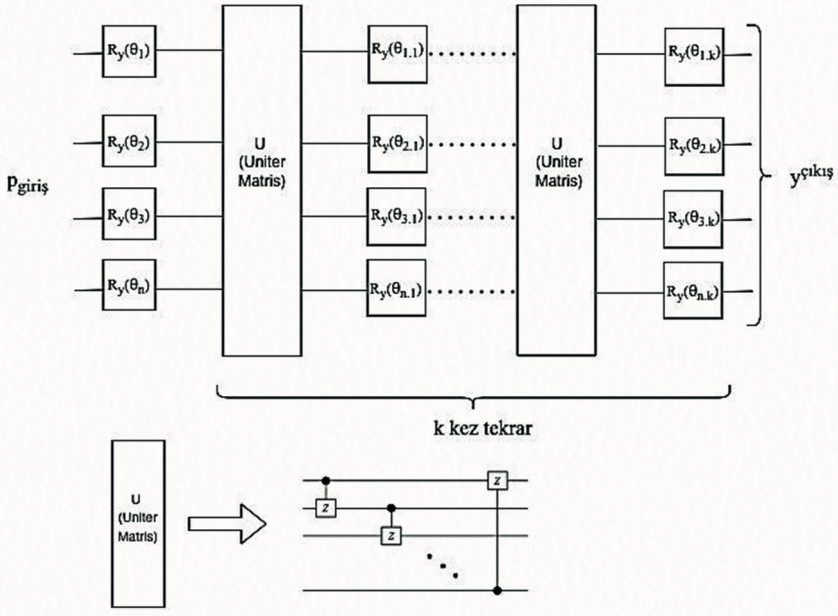
Huang ve ark., kuantum cihazlarında uygulanan kuantum çekişmeli üretici ağların gerçek dünyadaki öğrenme görevlerini çözüp çözmeyeceğini analiz etmek amacıyla bir kuantum çekişmeli üretici ağ modeli önermişlerdir [46]. Oluşturulan model, yüksek boyutlu özelliklerle görüntü oluşturma yeteneğine sahiptir ve birden fazla örneği eş zamanlı olarak eğitmek için kuantum süperpozisyon özelliğini kullanabilir. Çalışmada süper iletken bir kuantum işlemci üzerinde gerçek dünyadaki el yazısı rakam görüntülerinin öğrenilmesi ve üretilmesi gerçekleştirilmiştir. Kuantum çekişmeli üretici ağ modeli klasik çekişmeli üretici ağ modeliyle kıyaslanmıştır. Çalışma, yakın

zamanda kuantum cihazlarında gelişmiş kuantum üretici modeller geliştirmek için bir rehber niteliğindedir ve çekişmeli üretici ağlar üzerindeki kuantum avantajlarını göstermektedir.

Situ ve ark., üreteç olarak parametreli bir kuantum devresinden, ayırıcı olarak ise klasik bir sinir ağından oluşan klasik-kuantum hibrit mimarisine sahip bir model önermişlerdir [47]. Parametrelendirilmiş kuantum devresi, yalnızca mevcut kuantum cihazlarında bulunan basit bir bitlik dönüş kapılarından ve iki bitlik kontrollü faz kapılarından oluşan bir devredir.

Zoufal ve ark., genel olasılık dağılımlarının, veri örnekleri tarafından verilen kuantum durumlarına verimli bir şekilde öğrenilmesini ve yüklenmesini kolaylaştırmak için kuantum-klasik hibrit modeline sahip kuantum çekişmeli üretici ağ modelini önermişlerdir [48]. Kuantum çekişmeli üretici ağ modeli Qiskit kütüphanesi kullanılarak gerçekleştirilmiştir ve IBM-Q platformunda test edilmiştir. Ayrıca bir kuantum finans uygulamasında eğitilmiş kuantum kanalının kullanımı kuantum simülasyonu kullanılarak gösterilmiştir.

Kuantum çekişmeli üretici ağ modeli genel hatlarıyla Şekil 4.4'te görüldüğü gibidir.



Şekil 4.4. Kuantum çekişmeli üretici ağ modeli [48].

4.4. KUANTUM DERİN ÖĞRENME PLATFORMLARI

Kuantum derin öğrenme uygulamalarının her zaman kuantum bilgisayarlar üzerinde gerçekleştirilme imkânı yoktur. Bu probleme çözüm olarak özellikle akademik alandaki çalışmalar klasik bilgisayarlar üzerinde kuantum devrelerini simüle eden simülatör platformlarına yoğunlaşmışlardır.

Bu platformlardan biri PennyLane'dir. PennyLane, kuantum ve hibrit kuantum-klasik hesaplamaların optimizasyonu ve makine öğrenimi için geliştirilmiş bir Python 3 yazılım kütüphanesidir [49]. Kütüphane hem kübit hem de sürekli değişken paradigmaları destekleyen, yakın vadeli kuantum hesaplama cihazları için birleşik bir mimari sağlar. PennyLane'in temel özelliği, geri yayılım gibi klasik tekniklerle uyumlu bir şekilde varyasyonel kuantum devrelerinin gradyanlarını hesaplama yeteneğidir. Böylece optimizasyon ve makine öğreniminde yaygın olan otomatik farklılaşma algoritmalarını kuantum ve hibrit hesaplamaları içerecek şekilde genişletir. Bir eklenti sistemi, kütüphaneyi herhangi bir kapı tabanlı kuantum simülatörü veya donanımı ile uyumlu hale getirir. Ayrıca TensorFlow, PyTorch, IBMQ ve autograd gibi hızlandırılmış makine öğrenimi kütüphaneleriyle arayüz oluşturur.

Bir diğer platform ise; TensorFflow-Quantum platformudur. Bu platform, TensorFlow kapsamında hem ayırt edici hem de üretken kuantum modellerinin tasarımı ve eğitimi için üst düzey soyutlamalar sunar ve yüksek performanslı kuantum devre simülatörlerini destekler [50]. TensorFflow-Qquantum platformu kuantum sınıflandırması için denetimli öğrenme, kuantum kontrolü, gürültülü kuantum devrelerini simüle etme ve kuantum yaklaşık optimizasyonu dahil olmak üzere temel uygulamalar için kullanılabilir. Ayrıca, meta-öğrenme, katmanlı öğrenme, Hamiltonian öğrenme, termal durumları örnekleme, değişken kuantum öz vektör çözücüler, kuantum faz geçişlerinin sınıflandırılması, üretken rakip ağlar ve pekiştirici öğrenme dahil olmak üzere gelişmiş kuantum öğrenme görevlerini de destekler. Bu kütüphanenin, kuantum hesaplama ve makine öğrenimi araştırma topluluklarının hem doğal hem de yapay kuantum sistemlerinin modellerini keşfetmesi ve nihayetinde potansiyel olarak bir kuantum avantajı sağlayabilecek yeni kuantum algoritmaları keşfetmesi için gerekli araçları sağladığı düşünülmektedir.

Bir diğer platform IBM'in kuantum servisinin sağladığı simülatörler ve QuantumLab platformudur. 2021 yılında, IBM Quantum 127 kübitlik bir kuantum işlemcisi olan Eagle'ı tanıtmıştır [51]. Eagle, haritası çıkarılmamış hesaplama alanlarını keşfetmeyi sağlayan ve 100 kübitin ötesindeki kuantum hesaplamaları gerçekleştirmeyi sağlayan bir kuantum işlemcisidir. IBM-Quantum, kuantum işlemcileri büyüdükçe her bir ek kübit, klasik bilgisayarların kuantum devrelerini güvenilir bir şekilde simüle etmesi için alan karmaşıklığını azaltır ve algoritmaları yürütmek için bellek alanını iki katına çıkarır. IBM Quantum servisleri farklı kübit sayısında çeşitli kuantum cihaz simülatörlerine çevrim içi bağlanarak QuantumLab ortamında Python 3 ile uygulamalar geliştirmeyi sağlar. QuantumLab Qiskit 0.34.2 versiyonunu desteklemektedir.

4.5. SONUÇ VE DEĞERLENDİRMELER

Günümüzde kuantum teknolojisi ve derin öğrenme alanındaki gelişmeler kuantum derin öğrenmeyi popüler bir çalışma alanı haline getirmiştir. Derin öğrenme algoritmaları makine öğrenimi ve yapay zekâ çalışma alanlarından ortaya çıkmış ve içerisinde birçok gizli katman bulunduran yapay sinir ağlarını içermektedir. Günümüzde kuantum bilgisayarlar, klasik bilgisayarlarda çözülemeyecek karmaşık problemleri çözmek amacıyla derin öğrenme algoritmalarını zaman karmaşıklığı ve performans açısından daha ileriye taşıyacak çalışmalar konusunda umut vadetmektedir. Kuantum algoritmalarının klasik versiyonlarına kıyasla avantajlarını çalışma zamanı karmaşıklığı, örnek karmaşıklığı, gürültüye karşı dayanıklılık ve model karmaşıklığı açısından ele almak mümkündür. Kuantum algoritmaları, klasik yöntemlere kıyasla daha hızlı hesaplamalar sağlar. Ayrıca bozuk veri örnekleri üzerinde kuantum algoritmalarının klasik yöntemlere kıyasla veri sayısını arttırarak daha iyi öğrenme gerçekleştirdiği gözlemlenmiştir. Kuantum algoritmalarının karmaşık modeller üzerinde meydana gelen aşırı uyum sorununu çözmek için de umut verici bir yaklaşım olduğu düşünülmektedir. Tüm bu avantajlara rağmen kuantum derin öğrenme alanında algılayıcı, mimari, optimizasyon algoritmaları ve kayıp fonksiyonları konusunda doğru bir genelleştirme yapmak en büyük zorluklardan biridir. Günümüzde büyük ölçekli genel amaçlı kuantum bilgisayarlar henüz mevcut olmadığından, verimli kuantum algoritmaları genel olarak araştırmacılar tarafından kuantum bilgisayarların

simüle edildiği platformlarda araştırılmaktadır. Bu nedenle kuantum alanındaki uygulamalar bu platformların kısıtlarına bağlı olarak gerçekleştirilmektedir. Son zamanlarda kuantum sinir ağlarının ses tanıma, görüntü tanıma ve finans alanlarında gerçek dünya problemleri üzerindeki bazı uygulamaları bulunmaktadır. Ancak gelecekte kuantum bilgisayarların daha dada geliştirilmesi ve yaygınlaştırılmasıylae birlikte zaman karmaşıklığı ve performans açısından iyileştirmelerin daha net gözlemlenebilir olduğu sonuçlar elde edilmesi beklenmektedir. Böylece daha verimli kuantum algoritmalarının bulunması ve gerçek dünya problemleri üzerinde uygulanılması hedeflenmektedir.

KAYNAKLAR

[1] P. Ongsulee, "Artificial intelligence, machine learning and deep learning,"IEEE *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, Nov. 2017, doi: 10.1109/ICTKE.2017.8259629.

[2] L. W. Hosch, "Machine Learning" *Encyclopedia Britannica*, britannica.com. http://www.britannica.com/EBchecked/topic/1116194/machinelearning. (Erişim Tarihi: 2017-06-01)

[3] F. Provost, and R. Kohavi. "Glossary of terms."*Journal of Machine Learning* 30.2-3: 271-274, 1998.

[4] N. Wernick et al, "Machine learning in medical imaging," *IEEE signal processing magazine* 27.4: 25-38, July 2010, doi: 10.1148/rg.2017160130 .

[5] R. Collobert, "Deep learning for efficient discriminative parsing," *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, JMLR Workshop and Conference Proceedings, Jun. 2011.

[6] L. Gomes, "Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts," IEEE Spectrum. from http://spectrum.ieee.org/robotics/ artificial-intelligence/machinelearning-maestro-michael-jord, Oct. 2014.

[7] A. Montanaro, "Quantum algorithms: an overview."*NPJ Quantum Information* 2.1: 1-8, Jan, 2016, doi: 10.1038/npjqi.2015.23.

[8] C. H. Bennett, and D. P. DiVincenzo, "Quantum information and computation," *NATURE* 404.6775: 247-255, 2000.

[9] P. R. Feynman, "Simulating physics with computers,"*Feynman and computation*. CRC Press, 133-153, 2018.

[10] K. Zhang et al. "Toward trainability of quantum neural networks," *arXiv preprint arXiv:2011.06258,* 2020.

[11] L. Lamontagne, and M. Mario, "Advances in Artificial Intelligence," *19th Conference of the Canadian Society for Computational Studies of Intelligence, Canadian AI 2006, Quebec City, Quebec, Canada, June 7-9, Proceedings*. Vol. 4013. Springer Science & Business Media, 2006.

[12] L. Alchieri et al. "An introduction to quantum machine learning: from quantum logic to quantum deep learning." *Quantum Machine Intelligence* 3.2: 1-30, 2021.

[13] S. Boixo et al. "Characterizing quantum supremacy in near-term devices." *Nature Physics* 14.6: 595-600, 2018.

[14] F. Arute et al. "Quantum supremacy using a programmable superconducting processor." *Nature* 574.7779: 505-510, 2019.

[15] S. U. Khan, J. A. Ahsan, and G. Vall-Llosera, "K-means clustering on noisy intermediate scale quantum computers." *arXiv preprint arXiv:1909.12183,* 2019.

[16] J. Heredge et al. "Quantum support vector machines for continuum suppression in B meson decays." *Computing and Software for Big Science* 5.1: 1-9, 2021.

[17] S. Lloyd, M. Mohseni, and P. Rebentrost. "Quantum principal component analysis." *Nature Physics* 10.9: 631-633, 2014.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, 521(7553):436–444, 2015. doi: 10.1038/nature14539.

[19] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation," In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431– 3440, 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition," In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[21] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Van- houcke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

[22] A. Vaswani et al. "Attention is all you need," In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc., 2017.

[23] D. Silver, et al. "Mastering the game of go with deep neural networks and tree search," *Nature*, 529:484–489, Jan. 2016. doi: 10.1038/nature16961.

[24] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," IEEE international conference on acoustics, speech and signal processing, pp. 8599-8603, 2013, doi: 10.1109/ICASSP.2013.6639344.

[25] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM J. Comput.*, 26(5):1484–1509, Oct. 1997. doi: 10.1137/S0097539795293172.

[26] S. Takayuki. "Application of deep quantum neural networks to finance," *arXiv preprint arXiv:2011.07319,* 2020.

[27] Y., Dong, M. Wu, and J. Zhang. "Recognition of pneumonia image based on improved quantum neural network." *IEEE Access* 8: 224500-224512, 2020.

[28] F. Li, and G. Xu, "Quantum BP neural network for speech enhancement," *Asia-Pacific Conference on Computational Intelligence and Industrial Applications, PACIIA,* Vol. 2, pp. 389-392, IEEE. Nov. 2009.

[29] F. Edward, and H. Neven. "Classification with quantum neural networks on near term processors," *arXiv preprint arXiv:1802.06002,* 2018.

[30] J. Allcock, C. Hsieh, I. Kerenidis, and S. Zhang, "Quantum algorithms for feed-forward neural networks," *CoRR*, abs/1812.03089, 2018.

[31] M. Dobson, G. S. Uehara, and A. Spanias, "Quantum Neural Network Benchmarking with MNIST Dataset".

[32] C. Bu-Qing, and X. Niu. "A novel neural network based on quantum computing," *International Journal of Theoretical Physics* 59.7: 2029-2043, 2020.

[33] S. Ashkan, and A. A. Ghavifekr. "Quantum Neural Networks (QNN) Application in Weather Prediction of Smart Grids," *2021 11th Smart Grid Conference (SGC)*. IEEE, 2021.

[34] S. Kinshuk, and P. R. Srivastava. "Quantum algorithm for quicker clinical prognostic analysis: an application and experimental study using CT scan images of COVID-19 patients," *BMC Medical Informatics and Decision Making* 21.1:1-14, 2021.

[35] I. Cong, S. Choi, and M. D. Lukin, "Quantum convolutional neural net- works," *Nature Physics*, 15(12):1273–1278, Aug. 2019, doi: 10.1038/ s41567-019-0648-8.

[36] X. Zhang, et al. "PolSAR Image Classification via Learned Superpixels and QCNN Integrating Color Features." *Remote Sensing* 11.15: 1831, 2019.

[37] S. Oh, C. Jaeho, and K. Joongheon, "A tutorial on quantum convolutional neural networks (QCNN)," *2020 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2020.

[38] V. Rajesh, and U. P. Naik, "Quantum Convolutional Neural Networks (QCNN) Using Deep Learning for Computer Vision Applications," In *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT),* pp. 728-734, IEEE, Aug. 2021.

[39] Y. Cao, G. G. Guerreschi, A. Aspuru-Guzik, "Quantum neuron: an elementary building block for machine learning on quantum computers," *arXiv preprint arXiv:1711.11240*, 2017.

[40] M. Hibat-Allah et al. "Recurrent neural network wavefunctions," *Physical Review Research, 2(2),* 023358*,* 2020.

[41] C. Roth. "Iterative retraining of quantum spin models using recurrent neural networks," *arXiv preprint arXiv:2003.06228,* 2020.

[42] J. Bausch, "Recurrent quantum neural networks." *Advances in neural information processing systems* 33: 1368-1379, 2020.

[43] S. M. Taha, and Z. K. Taha, "EEG signals classification based on autoregressive and inherently quantum recurrent neural network." *International Journal of Computer Applications in Technology* 58.4: 340-351, 2018.

[44] I. Goodfellow et al. "Generative adversarial nets." *Advances in neural information processing systems* 27, 2014.

[45] P. Dallaire-Demers, and K. Nathan, "Quantum generative adversarial networks." *Physical Review A* 98.1: 012324, 2018.

[46] H. He-Liang et al. "Experimental quantum generative adversarial networks for image generation." *Physical Review Applied* 16.2: 024051, 2021.

[47] S. Haozhen et al. "Quantum generative adversarial network for generating discrete distribution." *Information Sciences* 538: 193-208,2020.

[48] C. Zoufal, L. Aurélien, and S. Woerner, "Quantum generative adversarial networks for learning and loading random distributions." *npj Quantum Information* 5.1: 1-9, 2019.

[49] V. Bergholm, et al. "Pennylane: Automatic differentiation of hybrid quantum-classical computations." *arXiv preprint arXiv:1811.04968,* 2018.

[50] M. Broughton, et al. "Tensorflow quantum: A software framework for quantum machine learning." *arXiv preprint arXiv:2003.02989,* 2020.

[51] IBM-Quantum, "IBM Quantum breaks the 100-qubit processor barrier," ibm.com. https://research.ibm.com/blog/127-qubit-quantum-processor-eagle, (Erişim Tarihi: 2022-01-01)

Bölüm 5

ROBOT İŞLETİM SİSTEMİNDE (RİS) YAPAY ZEKÂ TEMELLİ ANORMAL DURUM TESPİT YAKLAŞIMLARI

Hande Çavşi Zaim - Esra Nergis Yolaçan

Bu kitap bölümünde; Robot İşletim Sistemi (RİS) mimarisi, RİS'te görülen anormallikler, yapay zekâda kullanılan anormallik algılama yöntemleri ve literatürde robotik sistemler üzerinde uygulanan anormallik algılama sistemleri incelenmiş, robotik sistemlerin kullanıldığı alanlar ve robotik sistemlerde anormallik algılama yöntemlerinin önemi vurgulanmış, RİS'lerin güvenliği gözden geçirilmiş ve yapılan saldırılar incelenmiş ve çözüm önerileri sunulmuştur.

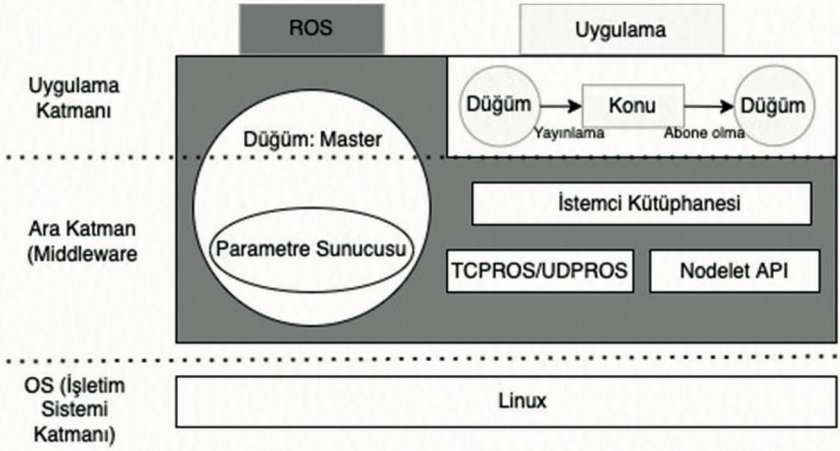
5.1. GİRİŞ

Yapay Zekâ, bilgisayarların zekâ sergilemek için hangi bilgilere ihtiyaç duyduğu, bu bilgileri nasıl temsil ettiği ve kullandığı sorularına cevap vermeyi amaçlar [1]. Robotlar ise muhakeme, algı veya eylem (motor kontrolü) için mekanik efektörleri, sensörleri ve bilgisayarları birleştirir. Bu durumda yapay zekâ, robotların her bileşenine önemli katkı sağlar. Robotik, yapay zekâyı gerçek dünya problemleri üzerinde uygulamayı sağlayan bir araştırma alanıdır. Bu nedenle yapay zekâ ve robotik birbirlerinin bir parçası olarak düşünülebilir. Robot İşletim Sistemi (RİS), modüler robotik sistemlerin geliştirilmesi için fiili bir standart olarak görülür. Geleneksel süreç yönetimi ve zamanlama anlamında bilinen işletim sistemlerinin aksine robot yazılımları yazmak için esnek bir yapı (framework) olarak üretilmiştir [2]. RİS, ilk olarak 2007 yılında Stanford Yapay Zekâ Laboratuvarı tarafından geliştirilmiştir. 2008 yılı itibariyle Willow Garage tarafından geliştirilmeye devam etmiştir ve 2013 yılından itibaren geliştirme ve bakımı Open Robotics tarafından yapılmaktadır [3]. Anormallik tespiti verilerde beklenen davranışa uymayan kalıpları bulma problemini ifade eder [4]. Robotlarda anormal veriler robotu istenmeyen şekilde eyleme geçirebilir. Klasik bilgi sistemlerinde herhangi bir anormal veri sistemi bozduğunda ilgili hizmet çevrim dışı duruma getirilebilir. Ayrıca bu sistemlerde kaybolan veriler yedeklerden geri yüklenebilir. Ancakdrone, insansız hava araçları, hidrolik barajlar, üretim robotları söz konusu olduğunda bu sistemleri kapatmak geçerli bir çözüm olmaz. Robotlar güvenli bir duruma geçiş sağlayana kadar kullanılabilir olmalıdır [5]. Robotların önemli ve kritik zararlar vermesine engel olmak için anormallik tespiti uygulamaları oldukça önemlidir. Günümüzde robotların endüstri, sağlık, tarım, askeri ve sivil alanlarda sıklıkla kullanılması robotlarda anormallik tespitini önemli bir araştırma alanı haline getirmiştir [6].

Bu kitap bölümde; RİS mimarisi, robotlarda görülen anormallikler, yapay zekâda uygulanan farklı anormallik tespiti yaklaşımları ve literatürde robotlar üzerinde uygulanan anormallik tespiti yaklaşımları incelenmiştir. Çalışmanın robotlarda anomali tespitinin önemi vurgulanarak bu alanda çalışma yapmak isteyen araştırmacılara bir rehber niteliğinde olması amaçlanmıştır.

5.2. ROBOT İŞLETİM SİSTEM (RİS) MİMARİSİ

RİS çalışma mantığı; düğümler, mesajlar, konular ve hizmetlere bağlı olarak işler [7]. RİS üzerindeki hesaplamalar düğümler tarafından gerçekleştirilmektedir. Bu bağlamda düğümlere RİS yazılım modülleri de denilebilir. Düğümler birbirleriyle mesajlar aracılığıyla iletişim kurmaktadır. Bir mesaj integer, float, boolean gibi ilkel veri türlerini destekleyen veri yapılarını ifade eder. Her bir düğüm "odometri", "harita" gibi bir diziden oluşan belirli konular hakkında mesajlar yayınlar. Mesaj yayınlayan düğüm yayıncı düğüm bir konudan gelen mesajları alan düğüm abone düğüm olarak adlandırılır. Tek bir konu için birden fazla eş zamanlı çalışan abone ve yayıncı düğüm olabilir ve tek bir düğüm birden fazla konuda mesaj yayınlayabilir. RİS üzerinde çalışan abone ve yayıncı düğümler genel olarak birbirinden habersizdir. Yayınlama-abone olma işlemlerine biri istek biri yanıt olmak üzere dizi (string) veri tipine sahip hizmetler denir. Bu hizmetler web hizmetlerinde bulunan ve URL'ler tarafından tanımlanan istek ve yanıt hizmetlerine benzemektedir. RİS, Linux tabanlı işletim sistemlerinin desteklediği TCP/RİS - UDP/RİS veri taşıma protokollerini kullanır. Tüm düğümlerin çalışması RİS-Master olarak adlandırılan ana düğümün çalışmasına bağlıdır. RİS'in, Linux işletim sistemi üzerine kurulu katman mimarisi Şekil 5.1'deki gibidir.

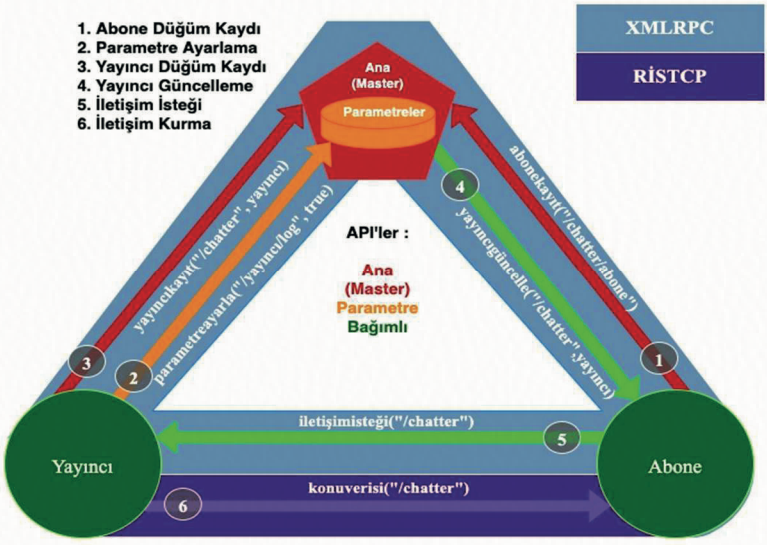


Şekil 5.1. Robot İşletim Sistemi Mimarisi.

RİS, endüstriyel üretim robotlarında, sağlık sektöründeki ameliyat robotlarında, tarım sektöründe toprak gübrelemesi, haşere kontrolü, toprak yönetimi, hasat ve nakliye robotlarında, askeri alandaki insansız hava araçlarında kullanılmaktadır.

5.3. ROBOT İŞLETİM SİSTEMİNDE (RİS) GÖRÜLEN ANORMAL DURUMLAR

RİS istemci kütüphanesi arasında paylaşılan ve RİS düğümlerini birbirine bağlı bir hesaplama grafiğinde bir araya getirmek için kullanılan ortak ve yerleşik alt sisteme API kümesi denir [8]. API'ler temel mesaj taşıma protokollerini içermektedir ve Ana API (Master API), Bağımlı API (Slave API) ve Parametre Servis API (Parameter Service API) olmak üzere üç kısımdan oluşmaktadır [9]. Şekil 5.2'de RİS API'lerin örnek bir diyagramı verilmiştir.



Şekil 5.2. RİS API'lerin örnek diyagramı [10].

Şekil 5.2'de verilen diyagram, abonelerin kayıt olma işleminden sonra yayıncıların konu sohbetini yayınladığı klasik konuşmacı/dinleyici RİS API senaryo örneğini temsil etmektedir. Bir yayıncı düğümü belirli bir konuya

yayın yapmak istediğinde ilk önce amacını XML-RPC aracılığıyla Ana API'ye bildirmelidir. Ana API verileri belirli bir konuya yayınlama niyetini belirten mesajı aldıktan sonra zaten bu konuya abone olan aktif bir düğüm olup olmadığını kontrol eder. Eğer aktif bir düğüm mevcutsa bu abone düğüme yeni bir yayıncı düğümü olduğunu Bağımlı API aracılığıyla bildirir. Yeni bir abone düğümü de yeni bir konuya abone olma niyetini önce Ana API'ye XML-RPC aracılığıyla bildirmelidir. Eğer bu konunun aktif bir yayıncısı mevcutsa Ana API bu durumu Bağımlı API aracılığıyla ilgili abone düğümüne liste şeklinde iletir. Abone daha sonra yayıncıların her biri ile iletişim protokolünün müzakeresini başlatacak ve başarılı bir anlaşmadan sonra yayınlanan bilgileri almaya başlayabilecektir. RİS düğümleri arasındaki iletişimde mesajları taşıyan yaygın protokol TCP/IP soketlerinden oluşan RİS-TCP'dir. Gelen bağlantılar masaj türü ve konunun adı gibi belirli alanları içeren bir bağlantı başlığı belirtmelidir.

Ana API, Parametre API ve Bağımlı API, Ana düğüm ile diğer düğüm arasında olan olayları yansıtır ve durum bilgisi bulundurmayan http tabanlı bir uzaktan yordam (remote) protokolü olan XML-RPC aracılığıyla uygulanır. RİS'te bulunan XML-RPC'nin uygulanması çok dilli istemci kitaplıklarının işini kolaylaştırsa da bazı dezavantajlara sebep olmuştur. İstemcilerin yetkilendirme amacıyla tanımlanması https güvenlik yöntemleri kullanılarak gerçekleştirilebilirken, RİS henüz temel erişim kontrolünü uygulamak için gereken doğrulama özelliklerini desteklememektedir [11]. Bu durum API'nin tamamını istenmeyen ve kötü niyetli değişikliklere karşı savunmasız hale getirmektedir. Her XML-RPC API çağrısı içerisinde çağrı ID'si, yöntemin tamamlanma koşulunu gösteren bir tamsayı, bir durum mesajı, hata ayıklamaları için durum bilgisi açıklayan ve anlaşılması son derece kolay olan bir dizi ve bazı veri türlerinin yanıt değeri bilgileri döndürülmektedir. Bu parametre değerlerine, RİS'in temel erişim kontrolünü uygulamak için gereken doğrulama özelliklerini desteklememesinden dolayı ulaşılması çok kolaydır.

Kötü niyetli kişiler RİS'in temel erişim kontrolü zafiyetini sömürerek mesaj verileri üzerinde değişiklikler yapabilir. Bu değiştirilmiş anormal veriler endüstriyel robotların normal çalışmasını ve performansını etkiler. Endüstriyel süreçleri aksatarak finansal kayıplara neden olur. Sistem tıkanması, veri kesintisi, fiziksel hasar gibi sonuçlara neden olabilir. Buna ek olarak robotik siste-

me uzaktan erişim noktaları olan kablosuz iletişim portlarının kullanılmadığı zamanlarda açık bırakılması robotik sistemi saldırganlara karşı savunmasız bir hale getirir. Bu durumda saldırganlar robotik sistemlere uzaktan erişim sağlayarak anormal veriler enjekte edebilir [12]. Robotlarda güvenli olmayan yedekleme ve veri depolama işlemleri veri kaybına veya bozulmasına neden olabilir. Bu şekilde oluşan anormal veriler endüstriyel kuruluşların güvenli bir şekilde çalışma yeteneğini zayıflatır ve robotik sistemlerin performansını etkiler. Robotik sistemlere saldırıldığında meydana gelen sistem arızaları, güç kaybı ve operasyonel kullanım eksikliğine neden olur. Bu sistem arızalarından bazıları aşırı pil tüketimi ve sınırlı kaynaklar nedeniyle pil gücünün bitmesi gibi sorunlardır. Bu durumlarda robotik sistemin aktivite eşiğinin olmaması, robotların anormal ve zararlı faaliyetler gerçekleştirmesine neden olur. Sonuç olarak operatörlerin hayatı riske girebilir ve operasyonel güvenlik prosedürleri etkilenebilir. Robotlarda anormal verilerin oluşması robotun devre dışı kalmasına ve maddi kayıplara neden olan üretim kesintilerine neden olur. Anormallik tespiti için gerekli olan anormal davranış verilerini toplamak oldukça zorlu bir görevdir. Robotların çalışma esnasında toplanan veriler çoğunlukla normal davranış verilerinden oluşur. Greenize isimli otonom çim biçme makinesinde bu anormal verileri kolaylıkla toplamayı sağlayan bir anormal veri alma sistemi mevcuttur [13]. Bu sistem sayesinde bir RİS geliştiricisi veya destek mühendisi RİSbag içerisindeki tüm verileri kolaylıkla çekebilir ve bu veriler rsync aracılığıyla bulut ortamına yüklenir. Daha sonra kullanıcı, verileri bir tarayıcı veya masaüstü üzerinde Foxglove Studio aracılığıyla görüntüleyebilir ve kullanabilir.

5.4. YAPAY ZEKÂ TABANLI ANORMALLİK TESPİTİ YAKLAŞIMLARI

Anormallik tespiti için genellikle model tabanlı ve veriye odaklı teknikler kullanılır [14-18]. Model tabanlı yaklaşımlar anomali tespiti için ölçülen ve tahmin edilen değerler arasındaki sapmaları kullanır ve sistemlerin ön bilgisine ihtiyaç duyar. Doğru bir model elde etmek için ürün bilgileri, sistem konfigürasyonu ve sistem durumları gibi alan bilgilerinin net olarak bilinmesi gerekir. Bu bilgilerin çıkarılması oldukça zaman alan bir iştir. Veri odaklı

teknikler model tabanlı tekniklere göre daha az sınırlama içerir. Burada sistemlerin ön bilgisine ihtiyaç duyulmaz ve anormallikler tamamen toplanan verilerden tespit edilebilir. Bu veri odaklı yaklaşımların anormallik tespitlerinde daha çok tercih edilmesine sebep olur [19]. Veri odaklı yaklaşımlar Şekil 5.3'de görüldüğü gibi sınıflandırılabilir. İstatistiksel yöntemlerde sistemin geçmiş verileri toplanır ve bu verilerin dağılımından sapan veri noktaları anormallik olarak tanımlanır. Bu yöntem anormallik tespiti için tüm geçmiş verilerin toplanmasını gerektirir. Gerçek zamanlı anormallik tespiti için uygun bir yöntem değildir. Sinyal analiz yöntemlerinde sistemlerin ek sensörlerinden alınan sinyal verilerine Fourier ve Wavelet gibi integral dönüşümleri uygulanarak özellik çıkarma işlemi gerçekleştirilir. Bu yöntemde yalnızca bir boyutlu sinyal verileri işlenir. Bu çok değişkenli sinyallerde her sinyalin ayrı ayrı işlenmesi anlamına gelmektedir.

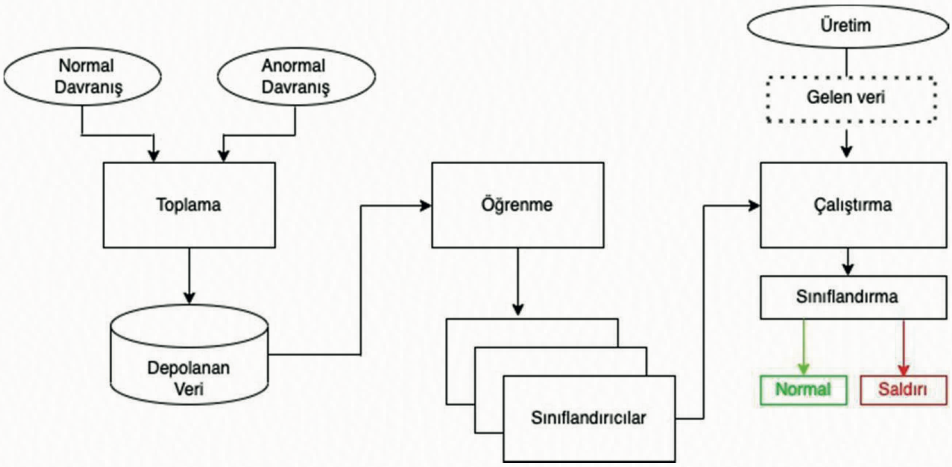
Şekil 5.3. Anormallik tespiti için kullanılan veri odaklı yaklaşımlar.

Bu yöntemdeki ek sensör kullanımı maliyet ve karmaşıklığı arttırır ve fazladan alan gerektirir. Endüstriyel ortamlar için verimli bir yöntem değildir. Makine öğrenimi tabanlı yöntemler denetimli ve denetimsiz olmak üzere iki şekilde gerçekleştirilebilir. Endüstriyel robotlar söz konusu olduğunda toplanan veriler dengesiz zaman serisi verilerinden oluşur. Bu verilerin her birinin etiketlenmesi oldukça zaman gerektiren bir işlemdir. Bu nedenle endüstriyel robotlarda denetimsiz makine öğrenimi tabanlı yaklaşımlar uygulamak daha faydalıdır.

5.5. ROBOT İŞLETİM SİSTEMİNDE UYGULANAN ANORMAL DURUM TESPİT YAKLAŞIMLARI

Robotların birçok alanda kullanılması, robotları kötü niyetli kişilerin hedefi haline getirmektedir. Günümüzde klasik sızma testi gibi yöntemler gelişen saldırı yöntemleriyle baş etmede yetersiz kalmaktadır. Anomali tespit yöntemleri, bu saldırılara aynı hızda etkin bir şekilde yanıt verilmesinde önemli rol oynamaktadır. Robotlarda bir anormallik söz konusu olduğunda robot güvenli bir duruma geçene kadar çalışır halde olmalıdır. Bu durum anormallik tespitini robotlar için daha kritik hale getirmektedir. Anormallik tespiti, bir veri kümesinin normal davranışından sapan veri noktalarını, olayları veya gözlemleri tanımlamak için kullanılır [20].

Naryanan ve ark., akıllı üretim ortamlarında RİS'e sahip robot kolları üzerinde anormallik tespiti yapan bir çerçeve önermişlerdir [21]. Anormallik tespiti için Destek Vektör Makinaları (Support Vector Machine - SVM) ve Tek Sınıf SVM makine öğrenme modelleri kullanılmıştır. Tek bir robot kolunun herhangi bir düşman saldırısı altında bulunmadığı standart yerleştirme hareketi normal davranış olarak kabul edilir. Anormallik tespit sistemi bilgi toplama, öğrenme ve çalıştırma modüllerinden oluşmaktadır. Bilgi toplama modunda RİS'e '/joint_states' konusuna abone olan düğümlerden gelen ortak durum değerleri toplanır. Ortak değer verilerinin etiket değerleri kullanıcı tarafından belirlenir. Öğrenme modunda öncelikle veriler işlenir daha sonra uygun veri kümesi kullanılarak eğitilir. Eğitimin sonunda sınıflandırıcılar depoya kaydedilir. Eğitim sonunda Gerçek Pozitif (TP), Yanlış Pozitif (FP), Gerçek Negatif (TN), Yanlış Negatif (FN), Doğruluk, Geri Çağırma (Recall), Kesinlik ve F1 puanından oluşan bir matris oluşturulur. Çalıştırma modunda kullanıcı bu değerlere bakarak en uygun sınıflandırıcıyı seçer. Gelen ortak değerler robotun çalışması sırasında toplanır ve uygun ön işleme tabi tutulur. Tahmin değerlerini elde etmek için veriler seçilen sınıflandırıcıdan geçirilir. Öngörü değeri belirtilen eşik değerinden yüksekse, sınıflandırıcı davranışı anormal kabul eder ve bir alarm verir. Eşik değeri, modelin eğitim ve doğrulama aşamasında belirlenir. Bu çalışmada geçen makine öğrenimi tabanlı anormallik algılama sistemi Şekil 5.4'de görüldüğü gibidir.



Şekil 5.4. RİS için makine öğrenimi tabanlı anormal durum algılama sistemi [20].

Utkin ve ark., robot sistem elemanlarının anormal davranışları için bir ön işleme prosedürü önermişlerdir [22]. Robotik sistemlerde anormallik tespitinin en etkili yolunun derin öğrenme yöntemleri olduğu tartışılmıştır. Bu nedenle, transfer öğrenme problemini çözmek için özel bir sinir ağı AutoEncoder modeli kullanılmıştır. AutoEncoder, modelin giriş ve çıkış verilerinin aynı olması amacıyla hata oranı geri yayılımını kullanmayan bir modeldir. AutoEncoder kullanmanın ilk amacı, veri boyutunu küçültmek ve yüksek boyutlu veri durumunda Mahalanobis mesafesini kullanarak sorunu çözmektir. Ancak AutoEncoder sadece veri boyutunu küçültmek için değil, aynı zamanda transfer öğrenme yöntemini uygulamak için de kullanılmıştır. AutoEncoder kullanmanın ikinci amacı; Mahalanobis mesafesi için karar eşiğini tanımlamaktır. Kaynak veri kümesinden Otomatik Kodlayıcı kullanılarak yeni bir hedef veri kümesi oluşturulur. Elde edilen hedef veriler yalnızca robotun normal davranışına karşılık gelen durumlardan oluşur. Kaynak veri kümesi, operatörler tarafından sınıflandırılan hem normal hem de anormal veri kümelerini içerir. Yeniden oluşturulan normal ve anormal kaynak verileri arasındaki uzaklık Mahalanobis uzaklık yöntemi ile hesaplanmıştır. Böylece anormal davranış tespit edilmiştir.

Vuong ve ark., robotik sistemlerde DOS ve komut enjeksiyon saldırılarını hızlı ve doğru bir şekilde tespit eden bir mekanizma önermişlerdir [23]. Bu mekanizmada bilgi tabanlı bir makine öğrenme algoritması kullanılmış-

tır. Test ortamı olarak, Linux işletim sistemi üzerinde çalışan yerleşik bir Intel Atom bilgisayarı tarafından kontrol edilen 4 tekerlek çekişli bir robot kullanılmıştır. Robotun uzaktan kumandası, bir TCP soketi üzerinden alınan komutları Ethernet kablosu veya WiFi kullanarak robot kontrol paneline iletebilir. Bilgi tabanlı makine öğrenmesi için Karar Ağacı C5.0 kullanılarak sınıflandırma yapılmıştır. İlk aşamada 4 tanesi iletişim ve işleme, diğer dördü ise robotun fiziksel özellikleri olmak üzere sekiz girdi özelliği için veri toplanmıştır. Her özellik, bir döngüde belirli bir zaman dilimi içinde kaydedilir. Bu değerler, alınan ve iletilen ağ trafiği oranlarını, CPU kullanımını, dosyaların diske yazılma oranlarını, tekerlek hızını, kasa titreşimini ve güç tüketimini içerir ve çevrim dışı olarak izlenir. Toplanan veri kümesi, her bir özellik için 40.454 veri noktasına sahiptir. Verilerin %30'u eğitim, %70'i test için ayrılmıştır. Öznitelik seçimi aşamasında tüm veri öznitelikleri 1. Küme olarak, 4 siber öznitelik 2. Set olarak ve 4 fiziksel öznitelik 3. Set olarak seçilmiştir. Karar Ağacı C5.0 modeli, her bir özellik kümesine uygulanır. Model, test verileri üzerinde oluşturulan kural setinin doğruluğu ölçülerek değerlendirilmiştir. Sınıflandırmanın doğruluğu, Receiver Operating Characteristic (ROC) ve Area Under the Curve (AUC) metrikleri ile değerlendirilmiştir.

Hsiao ve arkadaşları, açık kaynaklı Micro Aerial Vehicles (MAV'ler) için güvenilirlik analiz çerçevesi olan MAVFI'yi sunmuşlardır [24]. Ayrıca MAV sisteminin güvenilirliğini artırmak için anormallik algılama ve kurtarma şeması önermişlerdir. Anormallik algılama sisteminde geçici arıza tespiti için hem Gaussian tabanlı hem de AutoEncoder (AE) tabanlı teknikler uygulanmıştır. Anormallik algılama sistemi bir kapalı döngü sistemi oluşturan "Algı-Planlama-Kontrol (PPC)" hattı ile entegre olarak çalışır. PPC hattından izlenen çapraz aşama değişkenleri, anormallik tespit bloğuna gönderilmeden önce veri ön işleme bloğunda işlenir. Veri ön işlemede, veri biçimi dönüştürme ve delta hesaplama dahil olmak üzere iki adım vardır. İlk olarak, veri formatı dönüşümü için, float64 değişkenlerinin işaret ve üs bitleri 16 bitlik tamsayı değişkenlerine dönüştürülür ve çift kesinlikli kayan nokta işlemleri 16 bitlik tamsayı işlemlerine indirgenir. İkinci olarak, gelen değişkenlerin deltaları hesaplanır. Çoğu değişken için değer, tek tip dağılıma veya Gauss dağılımına sahip olabilir. Veri ön işlemesinden sonra, işlenen değişkenler önerilen anormallik tespit tekniklerinden birine girer. Anormal

bir davranış, yani veri bozulması algılandığında, ilgili aşamanın yeniden hesaplanmasını tetikleyen bir alarm sinyali üretilir.

Ahn & Hyojung, gerçek uçuş verilerini kullanarak sürü drone uçuşlarındaki anormallikleri gösteren ve her bir drone'un anormallik algılama modelinin gerçek doğrulamasını sağlayan bir model tabanlı anormallik algılama sistemi önermişlerdir [25]. Bu anormallik algılama sisteminde drone'ların normal davranış verileri ile bir eğitim modeli oluşturulmuştur. Anormal veriler test aşamasında kullanılmıştır. Eğitim için derin sinir ağları (DNN) kullanılmıştır. Sonuç olarak üç vakanın tanı sonuçları türetilir ve geçerlilik için analiz edilir. Önerilen anormallik algılama yöntemi, sürü drone uçuş operasyonlarının güvenilirliğini artırmak için kullanılabilir ve sürü drone uçuşlarının anormal davranışlarını anında tespit edebilen bir makine öğrenimi tabanlı çerçeve ile entegre edilebilir.

Bahavan ve arkadaşları, ön kamera videosu ve IMU okumalarını kullanarak heterojen otonom sistemlerdeki anormallikleri tespit etmek amacıyla kendi kendini denetleyen bir derin öğrenme tabanlı anormallik tespiti önermişlerdir [26]. Koşullu bir GAN (Generative Adversarial Networks) kullanan görüntü tabanlı sistem, anlık geçen üç kareyi analiz eder ve bir sonraki kareyi tahmin etmeye çalışır. Tahmin hatası ve bir eşik değeri kullanılarak tahmin edilen farkın derecesine göre anormal bir durum veya normal bir durum olarak sınıflandırma yapar. IMU tabanlı sistem, zaman damgalarını sınıflandırmak için üç ardışık IMU vektörünü yeniden yapılandıran bir Long Short Term Memory (LSTM) otomatik kodlayıcı ve önceki üç IMU vektörünü kullanarak bir sonraki vektörü tahmin etmek için kullanılan bir LSTM tahmincisi olmak üzere iki farklı yaklaşım kullanır. Yeniden oluşturma hatasına, tahmin hatasına ve bir eşik değerine bağlı olarak, zaman damgası anormal bir durum veya normal bir durum olarak sınıflandırılır. Yeniden yapılandırma tabanlı yaklaşım, normal örnekleri otomatik olarak kodlamayı öğrenir, dolayısıyla anormal bir girdi örneği için yeniden yapılandırma hatasının artması beklenir. Tahmine dayalı yaklaşım, önceki sıralı normal veri örneklerini verilen bir sonraki örneği tahmin etmek için öğrenir. Bu nedenle girdiler anormal ise yüksek bir tahmin hatası döndürmesi beklenir.

Literatürde robotik sistemlerde gerçekleştirilen anormallik algılama sistemleri Tablo 5.1'de görüldüğü gibi özetlenir.

Tablo 5.1. Anormallik algılama sistemleri

Kaynak	Yöntem	Uygulama Alanı	Veri Kaynağı	Algoritma
Vuong ve ark. (2015)	Makine öğrenimi tabanlı anormallik tespiti	Robotik Araç	Ağ trafiği, CPU, Disk, Kontrol Unitesi, WattsUp	Decision Tree C5.0
Utkin ve ark. (2016)	Derin öğrenme tabanlı anormallik tespiti	Robot Sistem	Sensör	AutoEncoder
Narayanan ve ark. (2018)	Öğrenme tabanlı anormallik tespiti	Robot Kol	Sensör	SVM & One-Class SVM
Ahn & Hyo-jung (2020)	Derin öğrenme tabanlı anormallik tespiti	Hava Aracı	Sensör	ANN
Bahavan ve ark. (2020)	Derin öğrenme tabanlı anormallik tespiti	Otonom Sistem	Kamera ve Sensör	LSTM, GAN
Hsiao ve ark. (2021)	İstatiksel ve derin öğrenme tabanlı anormallik tespiti	Hava Aracı	Sensör	Gaussian, AE

5.6. SONUÇ VE DEĞERLENDİRMELER

Bu bölüm incelenen ve özet olarak sunulan çalışmalarla, robotlarda anormallik tespitinin önemini anlatan, yapılan değerlendirmeler ile bu alanda çalışma yapmak isteyen araştırmacıları bilgilendirme mevcut sorunlar ve çözüm önerileri kapsayan bir bölüm olarak sunulmuştur. Bu kitap bölümünde;

- RİS mimarisi, RİS'te görülen anormallikler, yapay zekâ temelli kullanılan anormallik algılama yöntemleri ve literatürde robotik sistemler üzerinde uygulanan anormallik algılama sistemleri incelenmiş,
- RİS'ler, robotik sistemlerin kullanıldığı alanlar ve robotik sistemlerde anormallik algılama yöntemlerinin genellikle bu sistemlerin güvenlik düşünülerek tasarlanan bir işletim sistemine sahip olmadığı, birçok saldırı türüne karşı savunmasız oldukları, bunun sonucunda oluşan anormalliklerin ciddi hasarlara neden olduğu görülmüş,
- RİS'in XMLRPC API'si kimlik doğrulama veya yetkilendirme ile güvence altına alınmadığından ve abone düğüm yayıncı güncelle prosedürünün geçerli bir istemci tarafından çağrılıp çağrılmadığı kontrol edilmediğinden, bir saldırgan her düğüm ve konu bilgisi için

iletişim parametrelerinin yanı sıra RİS grafiğini elde etmek için yöneticiye (master) yapılan çağrıları kullanabileceği anlaşılmış,

- RİS güvenliğini arttırmak için ana bilgisayarın bir RİS yöneticisinin ana bilgisayarı olduğu kontrol edilmesi, ancak bu kontrol uygun bir kimlik doğrulama ve yetkilendirme kullanılarak gerçekleştirilmesi gerektiği raporlanmış,
- Bir dizi yayıncının değiştirilmesi gerektiğinde abone düğüm yayıncı hakkında yöneticiyi (Master) bilgilendirmediği için abone düğümün geçerli yayıncı düğümlerinden izole edildiği anda RİS grafiğinde görülmediğinden, bir saldırganın kendini abone düğüm gibi gösterip kendini gizlemesinin mümkün olabileceği belirtilmiş, ve

  RİS'te belirli bir düğümün parametrelerinin ilgili sunucudan değiştirilmesinin kolay olduğu ve saldırganın bu parametreleri kolaylıkla değiştirebileceği vurgulanmıştır.

Bu gibi durumlar robot hareketlerinde anormal davranışlara neden olabilecektir. RİS'lerde bir anormallik olması durumunda, klasik bilgi işlem sistemlerinde olduğu gibi sistemi kapatmak etkili bir çözüm değildir. Çünkü robotik sistemlerin güvenli bir duruma ulaşana kadar çalışması gerekir. Bu durum robotik sistemlerde herhangi bir anormallik durumu olduğunda anında müdahale etmeyi gerektirir. Günümüzde robotik sistemlerin ve araçların endüstride yaygınlaşmasının yanında, neredeyse her eve girmeye başladığı görülmektedir. Robotlarda sıklıkla kullanılan işletim sistemi RİS'in bu trendin itici gücü olması beklenmektedir [27,28]. Bu sistemlere yapılan bir saldırı veya yazılımsal/donanımsal hata sonucu oluşabilecek anormalliklerin tespiti ise istenmeyen durumların önüne geçebilmek için oldukça önem kazanmıştır. Ancak yeni bir araştırma alanı olmasından dolayı, RİS özelinde gerçekleştirilen anormallik tespiti çalışmalarının kısıtlı olduğu belirlenmiştir.

Genel olarak çözümler değerlendirildiğinde;

- Penetrasyon testi gibi bilinen geleneksel güvenlik yöntemleri anormallik durumunda ihtiyaç duyulan hızda önlem almayı sağlamadığı,
- Yapay zekâ yöntemlerinin anormalliklere karşı istenen hızda cevap vermesini sağlayan çözümler sunduğu,
- Anormallik tespitinde kullanılan AE, LSTM, DNN, GAN, SVM, Karar Ağacı gibi yapay zekâ algoritmalarının üretim robotlarında,

cerrahi robotlarda, toprak gübrelemesi, haşere kontrolü, toprak yönetimi, hasat ve nakliye robotlarında, askeri alandaki insansız hava araçlarında anormallik tespiti gibi gerçek dünya problemlerini çözmek için kullanıldığı,

- Bu robotların anormal davranışlarının hızlı bir şekilde tespit edilmesi ve önlem alınması zorunluluğu sebebiyle robotik sistemlerde anormallik tespitinin kritik bir konu haline geldiği

anlaşılmaktadır. Literatürde RİS için günümüze kadar gerçekleştirilen anormallik tespiti yöntemlerine bakıldığında;

- Öğrenme tabanlı, bilgi tabanlı, sinyal işleme tabanlı, derin öğrenme tabanlı anormallik ile istatiksel ve derin öğrenme tabanlı anormallik tespit modellerinin bulunduğu,
- Yukarıda kısaca tanıltılsa da; Gaussian, AE, LSTM, DNN, GAN, SVM, One-Class SVM, Karar Ağacı gibi yöntemlerin çözüm için kullanıldığı, ve
- Veri olarak sensör verilerinin yanında ağ trafiği veya kamera görüntüsü gibi verilerin de çözümlerde tercih edildiği

görülmüştür.

- Gelecekte bu alanda yapılacak çalışmalarda;
- uygulama alanı ve anormallik tipi üzerine spesifik çözümler üretilebileceği,
- anormalliklerin sebeplerinin (saldırı/hata) tespitinin yapılabileceği,
- veri tiplerinin ve kaynaklarının çeşitlendirilebileceği,
- yapay zekâ yaklaşımları ile anormalliklerin gerçek zamanlı tespit edilebileceği,
- yeni yapay zekâ yaklaşımlarının uygulanabileceği

değerlendirilmektedir.

Teşekkür

Bu çalışma İleri Robotik Sistemler ve Mekatronik alanında YÖK 100/2000 doktora bursu kapsamında desteklenmiştir. Yazarlar, ilgili birimlere teşekkür eder.

KAYNAKLAR

[1] Brady, Michael. "Artificial intelligence and robotics." *Robotics and Artificial Intelligence*. Springer, Berlin, Heidelberg, 1984. 47-63.

[2] https://robocademy.com/2020/07/01/what-is-RİS/ , Aleena Rentin, 01/07/2020

[3] Mazzeo, G., & Staffa, M. (2020). TRİS: Protecting Humanoids RİS from Privileged Attackers. *International Journal of Social Robotics*, *12*(3), 827-841.

[4] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." *ACM computing surveys (CSUR)* 41.3 (2009): 1-58.

[5] Pasqualetti, F., Dörfler, F., & Bullo, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, *58*(11), 2715-2729.

[6] Yaacoub, Jean-Paul A., et al. "Robotics cyber security: Vulnerabilities, attacks, countermeasures, and recommendations." *International Journal of Information Security* (2021): 1-44.

[7] Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., & Ng, A. Y. (2009, May). RİS: an open-source Robot Operating System. In *ICRA workshop on open source software* (Vol. 3, No. 3.2, p.5).

[8] White, Ruffin, et al. "SRİS: Securing RİS over the wire, in the graph, and through the kernel." *arXiv preprint arXiv:1611.07060* (2016).

[9] Keppmann, Felix Leif, Maria Maleshkova, and Andreas Harth. "Building REST APIs for the Robot Operating System-Mapping Concepts, Interaction." *SALAD@ ESWC*. 2015.

[10] Dieber, Bernhard, et al. "Penetration testing RİS." *Robot operating system (RİS)*. Springer, Cham, 2020. 183-225.

[11] Jallali, N., et al. "Vulnerability to RİS-induced cell death in ageing articular cartilage: the role of antioxidant enzyme activity." *Osteoarthritis and cartilage* 13.7 (2005): 614-622.

[12] Noura, Hassan, et al. "LoRaWAN security survey: Issues, threats and possible mitigation techniques." *Internet of Things* 12 (2020): 100303.

[13] Derek Haynes, URL: https://www.greenzie.com/post/effortlessly-investigating-robot-anomalies-at-greenzie

[14] Brambilla, D., et al. "Fault detection for robot manipulators via second-order sliding modes." *IEEE Transactions on Industrial Electronics* 55.11 (2008): 3954-3963.

[15] Capisani, L. M., et al. "Manipulator fault diagnosis via higher order sliding-mode observers." *IEEE Transactions on industrial electronics* 59.10 (2012): 3979-3986.

[16] Hornung, R., et al. "Model-free robot anomaly detection." *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014.

[17] Jaber, A. A., and Robert Bicker. "Industrial robot fault detection based on statistical control chart." *Am. J. Eng. Applied Sci* 9 (2016): 251-263.

[18] Khalastchi, E., et al. "Online data-driven anomaly detection in autonomous robots." *Knowledge and Information Systems* 43.3 (2015): 657-688.

[19] Chen, T., et al. "Unsupervised anomaly detection of industrial robots using sliding-window convolutional variational autoencoder." *IEEE Access* 8 (2020): 47072-47081.

[20] Zenati, H., et al. "Adversarially learned anomaly detection." *2018 IEEE International conference on data mining (ICDM)*. IEEE, 2018.

[21] Narayanan, V., and Rakesh B. Bobba. "Learning based anomaly detection for industrial arm applications." *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy*. 2018.

[22] Utkin, L. V., V. S. Zaborovskii, and Sergey G. Popov. "Detection of anomalous behavior in a robot system based on deep learning elements." *Automatic Control and Computer Sciences* 50.8 (2016): 726-733.

[23] Vuong, T. P., et al. "Decision tree-based detection of denial of service and command injection attacks on robotic vehicles." *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2015.

[24] Hsiao, Y. S., et al. "Mavfi: An end-to-end fault analysis framework with anomaly detection and recovery for micro aerial vehicles." *arXiv preprint arXiv:2105.12882* (2021).

[25] Ahn, H.. "Deep learning based anomaly detection for a vehicle in swarm drone system." *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2020.

[26] Bahavan, N., et al. "Anomaly detection using deep reconstruction and forecasting for autonomous systems." *arXiv preprint arXiv:2006.14556* (2020).

[27] Sezgin. "The Third Industrial Revolution: How Lateral Power is Transforming Energy, the Economy, and the World." *Turkish Journal of Business Ethics* (2018).

[28] Gates, Bill. "A robot in every home." *Scientific American* 296.1 (2007): 58-65.

Bölüm 6

EĞİTİMDE YAPAY ZEKÂ, SANAL GERÇEKLİK VE SANAL EVREN (METAVERSE)

Betül Ersöz - Halil İbrahim Bülbül

İnternet ve ilgili teknolojilerin hızla gelişmesi toplumda çeşitli dönüşümler sağlamıştır. Bu doğrultuda, dijital dönüşüm ve buna uygun ekipmanların ilerlemesi ile çeşitli kavramlar ortaya çıkmıştır. Bu kavramlardan olan yapay zekâ (YZ), sanal-artırılmış gerçeklik (VR/AR) ve sanal evren (metaverse) kavramları teknolojinin getirdiği yenilikler arasında yer almış, toplumun ihtiyaç duyduğu eksikliklere yönelik olarak geliştirilmiştir. Bu bölüm, son yıllarda çeşitli alanlarda kullanılan ve özellikle eğitimde öğrenenleri desteklemek adına etkili olduğu görülen YZ, VR/AR ve sanal evren kavramlarının öğrenme süreçlerine etkisini incelemek amacıyla oluşturulmuştur. Bununla birlikte, sanal evrenin getirdiği olası tehditler ve siber güvenlik sorunlarına yönelik risklere değinilmiş, bu bağlamda öneriler sunulmuştur. Özetle, YZ, VR/AR ve sanal evren bileşenlerinin eğitim-öğretimde önemi, yansımaları, sunduğu olanaklar, getirdiği yenilikler ve geleceğine yönelik ön görüler yer almaktadır.

6.1. GİRİŞ

21. yüzyıl ile birlikte dijitalleşen toplum, teknolojinin çeşitli alanlarına uyum sağlayabilme potansiyeline sahiptir. Yaşamı kolaylaştırmaya yardımcı olan teknoloji alanlarında, sayısız tarihsel değişime ve dönüşüme tanık olunmaktadır. Toplumun dijitalleşmesini ve çağa uyum sağlamasında en büyük role sahip olan internet, 1990'lı yıllarda dünya çapında yaygınlaşmaya başlamış, sonrasında gelişen teknolojik araçların işlevleri de bu yönde ilerletilmiştir. Geçmişten bu güne internet ve ilgili ekipmanların gelişmesiyle toplum dijital dünya ile iç içe yaşamaya başlamıştır. Tim Barnes Lee'nin Html işaretleme dilini geliştirmesi ve World Wide Web (W3) bilgi paylaşım sistemini kurmasından sonra web teknolojileri de bu doğrultuda güncellenerek dijitalleşmeye yön vermiştir. 1990'lı yıllarda internet dünyasına adım atan web teknolojileri öncelikli olarak Web 1.0 ile statik olarak (makineden-insana tek yönlü aktarım) başlamış ve bilgi aktarımı tek yönlü yapılmıştır. 2000'li yıllarda Web 2.0'ın gelişmesi ile birlikte zenginleştirilen web platformu, kullanıcılara bilginin aktarılmasının yanında kullanıcılardan makineye bilginin aktarıldığı çift yönlü bilgi akışını da başlatmış oldu. Bu noktada Web 2.0 ile entegre olan web araçları aktif rol oynayarak toplumun sanal ortamda sosyalleşmesine olanak sağlamıştır. Sosyal medya, wikiler, bloglar ve youtube gibi medya araçları ile görsel, işitsel ve metin olarak insanların oluşturduğu veriler internet ortamına aktarılmaya başlamıştır. Toplum sanal ortamda sosyalleşirken anlamlı-anlamsız oluşan verilerini web ortamında dolaşıma sunmuştur. Verilerin her geçen gün artmasıyla 2010 yılından sonra büyük veri kavramı ortaya çıkmıştır. Büyük veriyi besleyen veriler; arama motorları, bloglar, e-postalar gibi etkileşimli ortamlardan alınarak büyük veri yığınlarını ortaya çıkarmaktadır. Web 3.0 (Semantik Web), teknolojisinin gelişmesi ile birlikte elde edilen veriler anlamlı ve işlenebilen veriler haline gelmiştir. Dijitalleşen dünyada toplum, internet bağlantılarını bilgisayar, tablet, akıllı cep telefonları gibi araçlarla kurarken, dünyada bir günde onlarca terabayt veri depolanmaktadır. Depolanan düzensiz veriler yapay zekâ ve derin öğrenme algoritmalarının gelişmesi ve işe koşulması ile anlamlı hale dönüştürülerek sağlık, endüstri, bankacılık, ulaşım, enerji gibi çeşitli alanlarda stratejik amaçlı kullanılmaktadır.

Teknolojinin son 30 yılda geldiği nokta, bundan sonra da gelişen teknolojiler ile topluma yön veren farklı uygulamalar ve gelişmelerin hızla ilerle-

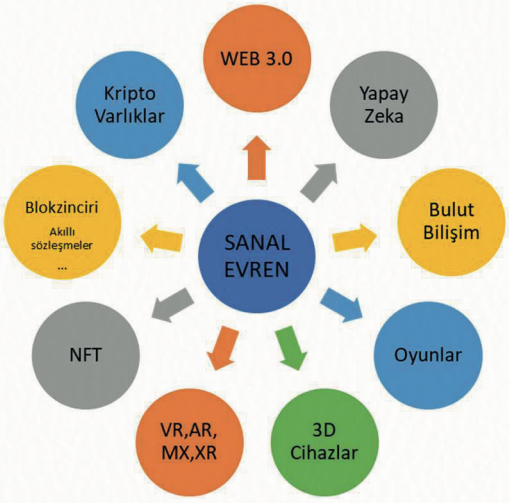
mesini mümkün kılmaktadır. Son yıllarda sanal gerçeklik (VR), artırılmış gerçeklik (AR), karma gerçeklik (MR) ve genişletilmiş gerçeklik (XR) teknolojilerinin hızla gelişmesi ile birlikte, teknoloji ile ilişkili çeşitli alanlarda sanal dünya teknolojilerinin kullanımı da giderek popüler hale gelmiştir. Toplum, 2006 yılında Second Life isimli oyunla sanal gerçeklik (VR) ile tanışmış, bu ve buna benzer oyun platformlarıyla ortaya çıkan sanal evren (metaverse), kişisel avatar oluşturma imkânı sunmuştur. 2003 yılında başlayan ve yaklaşık bir milyon düzenli kullanıcıya ulaşan Second Life ortamında kullanıcılar, kendi para birimi oluşturabilir, mülk inşa edebilir, etkinlikler düzenleyebilir ve kendi avatarlarını tasarlayabilmektedir [1]. VR, gerçek zamanlı etkileşime sahip, sürükleyici, tamamen yapay bilgisayar simülasyonlu bir görüntü ve ortam sunmuştur. Daha sonra Minecraft ve Roblox gibi sanal gerçeklik oyunları, aktif ve sayısı milyonları aşan kullanıcı profiline sahip olması ile genç neslin gözdesi olmuştur. Minecraft ve Roblox sanal alan oyunları kullanıcılara içinde oynadıkları ortamı oluşturabilecekleri ve şekillendirebilecekleri sanal evren sunmaktadır. Minecraft, ortak bir dünya oluşturmak için basit bloklar kullanan genç oyuncular için tasarlanmıştır. Roblox, oyun içinde çok sayıda oyuncunun dünyasını oluşturan daha programlanabilir bir ortam vadetmektedir [2]. Bu ortamlara bilgisayarlar, cep telefonları ve tablet gibi ekipmanlar ile bulunmanın yanı sıra sanal gerçeklik (VR) gözlükleri ve dokunsal (giyilebilir vb.) vücut ekipmanları ile gerçekçiliği yaşamak adına giriş yapılabilmektedir. Sanal evrende yaşama hissi veren kasklar, VR kulaklıkları ve çok duyulu ekipmanların yardımıyla, bu deneyim görme, ses, dokunma, hareket ve sanal nesnelerle doğal etkileşim yöntemleriyle güçlendirilmiştir [3].

2021 Aralık ayında Mark Zuckerberg'in Facebook'un adını META olarak değiştirmesi ile başlayan metaverse kavramı oldukça popüler hale gelmiştir. Meta's Reality Lab'ın baş bilimcisi Michael Abrash, avatarları foto gerçekçi hale getirecek, gerçekçi etkileşimli ortamlar yaratacak ve kullanıcıların ince el hareketleriyle girdileri kontrol etmesine izin verecek prototip teknolojisini göstermiştir [4]. Aslında sanal gerçeklik uygulamaları ile 2010 yılından sonra tanışan toplum, sanal evren olarak ta bilinen metaverse dünyası ile tanışmıştır. Sanal evren ile gerçek dünyada yaşanılan eylemler, kişisel avatarlar aracılığıyla etkileşim halinde olunarak çeşitli eylemler gerçekleştirilebildiği bir ortam haline gelmiştir. Bu kavram ile birden fazla sanal

evrenlerin oluştuğu ve bu evrenler arası geçişlerin yapılabileceği ön görülmektedir. Bu evreni oluşturan bileşenler, şeffaf ve güvenilir bir web ortamının oluşmasında ve blok zinciri sisteminde çalışan merkezi olmayan uygulamalar DApp'lerin tek bir sunucudan ziyade düğüm (node) olarak isimlendirilen binlerce bilgisayarla çalışmasını sağlamıştır. Web 3.0 ile gelen web3 cüzdanları ve sanal ekonomi, merkeziyetsiz hesapların oluşmasını sağlarken blok zincir defterlerinde dijital varlıkların korunduğu bir platform olarak yer almaktadır. Dijital paralar ile arsa alım-satım, dijital varlıklar alınması söz konusu iken dijital paralara (kripto para) rağbet artmaya başlamıştır. Merkeziyetsiz olan Bitcoin ve Etherium gibi kendi blok zincirine sahip kripto paraların yanı sıra sanal evren koinleri (coin) olarak bilinen ve oyun platformlarından adlarını alan çeşitli kripto para birimleri de sanal evrende kullanılmaktadır. İşlemlerin gerçekleştirilmesi adına kripto varlıklar ile parasal işlemlerin sürdürüldüğü bir platform oluşturulmuştur. Bununla birlikte benzersiz ve yerine tekrar konulması mümkün olmayan Nitelikli Fikri Tapular (NFT-Non Fungible Token), dijital varlıklar olarak sanal evrende yer almaktadır. NFT, sanat, müzik, oyun içi öğeler ve videolar gibi gerçek dünyadaki nesneleri temsil eden dijital bir varlıktır. Sanal evrende NFT'ler, işlemleri kaydeden dağıtılmış bir dijital defter olan blok zincirinde bulunmaktadır.

Sanal evrende, kullanıcıların verilerinden oluşan meta veri deposu sanal dünyaları bir araya getirmeyi amaçlamaktadır. Meta veri deposu, IoT (Nesnelerin İnterneti), yapay zekâ (YZ) ve makine öğrenimi (ML) aracılığıyla sanal alanlar ile gerçek dünya arasındaki etkileşimlerin yanı sıra tüm sanal ve artırılmış gerçeklik deneyimlerinin gelecekteki koleksiyonu olarak da düşünülebilinir. Öncelikli olarak oyun platformları olmak üzere çeşitli sektörlerde meta veri depolarının geliştirilen uygulamaları, eğitim için meta veri deposu kullanılmasına da olanak sağlamıştır. Dolayısıyla, eğitimde sanal evren uygulamalarının gelişiminde, uygulamaları besleyebilecek meta veri tabanı kavramlarının önemi artmaktadır. Eğitim için meta veri deposunun kullanımı, büyük ölçüde, farklı sanal alanlar arasında sorunsuz bir şekilde hareket etmek için meta veri deposundaki beklentilere de bağlı olacaktır. Örneğin, insan vücudunda iletkenliğin etkisini görmek için elektronların nasıl aktığına dair endişelenmenize gerek yoktur. Meta veri deposu, bütünsel öğrenme deneyimleri için öğrenme ortamlarında birden fazla konunun birleştirilmesi için gelişmiş beklentilere izin verebilecektir. Meta veri deposu-

nun oluşumunda, sanal evren ile sanal oluşturulan dünyalar her geçen gün gelişmekte ve şekillenmektedir. Aşağıda yer alan Şekil 6.1'de sanal evren (metaverse) ve bileşenlerine dair şekil yer almaktadır. Bu kavramların sanal evrenle bağlantısı gün geçtikçe atmaktadır.



Şekil 6.1. Sanal Evren ve Bileşenleri

6.2. EĞİTİMDE YAPAY ZEKÂ

Yapay zekâ (YZ), akıllı makinelerin özellikle akıllı bilgisayar programları yapma bilimi ve mühendisliğidir. İnsan zekâsını anlamak için bilgisayarları kullanma görevine benzer, ancak yapay zekânın kendisini biyolojik olarak gözlemlenebilir yöntemlerle sınırlaması gerekmemektedir [5]. YZ, görsel algılama, konuşma tanıma, karar verme ve çeviri dâhil olmak üzere insan zekâsı gerektiren görevleri yerine getirme becerisi, insan yeteneklerine sahip bilgisayar sistemlerinin geliştirilmesi ve kullanımına rehberlik eden teorik bir çerçevedir. Bilgisayar ve bilgi iletişim teknolojileri yıllar içinde gelişmeye devam ederek yapay zekânın gelişmesini sağlamıştır. YZ, problem çözme ve örüntü tanıma gibi insan zekâsından esinlenerek oluşturulan, farklı bilişsel problemleri çözmeyi ve ardından uyarlamayı amaçlayan bilgisayar biliminde bir çalışma alanı olarak tanımlanabilir [6]. Genel olarak, bu tanımlar dan YZ, bilişsel, öğrenme, karar verme ve çevreye uyum sağlama dâhil olmak üzere insanlarla benzeri işlevleri yerine getirme becerisine sahip, belirli

bir zekâ düzeyine sahip makinelerin geliştirilmesini kapsamaktadır. Bununla birlikte YZ, minimum insan müdahalesi ile akıllı davranışı modellemek için bir bilgisayarın kullanımını gerektiren genel bir terimdir [7]. YZ, başlangıçta bilgisayar ve ilgili teknolojiler ile şeklini alarak web tabanlı ve çevrimiçi akıllı eğitim sistemlerine katkı sağlamıştır. Nihayetinde diğer teknolojileri ve gömülü bilgisayar sistemlerinin kullanımıyla, insansı ve web tabanlı sohbet robotlarının geliştirilmesinde rol oynamıştır. Eğitim alanında kullanımı incelendiğinde, eğiticilerin yapay zekâ ile ilgili platformları kullanarak, öğrencilerin ödevlerini daha etkili ve verimli bir şekilde gözden geçirme, not verme gibi farklı işlevleri yerine getirebildiği görülmüştür. Dolayısı ile öğretim etkinliklerinde daha yüksek kalite elde edilmesini sağlamaktadır. Öte yandan, sistemler makine öğrenimi ve uyarlanabilirlik gibi işlemlerden yararlandığından, müfredat ve içerikler öğrencilerin ihtiyaçları doğrultusunda özelleştirilmiş ve kişiselleştirilmiştir. Web Zekâsı (WI) ve YZ araştırma ve geliştirmesinin, dağıtılmış zekâ oluşturmak için makine öğrenimi ve web teknolojisi ile akıllı aracı teknolojisi arasında bir denge oluşturma, kendi kendine öğrenen platformlarda öğrenme dâhil olmak üzere farklı unsurlara odaklandığı gözlemlenmektedir. Örneğin, robotların yapay zekânın veya bilgisayarların ve destekleyici ekipmanların içine yerleştirilmiş olması, eğitimin en temel birimi olan erken çocukluk eğitiminden itibaren öğrencilerin öğrenme deneyimine katkı sağlayan robotların oluşturulmasında kullanılmaktadır. YZ tabanlı sohbet robotları, kişiselleştirilmiş çevrimiçi öğrenme ortamı sağlar ve eğitmeni sohbet ortamına taşır. Bu teknoloji ile öğrencilerin anlama düzeylerini etkilenmektedir [8]. YZ destekli eğitim ile akıllı eğitim ortamları oluşturulmasında, sanal öğrenme ve veri analizi tahminleri destek sağlamaktadır. YZ öğrenme sisteminde öğrenen modeli, bağımsız öğrenme yeteneklerini geliştirmek için kritik öneme sahiptir. Öğrenenlerin öğrenme sürecinden elde edilen davranış verilerine dayalı olarak kurulur. Öğrencilerin düşünce ve yetenekleri, öğrenme yeteneklerini değerlendirmek için analiz edilir. Daha sonra, öğrenenlerin bilgi edinme ve uzmanlaşmasını sağlamak için bilgi analizi haritalanmaktadır. Öğrenci modellemesi, öğrenme sonuçları ile öğrenme materyalleri, kaynaklar ve öğretme davranışları dâhil olmak üzere çeşitli faktörler arasında bağlantılar kurmaktadır [9].

YZ; bilgi teknolojisinin eğitime entegrasyonunu hızlandırmakta oldukça etkin olup insanların yapamayacağı işlemleri yapabilmenin yanı sıra per-

formansı artıran birçok uygulama sunmaktadır. Bununla birlikte, YZ tabanlı eğitim, eğitim kalitesini, verimliliğini ve eşitliğini artırarak öğrenciler, öğretmenler ve bölgesel eğitim sistemleri için oldukça fayda sağlamıştır. Bu doğrultuda yapay zekâ tabanlı eğitimin etkin olduğu bazı alanlar aşağıda verilmiştir.

- *Öğrenme Analitiği:* Öğrenenlerden elde edilen bilgi nesnelerinin özelliklerinden gelen verilere odaklanır. Öğrenme analitiği kavramı, teknik etkinlik içerisinde olmayan eğitim dünyasına uygulanan yeni teknolojileri ve makine öğrenimini entegre etmektedir. Amaç, eğitime dâhil olan problemli ya da riskli öğrencilere destek olmak, geri bildirim ve öğretim içeriği edinmelerini sağlama gibi bireysel olarak ihtiyaç ve yeteneklerine göre uyarlamaktır [10].
- *Makine Öğrenimi:* Makine öğreniminin özü, bilgi keşfi, "eğitim verileri" olarak bilinen örnekleme veri setine dayalı ayrıştırma, anlamlı modeller ve yapılandırılmış bir bilgi üretme sürecidir. Örneğin, makine öğrenimi, öğrenenlerin sınıflarını ve üniversitelerini seçerken onlara öneriler oluşturmaya yardımcı olabilmektedir. Öğrencilerin başarı verilerini, isteklerini, tercihlerini en iyi şekilde geliştirilebile- cekleri kurumlarla eşleştirerek yönlendirmektedir. Ayrıca bu teknoloji, eğitmenlerin her kavramın öğrenciler tarafından nasıl sindirildiğini anlamalarına yardımcı olabilmektedir [11].
- *Veri Madenciliği*: Eğitimsel olarak veri madenciliği, öğrencilere sistematik ve otomatik yanıtlar üretmeye çalışmaktadır. Yapay zekâ ta- banlı eğitimsel veri madenciliği, içsel birliktelik kuralları geliştirmeyi ve öğrencilere kişisel ihtiyaçlarını karşılamaları için bilgiyi sunmayı amaçlamaktadır. Örneğin, öğrencilerin demografik karakteristik verileri ve not verme verileri, az sayıda yazılı ödevden analiz edilebilmektedir.

YZ'nin eğitim sektöründe uygulandığı diğer alanlar arasında müfredat, içerik geliştirme, sanal gerçeklik, artırılmış gerçeklik, web tabanlı platformlar, robotik, video konferans, görsel-işitsel dosyalar ve 3 boyutlu teknoloji gibi teknolojiler yer almaktadır. Yapay zekâ, sosyal etkileşimlerin her yönünü değiştirebilen, gelişen bir teknolojik alan olarak öğrencilerin daha iyi öğrenmesini mümkün kılmaktadır. Eğitimde YZ günümüzde, farklı bağlamlarda test edilmekte olan yeni öğretme ve öğrenme çözümleri üretmeye baş-

lamıştır [12]. Bu platformların ve araçların kullanımı, öğretici etkililiğini ve verimliliğini etkinleştirmiş ve iyileştirmiştir. Böylelikle daha zengin veya gelişmiş öğretim kalitesi amaçlanmaktadır. Benzer şekilde YZ, öğrencilere gelişmiş öğrenme deneyimleri sağlamada öğrenme materyallerinin öğrencilerin ihtiyaçlarına ve yeteneklerine göre özelleştirilmesini ve kişiselleştirilmesini sağlamaktadır [8]. Geçmişten bu güne, YZ'nin eğitim alanında kullanılması, öğretim-öğrenimde uygulanması ve etkileriyle ilgili çeşitli çalışmalar yapılmıştır. Yapay zekâ ile öğrenenlerin bilişsel yeteneklerini artırma ve öğrenmelerinde insan benzeri zekâya sahip bilgisayarlar ve makineler ile yapılan çalışmalar sonucunda ortaya çıkan yenilikler ve gelişmeler son yıllarda oldukça yaygınlaşmıştır. Yapay zekânın çeşitli alanlarda sağladığı faydalarının etkin olması ile eğitim kurumları web destekli ve çevrimiçi özelliklere sahip öğrencilerin gelişimine yönelik uygulamalar kullanmaya başlamıştır. Eğitim deneyimlerini zenginleştirerek çeşitli web ortamına gömülü eğitmen-öğrenci davranışlarını öğrenen sistemler son yıllarda oldukça artmıştır. Eğitimde YZ ve Öğrenme Analitiği (ÖA) biçiminde eğitim ve araştırma alanlarına ilgi artmıştır. Eğitimde Yapay zekânın hedefi, öğrenenlerin yapay zekâ uygulamalarıyla insan-bilgisayar etkileşimi ve öğrenme bilimleri yoluyla öğrenmesini geliştirmek için insan-yapay zekâ destekleyici bir platform sağlamaktır [13]. Eğitimin yapay zekâ ile derinden dönüştürüleceği düşünülmektedir. Örneğin, UNESCO'nun "Eğitimde 2030 hedefine ulaşmak için yapay zekâ teknolojileri" gündemi, eğitim ve öğrenimde sürdürülebilir kalkınma hedeflerini geliştirerek destekleyen eğitim yönetim sistemleri, yapay zekâ destekli öğrenme yönetim sistemleri veya eğitim uygulamalarında yapay zekâ, öğretmenleri destekleyebilecek ve eğitimde zorlukların üstesinden gelebilecek yeni kişiselleştirilmiş öğrenme biçimlerini tanıtmaktır [12]. Bu gelişmeler doğrultusunda yapay zekânın gelecek yıllarda eğitim ve öğretimde daha aktif rol alacağı öngörülmektedir.

6.3. EĞİTİMDE SANAL VE ARTIRILMIŞ GERÇEKLİK

Son yıllarda oldukça popüler bir kavram olan sanal gerçeklik (VR), yeni bir teknoloji değildir. İlk sanal gerçeklik kulaklığını 1970'lerde Utah Üniversitesinden Daniel Vickers tasarlamıştır. İki ekrana sahip olan kulaklık, kullanıcıya kendisine sunulan sanal sahneyi başını çevirerek gözlemleme imkanı vermiştir. Birkaç yıl sonra veri eldiveni arayüzü tasarlanmıştır. 1982 yılında

ortaya çıkan bu cihaz, elin ve parmakların hareketini ölçerek bilgisayara iletmiştir [14]. Bugün, teknolojinin hızla ilerlemesiyle sürekli güncellenen ekipmanlarla sanal gerçeklik, kullanıcıların çeşitli donanımlarla sanal bir dünyada olduklarını hissettikleri ve gerçek dünya ile etkileşime girdikleri sistemler daha da artmıştır. VR, eğitimin yanı sıra birçok faaliyet alanı için oldukça etkindir. Eğitimciler, iyi tasarlanan sanal ortamlar sayesinde, sanal deneyimlerin öğrencilere sunabileceği öğrenme potansiyellerini anında ve içgüdüsel olarak algılamaktadır [15]. VR kullanımıyla öğrenme sürecinde öğrenenlerin performansı artırılabilinir, motive edilebilir ve teşvik edilebilmektedir. Soyut veya ulaşılması zor senaryolar ve sunulan sürükleyici ortamlar öğrenenleri cezbedebilmektedir. Eğitimde sanal gerçeklik uygulamaları, öğrencilerin gerçek hayatta edinmeleri tehlikeli veya imkânsız olan ve önceleri eğitimde ilgiyi, etkileşimi artırmak amacıyla bilgisayar ortamı kullanılarak eğitim ortamları dijital hale getirilmiştir. Bugün, bu ortamlar bir adım daha ileriye taşınarak üç boyutlu bir eğitim ortamı olan VR teknolojileri deneyimleri kazanmada, yaparak ve yaşayarak öğrenme yöntemi deneyimlenmektedir. Bu nedenle sanal gerçeklik, geleceğin eğitim ortamları için önemli bir yeniliktir [16]. VR, kullanıcının bilgisayar tarafından oluşturulan üç boyutlu bir model veya sanal ortam ile etkileşime girmesine olanak tanır. Örneğin, uygulanabilir bilimler, arkeoloji, tarih ve mimarlık dâhil olmak üzere birçok farklı eğitim alanında VR uygulanmıştır. VR'nin eğitimde avantajı, geleneksel yöntemlerle tasvir edilmesi imkânsız veya anlatılması zor olan birçok konuda öğrencilere deneyimleme fırsatı verilmesidir. Özellikle VR, deneyimsel doğası, diğer temel özelliği olan etkileşim ve geleneksel öğrenme paradigmalarından oldukça farklı olması sebebiyle öğrenenlere yeni bir boyut sunmaktadır. Soyut kavramları öğretmek için ortak mekanizmalar olan, metafor ve analoji kullanılmaktadır. Öğrenenler deneyimler yaşayıp, çevreleriyle etkileşime girerek sanal dünyadan bilgi elde etmek için duyu organlarını kullanarak öğrenirler. Sanal gerçeklik, gerçek dünyadan elde edilen duyusal girdiyi bilgisayar simülasyonu tarafından oluşturulan duyusal girdi ile değiştiren bir teknolojidir [17]. Öğrenenlerin hareketlerine ve doğal davranışlarına yanıt vererek etkileşim sağlar. Bu açıdan VR, öğrenenlerin senaryoları ve durumları hayal etmek yerine deneyimlemesine olanak tanıyan bir ortam sağlayarak öğretime yardımcı olabilecek güçlü bir teknolojidir. VR sistemlerinin deneyimsel doğası üç kaynaktan türetilir: daldırma (*öğrenenlerin sanal gerçeklik cihazı ve fiziki ortam ile olan bağı-*

nın bazı sınırlılıklarının kesilerek sanal dünyada bulundurma hissini vermek), etkileşim ve çoklu duyusal geri bildirimdir [18]. Bu yönü ile VR'nin amacı, gerçek dünyayı sanal bir dünyayla kullanıcının gerçek dünyadaymış gibi davranmasına izin vermektir. VR deneyimsel doğası gereği öğrenmeyi yapılandırmacı yaklaşım ile desteklemektedir. Yapılandırmacılık, insanların bilgiyi deneyimlerinden öğrenerek yapılandırdığını belirten bir bilgi edinme teorisidir [19]. Piaget tarafından popüler hale getirildiği gibi teori, öğrencinin hâlihazırda kurulmuş olan dünya modeli içinde yeni deneyimleri özümsemeye çalıştığını belirtmektedir. Öğrenen yeni ayrıntıları başarılı bir şekilde özümseyemezse, yeni deneyimlere uyum sağlamak için farklı bir boyuttan hareket etmektedir [17]. Bu noktada, gerçek dünyada yapılamayan ya da yapılmasının zor olduğu eylemlerin gerçekçi hisler ile öğrenenlere kazandırılması sanal gerçekliğin sunduğu gelişmiş bir teknoloji olarak ortaya çıkmaktadır. Ancak sanal gerçeklik deneyimlerinin geleneksel öğrenme tekniklerine göre dezavantajlarından biri, özellikle ilk benimsendiği ve popüler olduğu dönemlerde VR gözlük ve ekipmanların pahalı ve istenilen kolaylıkta deneyimlenememesi fırsat eşitsizliği olarak görülmüştür. Eğitimde sanal evrenin kullanımının maliyetli olduğu ve sürekli yeni teknolojilerle uyum sağlamak zorunda olma mecburiyeti bu alanda yapılan uygulamaların bir diğer zorluğu olarak düşünülmektedir. Yüksek performanslı bilgisayarlar veya oyun konsolu kullanılmasını gerektiren bazı üst düzey kulaklılar ve kasklar oldukça pahalıdır. K-12 ve Yükseköğretimde sanal evren dünyasına entegre platformların geliştirilmesi sanal evrene yapılacak olan maliyetin oldukça fazla olacağı yönündedir. Ancak bilgisayar ve internet teknolojilerinin evrimi de bu şekilde yapılmıştır. Sanal evrenin eğitimde kullanımında düşük performanslı bilgisayarlarla çalışan ekipmanlarda daha ucuz VR projeleri devam etse de fırsat eşitsizliğine sebebiyet vermemek için kurumların eğitim ortamlarında alt yapı ve donanım malzemelerine yatırımlarını yapmasını gerektirmektedir [14]. Çünkü, çoğu özel sanal gerçeklik sistemi, okullarda ve Yükseköğrenim Enstitülerinde yaygın olarak kullanım için pahalıdır [20]. Yakın zamana kadar, kullanıcıları farklı yönlere bakarak görünümü değiştirebilecekleri bir ortama "daldıran" başlıklar kullanan VR, hem pahalı donanım hem de özel uzmanlık gerektiren kaynak yoğun bir teknolojiydi. Şu anda, ortalama bir eğitimcinin erişebileceği çeşitli ucuz, kullanımı kolay VR donanım ve yazılım seçeneği bulunmaktadır. Sanal gerçeklik oldukça uzun bir süredir var olmasına rağmen, nispeten düşük maliyetli, tüketici sınıfı

sanal gerçeklik donanımı yeni bir olgudur. Bu düşük maliyetli veya ücretsiz VR seçenekleri, hem öğretim araçları hem de öğrencilerin sanal gerçeklik sunumlarının tasarımı ve üretiminde deneyim kazanmalarına ve uzmanlık geliştirmelerine olanak tanıyan medya üretim araçları olarak umut vaat etmektedir. Sanal gerçeklik üretim ve sunum yazılımlarının çoğu, eğitimciler ve öğrenciler tarafından ücretsiz olarak kullanılabilir. Örneğin, Unity açık kaynak bir yazılımdır ve VR üretimi için endüstri standardı olarak kabul edilmektedir. En ucuz VR görüntüleyicilerden biri Google Cardboard'dur. Google Cardboard'un arkasındaki temel dayanak, sabit bir görüntüleyici olarak hizmet veren bir cihazın içine yerleştirilmiş bir akıllı telefon veya küçük tablet cihazı (iPod Touch gibi) kullanmaktır. Şu anda mevcut olan en yaygın ve kolay uygulanan sürükleyici VR teknolojisidir [21]. Eğitimde kullanımı giderek artan VR teknolojileri, öğrenenlerin öğrenme biçiminde farklı deneyimler sunmaya ve güncellenen sürümleriyle gelişmeye devam etmektedir.

Son yirmi yılda büyük bir hızla değişen teknoloji dünyası, akıllı telefonlar, G-6, robotik, yapay zekâ, blok zincir teknolojileri, nesnelerin interneti (IoT), artırılmış gerçeklik/sanal gerçeklik (VR/AR), sanal evren (metaverse), uzaysal bilgi işlem, otonom araçlar, uçan arabalar, gökyüzü bağlantılar, sinir ağları, kuantum hesaplama, uzay yolculuğu ve diğer şaşırtıcı araştırma ve icatlar insanların hayal ettiği şeylerinin ötesinde gelişmiştir [22]. 1999 yılına kadar artırılmış gerçeklik (AR), bilimsel araştırma laboratuvarlarıyla sınırlıydı. Ekipman pahalıydı ve yazılımı normal kullanıcıyı böyle bir teknolojiyle baş edemez hale getirecek kadar karmaşıktı. ARToolKit2'nin Kato ve Billinghurst tarafından 1999'da açık kaynak topluluğuna sunulması, bu teknolojiye geçişte yeni bir dönem başlatmıştır [14]. Günümüzde çeşitli alanlarda kullanılan bir teknoloji olan AR, eğitim ve öğrenme şeklinde de etkilidir. Çünkü günümüze kadar eğitim-öğretim büyük değişimler yaşamıştır. Ele alınması gereken büyük bir dijital uçurum olmasına rağmen, inovasyon süreci bu aşamada durdurulamaz hale gelmiştir. VR, AR ve YZ, eğitim ve araştırmayı kökten değiştirmek için güçlü yetenekleriyle modern çağın teknolojileridir. Geleneksel öğretimde köklü bir dönüşümün yanı sıra gömülü teknolojilere de entegre bu dijital teknolojiler öğrenen-öğreticilerin bilgiyi özümseme ve aktarma becerilerini geliştirmekle kalmaz, aynı zamanda birbirleriyle simbiyotik bir ilişki içerisinde olmasını sağlamaktadır [23]. Örne-

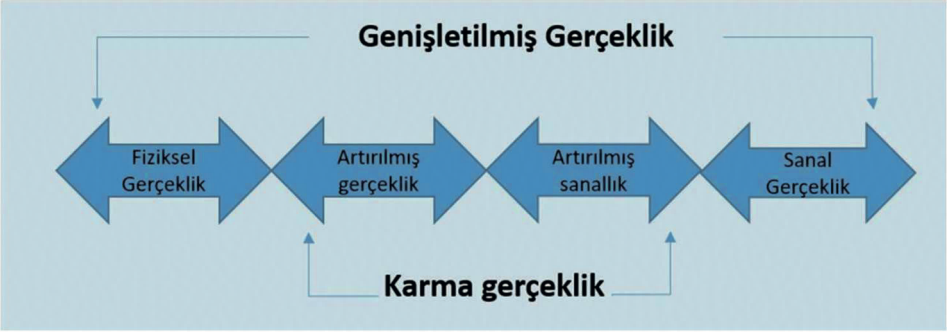
ğin, sanal bir görüntü, video veya animasyonla genişletilmesi olarak bilinen AR, öğrenenlerin makalelerini veya kitaplarını etkileşimli ve eğlenceli hale getirmek için etiketler olarak ses, video ve animasyonu yerleştirebilecekleri çalışmaları için hikâye anlatımı kitaplarında veya broşürlerinde kullanılmaktadır [22]. Eğitim alanlarında yapılan yatırımların değerinin artması ve sanal evrenlerin yaygınlaşmasıyla YZ, AR ve VR uygulamalarının eğitim dünyasında etkinliğinin her geçen gün artacağı kaçınılmazdır.

6.4. EĞİTİMDE SANAL EVREN (METAVERSE)

Sanal evren (metaverse) kavramı, Neal Stephenson adlı yazarın bilimkurgu Snow Crash romanında 1992 yılında gündem olmuştur. Romanda avatarların içerisinde yer aldığı sanal dünyaya "metaverse" denilmiştir. Metaverse kavramı "meta" ismini Yunanca "sonra, öte" anlamından ve İngilizcedeki "universe" kelimesinin birleştirilmesiyle türetilen metaverse "evrenötesi" ve "sanal evren" anlamından almıştır. Sanal evren kavramı hakkında aslında genel kabul görmüş bir tanım bulunmamaktadır. Ancak Sanal evren, hem yaşanılan hem de canlandırılan bir evrende paradokstur. Sanal evren, sürükleyici bir 3D sanal ortam ve avatarların içinde bulunduğu gerçek bir sanal yapay topluluktur [24]. Sanal evren, sadece artırılmış ya da sanal gerçeklik demek değildir. Bu teknolojilerin de desteklediği devasa bir sanal toplum inşası, küresel ama dijital bir köy olarak tanımlanabilir. Ayrıca diğer kavramlardan ayıran en önemli özelliğinin, sınırları-sınırlamaları ortadan kaldırması olarak söylenebilir. Oluşturulan teknolojilerin yazılım veya donanımları imkân verdiği ölçüde kullanılabilirken sanal evrende platform olarak sınırlar bulunmamaktadır [25]. Eğitim dünyasında da yankı bulan ve çalışmalarının sürdürüldüğü sanal evren uygulamaları oldukça önem kazanmıştır. Google, HTC ve Microsoft gibi şirketler, son on yılda Microsoft'un Hololens'i, Google'ın AR geliştirme yazılımı ARCore gibi karma gerçeklik alanına büyük yatırımlar yapmaktadır. Microsoft tarafından geliştirilmiş avatarlara sahip 3D gözlükler ve veri eldivenleri kullanarak Teams için Mesh adlı bir VR eklentisi Covid-19 küresel salgın döneminde karma gerçeklik uygulaması oldukça ses getirmiştir. Yeniden markalaşan Facebook'un adının Meta olarak değişmesiyle, tamamen sürükleyici VR toplantıları için tasarlanmış bir dizi iş odaklı VR aracı sunan Horizon Workrooms adlı Oculus Quest 2, sanal gerçeklik gözlükleri ile sanal evren dünyasında toplantılar

oluşturulmasına fırsat sunulmuştur. Akademisyenler, çalışanlar ve öğrenciler için sanal konferanslar, konferanslar ve sınıflar gün geçtikçe artmaktadır. Bu durum eğitim ve iş dünyasında kişisel avatarların oluşturularak katılım yapılmasına olanak tanıyan oldukça önemli bir gelişmedir. Seyahat maliyetlerinin azalmasının yanı sıra fiziksel olarak bir araya gelinmeden toplantıların sanal ortama taşındığını göstermektedir. Gelecekte eğitim-öğretim sanal evren uzayında olacağı ön görülmektedir. Örneğin, Sanat tarihi dersini sınıfta VR başlıkları kullanarak öğretmek ve müze parçalarını 3D sanal gerçeklik alanında deneyimlemek için öğrenenleri sanal olarak doğrudan sınıfa getirmek için çeşitli üniversitelerde hâlihazırda uygulamalar yapılmıştır. Öğrenciler sanat eserine girebilmekte ve ona sanal olarak dokuna- bilmektedir. Bu öğrenme deneyimsel olduğu için bilginin akılda tutulması daha kalıcıdır [22].

Geçmişten bugüne teknolojinin getirdiği yenilikler arasına VR, artırılmış gerçeklik (AR), karma gerçeklik (MR) ve genişletilmiş gerçeklik (XR) gibi sanal ve gerçek dünyanın bir arada kurgulandığı ortamlar dâhil olmuştur. Bu noktada teknoloji ve ilgili ekipmanların gelişmesi, VR, AR, MR ve XR teknolojilerinde ciddi ilerlemelerin önünü açmıştır [26]. VR, kullanıcının algısını tamamen gerçek dünyadan benzer veya tamamen farklı bir sanal dünyaya çeviren bilgisayar simülasyonlu bir deneyimdir. VR, bu teknolojilerin en yaygın olarak bilinenidir. AR, sanal nesnelerin ve ortamların gerçek dünyayla karıştırıldığı yerdir. Kullanıcı, bilgisayar tarafından oluşturulan algısal bilgilerle birlikte gerçek ortamını görüntülemektedir. AR, diğer dijital ayrıntılarla gerçek dünya deneyimini geliştirir, yeni bir algı katmanı oluşturur ve gerçek dünyayı merkezde tutarak kullanıcının gerçekliğini veya çevresini tamamlar. MR, gerçek ve sanal dünyaların karıştırıldığı bir sanal ortamdır. MR, gerçek dünya ile dijital unsurları birleştirir. Genellikle, kullanıcı başa bir ekran takmaktadır. MR'de kullanıcı, yeni nesil algılama ve görüntüleme teknolojilerini kullanarak fiziksel, sanal öğeler ve ortamlarla etkileşime girebilir ve bunları değiştirebilir. XR, insan ve makine arasındaki etkileşimin bilgisayar teknolojisi ve donanımı tarafından oluşturulan etkileşimler yoluyla gerçekleştiği, bir araya getirilmiş tüm gerçek ve sanal ortamları ifade etmektedir. XR teknolojileri VR, MR ve AR'den oluşmaktadır [27]. Şekil 6.2'de sanal evren ile bağlantılı olan XR, MR, VR ve AR teknolojilerine yönelik şekil yer almaktadır. Bu kavramlar, sanal evren ile ilişiklidir. Yakın gelecekte etkileşimlerin daha da artacağı düşünülmektedir.



Şekil 6.2. XR, MR, VR ve AR ile Gerçeklikler

Sanal evren ile kullanıcıların çalışabileceği, öğrenebileceği, oynayabileceği, satın alabileceği, tanıtabileceği ve kazanabileceği bir evren oluşturulması planlanmaktadır. Animasyon, oyun, AR/VR, sanal evren, robotik, interaktif sanat ve ilgili alanlarda daha iyi öğrenme, daha çeşitli dijital içerikler ve daha yeni vizyonlar sağlayacaktır [22]. Sanal ortamlar, dijital nesneler ve VR ve AR gibi insanlarla çok duyarlı etkileşimler sağlayan teknolojiler ile entegredir. Bu nedenle, sanal evren, kalıcı ve çok kullanıcılı platformlarda sosyal ağa bağlı sürükleyici ortamlardan oluşan birbirine bağlı bir ağdır [3]. VR ve AR teknolojileri bugün çoğunlukla oyun alanında aktif olarak yer almaktadır ancak sanal evren ile sadece oyun oynamaktan ziyade, çeşitli etkinliklerin oluşturulduğu yeni bir sanal evrenin oluşturulacağı söylenebilir. Toplum bu sanal dünyada etkileşimler kurabilecek ve bu evrende zaman geçirecektir. Web 4.0 teknolojisine doğru yol alırken, yapay zekâ ve artırılmış gerçeklik ile sanal evren dünyasının ilişkilendirilmesi söz konusudur. Sanal evren ile üç boyutlu avatarlar oluşturularak hem gezilebilir hem de çeşitli aktiviteler yapılabilme imkânı sunulmaktadır. Dolayısıyla sanal evren dünyası oldukça geniş bir alana hâkim olacak dahası yeni bir ekonomik pazar haline gelebilecektir. Beklenen sanal evren, doğrudan fiziksel etkileşimlere sahip gerçekçi bir toplum, ırk, cinsiyet ve hatta fiziksel engellilik kavramları zayıflatarak toplum için oldukça faydalı olacaktır [28]. Dijital olarak bir alan sunan Microsoft sanal evrenini, kullanıcıların ve yaşamların dijital birer temsilini yaşattığı yönünde tanımlamaktadır. Sanal evren dünyasında avatarlar sayesinde toplu halde buluşmalar, eğlence mekânlarına giderek sosyal deneyimler yaşamak VR ile sanal dünyada mümkün hale gelmektedir. Sanal evren ile geliştirilen yeni donanımsal ekipmanlar oldukça ger-

çekçi görünen bir hayatı yaşamayı mümkün kılmaktadır. Sanal evrenin sunduğu imkânlar sadece eğlence değil aynı zamanda eğitim alanlarında da kullanımlarının artacağıdır. Bu yönüyle sanal evrenin eğitimde kullanımına yönelik beklentiler öğrenmeyi daha kapsamlı, ayrıntılı, sosyal ve tarafsız hale getirerek eğitimi iyileştireceği yönündedir. Ancak geleneksel bir sınıfı sanal bir dünyaya dönüştürmek oldukça zorlu bir süreçtir. Microsoft ve Meta gibi büyük şirketler olmak üzere birçok şirket öğrenciler için sanal evrende eğitim deneyimleri üzerinde çalışmaktadır. Meta veri deposunun gerçek hayatı sanal bir ortama dönüştürmekle kalmayıp, eğitimi daha da iyileştireceği beklenmektedir. Yaşanılan sanal deneyimler sınıftaki her öğrenci için pahalı teknolojileri gerektirmektedir. Aileler ve eğitim kurumları bu maliyetleri öğrenme sürecinde bir engel olarak görebilmektedir. Ancak sanal dünyaların faydaları anlamlı ve ölçülü şekillerde gerçekleşirse muhtemelen kurumlar bütçelere öncelik verecektir. Sayısallaştırılmış benzerliklere sahip üç boyutlu avatarlarını oluşturan öğrenenler, fiziksel özelliklerini kendi isteklerine göre değiştirebileceği gibi kendilerine hiç benzemeyen fabrikasyon avatarlar seçebilmektedir. Bir sınıf ortamında bir araya gelerek bireysel avatarlar aracılığıyla öğrenenlerin fiziksel ortamdan ziyade sanal bir evrende buluşturan çalışmalar devam etmektedir. Örneğin, tekerlekli sandalye kullanan öğrenenler gerçek dünyadaki engellerinden sıyrılıp çoklu evrende tamamen işlevsel bacaklara sahip olabilirler. Sanal sınıflar veya eğitim için artırılmış gerçeklik, dijital oyunlar ve sosyal medya platformları yakın gelecekte meta veri tabanının büyük çoğunluğunu oluşturacaktır. Eğitim alanında yeni örneklerine rastlanılan ve sanal evrende çığır açan bir okul, eğitim için meta veri tabanına çoktan girmiştir. Optima Classical Academy 2022 yılında %100 sanal sözleşmeli okul açılmayı planlamıştır. Okul, 3-8. sınıflardaki ABD-Florida'da öğrencilerine ücretsiz olarak sanal evren sınıfları sunulması planlanmaktadır. Bununla birlikte, kurum dünyanın her köşesinden öğrenci kabul etmeyi planlamaktadır. Sualtı sınıflarından Yunan tapınaklarına kadar çok çeşitli temalara sahip sanal sınıflar planlanmaktadır [29]. Öğrenenler kendi evlerinden ya da istedikleri bir yerlerden çalışacak ve katılmak için Oculus gibi VR başlıkları kullanacaktır. Meta veri deposu ile sanal evren uzmanları kaliteli içeriklerin oluşturulmasını hedef edinmiştir. Eğitimde sanal evren etkinliğine ilişkin veriler hala karışıktır. Ancak eğitim alanında sanal evren ile bağlantılı olarak kullanımına ilişkin bütünleşik öğrenme, sorumlu kullanım değerlerin geliştirilmesi, ilgi çekici rehberlikler ve yarat-

ma-paylaşma özgürlüğünün sunduğu özellikler sanal evrende ilerlemelerin potansiyel yolların başında gelmektedir. Aşağıda yer alan Şekil 6.3'de bu doğrultuda oluşturulmuş şekil yer almaktadır.

Şekil 6.3. Eğitimde Sanal Evren ve Potansiyel Yöntemler [30]

Eğitim dünyası, geçmişten bugüne çeşitli dönüşümler içerisinde yer almıştır. Sanal gerçeklik ve eğitim ile ilgili çalışmalar özellikle son 15 yıl içerisinde daha fazla rağbet görmüştür. Çünkü VR, kullanıcılara görsel, dokunsal ve gerçeğe benzer his duygusunu gelişmekte teknolojiler ve uzaysal bilgi işlem teknolojisi sınıfı deneyimi sunar. Sanal gerçeklik kav- ramı, teknolojik ekipmanlar ile kurgulanan gerçek ve hayali bir yaşamın birleştirilmesi olarak ta tanımlanabilmektedir. Öğrenmeyi sanal olarak öğretmek, öğrenmeyi zenginleştirerek deneyimle öğrenmeye olanak sağla- yan bir platformdur [31]. VR, kullanıcıların etkileşimde bulunabileceği, bilgisayar tarafından oluşturulan üç boyutlu bir görselleştirmedir. Second Life adlı oyun ile tanışan 21. yüzyılın toplumu etkileşimli öğrenmenin daha fazla yoğun olduğu bir ortama geçiş sağlamıştır. Ancak gelişen teknoloji ve sanal evreni kullanma oranlarının artması, uzaktan eğitim ile kat edilen yolların son iki yılda oldukça fazla olması ile çocuklar ve genç yetişkinlerin öğretme ve öğrenmeye meta-öğrenme ile daha fazla yönelmeleri öngörülmektedir. Zoom ve Teams gibi platformlarda yapılan toplantılar, görüşmeler çalışan ve öğrenenler için yeni normal hale gelmiştir. Profil resimleri gibi temsili avatar seçme

olgusu da toplumsal olarak benimsenmiştir. Eğitimciler, öğrenenler ve sosyal varlıklar olarak sanal gerçeklik ve meta evrende buluşma davetleriyle giderek daha fazla karşılaşmak kaçınılmazdır. Meta veriye daha geniş bir şekilde erişildikçe, insan-bilgisayar etkileşimi ve bilgisayar destekli ortak çalışma için heyecan verici bir alan sunmaktadır. VR'nin yeni öğrenme ve öğretme biçimlerini nasıl mümkün kılacağı araştırmaları devam etmektedir. VR'yi benimsemek ve sanal evren ortamına girmek teknolojinin her noktasına ayak uyduran toplum için eğitim yaşamında da denemeleri ile devam edecektir. Pandemi ile fiziksel ortamdan ziyade sanal ortamlarda uzaktan eğitim çalışmalarına daha fazla yoğunlaşılmıştır. 2022 yılının başlarında ODTÜ araştırmacıları, öğrencilere yönelik geliştirilen her öğrenenin bir avatarla temsil edildiği ve üç boyutlu "metaverse" dünyasında, derslerin fiziksel gerçekliğe en yakın ortamda yapılabildiği "sanal okul" projesini geliştirmiştir. Sanal okul gibi üç boyutlu sanal evren ortamında ise öğrencilerin hem kendilerinin hem sınıf arkadaşlarının ve hem de öğretmenlerinin avatarı bulunmaktadır. Proje ile sanal evren ortamlarında bu bulunuşluk hissine odaklanılmaktadır. Gerçek dünyada nasıl hissediyorlarsa ve diğer kişilerle nasıl etkileşim halindelerse, sanal evren ortamında da fiziksel gerçekçiliğe en yakın ortam sağlanabilme hedeflenmiştir [32]. Bununla birlikte, Education (eğitim) ve Metaverse kelimeleriyle oluşturulan EduVerse kavramı şimdiden popüler hale gelmiştir. Hatta K-12 eğitimi için eğitimcilerin iş birliği yapabileceği onlarca etkileşimli Metaverse ortamı (Metaverse hub) oluşturulmuştur [25]. Sanal evren ile ilgili yapılan çalışmalar ve buna olan ilginin her geçen gün arttığı dijital çağda bir sınıfta fiziksel olarak bulunma, ayrıcalıklı bir eğitim deneyimi olmaktan çıkacaktır. Avatarlar ile beden dili ve yüz ifadesi doğruluğu, sanal katılımın eşit derecede etkili olmasını sağlayacaktır [3]. Gelişmiş sanal dünyanın bir özelliği, öğrencinin öğrenmesini kolaylaştırmak için içinde barındırdığı ve öğreticiler tarafından özel olarak yetiştirilen çeşitli BT kaynaklarını uygulayabilmektir. Benzer şekilde, yüksek öğretim kurumları ve okullara öğrenenlerin mobil cihazlarına kurulmak üzere Google Play Store'da bulunan uygulamalara erişmek için QR kodları dahil edilebilmesi gerekmektedir [33].

Meta verinin daha geniş bir sınıflandırması artırılmış gerçeklik, yaşam günlüğü, sanal gerçeklik ve ayna dünyaları gibi dört önemli kavramla bir araya getirmektedir. Bu kavramlar eğitim sektöründe kullanım örneklerini

yönlendirmek için sanal evren teknolojisini kullanma konusunda ayrıntılı bir izlenim sağlamaktadır. Bu kavramlar:

- *Arttırılmış Gerçeklik (Augmented Reality-AR):* Temel olarak konum tabanlı teknoloji ve ağları kullanarak akıllı bir ortamın yaratılmasını sağlamaktadır. İstenilen özelliklere göre gerçek dünyadaki nesneler üzerinde dijital bir görüntü oluşturmaya yardımcı olmaktadır. AR'nin önemli uygulamalarından bazıları, araç HUD'larının yanı sıra akıllı telefonlarda da mevcuttur. AR'nin en popüler örneği Pokemon Go oyunudur.
- *Yaşam Günlüğü (Lifelogging):* Eğitim ve öğrenim alanında sanal ev- ren kullanım senaryoları için yaşam günlüğü bir başka önemli kav- ram olarak görülmektedir. Bilgi ve günlük deneyimleri insanlar ve nesnelerle yakalamak, depolamak ve paylaşmak için kullanılan bir teknolojidir. Lifelogging, artırılmış teknolojinin kullanımına odaklanır ve bunun pratik örneklerinden bazıları kara kutulara ve giyilebilir teknolojilerde örnek olarak yer alabilmektedir. Örneğin, *Apple Watch Samsung Health*, yaşam günlüğünün gerçek dünyada öne çıkan kullanımlarından biridir. Lifelogging, günlük yaşam ve düşüncelerden oluşan üretken sonuçlarla belgelenmesine odaklanmaktadır. Kendi kendine öğrenme deneyimlerine odaklanarak eğitim için sanal evren kullanımlarını artırmaya yardımcı olabilmektedir. Bir bireyin hayatındaki günlük olayları gözden geçirebilir ve yansıtabilir. Meta veri deposu bu nedenle, geri bildirimi ve pratikte kişinin hayatından öğrenmeyi temsil etmek ve uygulamak için hayati bir araç sunabilmektedir. Aynı zamanda, farklı bilgi türlerini eleştirel bir bakış açısıyla keşfetmek için farklı yollar sağlamaktadır. Yaşam günlüğü, bilgilerin yaratıcı bir şekilde yeniden yapılandırılmasının yanı sıra farklı öğrenme deneyimlerini yansıtmaya da yardımcı olabilir [30].
- *Ayna Dünyası:* Bu kavram, eğitimde sanal evrenin çeşitli kullanım durumları için de önemli ölçüde önemlidir. Aslında gerçek dünyanın gelişmiş bir sanal karşılığı olan dış dünyanın belirli bir simülasyon türü olarak tanımlanabilmektedir. Ayrıca, gerçek dünyanın görünür özelliklerinin sanal gerçekliğe çevrildiği meta veri deposu

olarak ta düşünülebilmektedir. Google Earth gibi harita tabanlı hizmetler, ayna dünya kavramının en iyi örneklerindendir. Ayna dünyalar, öğrenme ve eğitim üzerindeki uzamsal sınırlamaların ve fiziksel kısıt- lamaların üstesinden gelinmesine yardımcı olabilir. Örneğin, çevrimiçi video konferans araçları, çevrimiçi gerçek zamanlı sınıfları yürütmek için temsili ayna dünyaları olarak hizmet edebilmektedir.

- *Sanal Gerçeklik:* Meta verinin eğitim uygulamalarının yönlendirilmesinde çok önemli bir rol oynayacak olan meta veri deposunun bir diğer önemli yönü de sanal gerçekliktir. Temel olarak, iç dünyaları simüle edebilen sanal evren türünü belirtmektedir. Bu noktada, sanal gerçeklik sanal evrende, avatarlar ve çevre arasındaki etkileşim etkinliklerinden elde edilen dijital veriler kullanılarak geliştirilen dijital bir alan olarak düşünülebilir.

Şekil 6.4. Eğitimde Sanal Evren Kavramları [30]

Yukarıdaki Şekil 6.4'de yer alan sanal evren ile bağlantılı ve meta veri deposunu etkileyen bu dört kavramın eğitim ortamlarını da etkileyeceği düşünülmektedir. Bu dört kavramın eğitsel sonuçları, eğitim ve öğrenmeye yönelik geleneksel yaklaşımların dönüştürülmesi için önemli çıkarımlar sağlamaktadır [30].

6.4.1. Web 3.0 ve Blok Zinciri

Blok zinciri teknolojisi ile desteklenen Web 3.0, dijital altyapının bir evrimidir; bu sayede protokol, kullanıcılar arasında doğrudan varlık alışverişini kolaylaştırır ve güvenilir aracılara olan ihtiyacı ortadan kaldırmaktadır. Gerçek dünyadaki iletişimi sağlamak için makine öğrenimi, yapay zekâ ve blok zincirinden yararlanan en son internet teknolojisidir [34]. Web 3.0, ademi merkeziyetçilik ve sunucu mimarileri kullanmadan kullanıcı merkezli olmayı taahhüt etmektedir [35]. Blok zinciri, Web 3.0 olarak adlandırılan web teknolojisinin üçüncü çağında ortaya çıkmıştır. Bu teknoloji varlık transferine izin vermektedir. Blok zinciri, kâr merkezi veya tekelci hizmet sağlayıcılar olmadan değer transferini sağlamaktadır. Sosyal medyanın gelişimi kullanıcılar arasında bilgi alışverişine izin verirken, kontrolü birkaç özel aktör arasında tutarken, blok zincir teknolojileri, merkezi kontrol noktaları olmadan merkezi olmayan ağlar oluşturmaktadır [36]. Blok zinciri kavramı, şifreleme bilimi olan kriptoloji tabanlı olup, 2008 yılında açık dağıtılmış bir defter ve dijital para birimi olan bitcoin uygulaması ile ortaya çıkmış bir teknolojidir. Blok zinciri, verilerin bütün bir şekilde korunması prensibine dayalı geliştirilmiştir. Blok zincir ağı, güvene dayalı olmayan işlemler üzerine kurulmuş bir eşler arası P2P ağıdır. Blok zincirinde, ağdaki her bilgisayarın (düğüm) paylaşılan bir defterin bir kopyası vardır ve bu kopya ağdaki tüm düğümler tarafından güncellenir ve korunur. Blok zincirine bağlı bir bloğa eklenene kadar bir işlem gerçekleştirilmez. Blok zinciri güven gerektirmeyen işlemler üzerine kurulduğundan, her düğümün bu isteğin gerçek bir kaynaktan geldiğini doğrulaması gerekir. Bunu yapmak için, mesaj isteği ve istek sahibinin özel anahtarı kullanılarak dijital bir imza üretilir. Bu dijital imza, mesajın gerçek bir kaynaktan geldiğini doğrulamak için ağdaki düğümler tarafından kaynağın ortak anahtarıyla kullanılır. İmza mesaja bağlı olduğundan, mesajda yapılacak herhangi bir değişiklik imzayı geçersiz kılacağından, hiç kimse mesajı iletirken değiştiremeyeceğinden mesajın bütünlüğünün sağlandığı anlamına gelmektedir [37].

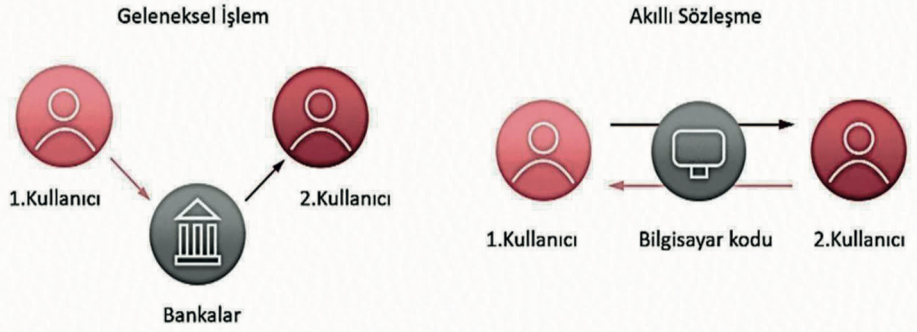
Her ne kadar üzerine kurulu olduğu uygulamaların güvenlik problemleri olsa da, kripto para ticareti ve akıllı sözleşmelerin yürütüldüğü bir teknolojidir. Akıllı sözleşmeler, blok zincirinde çalışan makine tarafından okunan yazılım kodlarıdır. Blok zinciri kriptografi teknikleri kullanılarak bir ağda birbirine bağlanan kayıtlı işlemlerin ve izlenen varlıkların bir listesini içeren bir dijital defter olarak tanımlanır. Blok zincirinin dağıtık defteri, ademi

merkeziyetçiliği, bilgi şeffaflığı, güçlü yapısı ve açık bir platform olması avantajlarından bazılarıdır [38]. Blok zincirinin evrimi ilerleyici bir süreç olmuştur. Blok zinciri, yalnızca izinli ağ üyeleri tarafından erişilebilen bir defterde depolanan verileri paylaşmaktadır [39]. Bugünün teknolojisi blok zincirine dayalı merkezi olmayan bir ekonomik model, oyun ve sanal evren arasındaki geniş boşluğu doldurarak yeni bir dijital-yerel varlıklar ve para kazanma fırsatları çağını başlatmaktadır. Tüm platform içi varlıkların bir geliştiriciye ait olduğu ve bir geliştirici tarafından kontrol edildiği bir oyun oynamanın aksine, merkezi bir sunucu sistemi içinde, meta veri deposundaki kullanıcılar tüm deneyimlerinin tek sahibidir. Buna sanat eserleri, müzik ve dans koreografileri gibi yaratıcı çalışmalar dâhildir. Meta veri deposu, bir bireyin herhangi bir görevi sanal olarak yapabileceği paralel bir evrene benzemektedir. Sanal evrende kullanıcılar arazi satın alabilir, alışveriş yapabilir, parti yapabilir vb. eylemleri gerçekleştirebilmektedir [40].

6.4.2. Akıllı Sözleşmeler

1990'lı yılların başlarında kriptoloji uzmanı ve bilgisayar bilimcisi olarak bilinen Nick Szabo tarafından ortaya çıkan akıllı sözleşme kavramı, ilk başlarda bilgisayarlı işlem protokolleri olarak tanımlanmıştır. Bitcoin ve blok zincir teknolojilerinin geliştirilmesiyle yeni bir boyut kazanan akıllı sözleşmeler, harici bir güvenilir otoriteye ihtiyaç duyulmadan, karşılıklı güven problemi olan aracılardan oluşan ve bir ağ tarafından yürütülen bilgisayar programlarıdır [41]. Önceleri yalnızca güvenilir bir aracı aracılığıyla çalışabilen bir blok zinciriyle artık merkezi bir otoriteye ihtiyaç duyulmadan ve merkezi ol- mayan bir şekilde çalışabilen uygulamalar, aynı işlevselliğe aynı miktarda ulaşabilmektedir. Blok zinciri ağlarının önemli bir özelliği olan kriptografinin yoğun kullanımı, ağdaki tüm etkileşimlerin arkasında otoriteyi beraberinde getirmektedir. Akıllı sözleşmeler bu noktada blok zincirinde bulunan ve kendi kendini yürüten komut dosyaları olarak ortaya çıkmaktadır [42]. Bitcoin'den sonra ortaya çıkan dijital para birimlerinden olan Ethereum gibi yeni ortaya çıkan kripto para sistemleri, blok zincirinde isteğe bağlı kullanıcı tanımlı programlar çalıştırma fikrini benimsemektedir. Böylece anlamlı bir merkezi olma- yan akıllı sözleşme sistemi oluşturulmaktadır [43]. Akıllı sözleşmeler, blok zincirine ya da benzer dağıtılmış altyapıya sabitlenmiş, önceden tanımlanmış iş süreçlerin şeffaf bir şekilde yürü-

tülmesine izin veren blok zincirler veya benzer altyapılar üzerinde merkezi olmayan şekilde sabitlenmiş komut dosyalarıdır. İşlemler blok zinciri tarafından tetiklenir ve ağ genelinde doğrulanarak daha önceden tanımlanmış eylemler yürütür. Akıllı bir sözleşmenin şartları blok zincirinde şeffaf bir şekilde saklı kaldığından, tüm tarafların istediği gibi çalışır ve bu da ilgili taraflar arasındaki güveni sağlar [44]. Bu sayede iş süreçleri otomatikleşebilir ve maddi varlıklar programlanabilir. Örneğin, patent kanununun düzenlemeleri akıllı bir sözleşme ile uygulanabilir. Bu, fikri mülkiyet haklarının yasal düzenlemelere uygun olarak yönetilmesini sağlayacaktır [45]. Sözleşmeler, saniyeler ve dakikalar içinde gerçekleştirilen çok basit işlemler veya tamamlanması saatler veya aylar sürebilen, belirli haklara ve yükümlülüklere sahip müzakereleri ve onlarca sayfa yazılı metni içeren nispeten karmaşık ve uzun işlemler olabilir. Bugün akıllı sözleşmeler nispeten basit işlemler kategorisine girmektedir [46]. Bununla birlikte kötü niyetli sözleşme tarafları finansal ödemeden kaçınmak için bir protokolden zamanından önce vazgeçebilir. Belirli bir saatte, bu tür iptalleri belirgin hale getirmek için zaman aşımları kullanılabilir. Bu noktada blok zinciri, teminat mevduatlarını dürüst, iptal etmeyen taraflara yeniden dağıtarak iptal eden tarafları finansal olarak cezalandırabilir. Şekil 6.5'de akıllı sözleşmelerin çalışma prensibine yönelik şekil yer almaktadır.



Şekil 6.5. Geleneksel ve Akıllı Sözleşme [47]

Şekil 6.5'de yer alan şekilde görüldüğü üzere geleneksel işlemlerle yapılan işlemlerde bankalar aracı olarak görülürken, akıllı sözleşmeler ile yapılan işlemlerde bilgisayar kodları aracı olarak işlemleri gerçekleştirmektedir.

6.4.3. Nitelikli Fikri Tapu (Non-Fungible Token- NFT)

Sanal evren ve Web 3.0 ile birlikte gelen bir diğer kavram ve İngilizcesi Non-Fungible Token ve Türkçesi Nitelikli Fikri Tapu, sanat eserleri olarakta ifade edilen NFT; VR gözlükler, dijital para birimleri, artırılmış ve sanal gerçeklik kavramları oldukça popüler hale gelmiştir. NFT, dijital sanat veya müzik gibi gerçek dünya varlıkların, evler veya arabalar gibi tokenize edilmiş fiziksel varlıklara bağlanan benzersiz bir blok zinciri tabanlı dijital varlıktır. Çoğu NFT, Ethereum tarafından izlenen merkezi olmayan dijital para birimi platformlarında listelenmektedir [48]. NFT, sanat ile ilgili varlıkları menşei, mülkiyeti ve erişimi doğrulamaktadır. NFT olarak satışa sunulan çeşitli dijital eserlerin taklit edilemezliği sanal dünyada blok zinciri teknolojisi kullanılarak sağlanmaktadır. NFT'lerin, önümüzdeki iki ila on yıl içinde iş dünyası ve toplum üzerinde önemli etkiye sahip olacağı ön görülmektedir. NFT'ler sanal evren ekonomisinin kalbinde yer alacaktır. Ancak bu durum, NFT dolandırıcılığında bir artış olasılığını artıracaktır. Sanal gerçeklik gözlüklerinin artan kullanımıyla, veri ihlalleri ve kötü amaçlı yazılımlar için yeni olasılıklar ortaya çıkabilecektir. Blok zincir ile ilgili dolandırıcılıklarda ve finansal kurumlara yönelik tehditlerde bir artış olacağı düşünülmektedir. Sanal evrendeki 3B ortamı, mühendislik yöntemlerinin daha etkin kullanımını kolaylaştıracaktır. Veri güvenliği daha hassas bir nokta olacak ve muhtemelen güvenlik açıkları artacaktır [49].

6.4.4. Kripto Varlıklar

Kripto para birimleri ve basit bir ifadeyle kriptografi, bilgiyi yalnızca gizli bir anahtara sahip biri tarafından deşifre edilebilen (veya şifresi çözülebilen) okunamaz bir biçime dönüştürerek (yani şifreleyerek) koruma tekniğidir. Bitcoin gibi kripto para birimleri, bu teknikle ustaca genel ve özel dijital anahtar sistemi kullanılarak dijital varlıkları güvence altına almaktadır [50]. 2010 yılından sonra Bitcoin ve alt coin gibi merkezi olmayan kripto para birimleri hız la popülerlik kazanmıştır. Ortaya çıkan bu kripto para birimi sistemleri, güvenliği sağlanan dağıtılmış blok zinciri teknolojisinin üzerine inşa edilmiştir [43]. Sanal evrende kripto paralar yapılan işlemlerin kayıt altına alındığı blok zinciri tabanlı para birimleri, meta veri deposu içindeki tüm varlıkların kolayca ve güvenli bir şekilde oluşturulmasına, değişti-

rilmesine, paylaşılmasına, izlenmesine ve potansiyel olarak öğelerin kullanıcıları tarafından dünyalar veya meta hedefler arasında sorunsuz bir şekilde taşınmasına olanak tanır. Kripto para birimleri içerisinde popüler olan sanal evren projeleri yatırımcıları oldukça popüler hale gelmiştir. Metaverse projeleri kapsamında aktif rol oynamaya başlayan Starlink (Starl), Axie Infinity (Axs), Ufo Gaming (Ufo), The Sandbox (Sand), Aavegotchi (Ghst) ve Decentraland (Mana) gibi alt coinler oldukça müşteri çekmektedir. Bu ortamlarda geçerli olan tokenlar çeşitli kripto para borsalarında işlem görebilmektedir. Kripto paralar, sanal dünyada ekonominin oluşmasında aktif rol oynayacaktır. Decentraland ve OVR gibi tokenlar ile arsa alımları ciddi oranda artmıştır. Ethereum tabanlı bir kripto para birimi olan OVR ile OVRLands platformlarında arsa alımları yapılmaktadır. Özellikle dolar bazlı olarak USDT (Tether) karşılığı olarak token alım-satım işlemleri gerçekleşmektedir. Sanal evrende maddi varlıkların kripto paralar ile gerçekleşmesi söz konusu olduğundan, dijital paraların popülaritesi her geçen gün artmaktadır.

6.5. SANAL EVREN (METAVERSE), SİBER GÜVENLİK VE OLASI TEHDİTLER

Eğitim alanında örneklerinin görülmeye başlandığı sanal evrende, siber güvenlik ve olası problemler de oldukça merak konusudur. Gerçek ve dijital dünyanın bir bütün olarak ele alındığı sosyal ekonomik sistemdir. Avatar, deneyimler ve bedensel dünyanın parçası olan her bir eylem bu evrende sorunsuz bir şekilde çalışabilir. Blok zincir, ardından kripto para ve tokenlerin, NFT'lerin ve sanal evrenin yaygınlaşmasıyla, dijital dünyada yeni bir dönüşüm yaşanmaya başlamıştır. Gerçek benliğin hiçbir kısıtlama olmaksızın keşfedilebileceği ve en çılgın hedeflerin kovalanabileceği başka bir yaşamdır. Ancak yapılan eylemler bununla sınırlı ve göründüğü kadar hep olumlu yönde ilerlememektedir. Bu sanal evren beraberinde siber güvenlikte tehditlerle gelmektedir. Fikri büyüleyici olsa da, kişisel güvenlik açıklarına sebep olmaktadır. Bu durumdan uzak durmanın bir yolu, bu alanda sağlam bir siber güvenliğe sahip olmak ve düzenlemeye tabi olmaktır [51]. Web 2.0 ile gelen sosyal medya veya e-posta ile kayıt olunan platformların olası güvenlik açığı ve saldırılara maruz kalmaları durumunda kullanıcı bilgileri tehdit altındadır. Web 3.0 ile gelen web3 cüzdanları kullanıcıların herhangi bir sosyal medya ağı ya da e-posta platformuna kayıt olmadan kendi web arayü-

zü ile bağlantının sağlanması temeline dayanmaktadır. Ancak MetaMask ve benzeri cüzdanlar Ethereum blok zinciri ile merkezi olmayan uygulamalara etkileşim kurmak için kullanılabilecek tarayıcı eklentisi web3 kimliği ile gelen bir yeniliktir. Mobil olarak tercihlerde MetaMask, Trust Wallet, MathWallet, ve SafePal cüzdanları Android ve iOS için kullanılabilmektedir. Blok zinciri teknolojisinde var olan özelliklerinden biri yapılan işlemlerin değiştirilmeden aynı şekilde blok zincirine ekleme özelliğidir. Güvenlik açığı olmasını engellemede önemli olan bu özellik, hatalı olarak yapılan işlemlerin, hırsızlık, kara para aklama işlemlerini temizleyebilmede dezavantaj olarak görülmektedir. Sanal evren içinde alınıp-satılan eşyaların veya yapılacak işlemlerin tümü, blok zincir üzerine inşa edilmiş akıllı sözleşmeler yoluyla gerçekleşecektir. Kişilerin güvenliğini sağlamak için en önemli çevre dostu yöntemlerden biri olan akıllı sözleşmelerin denetlenmesi gerekmektedir [51].

Gerçek dünyada büyük verilerin depolanması gibi meta depo adı verilen depolama alanların oluşturulması söz konusudur. Bununla birlikte, meta veri depolarının da siber güvenlik sorunları büyük zorluk olarak ortaya çıkmaktadır. Sanal evrende veri tabanı olarak meta veri deposu içinde yer alan tek portallar genellikle diğer daha büyük portallara giden yolları kapsamaktadır. Meta evrenine bağlantıyı sağlayan ekipmanların (VR Biyometrik verilerin) güvenlik açıkları oluşturması mümkün görülmektedir [52]. Teknolojiler her zaman yeni endişeler ve sorunlarla ortaya çıktığından, meta veri deposu da farklı değildir. Bu teknoloji birden çok boyutta yeni bir gelecek vaat ettiğinden, meta veri deposunun uygulanmasının kullanıcılar, eğitimciler ve öğrenenler için birçok güvenlik endişesi bulunmaktadır. Bunlar:

- Bireyin Kimliği: Sanal kimlik, meta veri deposundaki anahtar kimlik olsa da, kimliği kanıtlama mekanizması çok önemlidir. Yine de, meta veri deposunda kimlik sahtekârlığının nasıl önlenebileceği belirsizdir. Ayrıca sanal evrende bir bireyin fiziksel dünyadaki fiziksel bedenine benzer bir avatar kavramı vardır. Avatarın meta veri tabanında kimliği nasıl koruduğu konusunda birçok güvenlik endişesi bulunmaktadır.
- Kanunlar ve Yönetmelikler: Sanal evrende meydana gelen suçlarla ilgili net bir çözüm mevcut değildir. Bununla başa çıkmak için sanal

evrende mahkemeler ya da hukukla ilgili yönleri ele almak için net bir çözüm yoktur.

- Kötü Amaçlı Yazılımdan Koruma: Şu anda, sanal evreni kötü amaçlı yazılımlardan korumak için kullanılabilecek bir güvenlik çerçevesi yoktur. Bu teknoloji yeni olduğu için, meta veri aktarımını kötü amaçlı yazılımlardan korumak için giderek daha fazla güvenlik standardı geliştirilmelidir.
- Veri Koruması: Meta veri deposundaki verileri korumak için açık yönergeler bulunmamaktadır. Kötü amaçlı insanların hassas ve kişisel bilgileri çalması durumunda, sanal evren kavramı için oldukça olumsuz etkiye sebep olacaktır. Çünkü meta veri tabanının akademik özgürlük, dijital bölünme ve veri gizliliği üzerinde de etkileri vardır.
- Fikri Mülkiyet Hırsızlığı (IP): Fikri Mülkiyetin (IP) korunması kritiktir ve meta veri deposunda güvenliği son derece önemlidir. Fiziksel dünyada IP'yi korumak hala zor iken meta veri tabanında güvenliğini sağlamak daha da zor olacaktır.
- Teknolojisiyle İlgili Tehdit: Meta veri deposu, blok zinciri, kripto para birimleri ve NFT'ler gibi teknolojilere özgü güvenlik endişeleri vardır [40].

Sanal evreni kötü amaçlı yazılımlardan korumak için giderek daha fazla güvenlik standardı geliştirilmelidir. COVID-19 salgını sırasında sosyal medya platformlarında ve uzaktan çalışma ortamlarında birkaç büyük siber suç olayı yaşatmıştır. Uzmanlar, meta veri deposunun bir istisna olmayacağına ve siber suçların, çevrimiçi kötüye kullanımın ve dolandırıcılığın sanal evrende de artabileceğine inanmaktadır. Kullanıcı güvenliği ve her avatara özgü kullanıcı arayüzü oluşturulması amacıyla sanal evrenine girerken tek kullanıcı profili oluşturulması gerekmektedir. Aksi takdirde oluşturulan sahte hesaplar problemlere yol açacaktır. Gerçek dünyanın sanal olarak oluşturulması olan bu ortamda, siber güvenlik yatırımlarının gerçek dünyadaki gibi yapılmasını gerektirmektedir. Sanal evrende olası taciz durumuna karşı gerekli önlemlerin alındığı bir ortam da alınması gereken tedbirler olarak düşünülmelidir. Gerçek dünyanın sanal olarak biçimlenmesi söz konusu iken, olası suçlara karşı uluslararası yaptırımların aktif olması yani sanal

uluslararası mahkemelerin, hâkimlerin, güvenlik görevlisi gibi siber suç bürolarının da olması gerekmektedir.

Araştırmacılar sanal evren uygulamasını çevreleyen diğer riskler, VR/AR cihazlarına aşırı derecede bağımlılık, kullanıcıları gerçek dünyadan ayırabilmektedir. 3B sanal dünya aynı zamanda gerçek hayattaki mutsuzluğu da yaratabilir. Kullanıcılar sanal evrende daha fazla zaman harcayacak ve gerçek dünyadaki görevlerinde gevşeyecektir. Bu nedenlerle sanal dünyaya bağımlılık gelecek nesiller için gerçek bir tehdit olabilir. Öte yandan özellikle çocuk ve genç öğrenenlerden elde edilen görüntü ve ses gibi meta verilerinde manipüle edilmesi söz konusudur. Bu doğrultuda sanal evren ortamına erişimde bir yaş sınırının getirilmesi de gerekmektedir. Sanal evrende kullanıcıların giriş yapabileceği ortamların sınırlandırılması, NFT ve kripto paraların değerlerine göre ödeme ve girişlerinin yapıldığı bir sanal ortamın oluşturulması söz konusudur. Ayrıca, sanal evren ortamına erişim halinde iken kullanıcıların bu dünyada geçirdiği vakit oldukça olası sağlık sorunlarına karşı önem arz etmektedir. Fiziksel ve ruhsal olarak çeşitli hastalıkların da olması da gündemdedir. Sanal evrende VR gözlük ya da herhangi bir ekipman takılı iken kullanıcıların güvenli bir alanda olması gerekmektedir. Fiziksel olarak olası heyecan verici bir koşu ya da denge kaybında gerçek dünyada yaşanabilecek fiziksel yaralanma hatta ölümlere bile sebebiyet verilebilmektedir. Özellikle Z kuşağının günlük yaşamlarında oyun platformlarında geçirdiği sürede bağımlılığın yüksek olduğu ve ailelerin bu durumdan şikâyetçi olduğu bu çağda, sanal evrende olan bağımlılık ve bu ortamı yaşama süresi kullanıcıların fiziksel-ruhsal sorunlarına yönelik bir problem olarak ortaya çıkmaktadır. Tıpkı bedensel gerçeklik gibi, soyguncular da kılık değiştirmiş siber suçlulardır. Bu dolandırıcılar, bu düzenlenmemiş alanı doğrulamak için herhangi bir seçenek olmadığı için saygın ve güvenilir görünen öğeler veya sağlayıcılar da ayarlayabilmektedir.

Sanal evren bir kavram olmaya devam etse de, siber saldırılara son derece yatkındır ve korunması titizlikle düşünülmelidir. Sanal evrende sağlam düzenlemeler getirilmesi gerekmektedir. Her yeni teknoloji beraberinde güvenlik riskleri getirir, aynısı bu teknoloji için de geçerlidir. Şu anda, bu alanda siber güvenliğe yönelik yatırımların yapılması sanal evrende meydana gelebilecek olası tehditleri engellemede rol oynayacaktır.

6.6. SONUÇ VE DEĞERLENDİRMELER

Bu bölümde son yıllarda etkin olan teknolojiler YZ, VR, AR ve Sanal Evren kavramının eğitim amaçlı kullanımı ve bileşenleri ile ilgili genel bilgilendirme yapılmış, sanal evrende siber güvenlik ve olası tehditlere karşı alınabilecek yapay zekâ temelli çözümlere değinilmiş ve sonuç bölümünde genel değerlendirme yapılmış ve gelecekte sanal evrenin niteliği ile ilgili görüş bildirilmiştir. Bölüm ile ilgili genel bir değerlendirme yapıldığında; sanal evren ve bağlantılı teknolojilerin gelecek dönemlerde çeşitli alanlarda özellikle eğitimde oldukça aktif olacağı ön görülmektedir. Bu doğrultuda bazı çıkarımlar yapılmış, öngörülerden bulunulmuş ve aşağıda maddeler halinde verilmiştir. Bunlar:

- Daha hızlı bilgisayarlar ve genişletilmiş bant genişliği, YZ, IoT, uzamsal bilgi işlem, AR/VR, daha iyi akıllı gözlükler ve sürükleyici öğretme ve öğrenme gibi yeni alanlar daha etkileşimli hale gelecektir. Eğitim ortamlarına entegre olması için çalışmaların giderek artan çeşitli uygulamaların geliştirilmeye başladığı sanal evrende, YZ tabanlı yöntemler, doğal dil işleme, makine görüşü, blok zinciri, ağ oluşturma, dijital ikiz ve sinir arayüzleri potansiyel olarak görülmektedir.
- Sanal evrende avatarların hareket etmesi ve gerçek dünyada yapılan hareketlerin sanal dünyada aynı eylemlerin yerine geçmesi kullanıcıların ikinci kişiliği olarak görülmektedir. Bu bağlamda XR, sanal evren, avatarları ve gerçek dünyadaki kullanıcıları birbirine bağlayan ortamdır. İç dünyayı simüle eden bir sanal evren türü olan sanal gerçeklik teknolojisi, avatarlar, gelişmiş 3D grafikler ve anında etkileşim sunmaktadır. Kullanıcıların kendilerini tamamen sanal bir gerçeklik içinde olduklarını hissettikleri bir dünyadır. Sanal evren ile türü arasında en fazla uygulamalarının yapıldığı ve eğitime uyarlandığı görülen VR uygulamaları aktif olarak öğrenenlerin hayatında yer edinmiş- tir. Daha fazla duyu organına hitap edilmesi ve gerçeğe benzer hisleri yaşatma ve üç boyutlu ortam sunan sanal evren, yakın gelecekte daha fazla uygulamaları ile eğitim hayatında olacaktır. Sanal evrende, önceden belirlenmiş bir görev olmadan öğrenenlerin istediği her şey mümkündür. Sınırsız bir özgürlük derecesine dayanarak kullanıcılar, sanal dünyada çalışma, alışveriş, performanslar, sergiler ve tu-

rizm gibi çeşitli eylemleri deneyimleyebilirler. Gökyüzünde uçmak ve uzaya gitmek gibi gerçek dünyadaki fiziksel kısıtlamalar nedeniyle kolayca yaşanması zor olayları da paylaşabilirler. Alternatif olarak, kullanıcılar yoğun yaşamlarından uzaklaşabilir ve sanal evreni çözme görevi olmadan tüm gün balık tutma, meyve toplama veya bir arkadaşının adasına seyahat etme gibi sakin bir hayatın tadını da çıkarabilirler. Öğrenenler sanal evren özelliklerinden aktif olarak yararlanarak, özgürlüklerini ve deneyimlerini sonsuz ölçüde genişletebilecek öğrenme etkinlikleri tasarlamak mümkün olacaktır.

- Öğrenenler, sorularını sonsuz özerkliklerine dayalı olarak keşfetmeleri-ne olanak tanıyan kendi kendilerine öğrenmeyi yürüteceklerdir. Zaman ve mekandan bağımsız sayısız insanın fikirlerini edinebilir ve merak edilen bir konuya yanıt bulabilmek için inisiyatif alabilirler [53].
- 2021'de Meta, dijital akıllı gözlüklerdeki en son teknolojilerini tanıtımında Ray-Ban Stories gözlükler, Ray-Ban'ın ünlü olduğu klasik Wayfarer stili kullanılarak modaya uygun görünmenin yanında son teknoloji ve gizliği sağlamada uyum sağlamak üzere tasarlanmıştır. Gözlükler bir kamera içerdiğinden, gizlilik endişelerini gidermek için kamera açıldığında yanan bir ışıklı göstergeye de sahiptirler. 2022 yılı itibariyle maliyeti 300 dolardan başlamaktadır. Sadece yerleşik teknolojiler değil, aynı zamanda modaya uygun oldukları için VR/ AR gözlüklerinde yapılan önceki denemelere göre büyük bir gelişme olarak görülmektedir. VR, dijital ortamlarda pratik gösterilerin sanal simülasyonlarını sağlayarak eğitimde önemli gelişmelere hizmet edebilir. Örneğin, yangın kontrolüne ilişkin bir risk yönetimi eğitim oturumu gerçek dünyada maliyetli olabilir. Bununla birlikte, yangınla mücadele tatbikatlarının bir VR simülasyonu kolayca uygun maliyetli bir alternatif olarak hizmet etmesi mümkündür. VR/AR teknolojilerinin sanal evren ile yeniden şekillenmesi mümkün görülmektedir.
- AR teknolojisi, nesneleri doğada üç boyutlu hale getirmek için gerçek dünya nesneleri üzerinde dijital kaplamalar oluşturmaya yardımcı olabilir. Çeşitli eğitim çıkarımlarını yönlendirerek, artırılmış gerçeklikle eğitimde sanal evrenin farklı kullanım örneklerini keşfedilebilmek-

tedir. Örneğin AR, görünmez parçaların görsel olarak incelenmesine yardımcı olabilir ve sorunları daha hızlı çözebilir. Okuma-yazma gibi etkileşimli deneyimler de farklı, sürükleyici, ilgi çekici deneyimler yaratabilir [30].

- Meta veri deposunun yardımıyla daha sürükleyici öğrenme deneyimlerine erişebilinir. Ayrıca, öğrenmeyi güçlendirmek için gerçek dünyadaki süreçlerin, sistemlerin ve nesnelerin daha iyi bir simülasyonuna yardımcı olabilmektedir. Meta veri deposu öğrencilerin öğrenme deneyimlerini belgelemelerine ve eleştirel öğrenmenin değerinden yararlanmalarına da yardımcı olabilir. Uzun vadede eğitim sektöründe sanal evren uygulamalarının bu doğrultuda ilerleyeceği mümkün görülmektedir.
- Sağlık, üretim, akıllı şehirler ve oyun gibi destekli çeşitli uygulamaların sanal dünyalarda bulunması söz konudur. YZ, AR ve VR, blok zinciri ve ağ oluşturma gibi diğer teknolojilerle birleştirerek, meta veri deposunun sanal platformda daha güvenli, ölçeklenebilir ve gerçekçi sanal dünyalar oluşturabilmesi amaçlanmaktadır. Yedi katmanlı meta veri tabanı platformuna göre, altyapının güvenilirliğini sağlamak ve performansını iyileştirmek için yapay zekânın önemli rolünün farkına varmak şüphesizdir.
- 5G ve gelecekteki 6G sistemlerinde, otomatik kaynak tahsisi, kanal tahmini, trafiği boşaltma, saldırı önleme ve ağ gibi farklı zorlu görevler için denetimli öğrenme ve pekiştirmeli öğrenme ile birçok gelişmiş makine öğrenmesi algoritması benimsenmiştir. Sensör tabanlı giyilebilir cihazlar ve diğer insan-makine etkileşim araçlarıyla, basit insan hareketleri ve karmaşık eylemler, makine öğrenmesi ve derin öğrenme modellerinin öğrenilmesine dayalı olarak analiz edilebilir ve tanınabilir [54]. Bu doğrultuda meta veri ve meta evren ilgili teknolojiler ile entegre halindedir. Kullanıcıların gerçek dünyadaki hareketleri sanal dünyalara yansıtılarak, kullanıcıların meta veri tabanındaki diğer nesnelerle rahatça etkileşime girmesi için avatarlarını tam olarak kontrol etmelerine olanak tanımaktadır. Bu avatarlar, doğruluk ve işlem hızı açısından YZ tarafından desteklenen konuşma tanıma ve duygu analizinin yanı sıra yüz ifadeleri, duygular, vücut hareketleri ve fiziksel etkileşimler gibi gerçek dünyada benimse- nen birçok modalite ile etkile-

şime girebilecektir. XR/VR bir şekilde başa takılan cihazlar gibi sürükleyici ekipmanlara sahip bir sanal ev- ren temsil etse de YZ, sahne arkasında yaratıcı ve güzel bir dünya inşa etmek için çalışan çok önemli bir teknolojidir. Dolayısıyla YZ ve ilgili teknolojilerin sanal evrende oldukça aktif rol oynayacağını söylemek mümkündür.

- Sanal evrenin oluşumunda rol oynayan VR, AR, MR, XR, YZ ve ilgili teknolojilerin meta veri deposunun oluşumunda oldukça etkin olacağı ön görülmektedir.
- Sanal evren, eğitim için önemli etkileri olan sürükleyici bir ortamdır. Gerçek dünyanın dijital bir ikizi olarak görünün sanal evrende, oyun platformları meta veri tabanından milyarlarca dolar kazanmaktadır. Elon Musk tarafından 2016 yılında kurulan ancak 2020 yılında popülerleşen NeuroLink projesi, beyin-bilgisayar arayüzü (BCI) insanların beyinlerine takılan çipler ile çeşitli hissiyatların kazanılması yönüyle sanal evren ile ilişkili olacağı ön görmektedir. Proje insanlar üzerinde yapılan çalışmalarda olumlu bir yönde giderse sanal evren ortamında kullanıcılar, karakter, nesneler ve dijital işlemlerin beyin sinyalleriyle kontrol edilmesi, koku, korku, heyecan, endişe vb. duyguları yaşama imkânı sağlayabilecektir.
- Facebook yaptığı açıklamalar ile sanal evren ve meta veri geleceğine yatırım yapmaya devam etmektedir. Sanal evren öğreniminin popülaritesi K-12 ve yükseköğretimde ayrışarak çalışmalar hızlandırabilir. Tamamen sanal müfredat bulmak ya da oluşturmak okullar için büyük bir zorluk yaşatacaktır. Ancak hibrit öğretim yöntemleri ile çalışmaların sürdürülmesi mümkün görülmektedir. Örneğin, 2022 yılının sonlarına doğru dokuzu halka açık olan 10 kolej, sanal gerçeklik aracılığıyla sürükleyici sınıflar ve kampüsler sunan VictoryXR aracılığıyla "metaversite" kampüslerini başlatacaktır. Her kolej, öğrencilerin yerinde veya uzaktan eğitim alırken katılabilecekleri bir “dijital ikiz” kampüs kopyası tanıtacaktır. Öğrenciler, eşzamanlı kursları sırasında kullanmak üzere bir Meta Quest-2 VR kulaklığı alacaklardır [55]. Bu durum sanal evren ile ilgili yatırımların hızla artacağı ve ilgili teknolojilerinde bu doğrultuda ilerleyeceğini göstermektedir. Web 3.0, blok zincir, NFT, Kripto varlıklar, IoT, YZ, VR, AR, MR, XR ve çeşitli teknolojiler ile entegre olan sanal evren kavramı insanların hayatlarında oldukça fazla yer alacaktır.

- Her yeni teknolojinin getirdiği zorluklar ve kolaylık olacaktır muhakkak ancak sanal evrenli çözümlere hızla uyum sağlanması ve maliyetlerin makul seviyelere çekilmesi ile sadece eğitimde değil başka alanlarda da etkinliğinin daha da artması ve günlük hayatımızda hızla kullanılması beklenmektedir.

KAYNAKLAR

[1] M. Sparkes, "What is a metaverse," ed: Elsevier, 2021.

[2] P. a. Rospigliosi, "Metaverse or Simulacra? Roblox, Minecraft, Meta and the turn to virtual reality for education, socialisation and work," *Interactive Learning Environments,* vol. 30, no. 1, pp. 1-3, 2022/01/02 2022, doi: 10.1080/10494820.2022.2022899.

[3] S. Mystakidis, "Metaverse," *Encyclopedia,* vol. 2, no. 1, pp. 486-497, 2022. [Online]. Available: https://www.mdpi.com/2673-8392/2/1/31.

[4] M. S. Smith and R. Brooks, "Meta Offers Nothing New to the Metaverse," vol. 59, ed: IEEE-Inst Electrical Electronics Engineers INC 445 Hoes Lane, Piscataway, NJ, 2022, pp. 22-22.

[5] J. McCarthy, "What is artificial intelligence," *URL:* http://www-formal.stanford.edu/jmc/whatisai/, 2004.

[6] M. Chassignol, A. Khoroshavin, A. Klimova, and A. Bilyatdinova, "Artificial Intelligence trends in education: a narrative overview," *Procedia Computer Science,* vol. 136, pp. 16-24, 2018.

[7] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," *Metabolism,* vol. 69, pp. S36-S40, 2017.

[8] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access,* vol. 8, pp. 75264-75278, 2020, doi: 10.1109/ACCESS.2020.2988510.

[9] J. T. Avella, M. Kebritchi, S. G. Nunn, and T. Kanai, "Learning analytics methods, benefits, and challenges in higher education: A systematic literature re- view," *Online Learning,* vol. 20, no. 2, pp. 13-29, 2016.

[10] Y.-S. Tsai and D. Gasevic, "Learning analytics in higher education---challenges and policies: a review of eight learning analytics policies," in *Proceedings of the seventh international learning analytics & knowledge conference*, 2017, pp. 233-242.

[11] D. Kucak, V. Juricic, and G. Dambic, "Machine Learning In Education-A Survey Of Current Research Trends," *Annals of DAAAM & Proceedings,* vol. 29, 2018.

[12] F. Pedro, M. Subosa, A. Rivas, and P. Valverde, "Artificial intelligence in education: Challenges and opportunities for sustainable development," 2019.

[13] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, "Intelligence unleashed: An argument for AI in education," 2016.

[14] N. Elmqaddem, "Augmented reality and virtual reality in education. Myth or reality?," *International journal of emerging technologies in learning,* vol. 14, no. 3, 2019.

[15] S. Helsel, "Virtual reality and education," *Educational Technology,* vol. 32, no. 5, pp. 38-42, 1992.

[16] H. Serin, "Virtual reality in education from the perspective of teachers," *Amazonia investiga,* vol. 9, no. 26, pp. 291-303, 2020.

[17] C. Christou, "Virtual reality in education," in *Affective, interactive and cognitive methods for e-learning design: creating an optimal education experience*: IGI Global, 2010, pp. 228-243.

[18] T. Morimoto *et al.*, "XR (Extended Reality: Virtual Reality, Augmented Reality, Mixed Reality) Technology in Spine Medicine: Status Quo and Quo Vadis," *Journal of Clinical Medicine,* vol. 11, no. 2, p. 470, 2022. [Online]. Available: https://www.mdpi.com/2077-0383/11/2/470.

[19] W. Winn, "A conceptual basis for educational applications of virtual reality," *Technical Publication R-93-9, Human Interface Technology Laboratory of the Washington Technology Center, Seattle: University of Washington,* 1993.

[20] A. S. Mathur, "Low cost virtual reality for medical training," in *2015 IEEE Virtual Reality (VR)*, 2015: IEEE, pp. 345-346.

[21] A. Brown and T. Green, "Virtual reality: Low-cost tools and resources for the classroom," *TechTrends,* vol. 60, no. 5, pp. 517-519, 2016.

[22] A. Sharafy, "Future of education/art: AI/AR/VR/MR, spatial computing & upcoming technos," *C2C Digital Magazine,* vol. 1, no. 16, p. 7, 2022.

[23] N. H. Gandedkar, M. T. Wong, and M. A. Darendeliler, "Role of virtual reality (VR), augmented reality (AR) and artificial intelligence (AI) in tertiary education and research of orthodontics: An insight," in *Seminars in Orthodontics*, 2021, vol. 27, no. 2: Elsevier, pp. 69-77.

[24] J. Y. Lee, "A study on metaverse hype for sustainable growth," *International journal of advanced smart convergence,* vol. 10, no. 3, pp. 72-80, 2021.

[25] "İnternet: Erkan Uçar, *Dijital Evren.* [Online]. Available: https://www.cyberpark.com.tr/content/upload/bulletins/documents/cyberspot25web-20220526104702.pdf, Son Erişim Tarihi: 06.06.2022.

[26] W. S. Khor, B. Baker, K. Amin, A. Chan, K. Patel, and J. Wong, "Augmented and virtual reality in surgery the digital surgical environment: applications, limitations and legal pitfalls," *Annals of translational medicine,* vol. 4, no. 23, 2016.

[27] S. Doolani *et al.*, "A review of extended reality (xr) technologies for manufacturing training," *Technologies,* vol. 8, no. 4, p. 77, 2020.

[28] F. Schroeder, C. Guedes, and L. Kuzmenko, "Mirage-a virtual environment for Ars Electronica's Metaverse 2020: A metaverse for sound art and music, Kunstuniversität Linz sound campus-> Ars Electronica Festival 2020. 9-13 Sept 2020. Metaverse Site: https://sound-campus. itch. io/metaverse, Ars Electronica Site: https://ars. electronica. art/keplersgardens/en/metaverse," 2020.

[29] Optima. [Online]. Available: https://www.optimaclassical.org/.

[30] "İnternet:James Howell," *Metaverse For Education – How Will The Metaverse Change Education?* [Online]. Available: https://101blockchains.com/metaverse-for-education/.

[31] M. Pesce, *VRML browsing and building cyberspace* (no. 006 P473). New Riders Publishing, 1995.

[32] "İnternet: Selma Kasap " *Türk araştırmacılardan, okulları "metaverse" ortamına taşıyan teknoloji*. [Online]. Available: https://www.aa.com.tr/tr/bilim-teknoloji/turk-arastirmacilardan-okullari-metaverse-ortamina-tasiyan-teknoloji/2446867, Son Erişim Tarihi:01.06.2022.

[33] J. Díaz, C. Saldaña, and C. Avila, "Virtual world as a resource for hybrid education," *International Journal of Emerging Technologies in Learning (iJET),* vol. 15, no. 15, pp. 94-109, 2020.

[34] J. Potts and E. Rennie, "Web3 and the creative industries: how blockchains are reshaping business models," in *A Research Agenda for Creative Industries*: Edward Elgar Publishing, 2019.

[35] J. Bambacht and J. Pouwelse, "Web3: A Decentralized Societal Infrastructure for Identity, Trust, Money, and Data," *arXiv preprint arXiv:2203.00398,* 2022.

[36] G. Ragnedda and G. Destefanis, *Blockchain and Web 3.0*. London: Routledge, Taylor and Francis Group, 2019.

[37] F. A. Alabdulwahhab, "Web 3.0: the decentralized web blockchain networks and protocol innovation," in *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*, 2018: IEEE, pp. 1-4.

[38] M. Xu, X. Chen, and G. Kou, "A systematic review of blockchain," *Financial Innovation,* vol. 5, no.1, p.27, 2019/07/04 2019, doi: 10.1186/s40854-019-0147-z.

[39] A. Cannavo and F. Lamberti, "How blockchain, virtual reality, and augmented reality are converging, and why," *IEEE Consumer Electronics Magazine,* vol. 10, no. 5, pp. 6-13, 2020.

[40] "İnternet: Aat Team," *Cyber Security and Metaverse*. [Online]. Available: https://allabouttesting.org/cyber-security-and-metaverse/, Son Erişim Tarihi: 15.05.2022.

[41] K. Chatterjee, A. K. Goharshady, and Y. Velner, "Quantitative analysis of smart contracts," in *European Symposium on Programming*, 2018: Springer, Cham, pp. 739-767.

[42] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the internet of things," *IEEE Access,* vol.4, pp.2292-2303, 2016.

[43] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, "Hawk: The blockchain model of cryptography and privacy-preserving smart contracts," in *2016 IEEE symposium on security and privacy (SP)*, 2016: IEEE, pp. 839-858.

[44] L. Ante, "Smart contracts on the blockchain–A bibliometric analysis and re- view," *Telematics and Informatics,* vol.57, p.101519, 2021.

[45] T. H. Meitinger, "Smart contracts," *Informatik-Spektrum,* vol. 40, no. 4, pp. 371-375, 2017.

[46] M. Kolvart, M. Poola, and A. Rull, "Smart contracts," in *The Future of Law and etechnologies*: Springer, 2016, pp.133-147.

[47] "İnternet: Anna Buczak" *Smart Contract For Fintech - Definition And Use Cases.* [Online]. Available: https://www.ulam.io/blog/smart-contract-definition-use- cases/, Son Erişim Tarihi: 30.05.2022.

[48] K. Vasan, M. Janosov, and A.-L. Barabási, "Quantifying NFT-driven networks in crypto art," *Scientific reports,* vol. 12, no. 1, pp. 1-11, 2022.

[49] "İnternet: Kasey Panetta," *3 Themes Surface in the 2021 Hype Cycle for Emerging Technologies*. [Online]. Available: https://www.gartner.com/smarter with gartner/3-themes-surface-in-the-2021-hype-cycle-for-emerging-technologies, Son Erişim Tarihi: 15.04.2022.

[50] Q. DuPont, *Cryptocurrencies and blockchains*. John Wiley & Sons, 2019.

[51] "İnternet: Seprameen," *Need of Cybersecurity in Metaverse*. [Online]. Available: https://aiexpress.io/need-of-cybersecurity-in-metaverse/, Son Erişim Tarihi: 10.04.2022.

[52] D. van der Merwe, "The Metaverse as Virtual Heterotopia," in *3rd world conference on research in social sciences*, Vienna, Austria, 2021.

[53] B. Kye, N. Han, E. Kim, Y. Park, and S. Jo, "Educational applications of metaverse: possibilities and limitations," *J Educ Eval Health Prof,* vol. 18, no. 0, pp. 32-0, 2021, DOI: 10.3352/jeehp.2021.18.32.

[54] T. Huynh-The, Q. V. Pham, X.-Q. Pham, T. T. Nguyen, Z. Han, and D. S. Kim, "Artificial Intelligence for the Metaverse: A Survey," *arXiv preprint arXiv: 2202.10336,* 2022.

[55] "İnternet: D. Drozdowski, Ed.," *Higher Education in the Metaverse*. [Online]. Available: https://www.bestcolleges.com/news/analysis/2022/03/15/higher-education- inthe-metaverse/, Son Erişim Tarihi: 02.06.2022.

Bölüm 7

BÜYÜK VERİ VE YAPAY ZEKÂ TABANLI KAPALI ALAN KONUM TESPİTİ: İNAVİ

Murat Akın - Emrah Yılmaz - Tuncay Küçükpehlivan - Şeref Sağıroğlu

Bu kitap bölümünde; yönlendirme ve rota oluşturma uygulamalarında konum noktalarının tespiti için kullanılan Küresel Konumlama Sisteminden (GPS) verim alınamadığı ve bu sistemlerin çalışmadığı durumlar için iç mekân konumlama ve yönlendirme yapabilen ve bu kapsamda geliştirilen iNavi adında yeni bir uygulama anlatılmıştır. iNavi uygulaması, iç mekanlarda konum tespitini manyetik alan verisine ek olarak wi-fi teknolojileriyle sağlarken günümüzde önemli olan büyük verinin yönetilmesi noktasında çözümler üretmektedir. iNavi uygulaması, Tübitak Teydeb tarafından desteklenen bir projenin bir kısmıdır. Projenin belirli bir bölümü anlatılan bu kitap bölümünde, öncelikle navigasyon ve öneminden bahsedilmiş, iç mekân konumlama çalışmaları anlatılmış, sonrasında iNavi sistem mimarisi, kullanılan veri türleri ve altyapılar açıklanmıştır. Sonraki başlıkta elde edilen deneysel sonuçlara yer verilmiş ve genel olarak yapılan çalışmalar değerlendirilmiştir.

7.1. NAVİGASYON VE NAVİGASYONUN ÖNEMİ

Türkçe'ye seyir ya da seyrüsefer olarak çevrilen navigasyon, Türkçeleşmiş bir kelime olarak dilimizde "Yol Kılavuzu" olarak kullanılmaktadır. Navigasyon, gidilecek veya hedefe en kısa ve hızlı şekilde erişilmesine yardımcı olan sisteme verilen bir isimdir. En yaygın olarak kullanım alanı ise gidecek olduğumuz bir adresi kimseye sormadan, kaybolmadan, vakit kaybetmeden en hızlı bir şekilde gitmede kullandığımız, bizleri sesli veya görüntülü olarak yönlendiren sistemlerdir.

Navigasyon sistemi, GPS referans sistemi, yazılım, donanım, harita bileşenlerinden oluşmaktadır. GPS sinyalleri kullanılarak, yeryüzü konumumuzu tespit eden ve bu bilgi ile gitmek istediğimiz noktaya ulaşmamızı sağlayacak olan yazılım ve harita ile desteklenen sistem ile bu işlemler sağlanmaktadır. GPS'den alınan koordinat bilgisini harita üzerinde gösterme, gidilecek optimum rotayı belirleme, bunları görüntülü ve sesli olarak yönlendirme ve bunları farklı yolların yön ve hız bilgileri, yasaklar, kısıtlamalar, uyarılar, ilgi çekici noktalar gibi bilgilerin harita üzerinde gösterilmesi gibi hizmetleri içerisinde barındırmaktadır.

GPS sinyalleri kullanılarak; uydu fotoğraflarından ve karasal taramalarla elde edilen dijital haritalar üzerinde koordinat bilgilerini içeren GPS sinyalleri ile yerlerini bulabilir, haritada gideceğimiz yerin, ülke, il, ilçe, mahalle, cadde-sokak ve kapı nosu olarak girişlerini yapabilir, koordinat bilgisi kullanılarak yüzlerce hesaplama yaparak farklı tanımlamaları destekleyebilmektedir. IGO, Garmin, TomTom gibi navigasyon hizmetleri ile haritalar toplanmakta ve ayrıca bu hizmeti ülkemizde Başarsoft Bilgi Sistemleri A.Ş. kolaylıkla sunabilmekte ve hizmetler daha kolay ve hızlı verilmektedir.

Kent ve şehir merkezlerinde kırsal nüfusa oranla insanların kapalı mekanlarda zaman geçirme eğilimi fazladır ve bu eğilim birçok problemi beraberinde getirmektedir. Nüfusun yoğun kesimlerinde AVM, hastane, kamu binaları gitgide daha büyük oranlarda inşa edilmektedir. Yapılardaki bu büyüme insanların kullanımını hedef bulmalarını zorlaştırırken, kapalı alanlarda personel, donanım takibi de GPS'in kapalı alanlarda çalışmaması nedeniyle açık alanda olduğu gibi yapılamamakta, konum tabanlı çözümler kapalı alanlarda geliştirilememektedir. Açık alanlarda insanların alışageldiği konumlandırma servislerinin kapalı alanlarda da hizmet vermesi için donanım

tabanlı çözümler kullanılmakta, bu çözümler hem verimli sonuçlar sağlamamakta hem de donanımların tamamı yurtdışı (Çin vb. ülkeler) menşeli olduklarından döviz kaybına neden olmaktadırlar.

Günümüz şehirlerinde çok katlı, karmaşık ve geniş alanlara yayılmış binaların sayısı artış gösterirken, onlarca kat, yüzlerce koridor, oda ve geçitlerden oluşan bu binalar hem karmaşık yapıları hem de barındırdıkları nüfus itibarı ile birer küçük şehir halini almıştır. Dolayısıyla binaların büyüklüğü ve karmaşıklığına bağlı olarak çözülmeyi bekleyen birçok yeni problem karşımıza çıkmaktadır. Meydana gelebilecek acil durumlarda, binaların hızlı ve sorunsuz bir şekilde tahliye edilmesi gerekmektedir. Görme engellilerin yönlendirilmesi, güvenlik amaçlı uygulamalar, ziyaretçi takibi, adres bulma, hizmetlerin organizasyonu, otomatik turist rehberliği gibi uygulamalar çözümlenmesi gereken diğer problemlerdir. Ancak söz konusu problemleri çözebilecek düzeyde kullanılabilir sistemler yaygın değildir. Açık alanlarda konum belirlemede en yaygın olarak kullanılan GPS, alıcı ve uydu arasında doğrudan görüş gerektirdiği için kapalı ortamlarda başarılı sonuçlar verememektedir. Bununla birlikte Küresel Navigasyon Uydu Sistemi'nin (GNSS) de kapalı alanlarda doğrudan kullanılması mümkün değildir. Standart konumlama ve rotalama metotları kapalı alanlarda kullanılamadığından bina içlerinde de çalışabilecek etkin konum belirleme sistemlerine ihtiyaç vardır.

Bu çalışmada geliştirilen sistem ile gerçek büyük verinin oluşturulması için büyük veri toplama platformu, veri toplayıcı algoritmaların (ön hazırlıksız manyetik alan, wi-fi, barometre, görüntü verilerinin toplama algoritmaları) geliştirilmesi süreci tamamlanmış, sonrasında derin öğrenme modelleri kullanılarak pozisyonlama işlemi gerçekleştirilmiştir. Çalışma kapsamında elde edilecek çıktı uygulamayla geliştirilecek çözümün ülke ekonomisine girdi oluşturup ticarileşebilmesi için değer katan servisler sisteme eklenecek, dünya pazarına açılma yolunda çalışmalar gerçekleştirilecektir.

7.2. İÇ MEKAN KONUMLAMA ÇALIŞMALARI

GPS sistemlerinin ortaya çıkmasıyla uydu tabanlı dış mekân konumlama sistemleri, hayatı kolaylaştırmak ve bazı durumlarda güvenlik konularına katkı sağlamak için günlük yaşamımıza girmiştir. GPS sistemleri sayesinde havayolu araçları güvenli şekilde seyahat ederken, araç navigasyon sistemle-

ri kullanıcıları istenilen hedefe yönlendirmektedir. Dış mekân konumlama sistemleri tipik olarak GPS sinyalleri ile çalışabilirken iç mekânlarda GPS sinyali alınamadığından iç mekân konumlama sistemleri kullanıcıyı yerel alanda hareket sağlamak için farklı teknikler üzerinde yoğunlaşmıştır. Farklı teknikler ve farklı alanlarda iç mekân konumlama üzerine yapılan çalışmalar incelendiğinde çalışmalar dört ana başlık altında toplanmıştır:

7.2.1. Veri Toplama

Geçmiş dönemde yapılan birçok çalışma deneysel üretilen verilere dayanmaktadır ve bu çalışmalar yaya davranış karakteristiğini farklı geometrik şekiller ile ifade etmeye çalışmışlardır [1]. Ayrıca Havalimanları, şehir hastaneleri ve alışveriş merkezleri gibi günümüzde büyük ve karmaşık olarak büyük yapılar inşa edilmekte olup gezinme servislerinin oluşturulması ve yaya davranışlarının anlaşılması için güvenilir veriye ihtiyaç duyulmaktadır [2][2].

Yapılan çalışmada, gezinme yapılacak bir yapı ile ilgili herhangi bir bilgi gerekmeksizin konumlama ve harita çıkarımı için gereken gerçek veriler toplanarak büyük veri altlığı oluşturulmuştur. Veriler, Android 7, 8, 10 ve 11 işletim sistemine sahip pusula, ivmeölçer ve barındıran mobil akıllı telefonlar ile toplanmıştır. Bu yazılıma sahip cihazların seçilme nedeni, Android 9 işletim sisteminde Wi-Fi ağlarının taranmasında kısıtlama olması ve ayrıca iOS yüklü cihazlar Wi-Fi ağlarının taranmasına izin vermemesidir. Veri toplama işlemi, kullanıcıların mobil uygulamadan ilgili binayı seçerek mobil uygulamada başlat düğmesi ile tetiklemesi sonrasında gerçekleşmeye başlayacaktır. Kullanıcı ilgili binayı bir desen izlenmeden rastgele dolaşırken, dolaşma esnasında mümkün olduğunca kameradan da kayıt alınmıştır. Uygulamada bulunan durdurma butonu tıklandığında; kullanıcı, toplanan verilerle binanın veri toplanma yüzdesi ve binanın hangi bölümlerinin eksik, hangi bölümlerinin tamamlandığını görülebilmektedir.

7.2.2. Harita oluşturulması

Bir iç mekân gezinme sisteminin asıl amacı akıllı telefon vb. cihazlarla birlikte sürekli, güvenilir ve doğru konumlama sağlamaktır. Haritalar ise doğru bir iç mekân konumu sağlamada anahtar bir rol üstlenmektedirler ve harita sağlayan firmalar için harita çıkarımı mutlak bir önceliktir. Bir harita,

kısıtlamalar sağlayan uzamsal bir grafik görüntüsü olarak ifade edilirken ve en yalın haliyle bir binanın kat planı şeklinde çizilebilmektedir. Google ve Open Street haritalar gibi birçok harita platformu yeterli coğrafik ve grafik ara yüzü dış mekanlar için sağlarken çok katlı iç mekanlar için benzer coğrafik bilgi sistemi ve ara yüz bulunmamaktadır [3]. Konum ismi veya oda numarası gibi gösterimlerle kıyaslandığında grafiksel gösterimler içeren bir koordinat sistemi, iç mekân konumunu ve ortamı açık bir şekilde ortaya koyabilmektedir. Yapılan bu çalışma ile konumlama yapılacak yerin daha iyi analiz edilmesi ve daha doğru gezinmek için gezinilecek yapıların 2, 2.5 veya 3 boyutlu grafiksel olarak gösterilmesi amaçlanmıştır.

Standart harita ve navigasyon uygulamalarında, haritanın oluşturulma hızı ve rota hesaplama süresi genel performans kriterleri olarak düşünülmektedir, fakat kapalı alan gezinmede farklı kriterler de göz önünde bulundurulmaktadır. Coğrafik olarak dünya düzleminin vektörel verilerinin dışında kapalı alana ait detaylı koordinat bilgileri ve düzlemi harita üzerinde gösterilmektedir. iNavi ilk sürümünde (iNavi V1.0) aynı anda sadece tek bir binanın çizim ve pozisyonlama verileri kullanılabilirken, mevcut verinin çok büyük boyutta olmasından dolayı bina verisinin açılma ve işlenme süreci oldukça zaman almaktadır. Çalışma kapsamında elde edilen büyük verinin bir şehre ait çizim ve pozisyonlama verileri ile birlikte kapalı alanların aynı anda işlenebilmesi ve kullanıma hazır olması planlanmakta, kullanıcıların topladıkları yön, mesafe ve görüntü verileri kullanılarak binanın çizimi çıkarılarak ve sonrasında bu çizimlerden üç boyutlu modelleri oluşturulabilmektedir.

7.2.3. Konum Tespiti

İç mekân gezinme, yerelleştirme ve konum tespiti olarak iki kavramı kapsamaktadır ve bu iki kavram sıklıkla birbirinin yerine de kullanılabilmektedir [4]. Yerelleştirme, oda veya kapalı bir alanı ifade ederken konum tespiti ise hedefin haritaya göre konumunu işaret etmektedir. Konum tabanlı akıllı uygulamaların ve servislerin iç mekanda kullanılabilmesi için konum tespiti, üzerinde yoğun çalışılan bir konudur ve duvar veya katı cisimlerin engellemesinden dolayı dış mekanda kullanılan GPS sinyalleri iç mekan için yüksek doğruluk sağlamamaktadır [5]. Çalışmamızda konum tespit doğruluğunu arttırmak için; manyetik alan, Wi-Fi, görüntü verisi ve hibrid veri kullanılmaktadır.

Yapılan çalışmada iNavi V1.0 için iyileştirilmesi gerekenler, ilk konum tespit hızı ve konum doğruluğudur. İlk konum tespit hızının Wi-Fi desteği ile ciddi oranda artırılması planlanmıştır. Ayrıca, kullanıcının kamera kullanımına izin verdiği senaryoda cihazın kamerasından görüntü alınması ile kapalı alanın önceden işlenmiş görüntü verileri üzerinden ilk konum tespiti yapılabilecek ve böylelikle uygulama performansı arttırılacaktır. İç mekânda konumlama yapılırken kullanıcının adım tespiti de gerekmektedir ve azami hata ile adım tespiti, konumlamanın kalitesini oldukça arttırmaktadır. Buna paralel olarak kullanıcının cihazı tutuş açısı, yönden de adım tespiti ve pozisyonlamanın etkilenmemesi hedeflenmiştir. Bu verilerin gerçeğe yakınlığı önemli bir performans parametresi olarak düşünülmüştür. Bunun dışında kapalı alanlarda kullanıcının kat değişiminin tespit edilmesi de iNavi projesinin ilk versiyonunda dikkate alınmamıştır. Barometre ve ivme sensörü kullanarak elde edilen veriler incelendiğinde uygulanan filtreler yardımıyla "benzersiz" değişimlere sahip olduğu gözlenmiş, bu bilgiler ışığında kat değişim tespitinin eldeki veriler ışığında mümkün olduğuna karar verilmiştir. Bina kat değişiminin doğru tespiti (çıkılan/inilen kat sayısı) ve hangi araç ile yapıldığı önemli bir performans kriteri olarak belirlenmiştir. Bu kriterler göz önünde bulundurularak yapılan iç mekân gezinme sistemi için toplanan manyetik alan, Wi-Fi ve görüntü verileri zaman damgası ile belirlenen konum bilgileriyle eşlenmektedir. Eşleme işlemi sonrasında veriler makine öğrenme yöntemi ile eğitilerek daha önce verisi toplanmamış bina bölümlerinin tahmini, yapay sinir ağları kullanılarak yapılmış ve sonuçta geliştirilen tahmin modeli ile konum tespit edilmiştir.

7.2.4. Rota Oluşturma ve Yönlendirme

Konum tespiti yapılan iç mekanlarda kullanıcının mevcut konumundan hedef konumuna ulaştırılması gerekmektedir. Hedef konuma ulaşmak için bir güzergâh (rota) belirlenmekte ve kullanıcının anlayacağı şekilde basit yönlendirmeler yapılmaktadır. Rota belirlenmesi bir gezinme gerçekleştirmek için son derece önemlidir ve sistemin genel performansını da etkilemektedir [6]. Akıllı bir rota planlama, kullanıcı isteklerini göz önünde bulundururken rotayı da buna göre şekillendirmelidir. Birçok kullanıcının beklentisi en kısa yoldan ve en kısa sürede hedefe ulaşacak rota oluşturulmasıdır [7]. Birçok çalışma, en kısa yol üzerine rota oluşturma üzerine yapılmışsa da farklı kullanıcıların farklı rota beklentileri olabilmektedir. Örneğin; görme engellilerin bir iç mekânda

rahatça hareket edebilecekleri sistem tasarlanabilmekte [8] veya otomatik yönlendirilmiş personel aracı (forklift) için gezinme sağlanabilmektedir [9]. Tüm bu hususlar dikkate alındığında yapılan bu çalışmada; kullanıcıların arzu ettikleri hedefe gidebilmeleri için en uygun yolun gösterilmesi, rota oluşturma hizmetinin çevrimdışı mobil cihazla yapılabilmesi ve farklı kullanıcı profillerine göre rotalama hizmetinin oluşturulması hedeflenmiştir.

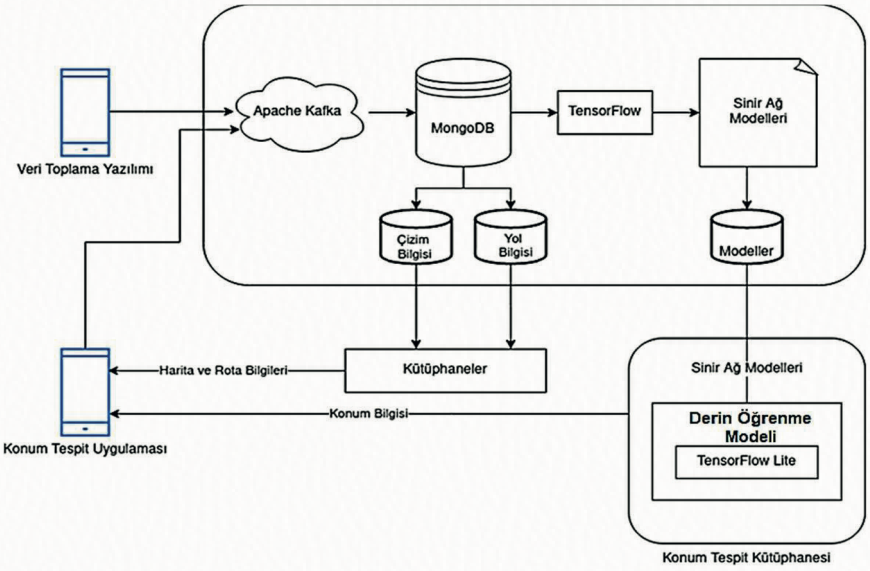
Harita oluşturma sürecinde elde edilen ve kullanıcılardan toplanan veriler (ivme, manyetik alan, görüntü verisi vb.) ile birlikte kapalı alan içerisindeki yürüme yolları ve kat değişim araçlarının (asansör, merdiven vb.) konumları belirlenmekte ve rota oluşturma algoritması bu belirlenen konum verilerini kullanılarak rota oluşturulmaktadır. Mevcut konumdan belirlenen bir noktaya doğru oluşturulan rotalar, kullanıcı tercihlerine alternatif yolları sunmaktadır. Farklı katlar için oluşturulacak rotalar kullanıcının tercihine sunulmakta, kullanıcı isteğine göre asansör veya yürüyen merdiven kullanılabilmektedir. Örneğin engelli bireyler için oluşturulan farklı profiller ile tekerlekli sandalyedeki bir bireyin kat değişimi yapması gerektiğinde oluşan rota merdiven kullanmadan asansör üzerinden oluşturulmaktadır. Farklı bir örnekte; hastane içerisinde sadece yetkili personelin geçiş izni olan alanlar ve ziyaretçilerin kullanamayacağı alanlarda bulunmaktadır. Dolayısıyla ziyaretçi profili için oluşturulacak rotalar bu alanları kullanmadan oluşturulmaktadır.

Araştırmaların bu şekilde çalışma yöntemlerine göre sınıflandırılmasının yanı sıra kullandıkları tekniklerle de ilgili bir ayrıştırma yapılabilmektedir. İç mekân konumlama sistemleri mesafe ölçüm safhası ve konum tahmin safhası olarak iki aşamadan oluşan ve GPS sinyallerinden bağımsız sınırlandırılmış alanlarda nesnelerin konumunu tahmin eden sistemler olarak ifade edilmektedir [10]. Kapalı alandaki nesnelere ait konum bilgi tespiti için genellikle alınan sinyalin güç indikatörü (RSSI), sinyal varış süresi (ToA), varış süre farkı (TDoA), varış açısı (AoA) ve kanal durum bilgisi (CSI) gibi bilgiler kullanılırken çoklu noktalama, üçgen yöntemi ve fingerprint vb. geliştirilen yöntemler de yerelleştirme ve konum bilgileri için kullanılmaktadır [11]. İç mekân sistemlerde önemli olan safha ilk konum tespitidir ve yapay zekâ teknikleri (makine ve derin öğrenme) iç mekan konumlamada geleneksel yöntemlerin kısıtlarını ve dezavantajlarını ortadan kaldırmak için sıklıkla kullanılmaktadır. Örneğin; iç mekân konumlama için görüş açısı dışında kalan alanlarda tanımlama ve görüş açısından dolayı kaynaklanan hesaplama hatalarını azaltmak için makine öğrenme [12], [13] ve derin öğrenme [14] yöntemlerinden fayda-

lanılmıştır. Başka bir çalışmada; RSSI dalgalanması ve dolayısıyla konum tespitinde hataya sebep olan etkileri azaltmak adına zaman aralıklarında RSSI ilişkileri keşfedilerek ve benzer örüntüler dikkate alınarak RNN yöntemi kullanılmıştır [15]. Ayrıca, parmakizi (fingerprint) verisinin saklanması için gerekli alanı ve karmaşıklığını azaltmak için boyut indirgeme yöntemleri [16] veya parmakizi verisini depolamayı azalmak için RSSI özniteliklerini çıkarmak için boyut arttırma yöntemi de kullanılmıştır [17].

7.3. GELİŞTİRİLEN SİSTEM MİMARİSİ

Bu bölümde geliştirilen sisteme ait oluşturulan mimarisi Şekil 7.1'de verilmiştir. Geliştirilen modelde iki önemli unsur, veri toplama ve konum tespitidir. Veri toplama ekibinden ve son kullanıcılardan gelen akan veri Apache Kafka ile karşılanıp belirlenen formatta MongoDB veri tabanına yazılmaktadır. Belirli zaman aralıklarında, MongoDB üzerindeki veriler işlenerek harita ve yol verileri için veri tabanları oluşturulmuştur. Konum tespiti için geliştirilen özel bir derin öğrenme algoritması kullanılarak TensorFlow yardımıyla MongoDB üzerindeki ham veriler kullanılara model eğitilmekte ve bu sayede konum tespit modeli oluşturulmakta ve tespiti yapılabilmektedir. Bütün bu süreçlerde veriler toplanmış ve toplanan veriler kullanıma hazır hale getirilmiştir.



Şekil 7.1. Geliştirilen iNavi Sistem Mimarisi

Geliştirilen uygulamanın yüklü olduğu mobil uygulamanın kullanımı esnasında, sunucuda hazır hale getirilmiş veri tabanları ve konum tespit kütüphanesi kullanıcının mobil cihazına indirilebilmektedir. İlgilenilen kapalı alan, harita üzerinde harita kütüphanesi vasıtasıyla gösterilmekte, konum tespit modeli kullanılarakta kullanıcının konumu tespit edilmektedir. Her iki işlemin ardından kullanıcı için rota oluşturulmakta ve iç mekân içinde gezinme sağlanmaktadır. Ayrıca, kullanıcının konumu tespit edildikten sonra konum bilgisi ve ölçülen sensör verileri sunucuya aktarılarak konum tespit modeli oluşturulmakta ve tespit yapılabilmektedir.

Daha önce de belirtildiği gibi kapalı alan içerisinde rastgele yürünerek toplanmış veriler ile harita ve parmakizi (fingerprint) örüntüsü oluşturulduğundan, veri toplayıcı kişinin uyması gereken standart bir kural yoktur. Burada, dikkat edilmesi gereken husus, kapalı alan içerisindeki yürünebilir her alandan veri toplanmasıdır.

7.3.1. Sistemde Kullanılacak Veri Türleri

Kapalı alanlarda gezinme esnasında okunan veya toplanan veriler, mobil cihazda aşağıdaki biçimde depolanmaktadır:

- İvme-Jireskop : x, y, z
- Manyetik alan verisi: mx, my, mz, zaman damgası, yapı bilgisi
- Wi-Fi: bssid, freq, seviye bilgisi, zaman damgası, yapı bilgisi
- Görüntü: video (MPEG-4 h.265), zaman damgası
- Barometre: pa, zaman damgası

7.3.2. Alt Sistem ve Yapılar

Tasarlanan sistemde kullanılacak teknolojilerin belirlenmesinin ardından geliştirme ortamı için hazırlanan alt sistemler aşağıda maddeler halinde verilmiştir;

Apache Kafka kurulumu yapılmıştır.

- Üç farklı sunucuya MongoDB kurularak, biri ana diğerleri yedek olacak şekilde yapılandırılmıştır. Böylece beklenmeyen durumlara karşı önlem alınarak, oluşacak kayıt dosyalarının yedeklenmesi ile

istenilen bir zaman dilimine geri dönüş mümkün hale getirilmektedir. Bina yapısı, ızgara çizgileri, kat yapıları, ham veri, veri, kenar çizgileri, oda kümeleri oluşturulmuş, ham veri içerisinde kat, kullanıcı, zaman bilgisi, manyetik, Wi-Fi, kamera alanları, veri içerisinde ise kat, konum, manyetik, Wi-Fi, kamera alanları belirlenmiştir.

- Mevcut veriyi MongoDB üzerine taşımak için özel bir betik yazılmıştır.
- MongoDB üzerindeki harita verilerinden veri tabanı oluşturmak için özel bir betik yazılmış ve Sqlite veritabanı kurulmuştur.
- MongoDB üzerindeki rota oluşturma için gereken verilerden yol verileri oluşturmak için betik yazılmıştır.
- MongoDB üzerindeki manyetik alan ve Wi-Fi verilerinden RBF modelleri oluşturmak için uygulama yazılımı geliştirilmiştir.
- Yazılan betik ve uygulamalar test edilmiştir.

Bir üst başlıkta belirtilen veri standartlarını kullanılarak, mevcut sistemlerin verilerini yeni oluşturulan sisteme aktarabilmek için API geliştirilmiştir. Bu API sayesinde, mevcut mobil ve web uygulamasından gelen istekler yakalanarak ilgili cevaplar verilebilmektedir. Bu sayede mevcut verilerin kullanılması için toplanılan tüm verilerin belirlenen format ve standartlara uygun hale getirilmesi sağlanmıştır. Verilerin dönüşümü sırasında uygun olanlar akan veri besleyicileri vasıtasıyla büyük veri sistemine entegre edilerek modellere katkı sağlayacak hale getirilmiş ve sistem için kurulan büyük veri sistemine aktarılması sağlanmıştır.

Planlanan sistemde ön hazırlıksız olarak:

- Manyetik Alan Verileri
- Wi-Fi Verileri
- Jiroskop Verileri
- Barometre Verileri ve
- Görüntü Verileri

toplanmıştır. Bahsedilen ön hazırlıksız veri toplama algoritmalarının geliştirme aşamasında mevcutta konum bilgisi olmayan verilere ihtiyaç duyulmakta fakat mevcuttaki verilerin tamamı konum bilgisi barındırmaktadır. Dolayısıy-

la, mevcuttaki verilerin konum bilgileri kaldırılarak büyük veri sistemine aktarılması sağlanmıştır. Sistem mimarisi temel alınarak, veri toplama ekibinden ve son kullanıcılardan gelen akan veri Apache Kafka ile karşılanacak şekilde tasarlanmıştır. Bu tasarıma uygun olarak; elde edilecek akan veri için sunucuya Apache Kafka kurulumu yapılırken mevcuttaki son kullanıcı uygulamasına da Kafka kütüphanesi eklenmiştir. Böylelikle Kafka kütüphanesi ile mobil uygulamada belirlenen konumların kurulmuş olan Kafka sunucusuna gönderilmesi sağlanmıştır. Ayrıca, Kafka sunucusuna gönderilen veride bina, cihaz, koordinat ve zaman damgası bilgileri de bulunmaktadır.

7.3.3. Veri Depolama ve Gereklilikler

Konum verisinin depolanmasındaki en önemli hususlardan bir tanesi verinin anonim olmasıdır. Veri alanlarından biri olan AndroidID bu konuda oldukça önemli bir yer teşkil etmektedir. Anonimleştirme için daha önceleri akıllı telefonlarda yakın zamana kadar herhangi bir lisanslama işleminde veya cihaza özel bir tanımlayıcı bir numaraya ihtiyaç duyulduğunda cihazın IMEI numarası kullanılırken, IMEI numarasının kullanılması tamamen cihaza özel ve değişemez bir ID olması sebebiyle veri gizliliği konusunda endişeleri beraberinde getirmektedir. Gizliliğin ihlal edilme ihtimalinden dolayı akıllı telefon işletim sistemleri (Android, iOS) geliştiricilerin IMEI numarasına erişmesini engellerken geliştiricilere alternatif olarak birçok seçenek sunmaktadır. Fakat hiçbiri tam olarak değişemez ve cihaza özel olma özelliklerine sahip değildir. Bu kapsamda sunulan alternatifleri değerlendirdiğimizde, cihazda açık olan oturum ve cihaz eşleştirme işlemi sonucunda Google tarafından oluşturulan AndroidID yetenekleri bakımından sistem gereksinimlerini karşılamaktadır. Tersine mühendislik ile cihazın kendisine veya cihazda açık olan oturuma erişilemeyecek şekilde oluşturulan AndroidID bilgiler, toplanan verinin anonim olmasını sağlarken aynı zamanda toplamış olduğumuz veriyi de anlamlandırmamıza katkı vermektedir.

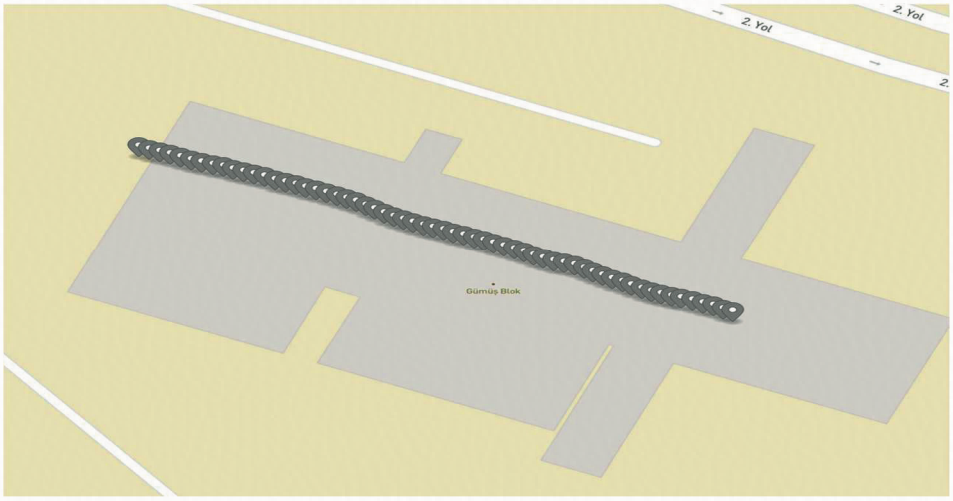
7.4. DENEYSEL SONUÇLAR

Bu bölüme kadar üst başlıklarda iNavi uygulamasının sistem mimarisi ve kullanılan veri türleri ve alt yapılar anlatılmıştır. Bu başlıkta ise elde edilen sonuçlar sunulmuştur. Deneysel sonuçlar, literatür çalışmalarının anlatıldığı

bölüme göre, bu alanda yapılan çalışmaların temel olarak ayrıştığı haritanın elde edilmesi, verinin toplanması ve konum tespiti olarak alt başlıklarda açıklanmıştır.

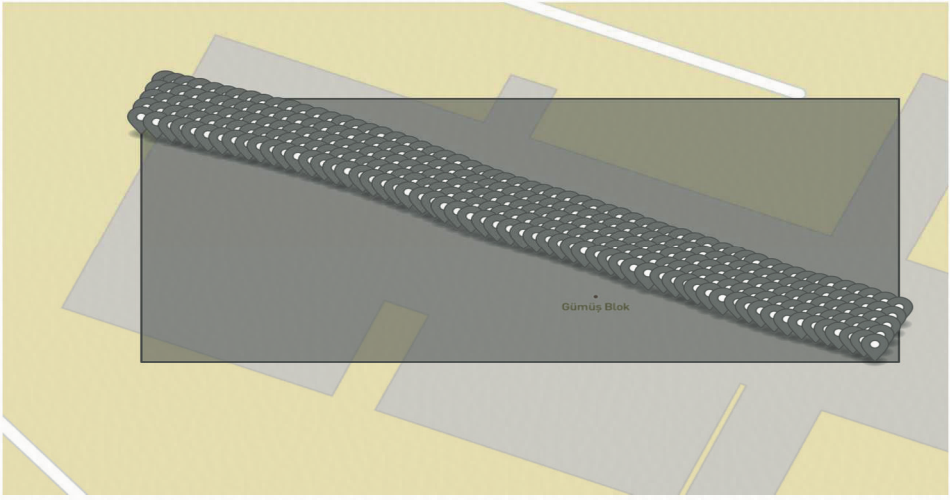
7.4.1. Haritanın Elde Edilmesi

Harita elde edilme sürecinde; tüm veri tipleri için önceden elde edilen koridor planlarının sisteme tamamen yüklenmesi, manyetik alan verilerinin toplanması gibi işlemler yerine yürünebilir alanlara veri çizgileri çizilip bu çizgiler üzerinde tek tek yürünerek, daha basit hale getirmek hedeflenmiştir. Bu bölümde anlatılan bütün işlemler, tüm veri tipleri için geliştirilecek olan algoritmaların temellerini oluşturmaktadır. Öncelikle veri toplama uygulamasında yapılan değişiklik ile kullanıcının sistemine hiçbir plan yüklemesi yapmadan veri toplama işlemine başlamasını sağlayacak şekilde geliştirme yapılmıştır. Bu geliştirme sırasında, önceden kullanılan veri çizgileri üzerindeki ızgara (grid) noktalarına verileri kaydetmek yerine tüm veri ham halde kaydedilmiştir. Ham haldeki bu verilerin uzaktaki sunucuya otomatik olarak iletilmesi sağlanmış, sunucu tarafında geliştirilen algoritma ile ilk olarak kullanıcının mobil cihazı üzerinden okunan kullanıcının yönelim açı verisi (orientation) ve başlangıç noktası ile yürüme yolu oluşturulmuştur. Bu işlemin örnek gösterimini ve dünya koordinat sistemine dönüştürülmüş hali Şekil 7.2'de verilmiştir.



Şekil 7.2. Örnek bir yürüme yolu

Kullanıcının izlediği yolun belirlenmesinden sonra bu yol üzerindeki yürünebilir alanların tespiti yapılmıştır. Bu alanın tespiti yapılırken yine oryantasyon verisi kullanılmış olup koridor genişliği boyunca rota genişletilerek Şekil 7.3'de görülebilen sınırları belirlenmiş bir yüzey alanı elde edilmiştir.



Şekil 7.3. Sınırları belirlenen yüzey alanı

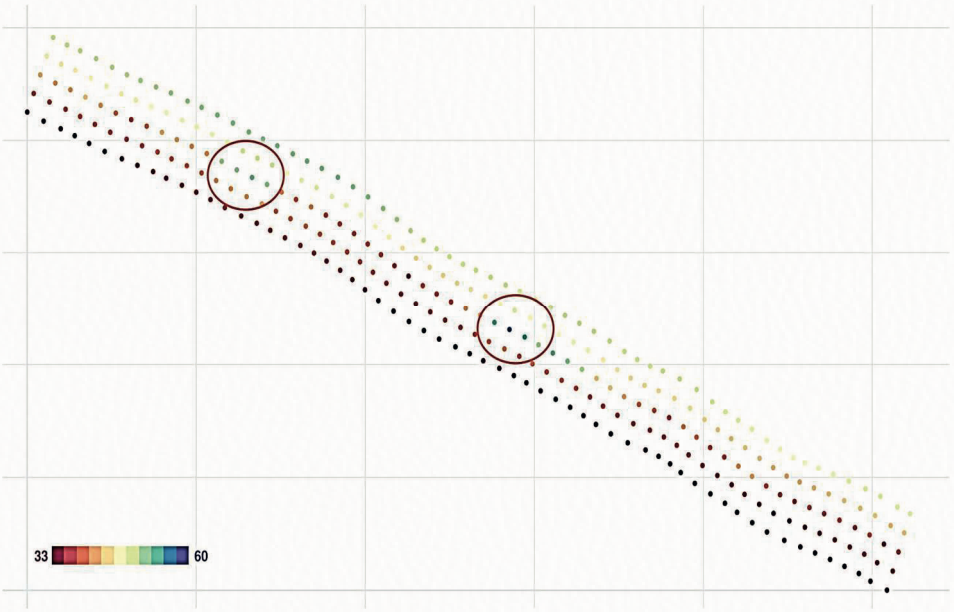
7.4.2. Veri Toplanması

Yüzey alanı oluştuktan sonraki aşama; oluşan yüzeyde her bir noktaya sensör verilerinden okunan değerlerin tahmini ve atanması işlemidir. Bu işlem sırasında sadece doğrusal yürümede elde edilen sensör verileri kullanılmıştır. Bu tahminleme sırasında Radyal-Tabanlı model kullanılmış, bunun sonucunda tüm sensör verileri yani manyetik alan, oryantasyon, ivme, barometre, Wi-Fi sinyalleriyle yüzey alandaki tüm noktalar tahminlenmiştir.

Manyetik alan verilerinin toplanmasında ön hazırlık olsun veya olmasın, mobil işletim sistemlerinde sensörlerden elde edilen veri hassasiyeti oldukça düşük olup, göreceli hassasiyet sağlayan bir biçimdedir. Bu sebeplerden dolayı, manyetik alan verilerinin okunup işlenmesinden önce cihazın 3 eksende döndürülerek kalibre edilmesi gerekmektedir. Bu kalibrasyon işlemi için mobil işletim sistemlerinin sağladığı sistemler mevcuttur. İşletim sistemlerinin sağladığı kalibrasyon sistemi gerekli hassasiyette olmadığı için

geliştirilen bir kalibrasyon sistemi halihazırda kullanılmaktadır. Bu sistem kullanıcıları zorlayabilecek bir seviyede olduğundan dolayı öncelikli amacımız kalibrasyonu tamamen kaldırırken aynı zamanda başarısız olunan senaryoda da istenilen hassasiyette kalibrasyonu sağlayan sistemi kullanıcı açısından basitleştirmektir. Yapılan çalışmalarda, kalibrasyonun tamamen kaldırıldığı senaryo için istenilen sonuçlar alınamamış ve elde edilen veri hassasiyetini istenilen seviyeye getirilememiştir. Bundan dolayı mevcut kalibrasyon sistemi kullanıcı tarafında iyileştirmeye gidilerek hesaplama algoritmamız değiştirilmiş, kullanıcı tarafında yapılan geliştirme kısmında işlemi oyunlaştırılarak süreç daha basit hale getirilmiştir. Böylelikle hem istenilen seviyede bir kalibrasyon hassasiyeti sağlanırken hem de kullanıcı tarafından daha anlaşılır ve kolay bir kalibrasyon işlemi elde edilmiştir.

Tüm bu bilgiler ışığında, herhangi bir plan sisteme yüklenmeden, koridorun sadece orta kısmında yürünerek geliştirilen algoritma ile koridorun tamamının modelinin minimum hata ile oluşturulması sağlanmıştır. Geliştirilen algoritma ile toplanan örnek veri setinin daha anlaşılabilir olması için bir kütüphane yardımıyla oluşan manyetik modelin görseli Şekil 7.4'deki gibidir.

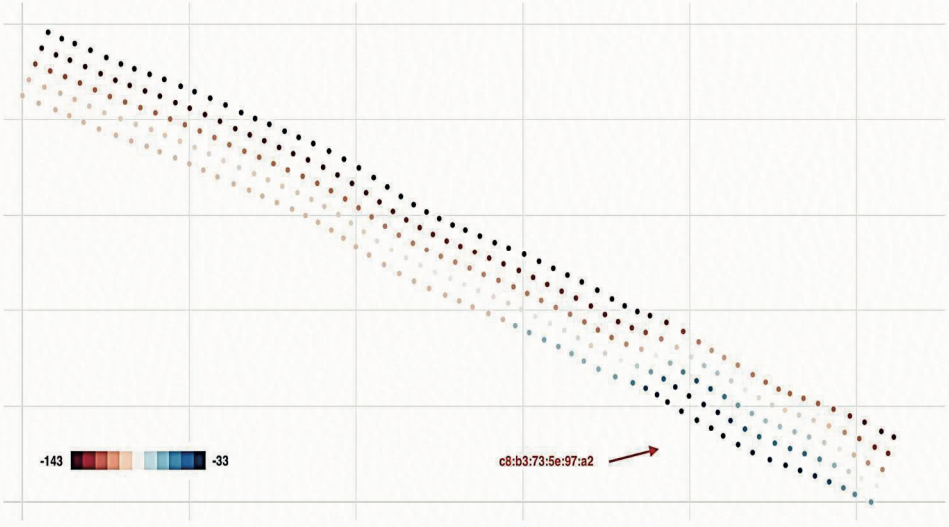


Şekil 7.4. Oluşturulan manyetik model

Şekil 7.4'de görselleştirilen modelin grafiği incelendiğinde, özellikle önemli noktalardan olan merdivenlerdeki (kırmızı daire işaretli) manyetik bozunumun doğru bir şekilde tahminlendiği, testler öncesi manyetik modelin istenilen seviyeye ulaştığı görüntülenmektedir.

Ön hazırlıksız Wi-Fi verisinin toplanmasında belirlenen veri toplama formatı esas alınmıştır. Sinyalin yayınını yapan cihazı belirten tanımlayıcı numara, yapılan sinyalin frekansını belirten değişken, sinyalin cihaza olan uzaklığı ve yapılan sinyal seviyesine bağlı olarak değişen sinyal gücünü gösteren seviye verisi dışında zaman göstergesi için zaman damgası ve verilerin ait olduğu binayı belirten bina numarası verileri toplanacak şekilde geliştirmeler yapılmıştır.

Veri toplama algoritmaları ile hazırlanan yüzey üzerine Wi-Fi verilerinin de tüm yüzey noktalarına tüm eşsiz ağların tahminlemeleri yapılmıştır. Bu tahminlemenin daha anlaşılır olması açısından Şekil 7.5'de konumu bilinen bir ağ geçidinin sadece koridorun orta kısmından yürünmesi ile gerçek konumunun eşit yaklaşık şekilde tespit edilebildiği, yüzeyde tahminlenen sinyal seviyelerinin gerçeğe çok yakın olduğu gözlemlenmiştir.



Şekil 7.5. Örnek ağ geçidi

Şekil 7.5'deki görselde, konumu belirtilen ağ geçidinin tüm modelde eşsiz numara ile filtreleme yapılarak elde edilen grafik verilmiştir. Bu grafikteki renklendirme tablosundan da anlaşılacağı üzere, ağ geçidinin gerçek konumu yüksek ölçüde tespit edilmiştir. Buna istinaden konumlama sırasında okunacak veriler ile bu modeli kullanarak çalışacak olan algoritmanın ihtiyacı olan seviyede bir doğrulukta bir modelin geliştirildiği gözlemlenmiştir. Bu aşamada, Wi-Fi modelinin oluşturulmasında da aynı algoritma kullanılmıştır. Wi-Fi verisinin Android işletim sisteminde (geliştirici seçenekleri açık halde) ortalama 4 saniyede bir okunmasını sağlanmaktadır. Bu veri ham bir veri değildir ve gelen sinyalin işletim sisteminin bir filtresinden geçmiş halidir. Kısacası Wi-Fi'de okunan sinyal seviyesi işletim sisteminden iletilirken ekstra bir işlem ile doğruluğu arttırılmaktadır.

Ön hazırlıksız ivme ve jiroskop verisinin toplanmasında 3 eksendeki x, y, z dışında zaman göstergesi için zaman damgası ve verilerin ait olduğu binayı belirten bina numara verisini toplayacak şekilde geliştirmeler yapılmıştır. Ön hazırlıksız barometre verisinin toplanmasında veri toplama formatı esas alınmıştır. PA (pascal) dışında zaman göstergesi için zaman damgası ve verilerin ait olduğu binayı belirten bina numara verilerini toplanacak şekilde geliştirmeler yapılmıştır.

Ön hazırlıksız görüntü verisinin toplanmasında belirlenen veri toplama formatından farklı olarak video biçiminin değil, fotoğraf biçiminin kullanılmasına karar verilmiştir. Fotoğraf biçiminin kullanılmasına karar verirken; video biçiminin en sıkıştırılmış halinin dahi cihazda yüksek depolama alanı gerektireceği, sunucuya gönderimi sırasında internet kullanımının yüksek ve bu görüntülerin işlenmesinin maliyetinin yüksek olacağı düşünülmüş, ek olarak mobil cihazın video kaydederken harcayacağı enerji tüketimi de göz önüne alınmıştır. Görüntü verisinin toplanması sırasında kullanıcı adımları baz alınırken kullanıcı adım attıkça kare (frame) şeklinde görüntüler ve bu görüntülerle birlikte zaman damgası için zaman damgası değeri kaydedilmiştir.

Yeni geliştirilen ön hazırlıksız veri toplama yeteneğinin en verimli şekilde test edilebilmesi için halihazırda verisi eski yöntem ile toplanmış olan binalar esas alınmıştır. Bu tercihteki amaç ön hazırlık yapmadan toplanan

verilerin ön hazırlıklı veriler ile örtüşüp örtüşmediğinin belirlenmesidir. Eski verilerde bulunmayan Wi-Fi, Barometre verilerinin toplanması ise test kapsamı dışında bırakılmıştır.

Veri standartları dahilinde geliştirilen manyetik alan verilerinin toplama algoritmaları sayesinde oluşan modeller üzerinden manyetik alan verisi ile pozisyonlama algoritması geliştirilmiştir. Oluşturulan modelde literatürde parmakizi olarak adlandırılan ve bir benzerlik sistematiğine dayanan model örnek alınmıştır. Parmakizi metodunda, alan içerisindeki (ızgara üzerindeki) her bir noktanın manyetik alan değerleri mobil cihazdan anlık olarak okunan değer ile benzeştirmeye çalışılarak bir olasılık değeri hesaplanmaktadır. Manyetik veriler mobil cihaza aktarılmaya devam ettikçe olasılık değeri düşük olan noktalar elenirken tahminlenen konum olasılık değeri yüksek olan noktalar üzerinde yoğunlaşılmakta ve böylece konum tahmininin çapı gittikçe küçülmektedir. Algoritmanın oluşturulan modeller üzerinde yapılan testlerinde verimli bir şekilde çalıştığı gözlemlenmiş, ilk konum tahminleme süresinin 8-10 adıma kadar çıkabildiği fakat konumu tahminledikten sonra mobil cihazdan veriler akmaya devam ettikçe kullanıcının konumunun takip edilmesinde herhangi bir sapma gözlemlenmemiştir.

Wi-Fi ile konumlama yapacak olan algoritmanın geliştirmesi, Android işletim sistemindeki kısıtlar baz alınarak yapılmıştır. iOS tarafında Wi-Fi tarama işlemi işletim sistemi tarafından tamamen kapatıldığından bu pozisyonlama algoritması sadece Android tarafı için geçerli olacaktır. Wi-Fi sinyal verilerinin toplanma algoritmaları sayesinde oluşan modeller üzerinden öncelikle üçgenleme algoritması geliştirilmiştir. Aynı anda en az 2 ayrı Wi-Fi sinyalinin mobil cihaza ulaştığı senaryoda kullanılabilen bu yapıda en önemli nokta ulaşan eşsiz sinyal verisidir. Bu yapının kullanımı için kapalı alanda gerekli Wi-Fi sinyali yayan cihaz sayısı fazla olmasından dolayı manyetik alan pozisyonlaması için kullanılan parmakizi ile konumlama algoritması referans alınarak üçgenlemeden farklı bir pozisyonlama algoritması da geliştirilmiştir. Yapılan çalışmalarda bu yapının daha az eşsiz Wi-Fi sinyali gerektirmesi dışında verimlilik olarakta üçgenleme algoritmasından daha yüksek doğrulukla konum tespiti sağlamasından dolayı parmakizi algoritması tercih edilmiştir.

7.4.3. Konum Tespiti

Konum tespit algoritması ile yapılan konum tespitinin verimliliği istenilen seviyede olmasına rağmen oluşabilecek anlık sapmalar ve mobil cihazdan akan verilerin herhangi bir dış etken ile bozunuma uğraması ihtimali göz önünde bulundurularak konum-teyit algoritması için çalışmalar gerçekleştirilmiştir. Gelinen bu aşamaya kadar geliştirilen algoritmalardan görece farklı olarak geliştirilen bu algoritmada akan Wi-Fi sinyallerinin yorumlanmasında farklı bir yöntem kullanılmıştır. Konum tespiti başladığında mobil cihaz üzerinde tamamen ayrı şekilde akan Wi-Fi sinyal verileri parmakizi yöntemiyle eş zamanlı olarak konum teyidi için kullanılmıştır.

Konum teyidi için algoritma aşamaları şu şekildedir.

- Akan veriler biriktikçe her seferinde parçacıkların tamamen yeniden dağıtıldığı yöntemden elde edilen konum verisi ile asıl modelden üretilen konum verisi karşılaştırılarak fark tespiti yapılmaktadır.
- Bu fark eğer belirlenen bir çapın dışındaysa ve yürüme düzeni farklılık gösteriyor ise ikinci bir teyit gerçekleştirilir.
- Akan veri tekrar biriktiğinde yine asıl modelin çıktısı ile karşılaştırılır ve eğer tekrar çap dışında kalır ise sapma teyit edilmiş olur.
- Bu senaryoda belirlenmiş olan çap içerisinde konum tespiti tekrar başlamaktadır.

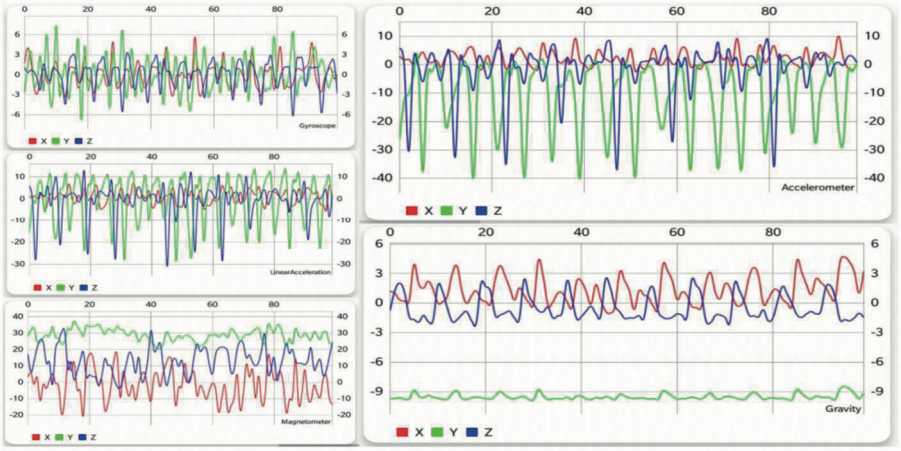
Halihazırda kapalı alan pozisyonlaması için geliştirilen sistemler [18] ve kullanılan eski algoritma için her iki durumda da pozisyonlamanın yapılacağı cihazın yönünün kullanıcının yönü ile aynı olmasını gerektirmektedir. Mobil cihazın cep veya çantada bulunması durumunda akan veriler gürültü içerdiğinden pozisyonlama tahminlemesini yapmak oldukça zorlaşmaktadır. Bu yönelimden bağımsız olarak konum tespiti yapabilmek için geliştirilen algoritma ile bu gereksinimin ortadan kaldırılması hedeflenmiştir. Bu amaç doğrultusunda, öncelikle mobil cihazın cep veya çantada olduğunun tespitinin yapılması için araştırmalar [19] ve çalışmalar incelenmiştir. Mobil cihaz cep veya çantada olduğu durumda yakınlık (proximity) sensöründen akan veriler üzerinde bir inceleme yapılmış, farklı cihazlar ve eşyalar ile olan yakın etkileşimden dolayı mobil cihazdan akan manyetik verilerde yüksek bir gürültü olduğu tespit edilmiştir. Bu olumsuz durumu ortadan kaldırmak amacıyla,

kapalı alan içerisindeki Wi-Fi cihazlarından gelen sinyalleri kullanarak kullanıcı yöneliminden bağımsız bir şekilde pozisyonlamanın devam etmesi için bir algoritma geliştirilmiş, geliştirilen algoritmayla konumlamanın başarılı bir şekilde yapılabildiği gözlemlenmiştir. Pozisyonlama bu şekilde devam ederken mobil cihaz çanta veya cepten çıkarıldığında tekrar yönelimli pozisyonlamaya geçilecek şekilde konumlamanın devam etmesi sağlanmıştır.

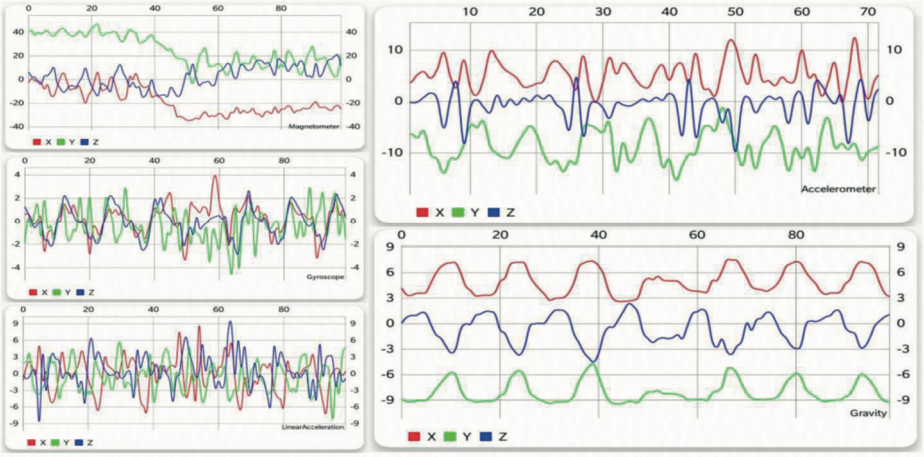
Manyetik parmakizi ile konum tespiti yapan algoritmaların en büyük eksilerinden biri ilk konum tespiti için kullanıcının ilerlemeye başlamasını gerektirmesidir. Veriler farklı konumlardan mobil cihaza akmaya başladıkça algoritma çalışarak oluşan farklılıklar sayesinde konum tespiti yapabilmektedir. Bu bağlamda, ilk konum tespiti için kullanıcı hareket etmese bile bulunduğu noktada konum tespiti yapabilmek büyük önem arz etmektedir. Geliştirilen ilk konum tespit algoritması ile algoritma kullanıcının sabit durduğu anlarda manyetik alan verisini kullanmadan, mobil cihaza sinyal yayan Wi-Fi cihazlarının sinyal güçlerinden yola çıkarak ilk konum tespiti yapabilmektedir. Anlık olarak mobil cihaza akan sinyal verileri sayesinde tek bir tarama sonucunda dahi büyük bir yapının içerisinde kullanıcının bulunabileceği alan oldukça küçültülebilmiştir. Yapının büyüklüğünden bağımsız olarak, kapalı alanda sinyal yayan cihazların sayısına bağlı olmak üzere tek bir tarama sonucunda ortalama 15 metre çapında bir konum tespiti yapılabilmektedir. Veriler akmaya devam ettikçe, bu alan küçülmeye ve kullanıcı adım atmaya başladığı anlarda da manyetik alan verisi konumlama algoritmasına katkıda bulunmaya başlayacaktır.

Belirlenen standartlar dahilinde geliştirilen ivme-jiroskop verilerinin toplanma algoritmaları sayesinde elimizde kullanıcı hareketlerinin algılanmasını sağlayacak modeller bulunmaktadır. Bu modeller ve yapılan akademik çalışmalar [20] incelendiğinde, adım atma, merdiven çıkma-inme gibi aktivitelerinin algılanması için mobil cihazdan anlık olarak akan ivme-jiroskop verilerinin oldukça verimli sonuçlar ortaya çıkardığı gözlemlenmiştir. Yapılacak olan tespit mobil cihaz üzerinde ve offline olarak gerçekleştirileceği için hareketlerin algılanmasındaki performans ve doğruluk yüzdesi ana kriter olacaktır. TensorFlow ve Keras Kütüphanesi kullanılarak mevcut verilerin eğitilmesi sonucunda kontrol verisi de oluşturulmuştur. Eğitilen modelde kullanılan veri seti ile oluşan modele ait kullanıcı etkileşimlerinin tespitindeki doğruluk grafiği yürüme için Şekil 7.6'da, merdiven inme-

çıkma durumu için Şekil 7.7'de görülmektedir. Kullanıcı hareketlerinin tespit hızı ve doğruluğu istenilen seviyededir.



Şekil 7.6. Yürüme Verisi



Şekil 7.7. Merdiven Çıkma Verisi

Barometre ile kat değişim tespiti üzerinde yapılan çalışmalar ve model üzerindeki veriler incelendiğinde, geliştirilecek algoritmanın çok fazla girdiye ihtiyacı olduğu tespit edilmiş, mobil cihaz üzerinden okunan barometre

verilerinin dış dünyadaki değişimlerden çok fazla etkilendiği saptanmıştır. Havanın sıcaklığı, yağışlı olması vb. etkenler anlık olarak okunan değerleri etkilemektedir. Bu durumda katlar arası basınç farkından faydalanma yöntemine başvurulmuş ve basınç farkının dış etkenlere göre farklılık gösterdiği şekilde uygulanmıştır. İncelenen çalışmalar [21] dikkate alınarak, barometre ile kat tespiti için kapalı alanın her bir katına basınç ölçer yerleştirilerek sunucu ile iletişimi sağlanmış, bu sayede anlık olarak her katın basınç değerleri mobil cihaza beslenerek kat değişimlerin algılanması sağlanmıştır. Geliştirdiğimiz yapıda barometreye ek olarak, Wi-Fi cihazlarından yayılan sinyaller ile konum tespiti için geliştirilen algoritma dikkate alınarak bulunulan kat tespiti yapılmıştır. Ayrıca kat değişim yerlerinde güçlü sinyal yayan Wi-Fi cihazlarının tespiti yapılarak mobil cihaza akan veriler parmakizi yöntemiyle bu cihazlarla eşleştirilmeye çalışılmıştır. Eşleşme sağlandığında kullanıcının bulunduğu kat tespit edilmiş, konumlamanın tespit edilen katta devam etmesi sağlanmıştır. Sonuç olarak; kat değişim tespit algoritması, planladığı gibi Barometre verileri ve Wi-Fi sinyalleri ile geliştirilmiştir.

Son olarak; belirlenen senaryolar kapsamında algoritmaların testleri ve modeller üzerinden öncelikli olarak pozisyonlama testleri gerçekleştirilmiş ve sonuçları açıklanmıştır. İlk konum tespiti için yapılan testler sonucunda, ilk konumun doğru tespit edilme yüzdesi 94 olarak tespit edilirken verilecek olan değerler 10.000+ konum tespit noktası üzerinden hesaplanmıştır. İlk konum tespit süresi için, geliştirilen algoritma sayesinde kullanıcının ilerlemesine gerek kalmadan işletim sisteminin kısıtlarına bağlı olarak değişmekle beraber ortalama 3 saniye içerisinde 10 metre çapında bir konum tespiti yapıldığı gözlemlenmiştir. İlk konum tespitinin başarılı olarak yapıldığı senaryolarda 15 dakikalık ve 1 saatlik konum takiplerinde yüzde 6'lık bir sapma yaşandığı izlenmiştir. Bu sapmaların gerçekleştiği senaryolarda da devreye giren konum teyit algoritması %100 bir doğruluk oranı ile karşılaşılan her senaryoda doğru pozisyona getirmeyi başarmıştır. Bu işlem için gerekli olan süre yine işletim sisteminin kısıtlarına bağlı olarak değişmekle beraber Wi-Fi sinyal taraması için gerekli süre olan 6-8 saniye ortalama olarak tespit edilmiştir. Geliştirilen ivme-jiroskop kullanılarak ilerleme çıkarım algoritmasının da testleri belirlenen kapsamında yapılmıştır. Geliştirilen algoritmanın test sonuçları Şekil 7.8'de verilmiş olup, yüksek bir başarımla sonuçların elde edildiği gözlemlenmektedir.

Pred True	LAYING	SITTING	STANDING	WALKING	WALKING_DOWNSTAIRS
LAYING	537	0	0	0	0
SITTING	5	390	93	0	0
STANDING	0	96	436	0	0
WALKING	0	1	0	473	10
WALKING_DOWNSTAIRS	0	0	0	0	420
WALKING_UPSTAIRS	0	0	0	0	1

Şekil 7.8. İlerleme Tespit Doğrulukları

7.5. SONUÇ VE DEĞERLENDİRMELER

Bu bölümünde; yönlendirme ve rota oluşturma uygulamalarında konum noktalarının tespiti için kullanılan GPS sinyalinin alınamadığı ve bu sistemlerin çalışmadığı durumlar için iç mekân konumlama ve yönlendirme yapabilen ve iNavi olarak isimlendirilen bir çözüm tanıtılmıştır. Bu çalışma genel olarak değerlendirldiğinde;

- İç mekânlara yönelik olarak bu zamana kadar yapılan çalışmalarda kullanılan konum belirleme seçenekleri; ultrasonic, RFID, WLAN (Wi-Fi), Pseudo GPS (Uydusallar), Infrared gibi teknolojiler olsa da söz konusu teknolojilerin avantajları ve dezavantajları göz önüne alındığında, kapalı alan navigasyon sistemlerindeki gelişmelerin istenilen seviyelerde olmadığı veya ürünleştirilemediği,
- İç mekanlar veya GPS sinyali bulunmayan ortamlarda da navigasyonlara duyulan ihtiyacın arttığı,
- Yüksek maliyet, alıcı ve vericilerin birbirini doğrudan görme gerekliliği, düşük konum hassasiyeti, yüksek bakım maliyeti ve güvenlik açığı gibi sebeplerden dolayı ise yapılan çalışmalarda sıkıntılar veya eksikliklerin olduğu,
- Uluslararası düzeyde manyetik sensör kullanılarak kapalı alanlarda navigasyon yapabilen çalışmaların bulunduğu, ancak veri toplama işleminde kullanılan yöntemler yeterli düzeyde ve hızda olmadığı,
- Büyük veri gibi güncel teknolojiler kullanan yaklaşımların ve zeki algoritmaların kullanılarak kendini geliştirebilen ve öğrenebilen bir altyapının oluşturulamadığı,

- Haritalandırma maliyetleri genellikle çok yüksek olması, sistemlerin açık alan navigasyon sistemleriyle entegrasyonunun olmaması, donanım maliyetinin yüksek olması, ulusal düzeyde benzer bir sistem veya daha önce tasarlanmış ve geliştirilmiş bir sistem olmaması

bu kitap bölümünde sunulan çalışmalara ihtiyaç duyulduğunu göstermektedir.

Sonuç olarak bu kitap bölümünde detayları tanıtılan ve paylaşılan iNavi uygulamasının;

- İç mekanlarda konum tespiti için kullanılabileceği,
- Sistemin genelde kullanılabilecek bir ürün olduğu,
- Yukarıda belirtilen kısıtları aşmada başarılı olduğu,
- Kapalı alan navigasyonu için geliştirilmiş, donanım bağımsız, akıllı telefon ile yerli kapalı alan konum tespiti yapabilen bir ürün olduğu ve,
- Sonuçta hedeflenen çıktılar ile elde edilen model sayesinde geliştirilen ürünün global bir kullanıcı ağına ulaşılacağı öngörülmektedir.

TEŞEKKÜR

Bu kitap bölümde sunulan çalışma; 3192062 numaralı TÜBİTAK projesi kapsamında desteklenmektedir. Yazarlar; TÜBİTAK tarafından desteklenen, Başarsoft Bilgi Teknolojileri A.Ş. yürütücülüğü ve Cybertech Ar-Ge Limited Şirketi ortaklığıyla yürütülen bu proje kapsamında hazırlanan bu bölüme verilen katkılardan ilgili kurum ve kuruluşlara teşekkür ederler.

KAYNAKLAR

[1] R. Löhner, “On the modeling of pedestrian motion,” *Appl Math Model*, vol. 34, no. 2, pp. 366–382, Feb. 2010, doi: 10.1016/j.apm.2009.04.017.

[2] L. Ruggiero, D. Charitha, S. Xiang, and B. Lucia, “Investigating pedestrian navigation in indoor open space environments using big data,” *Appl Math Model*, vol. 62, pp. 499–509, Oct. 2018, doi: 10.1016/j.apm.2018.06.014.

[3] X. Du and K. Yang, "A Map-Assisted WiFi AP Placement Algorithm Enabling Mobile Device's Indoor Positioning," *IEEE Syst J*, vol. 11, no. 3, pp. 1467–1475, Sep. 2017, doi: 10.1109/JSYST.2016.2525814.

[4] C. T. Li, J. C. P. Cheng, and K. Chen, "Top 10 technologies for indoor positioning on construction sites," *Autom Constr*, vol. 118, p. 103309, Oct. 2020, doi: 10.1016/j.autcon.2020.103309.

[5] G. Oguntala, R. Abd-Alhameed, S. Jones, J. Noras, M. Patwary, and J. Rodriguez, "Indoor location identification technologies for real-time IoT-based applications: An inclusive survey," *Comput Sci Rev*, vol. 30, pp. 55–79, Nov. 2018, doi: 10.1016/j.cosrev.2018.09.001.

[6] N. Fallah, I. Apostolopoulos, K. Bekris, and E. Folmer, "Indoor Human Navigation Systems: A Survey," *Interact Comput*, vol. 25, no. 1, pp. 21–33, Jan. 2013, doi: 10.1093/iwc/iws010.

[7] S. Chumkamon, P. Tuvaphanthaphiphat, and P. Keeratiwintakorn, "A Blind Navigation System Using RFID for Indoor Environments," in *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, May 2008, pp. 765–768. doi: 10.1109/ECTICON.2008.4600543.

[8] M. Murata, D. Ahmetovic, D. Sato, H. Takagi, K. M. Kitani, and C. Asakawa, "Smartphone-based localization for blind navigation in building-scale indoor environments," *Pervasive Mob Comput*, vol. 57, pp. 14–32, Jul. 2019, doi: 10.1016/j.pmcj.2019.04.003.

[9] M. Cramer, J. Cramer, D. de Schepper, P. Aerts, K. Kellens, and E. Demeester, "Benchmarking low-cost inertial measurement units for indoor localisation and navigation of AGVs," *Procedia CIRP*, vol. 86, pp. 204–209, 2019, doi: 10.1016/j.procir.2020.01.044.

[10] K. Liu *et al.*, "Toward Low-Overhead Fingerprint-Based Indoor Localization via Transfer Learning: Design, Implementation, and Evaluation," *IEEE Trans Industr Inform*, vol. 14, no. 3, pp. 898–908, Mar. 2018, doi: 10.1109/TII.2017.2750240.

[11] A. Nessa, B. Adhikari, F. Hussain, and X. N. Fernando, "A Survey of Machine Learning for Indoor Positioning," *IEEE Access*, vol. 8, pp. 214945–214965, 2020, doi: 10.1109/ACCESS.2020.3039271.

[12] J. Fan and A. S. Awan, "Non-Line-of-Sight Identification Based on Unsupervised Machine Learning in Ultra Wideband Systems," *IEEE Access*, vol. 7, pp. 32464–32471, 2019, doi: 10.1109/ACCESS.2019.2903236.

[13] M. Ramadan, V. Sark, J. Gutierrez, and E. Grass, "NLOS Identification for Indoor Localization using Random Forest Algorithm," in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, 2018, pp. 1–5.

[14] C. Jiang, J. Shen, S. Chen, Y. Chen, D. Liu, and Y. Bo, "UWB NLOS/LOS Classification Using Deep Learning Method," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2226–2230, 2020, doi: 10.1109/LCOMM.2020.2999904.

[15] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent Neural Networks for Accurate RSSI Indoor Localization," *IEEE Internet Things J*, vol. 6, no. 6, pp. 10639–10651, 2019, doi: 10.1109/JIOT.2019.2940368.

[16] B. Jia, B. Huang, H. Gao, and W. Li, "On the Dimension Reduction of Radio Maps with a Supervised Approach," in *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*, 2017, pp. 199–202. doi: 10.1109/LCN.2017.53.

[17] D. v Le, N. Meratnia, and P. J. M. Havinga, "Unsupervised Deep Feature Learning to Reduce the Collection of Fingerprints for Indoor Localization Using Deep Belief Networks," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2018, pp. 1–7. doi: 10.1109/IPIN.2018.8533790.

[18] J. Yang, E. Munguia-Tapia, and S. Gibbs, "Efficient in-pocket detection with mobile phones," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, Sep. 2013, pp. 31–34. doi: 10.1145/2494091.2494099.

[19] W. He, "Recognition of human activities using a multiclass relevance vector machine," *Optical Engineering*, vol. 51, no. 1, p. 017202, Feb. 2012, doi: 10.1117/1.OE.51.1.017202.

[20] S. Kaiser and C. Lang, "Detecting elevators and escalators in 3D pedestrian indoor navigation," in *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Oct. 2016, pp. 1–6. DOI:10.1109/IPIN.2016.7743688.

[21] Z. Xu, J. Wei, J. Zhu, and W. Yang, "A robust floor localization method using inertial and barometer measurements," in *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sep. 2017, pp. 1–8. doi: 10.1109/IPIN.2017.8115952.

Bölüm 8

GÜVENLİ KODLAMADA YAPAY ZEKÂ KULLANIMI: KOD KARIŞTIRMA (CODE OBFUSCATION) YAKLAŞIMLARI

Çağrı Şahin

Bu bölümde, güvenli kodlama ve yapay zekâ yaklaşımlarına genel bir bakış açısı sunulmuştur. Güvenli kodlamanın bir parçası olan yazılım kodunu koruma ile birlikte siber saldırılarda da kullanılabilen kod karıştırma teknikleri ile bu tekniklerin nasıl uygulandıkları açıklanmıştır. Özellikle, kod karıştırmada yapay zekâ yaklaşımlarının kullanımı, bu özel konuda yapılan çalışmalar ile güvenli kodlamada kullanılan yaklaşımlar incelenmiş ve elde edilen bulgular değerlendirilmiştir.

8.1. GİRİŞ

Yazılım mühendisliğinin temel ilkelerinden bir tanesi geliştiricilerin yazdıkları kodların okunabilir ve anlaşılabilir olmasıdır. Okunabilirlik ve anlaşılabilirlik yazılımın kalitesini arttırmada büyük önem arz etmektedir. Kaliteli yazılımlarda hatalar daha kolay düzeltilebilir, yazılımın bakımı ve evrimi daha kolay gerçekleştirilebilir. Bununla birlikte geliştirilen yazılım son kullanıcılara teslim edilmeden ve kullanımına açılmadan önce güvenlik, ticari sır, fikri mülkiyet, dijital haklar, gizlilik gibi nedenlerden dolayı yazılımın kodunun korunaklı hale getirilmesi gerekebilir.

Bu doğrultuda güvenli kod geliştirme prensiplerine başvurulabilir. Güvenli kodlama temel olarak yazılımda ortaya çıkabilecek olası güvenlik açıklarını yazılım geliştirme sürecinin başından sonuna kadar izlenecek kurallarla önlemeyi amaçlamaktadır. Güvenli kodlama kapsamlı bir konu olmakla beraber, yazılımın kaynak kodunu daha güvenli hale getirmede en çok tercih edilen yöntemlerden bir tanesi kod karıştırmadır (code obfuscation).

Geliştiriciler yazılımın kodunu son kullanıcılarla paylaşmasa bile tersine mühendislik (reverse engineering) yöntemleri ile yazılımın çalıştırılabilir dosyalarından (örneğin; exe, class, apk) kaynak koda erişmek mümkün olabilmektedir. Kod karıştırma kısaca yazılımın koduna erişilebilirliği geliştiricileri dışındaki kişiler için imkansız hale getirmeyi veya mümkün olduğu kadar zorlaştırmayı hedeflemektedir. Bu amaç doğrultusunda yazılımın kodunda yazılımın işlevini etkilemeyecek şekilde değişiklikler yapılarak karmaşıklık sağlanır. Bir başka değişle yazılım kodunun işlevsel olarak eşdeğeri olan, anlaşılması ve analiz edilmesi daha zor bir versiyonu oluşturulur. Yapılan değişiklikler sonucu yazılıma tersine mühendislik yöntemleri uygulansa bile elde edilebilecek kodun okunabilirlik ve anlaşılabilirliği asıl koda göre çok daha düşük ve maliyetli olacağından yazılımdaki güvenlik açığı keşfine, zararlı kod eklenmesine, ticari sır hırsızlığına, yetkisiz erişime, lisanslama ve diğer denetimleri atlatma gibi problemlere karşı koruma sağlanabilir.

Her ne kadar kod karıştırma koruma amaçlı kullanılsa da aynı zamanda siber saldırılar için bir araç olabilmektedir. Örneğin bilgisayar korsanları, kötü amaçlı yazılımların (malware) virüsten koruma tarayıcılarına yakalanmaması veya elde ettikleri asıl kaynak kodun içerisine ekledikleri zararlı kodları gizlemek için kod karıştırmayı kullanabilmektedirler.

8.2. GÜVENLİ KODLAMA

Güvenli kodlama, geliştiricilerin yazılımı oluştururken veri ihlali ve sistem saldırılarına neden olabilecek güvenlik açıklarını en aza indirecek şekilde kod yazmalarını sağlamayı amaçlar. Örneğin, kullanıcıların erişim yetkisine sahip olduklarını gösteremediği sürece hassas verilere erişimlerinin varsayılan olarak ret edilecek şekilde kodda tanımlanması en temel güvenli kodlama prensiplerinden bir tanesidir. Günümüzde OWASP Güvenli Kodlama Pratikleri ve SEI Cert Kodlama Standartları dahil olmak üzere yaygın olarak kullanılan çeşitli güvenli kodlama standartları bulunmaktadır [1, 2].

OWASP'ın birkaç yılda bir yayınladığı ve en son 2021 yılında güncellenen web uygulamaları güvenlik açığı risk listesinde sırası ile bozuk erişim denetimleri, şifreleme hataları, enjeksiyon, güvensiz tasarım, yanlış güvenlik yapılandırması, savunmasız ve eski bileşen kullanımı, tanımlama ve kimlik doğrulama hataları, yazılım ve veri bütünlüğü hataları, güvenlik günlüğü ve izleme hataları, sunucu tarafı istek sahteciliği ilk onda yer almıştır [3].

Kullanıcıya yetki alanı dışında erişim ve işlev imkanı sağlayabilen bir güvenlik açığı olan bozuk erişim denetimleri, geçmiş yıllarda sıralamanın ortalarında yer alırken listenin en üst sırasına yerleşmiştir. Bu güvenlik açığına karşı önerilen bazı güvenli kodlama prensipleri şu şekildedir:

i. Yalnızca güvenilir sistem nesnelerinin kullanılması.
ii. Erişim yetkilendirilmesini kontrol etmek için tek bir bileşen kullanılması.
iii. Uygulama güvenlik yapılandırma bilgilerine erişemezde tüm erişimlerin ret edilmesi.
iv. Hesap denetimi ile kullanılmayan hesapların devre dışı bırakılması.
v. Uygulamaya yetki verildiğinde hesapları devre dışı bırakmayı ve oturumları sonlandırmayı desteklemesi.

Güvenli kodlama ile güvenlik açıkları minimize edilirken, aynı zamanda yazılım kodunun daha da korunaklı hale getirilmesi sağlanabilir. Yazılım kodunun işlevini değiştirmeden anlaşılmasını ve analiz edilmesini zorlaştıran kod karıştırma, kod güvenliği sağlamada yaygın olarak kullanılan yöntemlerden bir tanesidir.

8.3. KOD KARIŞTIRMA NEDEN ÖNEMLİ

Kod karıştırmanın önemini anlayabilmek için öncelikle koruma sağlanabilecek alanlardaki bazı örnekleri incelemeliyiz. Fakat kod karıştırmanın tek başına bahsedilen problemlerin çözümü veya nedeni olmadığı unutulmamalıdır.

ABD hükümetinin Yaygın Güvenlik Açıkları ve Etkileri listesi (CVE) kaynak alınarak oluşturulan Ulusal Güvenlik Açığı Veritabanı'na (NVD) göre 2022 yılının ilk çeyreğinde 8051 yazılım güvenlik açığı bildirilmiştir [4, 5]. 2021 yılının tamamında 22000 bildirim yapılmış olsa da tespit edilemeyen veya bildirilmeyen yazılım güvenlik açıkları da düşünüldüğünde bu sayının çok daha fazla olacağı öngörülmektedir.

2020 yılında bir bilgi teknolojisi şirketi olan SolarWinds'e yapılan siber saldırı sonucu bilgisayar korsanları şirket sunucularından bir tanesine istemcilerin verilerini elde etmek ve cihazlarına uzaktan erişebilmek için kötü amaçlı kod yerleştirdiler [6]. SolarWinds her bilgi teknolojisi şirketinin yaptığı gibi hata düzeltmeleri ve yeni özellikleri içeren güncellemeleri müşterine gönderdi. Şirket çalışanları bu güncellemelerin içerisindeki kötü amaçlı kodu fark edemedikleri için müşterileri içerisinde olan Fortune 500 şirketleri (Microsoft, Intel, Nvidia, Cisco, FireEye, vb) ve ABD hükümetindeki çok sayıdaki kurum (hazine bakanlığı, savunma bakanlığı, enerji bakanlığı, nükleer güvenlik daire, vb) güvenlik açıklarına maruz kaldılar. Daha da kötüsü SolarWinds'e yapılan siber saldırı sonucu bahsedilen şirket ve kurumlar 9 ay boyunca güvenlik açığından haberdar değildiler. Çok gizli bir şeklide yapılan bu siber saldırı aylar boyunca fark edilemediğinden bazı kurbanlar saldırıya uğrayıp uğramadıklarını asla bilemeyeceklerdir [6]. Diğer bir önemli nokta ise bu ihlali ilk fark edenin milyarlarca dolar fon alan ve ABD bilgisayar ağlarını korumakla görevli olan ABD Siber Komutanlığı değil de FireEye siber güvenlik şirketinin olmasıdır. Gerçi FireEye siber güvenlik şirketinin de ihlali fark etmesi kendi sistemlerinin hacklenmesi sonucu olmuştur.

Sahte ürün, ticari sır hırsızlığı ve korsan yazılım gibi konuları kapsayan fikri mülkiyet hırsızlığı ülke ve şirket ekonomilerine büyük zararlar verebilmektedir. Fikri Mülkiyet Komisyonunun 2017 yılındaki raporuna göre fikri mülkiyet hırsızlığının ABD ekonomisine 225 milyar ile 600 milyar dolar arasında bir maliyete neden olduğu tahmin edilmektedir [7].

Lapsus$ hacker grubu son zamanlarda büyük şirketlerin kaynak kodlarını ve hassas verilerini çalarak ismini duyurmuştur. Kaynak kodu veya hassas verileri çalınan şirketler arasında Microsoft, T-Mobile, Nvidia, Samsung gibi şirketler bulunmaktadır [8]. Lapsus$ 250'den fazla Microsoft projesinin toplamda 37 GB kaynak koduna sahip olduğunu öne sürmektedir.

2018 yılında BSA Global Software'in yayınladığı rapor [9], kişisel bilgisayarlarda kurulu olan yazılımların %37'sinin lisanssız olarak kullanıldığını ortaya koymaktadır. Lisanssız yazılımların küresel ticari değerinin ise 46 milyar dolar civarında olduğu tahmin edilmektedir.

8.4. TERSİNE MÜHENDİSLİK

Tersine mühendislik, kaynak koduna erişimi bulunmayan çalışan bir yazılımın analiz edilerek çözümlenmesi ile kaynak kodunu keşfetmeyi amaçlamaktadır. Kötü amaçlı kullanıldığı takdirde siber saldırılara, ticari sır ifşasına, fikri mülkiyet ihlaline, dijital hak mahrumiyetine ve gizlilik ihlaline sebebiyet verebilmektedir. Takip eden kısımda, tersine mühendisliğin çeşitli programlama dilleri ve makine öğrenmesi ile olan ilişkisi irdelenmiştir.

8.4.1. Programlama Dillerine Göre Tersine Mühendislik

Farklı programlama dillerinde oluşturulan yazılımlar farklı şekilde çalışabilirler. C ve C++ ile yazılan kodlar derleyici kullanılarak makine koduna çevrilirler ve yazılımın makine koduna çevrilmiş yürütülebilir hali son kullanıcılara dağıtılır. Makine kodundan kaynak koda dönüştürme nispeten zor olsa da yazılımın tam olarak nasıl çalıştığı ortaya çıkarılabilir. Örneğin, kaynak kodunuzda kullandığınız benzersiz bir algoritma başkaları tarafından keşfedilebilir. Bununla birlikte bazı derleyici optimizasyonları kodu daha verimli hale getirmek için kod karıştırmaya benzer teknikler kullanabilmektedirler.

Java ve C# ile yazılan yazılımlarda durum biraz daha farklıdır. Bu iki dilde yazılan kodlar doğrudan makine dili yerine ara bir dile çevrilirler. Örneğin, Java programlama dilinde yazılan kod önce bytecode'a derlenir ve oluşturulan bytecode JVM (Java Virtual Machine) üzerinde çalıştırılır. Ara dildeki kod makine koduna göre daha çok bilgi içerdiğinden, kaynak koda çevrilmesi daha mümkündür.

Javascript ve Python gibi yorumlayıcı (interpreter) kullanarak çalışan diller de ise yazılımın çalışması için kaynak koda ihtiyaç duyulmaktadır. Kaynak kod çalışma zamanında yorumlanarak makine koduna çevrilir. Python her ne kadar yorumlayıcı kullanan bir dil olsa da Java'da olduğu gibi kaynak kodun bytecode'a çevrilmiş halinin paylaşılması ve çalıştırılması mümkün olmaktadır.

8.4.2. Makine Öğrenmesi ve Tersine Mühendislik

Makine öğrenmesinin son yıllarda yaygınlaşması ve popüleritesinin artması ile birlikte ile tersine mühendislik yöntemlerinde de değişiklikler meydana gelmeye başlamıştır. Amazon, Microsoft, Google gibi büyük şirketlerin makine öğrenmesi ile sunduğu hizmetleri bir gelir aracı haline getirmesi ve büyük yatırımlar yapması, bu şirketlerin geliştirdiği ve kullandığı makine öğrenme modellerini odak noktası haline getirmiştir. Bulut üzerinde "kara kutu" hizmeti olarak çalışan bu modellerin kodu, parametreleri, ve eğitim veri setleri paylaşılmayarak hem rakip şirketlere karşı rekabet avantajı elde edilmeye çalışılmakta hem de sistemin güvenliği sağlanmaktadır.

Her ne kadar şirketlerin makine öğrenmesi modelleri bulut sistemleri üzerinde gizli şekilde tutulsa da gelişen tersine mühendislik yöntemleri ile modellerin elde edilmesi üzerine çalışmalar bulunmaktadır. Örneğin, model çıkarma saldırıları kullanılarak lojistik regresyon, yapay sinir ağları, ve karar ağaçları gibi popüler model sınıfları için yüksek doğrulukta makine öğrenmesi modellerinin çıkarılması sağlanmıştır [10]. Bir başka çalışmada, girdi-çıktı sorguları kullanılarak çeşitli kara kutu sinir ağı özniteliklerini tahmin edecek bir metamodel geliştirilmiştir [11]. Çalışmada, mimari yapı, kullanılan optimizasyon algoritması, ve veri seti ile ilgili özniteliklerin başarılı bir şekilde tahmin edebildiği rapor edilmiştir. Kara kutu modelin işlevselliğini, görüntü sorgu etkileşimlerine dayanarak model hırsızlığı olarak adlandırılan taklidini oluşturan çalışma da mevcuttur. [12]. Model hırsızlığı ile kara kutu makine öğrenimi modelindeki parametre, eğitim seti bilgisi ve karar sınırları gibi özellikleri elde etmek üzerine de çalışmalar yapılmıştır [13, 14, 15].

8.5. KODLARA YAPILAN SALDIRILAR

Bilişim teknolojilerindeki zorluklardan bir tanesi, yazılımların siber saldırılara karşı güvenilir olmasını sağlamaktır. Yazılım güvenliği; veri gizliliğini, veri mahremiyetini, kod gizliliğini ve yürütme doğruluğunu kapsamaktadır.

Sondaki adam saldırıları (mate attacks); yazılıma fiziksel erişimi olan kötü niyetli son kullanıcılar tarafından yapılmaktadır. Bu kişiler son kullanıcı oldukları için yazılımı hata ayıklayıcılar, kod çözücüler, ve statik-dinamik analizörler kullanarak inceleyebilir, değiştirebilir, ve çalıştırabilirler. Bunun sonucunda ise bireysel kayıtların çalınması, hassas kişi ve kurum verilerin ihlali, ticari sırların ifşa edilmesi ve hatta kritik altyapıların devre dışı bırakılması gibi istenmeyen durumlar oluşabilir.

Sondaki adam saldırılarına (mate attacks) karşı yazılımı koruma; saldırganların yazılımı kurcalamalarını, tersine mühendislik yapmalarını ve yasa dışı olarak yeniden dağıtmalarını engellemek olarak tanımlanmaktadır [16]. Bu kapsamda; saldırılar tümüyle engellenemese bile en azından saldırıların ne zaman yapıldığının tespit edilmesi ve raporlanması da büyük önem arz etmektedir.

Sondaki adam saldırılarının analiz edilmesinin, modellenmesinin ve değerlendirilmesinin zorluğunu 3 temel nedene dayandırılmıştır [17].

i. Saldırganların insan olması yaratıcılık ve yeteneklerin kullanılmasını sağlar.
ii. Saldırgan hedefe sınırsız ve yetkili erişimi elde eder.
iii. Koruma tedbirleri belirli bir süreye kadar belirli sayıdaki saldırgana karşı etkili olur.

Sondaki adam saldırılarına karşı koruma tedbirleri kod karıştırma, şifreleme, sunucu tarafında çalıştırma, ve güvenilir yerel kod gibi sınıflara ayrılabilir [18]. Kod karıştırma teknikleri bir sonraki kısımda ayrıntılı olarak incelenecektir. Diğer tedbirleri kısaca açıklamak gerekirse [18, 19]:

Şifreleme: Şifreleme ile yazılım güvenliği sağlama güvenilir hesaplama olarak da adlandırılırken, genellikle bir donanım tarafından şifreleme sağlanır. Örneğin, dünyanın önde gelen bilgisayar işlemci üreticisi olan Intel bu

kapsamda donanım tabanlı bir teknoloji olan Software Guard Extension'ı (SGX) geliştirmiştir. Intel SGX, verilerin gizliliğini ve bütünlüğünü platformdaki diğer tüm yazılımlara, hatta güvenliği ihlal edilmiş bir işletim sistemine karşı korumak için "enclave" olarak adlandırılan korunmuş yürütme ortamları sağlamaktadır. Geliştiriciler, uygulamaların "enclave" aracılığı ile çalışmasını sağlayarak kod gizliliğinin korunmasına katkıda bulunabilmektedirler.

Sunucu tarafında çalıştırma: JavaScript komutları (script) genellikle son kullanıcı olan istemci bilgisayarında kaydedilir ve web tarayıcıları tarafından yorumlanarak çalıştırılırlar. Dolayısıyla istemci, yazılımın koduna erişim sağlayabilir ve kod üzerinde kötü niyetli analiz ve inceleme yapabilir. Buna karşılık, komut dosyaları sunucu tarafında saklanıp yürütülebilir. Sunucu tarafında çalışan komutlar, istemcinin talep ettiği içeriği dinamik olarak oluşturur ve istemcinin web tarayıcısında görüntülenecek şekilde gönderir. İstemcinin kaynak koduna erişimi olmadığı için kod güvenliği sağlanmış olur. Sunucu tarafı komut dosyası PHP, Python, Ruby, NodeJS gibi programlama dilleri kullanılarak oluşturulabilir.

Güvenilir yerel kod: Yerel kod, belirli bir işlemci ve talimat seti ile çalışmak üzere derlenen kodlardır. Yerel kod çalıştığı platforma spesifik olarak oluşturulduğu için, yazılım kodunun yetkisiz olarak değiştirilip yeniden dağıtılmasına karşı önlem sağlar.

8.6. KOD KARIŞTIRMA TEKNİKLERİ

Kod güvenliğini sağlamak ve sondaki adam saldırıları gibi siber saldırılara karşı tedbir almak için çeşitli kod karıştırma teknikleri kullanılabilmektedir. Kod karıştırma, yazılım koduna koruma sağlayabilmek için birbirini tamamlayabilen birçok farklı teknikten oluşur. Yazılımın geliştirildiği programlama diline de bağlı olarak kod değişiklikleri çeşitli seviyelerde (kaynak kod, ara kod, makine kodu) gerçekleştirilebilir. Örnek vermek gerekirse, Javascript programlama dilini kullanan geliştiriciler sadece kaynak kodu seviyesinde kod dönüşümleri ile ilgilenirken, C programlama dilini kullanan geliştiriciler makine kodu seviyesindeki kod dönüşümlerine de başvurabilirler.

Kodda yapılan değişiklikleri sınıflandırmada kullanılan bir başka kriter ise, yapılan dönüşümlerin hedefidir. Genellikle kod üzerinde veri dönüşümü ve kod dönüşümü olarak ikiye ayrılır. Veri dönüşümlerinde yazılımda sabit kodlanmış değerlerin (sayılar, string, vb) temsili ve konumu ile birlikte yazılım tarafından kullanılan değişken bellek değerleri değiştirilebilir. Kod dönüşümlerinde ise yazılımda kullanılan veri yapıları, değişken adları, kullanılan algoritmalar, kontrol akışı gibi kısımlarda değişiklikler yapılabilir.

Aşağıdaki şekilde, JavaScript kodları için online kod karıştırma sağlayan obfuscator.io web sitesi kullanılarak oluşturulan bir örnek yer almaktadır [20]. Şekil 8.1(a) kaynak kodu gösterirken, Şekil 8.1(b) kod karıştırma sonrası elde edilen kodu göstermektedir.

```
function hi() {
  console.log("Hello World!");
}
hi();
```

(a) Öncesi

```
(function(_0x499d97,_0x23da63){var _0x113e70=_0x59a1,_0x1cd7f4=_0x499d97
();while(!![]){try{var _0x198e4a=parseInt(_0x113e70(0xba))/0x1+parseInt
(_0x113e70(0xbb))/0x2+-parseInt(_0x113e70(0xb4))/0x3+-parseInt(_0x113e70
(0xb5))/0x4*(-parseInt(_0x113e70(0xb8))/0x5)+-parseInt(_0x113e70(0xb6
))/0x6*(parseInt(_0x113e70(0xb7))/0x7)+parseInt(_0x113e70(0xb9))/0x8
+-parseInt(_0x113e70(0xbe))/0x9;if(_0x198e4a===_0x23da63)break;else
_0x1cd7f4['push'](_0x1cd7f4['shift']());}catch(_0x1935b1
){_0x1cd7f4['push'](_0x1cd7f4['shift']());}}}(_0x4288,0x226a9));function
_0x59a1(_0x83a33f,_0x216320){var _0x42882c=_0x4288();return _0x59a1
=function(_0x59a16d,_0x3f3968){_0x59a16d=_0x59a16d-0xb4;var _0x1f5189
=_0x42882c[_0x59a16d];return _0x1f5189;},_0x59a1(_0x83a33f,_0x216320
);}function _0x4288(){var _0x1511d4=['18nfpbYc','114835xfTdyB'
,'65HgtoYw','1187OOOVAbRJj','222241WcBfkR','137426Jwbyal'
,'Hello\x20World!','log','2167560MgJEuF','217536osGECU','19756PMejNF']
;_0x4288=function(){return _0x1511d4;};return _0x4288();}function hi
(){var _0x135529=_0x59a1;console[_0x135529(0xbd)](_0x135529(0xbc));}hi
();
```

(b) Sonrası

Şekil 8.1. Kod karıştırma örneği

Şekil 8.1'de görüldüğü üzere kodun anlaşılırlığı ve analizi zor bir hal almıştır. Yaygın şekilde kullanılan kod karıştırma teknikleri şunlardır [19, 21, 22]:

- Yeniden Adlandırma:

Method ve değişkenlerin adları değiştirilerek anlaşılması zor hale getirilir. Değiştirilen method ve değişken adlarında 'a', 'b' gibi tek karakter kullanımı tercih edilirken, anlamsız ifade ve harflerden de adlar oluşturulabilir. Ayrıca yazdırılamayan veya gizli karakterler de kullanılabilir.

- Aşırı yükleme**:**

Aynı method ve değişken isimleri farklı kapsamları (scope) olduğu sürece tekrar tekrar kullanılarak karmaşıklık arttırılır. Farklı for döngülerinde iteratör olarak aynı değişken isminin kullanılması basit bir örnek olarak düşünülebilir.

- Opak yüklemler:

Geliştiriciler tarafından önceden bilinen ve kodda kullanılan "doğru" ve "yanlış" karşılaştırma ifadeleri yerine doğruluk değeri değişmeyecek şekilde farklı ifadeler kullanılır. Şekil 8.2'de verilen C kodu örneğinde orijinal kodun çalışması sanki 'if' koşulunun içerisindeki ifadenin doğruluğuna bağlı gibi gösterilmiştir. Oysa

(a*a+a) % 2 == 0

koşulu a değişkeninin değeri ne olursa olsun her zaman doğru olacaktır ve orijinal kod çalışacaktır.

```
int a
........
if((a*a+a) % 2 == 0 ){
    // orjinal kod
} else{
    // şaşırtma kodu
}
```

Şekil 8.2. Opak yüklem örneği

- Statik verileri prosedürel verilere dönüştürme**:**

Sabit olarak tanımlanmış değişken değeri yerine çalışma zamanında aynı değeri değişken için bir fonksiyon veya metot tarafından oluşturması

sağlanır. Böylelikle, statik olarak yapılan kod analizinde sabit değişken değeri gizlenmiş olur.

- Boolean-aritmetik karıştırma**:**

Bu teknikte verinin mantık ve aritmetik operatörler kullanılarak oluşturulan doğrusal eşiti ters çevrilebilir polinom fonksiyonu şeklinde şifrelenir. Çalışma zamanında ise polinom fonksiyonunun tersi alınarak veri tekrar oluşturuşur ve kullanılır.

- String şifreleme**:**

Kodda veya çalıştırılabilir dosyalarda bulunan string şeklindeki ifadeler açıkca okunabilir ve bu ifadeler referans alınarak değişken ve metot isimleri değiştirilmiş olsa bile kritik kod bölümleri açığa çıkartılabilir. Örneğin kullanıcılarda görüntülenen hata mesajları kod içerisinde incelenerek, kodun hangi aşamasında ne yaptığı anlaşılabilir. Bu durumun önüne geçmek için string ifadeler şifrelenerek gizlenir ve çalışma zamanında sadece gerekli olduğunda şifresi çözülerek asıl string ifadesi oluşturulur.

- Tek yönlü dönüştürme:

Verilerin eşleşmesini kontrol etmekte kullanılabilen bu teknik, veriyi tek yönlü olarak değiştirir. Çalışma zamanında kullanıcı girdisi ile dönüştürülen değerin karşılaştırılması gerekirse, kullanıcı girdisi de aynı işleme tabii tutularak karşılaştırma gerçekleştirilir.

- Yorumları kaldırma**:**

Kod karıştırma tekniklerinin içerisinde uygulaması en kolay olan tekniklerden bir tanesidir. Yapılması gereken sadece geliştiricilerin koda eklemiş oldukları açıklamaların kaldırılmasıdır.

- Biçim değiştirme**:**

Tümleşik geliştirme ortaları (IDE) kodun okunabilirliğini arttırmak için otomatik biçim ayarlaması yapmaktadırlar. Bu teknikte tam tersine tüm boşluk, sekme ve satırsonu karakterleri kaldırılarak kodun anlaşılırlığı zorlaştırılır.

- Veri birleştirme-bölme:

Kaynak kodunda tanımlanmış olan bir değişken birden fazla değişkene bölünerek veya birden fazla değişken tek değişkene birleştirilerek dönüşümler yapılır. Tanımlanmış diziler için de aynı işlem yapılabilir. Şekil 8.3'de aynı elemanlara sahip olacak şekilde iki boyutlu dizi - tek boyutlu dizi arasındaki dönüşüm gösterilmektedir.

```
int degisken [3][3] = {{1,2,3},{4,5,6},{7,8,9}};

int degisken [9] = {1,2,3,4,5,6,7,8,9}
```

Şekil 8.3. İki boyutlu - tek boyutlu dizi dönüşümü

- Yeniden sıralama**:**

Program kodundaki talimatların (instructions), rutinlerin, değişkenlerin, dalların (branch) yerleri yazılımın işlevini etkilemeyecek şekilde değiştirilir.

- Veri akışı yönetimi**:**

Bellek (memory) dinamik olarak yönetilerek, değişken ve veri öbeklerinin hafızada belli aralıklarla yeni bir alana taşınması sağlanır. Bunu gerçekleştirmek için yazılım tabanlı bir Bellek Yönetim Birimi (MMU) kullanılır. Bu birim kullanılacak şekilde kaynak kodda değişiklik yapılır.

- Meta verilerini kaldırma**:**

Metadata verileri program hakkında ek bilgi ve açıklamalar içerebilmektedir. Bu nedenle metadata verilerinin kaldırılması kodun anlaşılabilirliğini azaltacağı için faydalı olacaktır.

- Talimat kalıbı değiştirme**:**

Derleyici tarafından oluşturulan yaygın talimatlar (instructions), daha karmaşık veya az bilinen talimatlara dönüştürülür. Şekil 8.4'de register olarak adlandırılan küçük hafıza birimlerinin içerisindeki değişken değerlerinin yer değiştirilmesi örneği verilmiştir. Derleyicinin makine dilinde mov tali-

matı olarak oluşturduğu kod, yığın (stack) tabanlı yapısından yararlanılan önbellek aracılığı ile push ve pop komutları kullanılarak aynı işlevi yerine getirecek şekilde değiştirilmiştir.

```
mov register3, register1
mov register1, register2
mov register2, register3
```

```
push register1
push register2
pop register1
pop register2
```

Şekil 8.4. Register işlemlerinde talimat kalıbının değiştirilmesi

- Aritmetik şifreleme**:**

Boolean veya aritmetik ifadeleri daha zor hale getirmek için uygulanan bu teknikte, bit düzeyinde (bitwise) operatörler aracılığı ile kod dönüştürme gerçekleştirilir. C programlama dilinde and, or, xor, left shift, ve right shift operatörleri yaygın olarak kullanılmaktadır. Şekil 8.5'teki örnekte C kodunda basit bir toplama işleminin bit düzeyinde operatörler ile işlevsel eşiti oluşturulmuştur. İki sayının toplamının bir değişkene atanması xor, and, ve left shift operatörleri ile sağlanmıştır.

```
int main() {
    int deger1 = 1;
    int deger2 = 2;
    int toplam = deger1 + deger2;
    printf("%d",toplam);
    return 0;
}
```

```
int main() {
    int deger1 = 1;
    int deger2 = 2;
    int toplam = (((deger1^deger2) + ((deger1&deger2)<<1)));
    printf("%d",toplam);
    return 0;
}
```

Şekil 8.5. Toplama işleminde aritmetik şifreleme kullanımı

- Sahte kod ekleme ve kod genişletme:

Kaynak koda yazılımın işleyişi ile alakası olmayan, hiçbir zaman çalıştırılmayacak veya çalıştırılsa da herhangi bir etkisi olmayacak kod parçaları eklenir. Örneğin if-else koşulları ile kod içerisinde farklı dallar (branch) oluşturulur ama bu dallar asla yürütülmez. Kodun boyutu basit talimatlar yerine daha karmaşık ve alışılmadık talimatlar kullanılarak genişletilir.

- Paketleme:

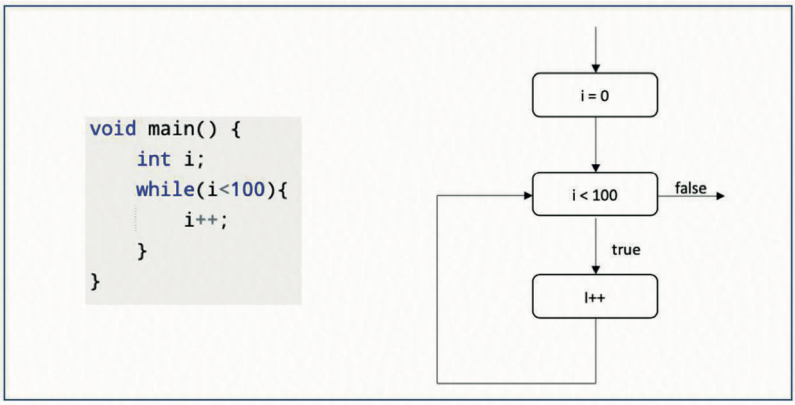
Yazılım kodu ve dosyaları sıkıştırılarak hem dosya boyutu küçültülür hem de tersine mühendislik yöntemleri ile elde edilebilecek bilgi minimize edilmiş olur. Paketlemede kodun içerisine paketlenmiş kısımları çözecek kod parçaları eklenerek, paketlenmiş kısımların çalışma esnasında çözümlenmesi sağlanır.

- Fonksiyon oluşturma-dağıtma:

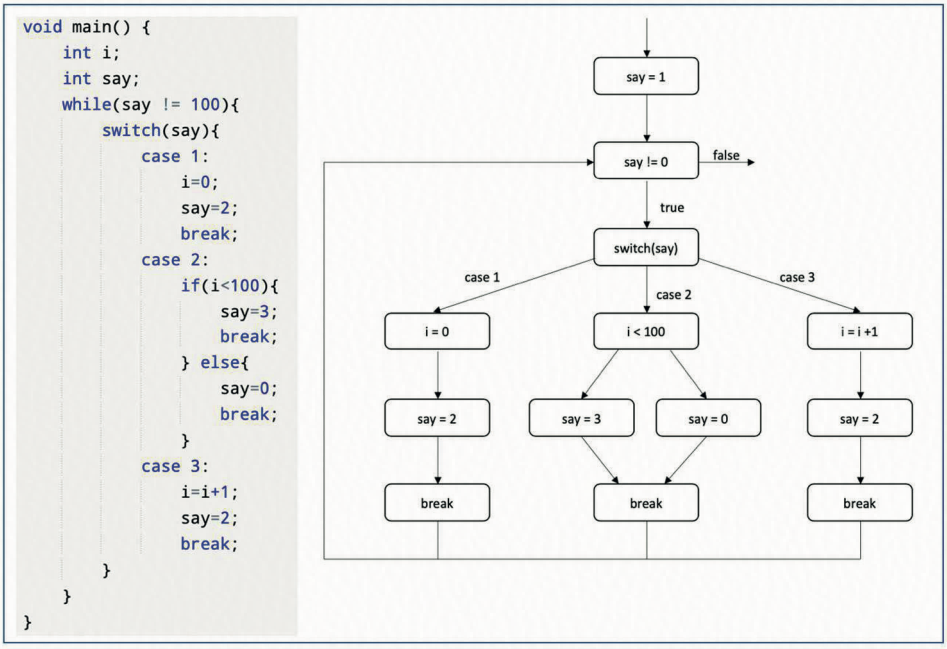
Fonksiyon tanımlama ve fonksiyon çağırma programlama dillerinin temel kodlama prensiplerindendir. Yazılım koduna bağlı olarak fonksiyonlar çağrıldığı kod parçacığı içinde uygulanabilir veya kod parçacığının içerisinde belli kısımlar ayrı bir fonksiyon olarak oluşturulabilir. Buradaki tercih tamamen yazılım karmaşıklığını arttırmaya yönelik olmalıdır. Benzer şekilde fonksiyonun aldığı ve fonksiyon çağrılarındaki parametreler güncellenerek farklı fonksiyonlar birleştirilebilir veya ayrılabilirler.

- Kontrol akışı zorlaştırma:

Kodun mantıksal yapısını daha az tahmin edilebilir ve izlenebilir hale getirmek için çeşitli değişiklikler yapılır. Geri derlenmiş kodun, geliştiricilerin sıklıkla kullandıkları bir terim olan “spaghetti logic” halinde olması hedeflenir. Bir başka deyişle geri derleme ile elde edilebilecek kod, yapılandırılmamış ve anlaşılması zor bir şekilde olur. Bu teknik aslında daha önce bahsettiğimiz birçok tekniği içerebilir. Örneğin gereksiz switch-case komutlarının eklenmesi yaygın olarak kullanılır. Şekil 8.6’daki örnekte hem kontrol akışındaki hem de koddaki değişiklikler öncesi ve sonrası olarak gösterilmiştir [23].



(a) Öncesi



(b) Sonrası

Şekil 8.6. Kontrol akışı zorlaştırma örneği

- Kod Sanallaştırma**:**

Makine dilinde olan kodun daha güvenilir hale gelmesi için özel bir sanal makine uygulamasında çalışacak bytecode'u oluşturularak sanallaştırılır.

Çalışma zamanında özel sanal makine, bytecode'u yorumlar ve talimatları yürütür. Makine kodunun tamamen değiştirilebilir olması, statik analiz araçlarının kodu tanımlamasını bile önleyebilir.

- Rastgele belirleyici isimleri kullanma**:**

Yeniden adlandırma yöntemine benzer şekilde değişken, sınıf, fonksiyon, ve benzeri tanımlayıcı isimleri değiştirilir. Farkı ise değişikliklerde karakter ve sayılardan oluşan uzun rastgele isimlerin kullanılmasıdır. Genellikle çeşitli araçlar kullanılarak otomatik olarak yapılır.

- Kütüphane çağrısı kaldırma**:**

Geliştiriciler sağladığı kolaylıktan dolayı kod yazarken çeşitli veri yapısı veya algoritma sağlayan kütüphaneleri kullanabilmektedirler. Kütüphane çağrıları yazılım çalışması hakkında bilgi verebileceği gibi kodun yapısının açığa çıkmasına neden olabilir. Kütüphaneleri çağırmak yerine, yazılımın içinde kütüphanelerin işlevini yerine getirecek kodun oluşturulması kod güvenliği açısından daha faydalı olacaktır.

- Karşı-Hata Ayıklama (Anti-debug):

Kodu satır satır incelemek ve hataları düzeltmek için geliştiriciler yazılım hata ayıklama araçları kullanırlar. Bu araçlar aynı zamanda tersine mühendislik için de kullanılabilir. Kod karıştırma tekniklerinden bir tanesi de karşı-hata ayıklama olarak adlandırılan kodun kendini koruma mekanizmasıdır. Kaynak koda eklenecek ek kod parçaları ile hata ayıklama tespit edildiği durumlarda, hassas verilerin bozulması, yazılımın çalışmasının sonlandırılması veya hata vermesi, geliştiricilere sinyal gönderilmesi, yazılım içinde farklı bir kod bloğunun çalıştırılması gibi seçenekler oluşturulabilir.

Kod karıştırma teknikleri incelendiğinde; bazı tekniklerin bir veya birden fazla tekniği kapsadığı görülmektedir. Örneğin, kontrol akışını zorlaştırma; karşı-hata ayıklama, sahte kod ekleme, ve kod genişletmeyi içerebilir. Bununla birlikte, bazı kod karıştırma teknikleri için farklı kaynaklarda değişik isimler kullanılabildiği unutulmamalıdır.

8.7. KOD KARIŞTIRMA VE YAPAY ZEKÂ

Kod karıştırma her ne kadar kod güvenliği amacı ile kullanılsa da aynı zamanda kod intihalini ve kötü amaçlı yazılımları (malware) gizlemek için de sıklıkla başvurulan bir araçtır. Kod karıştırma uygulanmış yazılımlarda insan tabanlı analizler yetersiz, zaman alıcı ve maliyetli olabilmektedir. Bu nedenle, kod intihali ve kötü amaçlı yazılım tespiti gibi zor ve karmaşık işlemlerde yapay zekâ ve makine öğrenmesi algoritmaları kullanımı tercih edilmektedir. Bu kısımda kod karıştırma ve yapay zekâ içeren çalışmalara örnekler verilmiştir.

8.7.1. Kod Karıştırma Teknikleri Tespiti

Kötü amaçlı yazılım, bilgisayarın işleyişini bozmak veya sistemi saldırganın isteği doğrultusunda manipüle etmek için kullanıcı bilgisayarında çalıştırılan kodu içeren bir yazılım olarak tanımlanmaktadır [24]. Bilgisayar korsanları kötü amaçlı yazılımlarını ve kod parçalarını gizlemek ve savunma mekanizmalarını atlatmak için kod karıştırma tekniklerine başvurabilmektedirler. Daha önce verdiğimiz SolarWinds örneğinde olduğu gibi, kötü amaçlı yazılım veya kod ne kadar geç tespit edilirse zararı da o kadar fazla olacaktır [6].

Kötü amaçlı yazılımlardan korunmada kullanılan en yaygın yöntemlerden bir tanesi, kötü amaçlı yazılım korunma imzalarıdır. Korunma imzaları, tespit edilebilen kötü amaçlı kod parçacıkları baz alınarak oluşturulur. Analiz edilen yazılım, bu imzalar aracılığıyla model eşleşmesine tabii tutulur ve kötü amaçlı olup olmadığı şeklinde bir karar verilir. Fakat koruma imzası tanımlanmış olsa bile kod karıştırma teknikleri kullanılarak kötü amaçlı yazılım varyantları oluşturulduğunda, tek bir koruma imzası ile bütün varyantları tanımlamak oldukça zordur.

Yakın zamanda yapılan bir çalışmada, kötü amaçlı Android uygulamalarını etkin bir şekilde analiz etmek için öncellikle hangi kod karıştırma tekniklerinin uygulandığının belirlenmesi gerektiğini öne sürülmüştür [25]. Bu doğrultuda, kod karıştırma tekniği uygulanmamış 148 Android uygulamasının, ProGuard, dashO, DexProtector ve DexGuard gibi kod karıştırıcıları ile 1.717 varyantını oluşturulmuştur. Varyantlar oluşturulurken yeniden isim-

lendirme, veri akışı zorlaştırma, string şifreleme, ve method çağrısı zorlaştırma teknikleri ayrı ayrı uygulanmıştır. Android analiz aracı olan AndroGuard kullanılarak APK (Android uygulamalarını yüklemekte ve dağıtmakta kullanılan dosya formatı) dosyaları geri derlenmiş, içerisindeki classes.dex dosyasından her bir Java sınıfının kodları elde edilmiştir [26]. Toplamada elde edilen 151.391 Java sınıf dosyası rastgele olacak şekilde %80/%20 oranında eğitim ve test seti olarak belirlenmiştir. Java sınıfları, Paragraf Vektörü kullanılarak (değişken uzunluktaki metinlerin sabit uzunluk temsilini öğrenen bir algoritma) vektörleştirildikten sonra Random Forest, AdaBoost, Extra Trees ve Linear SVM algoritmaları gibi makine öğrenme yaklaşımları kullanılarak makine öğrenmesi gerçekleştirilmiştir. Çalışmanın sonucuna göre makine öğrenmesi algoritmalarının performansları yakın olsa da en iyi sonucu Extra Tree algoritması sağlamıştır.

Benzer şekilde, yeniden isimlendirme, veri akışı zorlaştırma, string şifreleme gibi kod karıştırma teknikleri tespit edilmeye çalışılmıştır [27]. Geliştirilen Androdet aracı çevrimiçi öğrenme tabanlı olup, veri akışı madenciliği aracılığıyla makine öğrenmesinde kullanılan eğitim setinin sürekli güncel tutulmasını sağlamaktadır. Çevrimiçi öğrenme ile eğitim seti sürekli güncellendiğinden, çevrimdışı (batch veya offline) öğrenmede olduğu gibi eğitim setinin çıkarılması, sistemin eğitilmesi, eğitim setinin belirli süre sonra etkisiz kalabilmesinden dolayı tekrar eğitim setinin oluşturulması ve tekrar sistemin eğitilmesi gerekmemektedir. Çalışmada elde edilen sonuçlara göre, çevrimiçi öğrenme kod karıştırma tekniklerinin tespitini çevrimdışı öğrenme ile rekabet edebilecek doğrulukta, daha az zaman ve hesaplama kaynağı ile sağlamaktadır.

JavaScript, web teknolojisinin temelini oluşturan, neredeyse tüm web sayfalarında kullanılan, hem istemci hem de sunucu tarafında çalışabilen betik bir programlama dillidir. Kötü amaçlı yazılan JavaScript kodlarının imza tabanlı yaklaşımlarla tespit edilmelerini zorlaştırmak için sıklıkla kod karıştırma teknikleri uygulanır. Sınıflandırıcının tahmine dayalı modellerini hızlı bir şekilde oluşturmak, eğitmek, ve değerlendirmek için bulut tabanlı Microsoft Azure Machine Learning Studio kullanılabilmektedir [28]. Bu doğrultuda yapılan çalışmada, Azure ML'nin sağladığı sınıflandırma algoritmalarından ortalama algılayıcı, bayes nokta makinesi, güçlendirilmiş karar

ağacı, karar ormanları, yerel derin destek vektör makinesi, lojistik regresyon, sinir ağı, ve destek vektör makinesi algoritmalarının performansları karşılaştırılmıştır. Ayrıca, eğitim setinde bulunmayan kod karıştırma araçları tarafından kod karıştırma uygulanmış yazılımların da analizi denenmiştir. Kullanılan algoritmalar içerisinde en iyi kod karıştırma tespit performansı güçlendirilmiş karar ağacı algoritması ile elde edilmiştir. Eğitim seti içerisinde yer alan kod karıştırma araçları kullanılarak %99 oranında başarı sağlanırken, eğitim seti içerisinde olmayan kod karıştırma araçları için başarı oranı %47 olmuştur.

8.7.2. Kod Benzerliği Analizi

Kod benzerliği analizi, kod intihal tespitini ve yazılım fikri mülkiyet korumasını sağlamak amacı ile giderek daha fazla kullanılmaya başlanmıştır. Buna karşılık yetkisiz veya izinsiz şekilde başka geliştiricilerin kodlarını kullananlar, kod karıştırma teknikleri aracılığı ile orijinal kod üzerinde değişiklikler yapabilmektedirler. Yapılan değişiklikler sayesinde gizlenmeyi ve tespit edilmemeyi hedeflemektedirler. Bu nedenle, kod benzerliği analizi yapmak için kaynak kodundan ziyade ikili dosyalar (binary files) üzerinde analiz yapan yeni bir yaklaşım önerilmiştir [29]. Yürütülebilir dosyaları dikkate alan bu yaklaşımda, kodlar söz dizimsel olarak farklı görünse de benzer davranışa sahip yürütülebilir dosyaların tanımlaması ve kod karıştırma tekniklerinin ön bilgisi olmadan analiz yapılması hedeflenmiştir. Yapılan çalışmada, C ile yazılmış 47 programın her biri için Tigress kod karıştırıcı kullanarak artırımlı şekilde 200 varyantı oluşturulmuştur. Orijinal ve varyant kodlar derledikten sonra ikili dosyaları görüntüye dönüştürülmüş, ikili dosyalardan çıkarılan görüntüleri sınıflandırmak için evrişimli sinir ağları (CNN) yapısını kullanılmıştır. Veri seti %90/%10 oranında eğitim ve test seti şeklinde ayrılmış ve algoritma rastgele belirlenen eğitim ve test setleri ile birçok kez çalıştırılmıştır. Sonuç olarak, ikili dosya benzerlik tespitinde %88 doğruluk oranı elde edildiği belirtilmiştir.

Kod karıştırma tekniklerinin yazar kimliğini belirlenmesi üzerindeki etkisi de araştırılmıştır [30]. Evrişimli sinir ağlarına dayalı geliştirilen sistemde yazarları sınıflandırmak için K-en yakın komşular sınıflandırıcısı

kullanılmıştır. Bir tanesi orijinal diğeri kod karıştırma uygulanmış kod dosyalarını içeren, her bir yazar için 9 kod dosyası olan iki veri seti oluşturulmuştur. Python, C++, ve Java programlama dillerinde yazılmış kod dosyalarına sırasıyla PyObfx, Stunnix, JavaSource-CodeObfuscator kod karıştırma araçlarını kullanarak kod karıştırma tekniklerini uygulanmıştır. İki veri seti üzerinde yapılan testlerde, kod karıştırmanın uygulama yazar kimliğini belirlemede doğruluk oranını düşürdüğünü tespit edilmiştir.

8.8. GÜVENLİ KODLAMA VE YAPAY ZEKÂ

Yazılım geliştirme sürecinde yapay zekâ kullanımı gün geçtikçe artmaktadır. Geliştiriciler yapay zekâ tabanlı kod oluşturma araçları kullanabilmekte ve kod güvenliği için yapay zekâdan faydalanabilmektedirler. Bu kısımda yapay zekâ ile kod oluşturma ve kod güvenliği sağlama ile birlikte yapay zekâ tabanlı sistemlerde güvenlik konuları ele alınacaktır.

8.8.1. Kod Oluşturma

Bir programlama dilini öğrenmeye karar vermek ve etkin olarak kullanmak zaman alıcı ve maliyetli bir süreç olabilmektedir. Ancak, yapay zekâ modellerinin bir sorunu tanımlayabilmesi ve çözümü kodlayabilmesi ile bu sorun ortadan kalkabilir. Makine öğrenmesindeki ilerlemeler ile birlikte geliştirilmeye başlanan yapay zekâ tabanlı kod oluşturma araçları bir dizi kodlama görevini otomatikleştirmeyi mümkün kılabilmektedir. Bu görevler arasında otomatik kod tamamlama, programlama dilleri arası kod dönüşümü, doğal dil açıklamalarından kod üretimi bulunmaktadır. Kod oluşturma araçları geliştiricilerin daha kompakt, sürdürülebilir, ve yeniden kullanılabilir kod çözümleri oluşturmalarına imkan tanıyarak üretkenliklerini de arttırabilir.

Yapay zekâ araştırma laboratuvarı olan OpenAI tarafından geliştirilen ve Mayıs 2020 yılında tanıtılan GPT-3 (Üretken Ön İşlemeli Dönüştürücü 3), “transformer” olarak adlandırılan yapay sinir ağı tabanlı derin öğrenme sistemi ile insanların yazdığı metinlerden içerik üretmeyi sağlayan özbağımlı bir modelidir [31]. Eğitilmiş dil örnekleri ile doğal dil işleme (NLP) sistemini kullanan ve 175 milyar parametreye sahip GPT-3’e erişim API aracılığı ile sağlanabilmektedir.

```
import datetime

def parse_expenses(expenses_string):
"""Parse the list of expenses and return the list of triples (date, value,
    currency).
Ignore lines starting with #.
Parse the date using datetime.
Example expenses_string:
    2016-01-02 -34.01 USD
    2016-01-03 2.59 DKK
    2016-01-03 -2.72 EUR
    """
```

(a) Öncesi

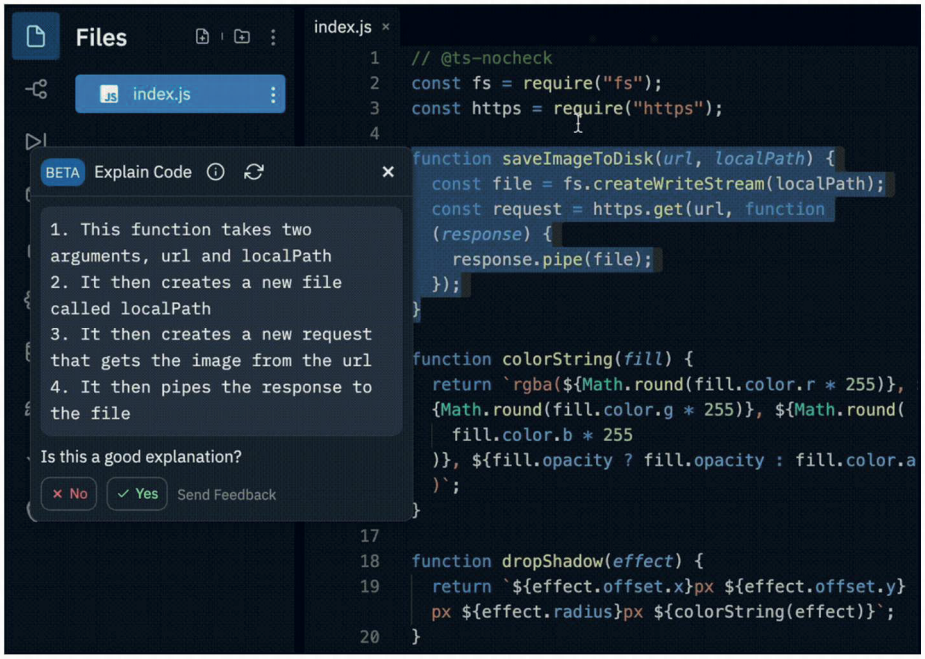
```
import datetime

def parse_expenses(expenses_string):
"""Parse the list of expenses and return the list of triples (date, value,
    currency).
Ignore lines starting with #.
Parse the date using datetime.
Example expenses_string:
    2016-01-02 -34.01 USD
    2016-01-03 2.59 DKK
    2016-01-03 -2.72 EUR
    """

    expenses = []
    for line in expenses_string.splitlines():
        if line.startswith("#"):
            continue
        date, value, currency = line.split(" ")
        expenses.append((datetime.datetime.strptime(date, "%Y-%m-%d"),float
            (value),currency))
    return expenses
```

(b) Sonrası

Şekil 8.7. GitHub Copilot ile Python kodu oluşturma

OpenAI'ın doğal dili koda çeviren GPT-3 tabanlı yapay zekâ modeli olan OpenAI Codex, GPT-3 modeline ek olarak 54 milyon GitHub havuzundan 159 GB Python kodunu da eğitim setinde kullanmaktadır [32, 33]. 2021 Ağustos yılından itibaren sağlanan API aracılığı ile kullanıma açılan OpenAI Codex, kod içerisindeki yorumlara ve dosyanın içeriğine dayalı eş zamanlı kod önerileri sağlayan GitHub Copilot'unda temelini oluşturmaktadır [34]. OpenAI, Codex'in en başarılı sonucu Python programlama dillinde verdiğini öne sürmekle birlikte JavaScript, Go, Perl, PHP, Ruby, Swift, Shell, ve TypeScript gibi birçok programlama dillinde de etkin şekilde çalışabildiğini iddia etmektedir. OpenAI Codex dışında yapay zekâ tabanlı kod üretebilen Tabnine, CodeT5, PolyCoder ve Cogram gibi kod oluşturucular da bulunmaktadır.

Şekil 8.7'de GitHub Copilot tarafından geliştiricinin doğal dille yazmış olduğu açıklamalara dayanarak eş zamanlı oluşturulan Python kodu örneği yer almaktadır [34]. Şekilden de görülebileceği üzere, geliştirici oluşturmak istediği kodu doğal dille belirtmiş ve kod bloğu otomatik olarak oluşturulmuştur.



Şekil 8.8. Replit kod açıklaması

Çevrimiçi tümleşik geliştirme ortamı sağlayan Replit ise OpenAI Codex'i JavaScript ve Python programlama dillerinde yazılmış kodun ne yaptığı açıklamak için kodu doğal dile çevirmektedir [35]. Şekil 8.8'de Replit web sitesinden alınan örnekte seçilen saveImageToDisk fonksiyonunun ne yaptığı kod açıklama kısmında belirtilmiştir.

Yapay zekâ ile kod oluşturma aynı zamanda programlama dilleri arasında kod dönüşümleri içinde kullanılabilmektedir. Eski bir programlama dilinde yazılmış ve hala çalışmaya devam eden bir yazılımın kodunu güncel bir programlama diline çevirmek gerekebilmektedir. Örneğin, Cobol (ana bilgisayar sistemlerinde bugün bile yaygın olarak kullanılmakta olan eski bir programlama dilidir) ile oluşturulan yazılımların bakımının yapılması zor olacağından, kodunun Java veya C++ koduna dönüştürülmesi tercih edilebilir. Programlama dilleri arasında kod dönüşümünü daha kolay ve verimli hale getirebilmek için MetaAI yapay zekâ laboratuvarı öz-denetimli öğrenme (self-supervised) tabanlı TransCoder modelini geliştirmiştir [36]. TransCoder önemli bir özelliği geleneksel denetimli öğrenme modellerin aksine modeli eğitmek için büyük ölçekli paralel veri setleri gerektirmeden kod dönüşümünü sağlayan ilk yapay zekâ sistemi olmasıdır.

8.8.2. Kod Güvenliği

Yapay zekâ destekli kod oluşturma günden güne gelişmekte olsa da henüz büyük ve karmaşık programları üretmekte yeterli değildir. Kod güvenliği açısından değerlendirdiğimizde ise yapay zekâ tarafından oluşturulan kodun çeşitli avantajları olabilmektedir.

Yazılım testi hem otomatik araçlarla hem de manuel analizlerle yapıldığı takdirde hatalar son kullanıcıya yazılım teslim edilmeden önce büyük oranda bulunup düzeltilebilir. Test otomasyonunun oluşturulması zaman aldığından ve uzmanlık gerektirdiğinden günümüzde ihmal edilebilmektedir. Kod oluşturan yapay zekâ araçlarına oluşturdukları her fonksiyon için aynı zamanda otomatik test üretme işlevi kolaylıkla eklenebilir ve kodun daha iyi test edilmesi sağlanabilir.

Yapay zekâ ile üretilen kodlarda tekilleştirme oluşturularak, geliştiriciler tarafından yeni oluşturulan kodun yerini alabilecek eşdeğerinin tespiti ve önerilmesi potansiyel siber saldırı alanını azaltabilir. Örneğin, daha önce iyi

test edilmiş, güncel, ve yaygın olarak benimsenen bir kütüphane geliştirici tarafından oluşturulan yeni kodun eşdeğeri ise, bu kütüphanenin kullanılması kodun daha güvenli olmasına imkan tanıyabilir.

Yapay zekâ kod oluşturma araçları, kütüphanelerin en güncel sürümünü kullanarak kod oluşturabildikleri için güvenlik açığı riskini minimize edebilirler.

Güvenlik açığı nedeni ile oluşturulan güncellemeleri geliştiriciler geriye dönük uyumluluk endişesi nedeni uygulamayabilirler. Bu durumda yapay zekâ kullanımı ile güncel sürüme entegrasyon hızlı bir şekilde sağlanarak güncelleme süresi kısaltılabilir ve güvenlik açığına karşı tedbir alınabilir.

Intel'in geliştirmiş olduğu ControlFlag ise kod güvenliğine koddaki hataları otonom olarak tespit ederek katkıda bulunmaktadır [37]. Makine öğrenmesi, derleyiciler, programlama dilleri, bilgisayar sistemleri, ve biçimsel yöntemleri birleştiren ControlFlag, anormallik algılama mantığı ile çalışmaktadır. Anormalliği tespit edebilmek için doğrulanmış örneklerden normal kodlama kalıplarını öğrenirken, denetimsiz (unsupervised) yaklaşım ile de geliştiricinin kodlama stiline uyum sağlayabilmektedir.

8.8.3. Sistem Güvenliği

Yapay zekâ tabanlı sistemlerin tasarımı yapılırken yazılım ve donanım güvenliği, veri bütünlüğü, model gizliliği, model sağlamlılığı, ve veri gizliliği gibi çeşitli güvenlik sorunları göz önünde bulundurulmalıdır. Tipik yapay zekâ sistemleri saldırıları ve savunma yöntemleri aşağıda özetlenmiştir [38].

- **Kaçınma saldırıları:** Yapay zekâ modelinin girdileri doğru tanımlayamaması için saldırgan tarafından girdi verileri üzerinde değişiklikler yapılır. Yapılan değişiklikler insan gözü ile ayırt edilemezken, modelin yanlış tahminde bulunmasına neden olur.
  - Model eğitim aşamasında birden çok derin sinir ağının birleştirilerek çalıştırılması,
  - Bilinen saldırı yöntemleri ile oluşturulan değiştirilmiş verinin, eğitim setine eklenerek model dayanıklılığının arttırılması,
  - Harici bir algılama modeli kullanılarak girdi üzerinde değişiklik olup olmadığının tespit edilmesi,

- Model oluşturma aşamasında girdi örneklerinin deforme edilerek çalıştırılması,
- Derin sinir ağları doğrulama teknolojisinin kullanılması, kaçınma saldırılarına karşı savunma yöntemlerindendir.

- **Zehirleme saldırıları:** Model eğitildikten sonra girdideki değişikliklere uyum sağlamak için toplanan yeni verilerle yeniden eğitilmesi gerekebilir. Bu aşamada saldırgan özenle hazırlanmış örnekleri eğitim verilerine enjekte ederek yapay zekâ sisteminin normal işlevini bozmayı amaçlar.
  - Eğitim veri seti kontrolüne odaklanılarak veri setinde filtreleme yapılması,
  - İstatiksel yöntemlerle verideki gürültü ve anormal değerlerin tespit edilmesi,
  - Sistemin farklı eğitim seti kullanan birden fazla alt modelden oluşturulması, zehirleme saldırılarına karşı savunma yöntemlerindendir.
- **Arka kapı saldırıları:** Yapay zekâ sistemlerine de geleneksel programlar gibi arka kapılar eklenebilir ve bu kapılar sadece belirli kişiler tarafından tetiklenebilir. Fakat bir sinir ağı modeli kaynak kodu yerine bir dizi parametreden oluştuğu için tespit edilmesi daha zordur. Arka kapı saldırılarının çoğu modellerin oluşturulması veya iletilmesi sırasında meydana gelir.
  - Arka kapıları tetikleyebilecek girdilerin filtrelenmesi,
  - Orijinal modelin sinir hücrelerinin, modelin işlevini etkilemeyecek şekilde budanması, arka kapı saldırılarına karşı savunma yöntemlerindendir.
- **Model çıkarma saldırıları:** Saldırgan verinin giriş, çıkış, ve diğer bilgilerini analiz ederek model veya eğitim verisini tespit etmeyi hedefler.
  - Model eğitimi sırasında eğitim verilerinin bağımsız derin sinir ağlarını eğitmek için bölünmesi ve bu ağların oylama yoluyla modeli ortaklaşa eğitmek için kullanılması,

- Model gizliliğini sağlamak amacı ile model eğitimi aşamasında verilere gürültü eklenmesi,
- Model eğitimi sırasında orijinal modelin çalınıp çalınmadığının kontrol edilmesini için özel tanıma sinir hücrelerinin oluşturulması, model çıkarma saldırılarına karşı savunma yöntemlerindendir.

8.9. SONUÇ VE DEĞERLENDİRMELER

Bu bölümde özetlendiği gibi kod karıştırma; güvenli kodlamanın bir parçası olmakla beraber yazılımın kodunu daha da korunaklı hale getirebilmek için en çok tercih edilen yöntemlerden bir tanesidir. Bu konu aşağıda farklı başlıklar altında değerlendirilmiştir.

Güvenli kodlama açısından değerlendirildiğinde;

- Geliştiriciler yazılımlarında oluşabilecek güvenlik açıklarını minimize edebilmek için, bu bölümde açıklanan güvenli kodlama prensiplerine başvurabilirler.
- OWASP Güvenli Kodlama Pratikleri ve SEI Cert Kodlama Standartları, başvurulabilecek önemli güvenli kodlama standartları sunmaktadır.
- Kod karıştırma teknikleri yazılımın davranışı etkilemeyecek şekilde kodun anlaşılırlığını ve analizini zorlaştıracak değişiklikleri içermekle beraber yazılım güvenliğine katkı sağlarlar.
- Kod geliştiriciler yazılımda yaptıkları değişiklikler ile güvenlik açığı keşfine, zararlı kod eklenmesine, ticari sır hırsızlığına, yetkisiz erişime, lisanslama ve diğer denetimleri atlatma gibi sorunlara karşı tedbir alabilirler.
- Tersine mühendislik yöntemleri uygulansa bile, kod karıştırma uygulanmış kodun analizini yapmak maliyetli olacaktır. Özellikle sondaki adam saldırılarına karşı kod karıştırmanın kullanılması önerilmektedir.
- Kullanılan programlama dillinin yapısına göre, kod karıştırma tekniklerinin hedefi değişebilmektedir. Örneğin, JavaScript gibi yorumlayıcı kullanan dillerde kaynak kodunun paylaşılması gerektiğinden,

kod karıştırma teknikleri ile direkt kaynak kod üzerinde işlem yapılması gerekmektedir.

- Kod karıştırma teknikleri kötü amaçlı yazılımları oluşturmada, kod intihalini gizlemede, koruma tarayıcılarını ve denetimleri atlatmada da kullanılabilmektedir.

Güvenli kodlamaya yapay zekâ uygulandığı durumda yapılan değerlendirmede;

- Makine öğrenme modellerinin elde edilmesi ve niteliklerinin ortaya çıkarılması üzerine çalışmalar yapılmaktadır.
- Kod karıştırma teknikleri uygulanmış yazılımlara karşı yapay zekâ ve makine öğrenmesi algoritmaları, kod karıştırma tekniklerinin belirlenmesi, kötü amaçlı yazılım tespiti, kod benzerliği, kod intihal tespiti gibi alanlarda kullanılabilir.
- Yapay zekâ hem otomatik kod tamamlama ve oluşturmada hem de kodun ne yaptığının açıklanmasında geliştiricilere yardımcı olmak için kullanılmaya başlanmıştır.
- Yapay zekâ tabanlı kod oluşturma araçları kod güvenliğine katkıda bulunabilmektedir.
- Kullanılan makine öğrenmesi ve derin öğrenme yapılarına bakıldığında; ortalama algılayıcı, bayes nokta makinesi, güçlendirilmiş karar ağacı, karar ormanları, yerel derin destek vektör makinesi, lojistik regresyon, sinir ağları, CNN gibi modeller kullanılmaktadır.
- Doğal dil işleme ve GPT-3 modeli, yapay zekâ ile kod oluşturmada yaygın olarak kullanılmaktadır.
- Yapay zekâ sistemleri de çeşitli siber saldırılara maruz kalabildiği için, sistem güvenliğini sağlamak adına savunma tedbirlerine başvurulmalıdır.

Sonuç olarak;

- Son dönemde güvenli kod ve kod karıştırma analizleri için yapay zekâ yaklaşımları kullanılmaya başlanmıştır.
- Günümüzde karşılaşılan tehditler ve kodlara yapılan saldırılar için pek çok yöntem kullanılsa da yapay zekâ sistemleri; anormal du-

rumları tespit etme, dinamik durumları izleme ve değerlendirme, kodların izinsiz kullanımını belirleme, kodlara zararlı kod enjeksiyonunu tespit etme gibi pek çok alanda kullanılabilecek çözümler sunma potansiyeline sahiptir.

- Bu bölümde sunulan yaklaşımların ve yapay zekâ çözümlerinin kodların güvenliğini ve kodlara olan güveni artırıcı çalışmalara katkılar sağlayacağı değerlendirilmektedir.

KAYNAKLAR

[1] "OWASP Secure Coding Practices-Quick Reference Guide," [Online]. Available: https://owasp.org/www-project-secure-coding-practices-quick-reference-guide/migrated_content. [Accessed Ağustos 2022].

[2] "SEI CERT Coding Standards," [Online]. Available: https://wiki.sei.cmu.edu/confluence/display/seccode. [Accessed Ağustos 2022].

[3] "OWASP Top Ten," [Online]. Available: https://owasp.org/www-project-top-ten/.

[4] "CVE," [Online]. Available: https://cve.mitre.org. [Accessed Ağustos 2022].

[5] "National Vulnerability Database," [Online]. Available: https://nvd.nist.gov. [Accessed Ağustos 2022].

[6] "Businessinsider," [Online]. Available: https://www.businessinsider.com/solarwinds-hack-explained-government-agencies-cyber-security-2020-12. [Accessed Ağustos 2022].

[7] "CNN Business," [Online]. Available: https://money.cnn.com/2018/03/23/technology/china-us-trump-tariffs-ip-theft/index.html. [Accessed Ağustos 2022].

[8] "BleepingComputer," [Online]. Available: https://www.bleepingcomputer.com/news/microsoft/lapsus-hackers-leak-37gb-of-microsofts-alleged-source-code/. [Accessed Ağustos 2022].

[9] "BSA," [Online]. Available: https://gss.bsa.org/wp-content/uploads/2018/05/2018_BSA_GSS_Report_en.pdf. [Accessed Ağustos 2022].

[10] Florian Tramer, Fan Zhang, Ari Juels, Michael K. Reiter, and Thomas Ristenpart, "Stealing machine learning models via prediction APIs," in *Proceedings of the 25th USENIX Conference on Security Symposium*, 2016.

[11] Seong Joon Oh and Maximilian Augustin and Mario Fritz and Bernt Schiele, "Towards Reverse-Engineering Black-Box Neural Networks," in *ICLR*, 2018.

[12] Orekondy, Tribhuvanesh and Schiele, Bernt and Fritz, Mario, "Knockoff Nets: Stealing Functionality of Black-Box Models," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[13] Nicolas Papernot and Patrick Mcdaniel and Ian J. Goodfellow and Somesh Jha and Z. Berkay Celik and Ananthram Swami, "Practical Black-Box Attacks against Machine Learning," in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, 2017.

[14] R. Shokri, M. Stronati, C. Song and V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," in *IEEE Symposium on Security and Privacy* , 2017.

[15] B. Wang and N. Gong, "Stealing Hyperparameters in Machine Learning," in *IEEE Symposium on Security and Privacy*, 2018.

[16] C. Collberg, J. Davidson, R. Giacobazzi, Y. X. Gu, A. Herzberg and F. Wang, "Toward Digital Asset Protection," *IEEE Intelligent Systems,* vol. 26, no. 6, pp. 8-13, 2011.

[17] Adnan Akhunzada, Mehdi Sookhak, Nor Badrul Anuar, Abdullah Gani, Ejaz Ahmed, Muhammad Shiraz, Steven Furnell, Amir Hayat, Muhammad Khurram Khan, "Man-At-The-End attacks: Analysis, taxonomy, human aspects, motivation and future directions," *Journal of Network and Computer Applications,* vol. 48, pp. 44-57, 2015.

[18] Christian S. Collberg and Clark D. Thomborson and Douglas Low, "A Taxonomy of Obfuscating Transformations," 1997.

[19] Sebastian Banescu and Alexander Pretschner, "Chapter Five - A Tutorial on Software Obfuscation," in *Advances in Computers*, Elsevier, 2018, pp. 283-353.

[20] "JavaScript Obfuscator Tool," [Online]. Available: https://obfuscator.io/. [Accessed Ağustos 2022].

[21] "PreEmptive-What is Obfuscation?," [Online]. Available: https://www.preemptive.com/obfuscation/. [Accessed Ağustos 2022].

[22] "GuardSquare-What is code obfuscation?," [Online]. Available: https://www.guardsquare.com/what-is-code-obfuscation. [Accessed Ağustos 2022].

[23] Blazy, Sandrine and Trieu, Alix, "Formal verification of control-flow graph flattening," in *5th ACM SIGPLAN Conference on Certified Programs and Proofs*, 2016.

[24] Skoudis, Ed and Zeltser, Lenny, Malware: Fighting Malicious Code, Prentice Hall PTRUpper Saddle River, NJUnited States, 2003.

[25] Mauro Conti and Vinod P. and Alessio Vitella, "Obfuscation detection in Android applications using deep learning," *Journal of Information Security and Applications,* vol. 70, 2022.

[26] "Androguard," [Online]. Available: https://androguard.readthedocs.io/en/latest/. [Accessed Ağustos 2022].

[27] O. Mirzaei and J.M. {de Fuentes} and J. Tapiador and L. Gonzalez-Manzano, "AndrODet: An adaptive Android obfuscation detector," *Future Generation Computer Systems,* vol. 90, pp. 240-261, 2019.

[28] Rennhard, Bernhard Tellenbach and Sergio Paganoni and Marc, "Detecting obfuscated JavaScripts from known and unknown obfuscators using machine learning," 2016.

[29] Marastoni, Niccolo and Giacobazzi, Roberto and Dalla Preda, Mila, "A Deep Learning Approach to Program Similarity," in *1st International Workshop on Machine Learning and Software Engineering in Symbiosis*, 2018.

[30] Zafar, Sarim and Sarwar, Muhammad Usman and Salem, Saeed and Malik, Muhammad Zubair, "Language and Obfuscation Oblivious Source Code Authorship Attribution," *IEEE Access,* vol. 8, 2020.

[31] "Wikipedia GPT-3," [Online]. Available: https://en.wikipedia.org/wiki/GPT-3. [Accessed Eylül 2022].

[32] "OpenAI Codex," [Online]. Available: https://openai.com/blog/openai-codex/. [Accessed Eylül 2022].

[33] "Wikipeida OpenAI Codex," [Online]. Available: https://en.wikipedia.org/wiki/OpenAI_Codex. [Accessed Eylül 2022].

[34] "GitHub Copilot," [Online]. Available: https://github.com/features/copilot. [Accessed Eylül 2022].

[35] "Replit Codex," [Online]. Available: https://blog.replit.com/codex. [Accessed Eylül 2022].

[36] "MetaAI," [Online]. Available: https://ai.facebook.com/blog/deep-learning-to-translate-between-programming-languages/. [Accessed Eylül 2022].

[37] "Intel ControlFlag," [Online]. Available: https://www.intel.com/content/www/us/en/newsroom/news/machine-programming-tool-detects-bugs-code.html. [Accessed Eylül 2022].

[38] "Huawei AI Security," [Online]. Available: https://www-file.huawei.com/-/media/corporate/pdf/trust-center/ai-security-whitepaper.pdf. [Accessed Eylül 2022].

Bölüm 9

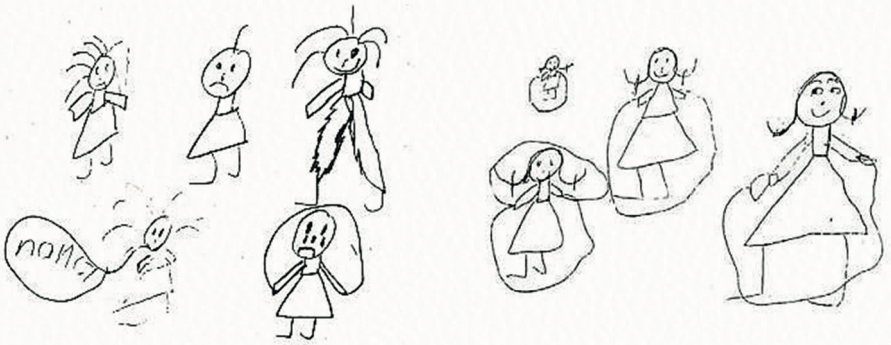
YAPAY ZEKÂ İLE ÇOCUK ÇİZİMLERİ ANALİZİ

S. Mine Okumuş, Pelin Canbay

Çocuk gelişimi ve psikolojisi üzerine yapılan araştırmalar çocukların sözlü iletişim yerine çizimleri ile kendilerini daha iyi ifade edebildiğini göstermektedir. Çocuklar öfke, korku, kıskançlık gibi duygularını yaptıkları çizimlerde kullandıkları renkler veya bazı özel şekiller ile aktarmaya çalışırlar. Resimlerde kullanılan renklerin ve şekillerin çocuk psikolojisinde ne anlama geldiğini belirleyen kalıplar ve resim temaları vardır. Bu temalar üzerine yapılan analizler sonucu çocukların psikolojik veya sosyal desteğe ihtiyacı olup olmadığı gibi önemli durumlar belirlenebilmektedir. Bir psikoloji uzmanı tarafından uygulanması gereken bu analizler oldukça maliyetli, zaman alıcı ve yorucudur. Diğer taraftan, coğrafi ve kültürel farklılıklar göz önüne alındığında her çocuğun bir psikoloji uzmanına erişimi mümkün olmamaktadır. Günümüz gelişen teknolojik araç ve uygulamaları yardımı ile bu analizlerin daha kısa sürede, daha az maliyetle ve daha geniş kitlelerce kullanılabilirliği sağlanabilir. Çocuk çizimlerinden analizler yapabilecek ve bu analizleri bir uzmana sunabilecek yapay zekâ tabanlı karar destek sistemlerine hem ulusal hem uluslararası alanda büyük ihtiyaç vardır. Bu ihtiyaç doğrultusunda hazırlamış olduğumuz çalışma, psikoloji ve pedagoji alanında çocuk çizimleri analizinde kullanılan bazı teknik ve kalıpları tanıtmakta, bu analiz işlemlerinde kullanılan yapay zekâ destekli çalışmaları farklılıkları ve eksiklikleri ile incelemekte ve çocuk çizimlerinin analizi ile ilgili yapay zekâ alanında yapılabilecek ileriki çalışmaların potansiyelini sunmaktadır.

9.1. GİRİŞ

İnsanlar, tarih öncesi çağlardan beri görsel dünyalarını tasvir etmek için çizimleri kullanmışlardır. Bugün bile, çizimler, muhtemelen tüm insanlar için hazır olan tek görüntü oluşturma tekniğidir [1]. Özellikle çocukluk dönemi gibi, sözlü iletişimde verimin düşük olacağı zamanlarda günümüzde hala bazı durumların tasviri için çizimlere başvurulmaktadır. Çocukların psikolojik ve sosyolojik durum analizleri için de en çok kullanılan yöntem çizimlerinin incelenmesidir [2]. Çocukların ruhsal olarak gelişme süreçlerini sağlıklı bir şekilde devam ettirebilmeleri için, kendilerini rahatlıkla ifade edebilmeleri gerekir. Belli kalıplar çerçevesinde incelenen çocuk çizimleri çocukların duygusal, zihinsel ve sosyolojik gelişimi hakkında önemli bilgiler verebilmektedir. Çizimler, davranış problemi olan çocuklar için de bir araç olup, sadece onların bilişsel gelişimin takibinde değil ayrıca sosyal çevrenin çocuk üzerindeki etkilerinin incelenmesinde de kullanılabilmektedir. Kendini aile içerisinde dışlanmış hisseden bir çocuk çizimlerinde kendisi ve ailesi arasına mesafe koyarken [3], bir başka çocuk sosyal çevresindeki akranları ile olan ilişkisini çizimlerle yansıtabilmektedir [4]. Örneğin; Avrupa'dan Amerika'ya ailesi ile göç etmiş, yedi yaşında bir kız çocuğunun sosyal ortama girdiği ilk zamanlarda çizdiği resim Şekil 9.1(a)'da, belli bir süre geçtikten sonra, ortama uyum sağladığı düşünüldüğünde çizmiş olduğu resim Şekil 9.1(b)'de gösterilmektedir [5].



a. Göç sonrası uyumsuzluk dönemi çocuk çizimi

b. Adaptasyon sonrası çocuk çizimi

Şekil 9.1. Göç etmiş bir ailenin 7 yaşındaki kız çocuğunun çizimleri

Yapılan uzman analizlerine göre; çocuğun katı disiplinli bir Orta Avrupa eğitiminden gelen bir çocuk olduğu için, göç ettiği ülkede okullarındaki daha serbest ortam karşısında uzun bir süre kendini toparlayamamış, bocalamıştır. Çocuğun, göç sonrası uyumsuzluk döneminde yaptığı çizimde kendisini iki kez çizdiği görülmektedir (Şekil 9.1(a)'da aşağıda en sağdaki ve yukarıda ortada yer alan çocuk figürü.) [5]. Çocuğun çizdiği ikinci resimde geçen süredeki farklılık görülebilmektedir. O dönemde çocuk, göç ettiği bu yabancı ülkede zamanla uyumsuzluk sorununu aşmış, diğer çocuklarla arkadaş olmaya başlamış ve bu nedenle figürler arasındaki uyumsuzluk da kaybolmuştur [5].

Günümüzde birçok alanda yapay zekâ destekli uygulamalar başarılı bir şekilde kullanılmaktadır. Bu uygulamalar sayesinde başta sağlık alanında olmak üzere birçok alanda, zaman, maliyet ve olası hatalar bakımından büyük oranda iyileşmeler gözlemlenmiştir [6], [7]. Bir uzman gözleminde bulunma fırsatı olmayan ama şiddet, taciz, zorbalık, dışlanma gibi istenmeyen durumlara maruz kalan çocukların kısa sürede tespit edilebilmesi için çocukların çizimleri üzerinde bir uzman gibi karşılaştırma yapabilen yapay zekâ ile desteklenmiş bir görüntü işleme sistemine olan ihtiyaç her geçen gün artmaktadır. Yapay zekâ destekli bir sistemin eğitilip konunun bilincinde olan rehber öğretmenler veya psikolojik danışmanların kullanımına sunulması, onların çocukların durumu ile ilgili kısa sürede kararlar vermelerine destek olunması çocukların yaşadığı zorluklar için bir çözüm fırsatı olabilir. Dolayısıyla çocuk psikolojisi alanındaki uzmanlara destek olacak, resim analizi sürecini kısaltarak sıkıntıları olan çocuklara ulaşmayı kolaylaştıracak bir sisteme ihtiyaç duyulmaktadır [8].

Görüntü işleme sistemi oluşturulması için öncelikle sistemi eğitebileceğimiz verilen toplanması gerekmektedir. Bu amaçla çocuklara çizdirilen resimlerin uzmanlar tarafından değerlendirilmesi ve ortak noktalarının belirlenmesi ile işe başlanmalıdır. Ardından yapay zekâ yöntemleri ile özellikle de günümüz yüksek başarılı derin öğrenme mimarileri ile kurulan bir görüntü işleme sistemine tanıtılan bu resimlerle sistemimizin görüntü analizi için eğitilmesi gerekmektedir. Eğitim verilerinin yeterli miktarda olması sistemin doğruluk oranının artırılması açısından önem arz etmektedir. Yeterince iyi eğitilen bir yapay zekâ sisteminde görüntü analizi sonuçlarının doğruluk oranlarının yüksek çıkması beklenir [9].

Literatürde çocuk çizimlerinin analizi ile ilgili çeşitli çalışmalar yapılmış ve çocuk çizimlerini içeren farklı veri setleri oluşturulmuştur. İnsan çi-

zimlerinin ilk büyük ölçekli araştırması kabul edilen Eitz ve arkadaşları tarafından oluşturulan 250'den fazla kategoride 20.000'den fazla benzersiz çizimden oluşan veri seti, çizim analizi çalışmaları için bir kilometre taşı niteliğindedir [1]. Görüntü sınıflandırması için Destek Vektör Makinesi (Support Vector Machine - SVM) yöntemi kullanılan veri setinde doğruluk oranı %56'ya kadar ulaşmıştır. Veri seti açık kaynak olarak kullanıma sunulduğundan bu alandaki bir çok araştırma için de öncülük ettiği görülmektedir, ancak bu veri seti dijital çizimlerden oluşmaktadır ve bu durum hem 4-9 yaş arası çocukların kullanımına uygun olmaması sebebi ile hem de psikologlar tarafından dijital çizimlerin analiz için kabul edilmemesi sebebi ile çocuk resimleri analizinde yetersiz kalmaktadır [10].

Çocuk çizimlerinin yaş gruplarına göre sınıflandırılmasını amaçlayan bir araştırmada 4-11 yaş grubu çocukların çizimleri üzerinde çalışma yapılmıştır. Googenough "Bir Kişi Çiz" testi [11] üzerinde yapılan çalışmaya 1000'den fazla çizim aktarılmış ve 4-11 arasındaki yaş grubu için 8 farklı sınıf oluşturulmuştur. Bu uygulamada farklı sınıflandırma metotları ve bunlara karşılık gelen doğruluk oranları karşılaştırılmış, elde edilen sonuçlara göre %62 oranında doğruluk elde edilmiştir [12].

Psikolojik analizlerde doğru sonuç üretmenin insanlar için bile zor bir görev olduğu kabul edilmektedir. Bu analizlerin, psikoloji veya pedagoji alanında uzman kişiler tarafından etiketlenmiş verilerle eğitilen, yapay zekâ destekli uygulamalar aracılığıyla yapılabilmesi hem alanın uzmanı kişilere kolaylık sağlayacak hem de daha kısa sürede daha fazla çocuğa destek olunmasında katkı sunacaktır. Birçok alanda insanlığa faydalı çözümler sunan yapay zekâ tabanlı bu tür karar destek uygulamalarının çocuk çizimlerinin analizinde de kullanılmasına olan ihtiyaç gün geçtikçe artmaktadır. Bu ihtiyaç doğrultusunda, psikoloji ve pedagoji alanında bu analizlerin uzmanlar tarafından nasıl yapıldığı, literatürde bu konuda hangi çalışmaların yürütüldüğü ve günümüz yapay zekâ teknikleri ile konuda nasıl fayda sağlanabileceği çalışmamızda ayrıntılı olarak ele alınmıştır.

Çalışmanın 2. bölümünde, çocuk çizimleri üzerine yapılan, psikoloji ve pedagoji alanındaki uzmanlar tarafından aktif olarak kullanılan bazı kalıplar, çıkarımlar, testler, analizler ve yorumlamalar sunulmuştur. 3. bölümde çocuk çizimleri üzerine yapılan bazı yapay zekâ tabanlı çalışmalar örneklendirilmiş, bu çalışmalarda kullanılan yöntemler, elde edilen çıkarımlar, çalışma-

ların pratikte nasıl kullanılabileceği ve avantajları ayrıntılı olarak ele alınmıştır. Çalışmanın sonuç bölümünde ise, çocuk çizimlerinin analizinde yapay zekâ tabanlı uygulamaların kullanılmasının önemi ve bu alanda yapılacak çalışmalara olan gereksinimler belirtilmiş, çalışmanın genelinde sunulmuş araştırma ve çıkarımların yorumlaması yapılmış, bu alanda yapılabilecek ileriki çalışmaların potansiyeli sunulmuştur.

9.2. ÇOCUK ÇİZİMLERİ ÜZERİNE YAPILAN ANALİZLER VE YORUMLAMALAR

Bireylerin bilişsel gelişimi üzerine çalışmalar yapan araştırmacılar, bireylerin sergilediği davranışların genel geçer, basit ve tesadüfi olmadığını, davranışların nedenlerinin araştırılması gerektiğini belirtmişlerdir [13]. Çocuk çizimleri analizinde yol haritası olarak kabul edilen "Sanatsal Gelişim Evreleri" teorisine göre "Çocuklar hazır olduklarında ya da bir gelişim döneminin özelliklerini yerine getirdiklerinde bir sonraki döneme geçmektedirler" görüşü ile çocukların çizimlerinde bir gelişim evresi / çizgisi izledikleri kabul edilmektedir [14]. Bu bakış açısıyla, gelişim evrelerine göre çocukların insan figürü çizimleri örnekleri Tablo 9.1'de gösterilmiştir [3].

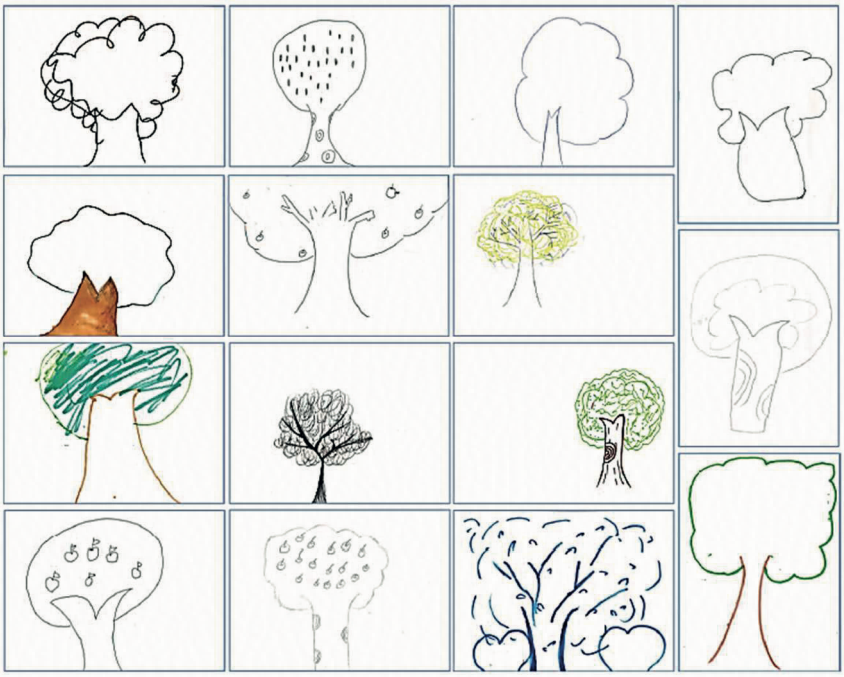
Tablo 9.1. Gelişim evrelerine göre çocuk çizimleri

Çizim Örnekleri					
Gelişim Evresi	Karalama Dönemi (2 - 4 yaş)	Şema Öncesi Dönem (4 - 7 yaş)	Şematik Dönem (7 – 9 yaş)	Gerçeklik Dönemi (9 – 12 yaş)	Mantık Dönemi (12 – 14 yaş)

Çocukların kendini nasıl algıladığı, duyguları, ailesi ve çevresindekiler hakkında görüşleri ile ilgili bilgi toplamak için yapılan testler, Projektif Testler olarak bilinir [15]. Çoğunlukla klinik ortamda bir uzman tarafından uygulanan çeşitli projektif testler bulunmaktadır. Aynı zamanda okul ortamında rehber öğretmenler ya da sınıf öğretmenleri tarafından da bu testler uygulanarak çocuğun ruhsal ve zihinsel durumu hakkında fikir sahibi olunabilmektedir. Yapılacak olan analiz bu testlerin temel yapısı üzerine kurulduğunda başarı oranı artmaktadır. Değerlendirme amacıyla çocuklara uygulanan "Goodenough-Harris İnsan Çizimi Testi", "Aile Çizim Testi", "Duygu ve Renk Çemberi Uygulaması" gibi projektif testler örnek olarak verilebilir. Çocukların ihtiyaçlarına ya da yaşadıkları problemlere göre hangi testin uygulanması gerektiğine uzmanlarca karar verilmektedir. Uzman gözetiminde seçilen teste uygun resim yaptırıldıktan sonra çocukla resim hakkında konuşulması büyük önem taşımaktadır. Böylece çocuğun sorunu daha rahat anlaşılabilmekte ve çözüm için gereken adımlar atılabilmektedir [3]. Bu testlerden bazılarının içeriği ve uygulanış biçimleri aşağıda alt başlıklar halinde verilmiştir.

9.2.1. Bir Ağaç Çiz Testi

Çocuk çizimleri analizi konusunda yapılan çalışmalarda kullanılan projektif testlerden biri "Bir Ağaç Çiz" testidir. Bu testin uygulamalarda seçilmesinin amacı ağaç çiziminde çocukların duygularını daha iyi yansıtması ve kendini daha iyi ifade edebilmesidir. Orijinal veriler, 2019 yılında, Hangzhou Jiangnan Deneysel İlköğretim Okulu, Hangzhou, Çin'den toplanmıştır. Okulun 5. sınıfa giden 300 öğrencisinden A4 boyutundaki boş bir kâğıda bir ağaç resmi çizmesi istenmiştir. Toplanan çizimler ilk olarak psikolojik analiz için uzman psikanalistlere gönderilmiştir. Psikanalistler çocukların çizim yoluyla kolayca yansıtabileceği 6 farklı problem olan kaygı, baskı, saldırganlık, çekingenlik, güvensizlik ve öğrenme güçlüğü sorunları göz önünde bulundurularak çizimleri etiketlemiştir [2]. Veri setinde bulunan çizimlerden bazıları Şekil 9.2'de örneklenmiştir.



Şekil 9.2. Çocukların ağaç çizimlerinden bazı örnekler [2]

9.2.2. Goodenough-Harris Bir Adam Çiz Testi

Goodenuogh – Harris Adam Çizme Testi; çocuğun yaptığı resimler aracılığı ile çocuğun zihinsel gelişimi hakkında bağlantı kurmayı amaçlayan bir testtir. Bu test, bireylerde zihinsel gelişmeyi ölçmek için kullanılan bir puanlama sistemini içerir. Bu test kültürden doğrudan etkilenmemektedir. Çocuğa sadece "Bir insan çiz, yapabileceğinin en iyisi olsun." talimatı verilir ve çizim sırasında konuşmadan, yönlendirme yapmadan çocuk takip edilir. Çocuğa çizim için verilen talimatların basit ve net olması, beklentinin sade olması testin uygulanabilirliğini artırmış ve yorumlanmasını kolaylaştırmıştır [16]. Bir insan çiz testinde, çocuğun motor becerilerinin gelişiminin yanı sıra duygusal bakımdan sorularının var olup olmadığı da izlenebilmektedir. Bu testte, ayrıca cinsel istismara uğramış çocuklar için de göstergeler bulunmaktadır [3].

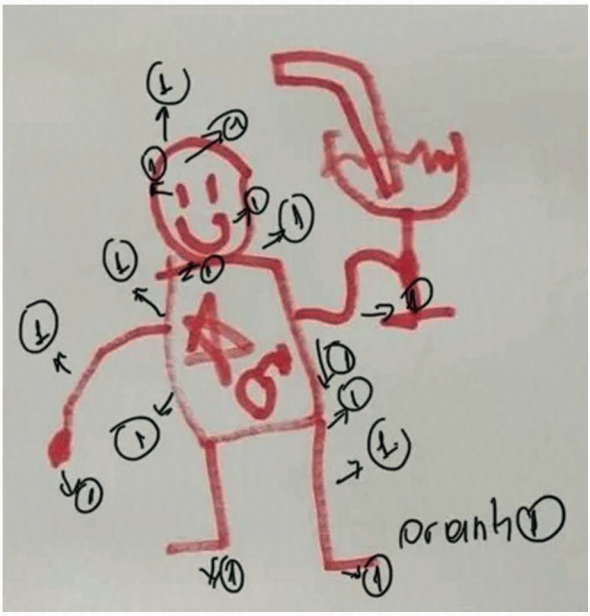
Bir İnsan Çiz Testi için çocuklardan bir insan figürü çizmeleri istenmektedir. Çocukların çizimlerine, çizimdeki öğelerin varlığı ya da eksikliğine göre bir puanlandırma yapılmaktadır. Resmin tamamı değerlendirildikten sonra toplam

puan hesaplanır. Hesaplanan toplam puan 0 ile 52 arasındadır. Alınan bu puan resmi çizen kişinin zihinsel yaşını simgeler. Bu ölçeğin altında yatan varsayım, çocukların yapacakları çizimlerin onların genel zihinsel becerilerini yansıtacağı yönündedir [17]-[19]. Ayrıca öğrencilerin motor koordinasyon becerilerinin de bu çizimlerde kendini belli edeceği de düşünülmektedir. Tablo 9.2'de Goodenough tarafından oluşturulan ve çocukların bedensel yaşına karşılık gelmesi beklenen zihinsel puanlar gösterilmektedir [20]. Zihinsel yaşını simgeleyen puan ile bedensel yaşını simgeleyen puan arasındaki farkın fazla olması çocuğun gelişimsel olarak ileri ya da geri olduğuna dair bize bir fikir verir.

Tablo 9.2. Goodenough "Bir Kişi Çiz Testi" Zekâ Yaşı - Puan Tablosu

Puan	Zekâ Yaşı	Puan	Zekâ Yaşı
1	3,3	26	9,6
2	3,6	27	9,9
3	3,9	28	10
4	4	29	10,3
5	4,3	30	10,6
6	4,6	31	10,9
7	4,9	32	11
8	5	33	11,3
9	5,3	34	11,6
10	5,6	35	11,9
11	5,9	36	12
12	6	37	12,3
13	6,3	38	12,6
14	6,6	39	12,9
15	6,9	40	13
16	7	41	13,3
17	7,3	42	13,6
18	7,6	43	13,9
19	7,9	44	14
20	8	45	14,3
21	8,3	46	14,6
22	8,6	47	14,9
23	8,9	48	15
24	9	49	15,3
25	9,3	50	15,6

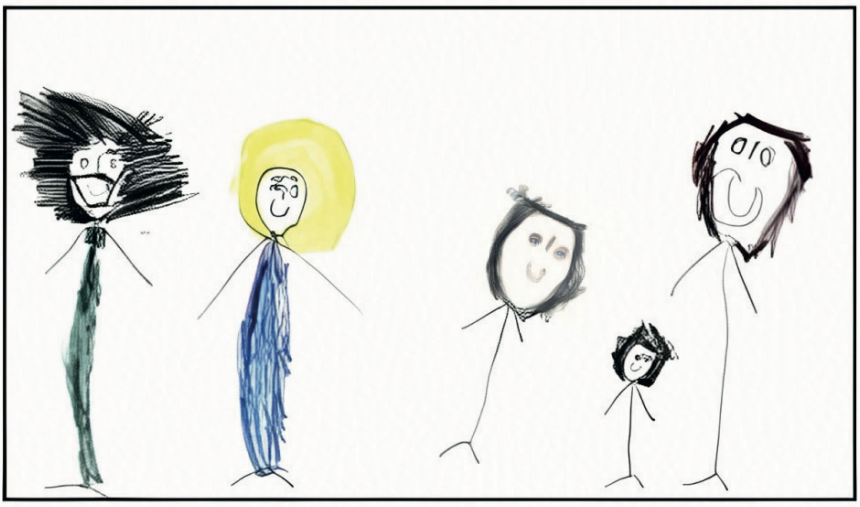
Şekil 9.3'te zihinsel gelişimi normal seyreden 7 yaşındaki erkek çocuğuna ait örnek bir çizim ve puanlaması gösterilmiştir [21]. Yapılan çizimin toplam puanı 17 olarak belirlenmiştir. Tablo 9.2'ye göre 17 puan 7,3 yaşa denk gelmektedir ve bu durum çocuğun zihinsel yaşının bedensel yaşına uygun olduğunu göstermektedir.



Şekil 9.3. Bir İnsan Çiz Testi Sonucu [21]

9.2.3. Bir Aile Çiz Testi

Bir aile çiz testi ile aile resmi çizimleri, çocuğun aile içindeki misyonu, diğer aile bireyleriyle olan ilişkileri, aile içindeki rolü ve duruşu, aile içinde kendini nasıl konumlandırdığı, algıladığı ve aile bireyleri ile yaşadığı sorunlarla ilgili bilgiler, ipuçları vermektedir [22]. Çizimlerde yer alan aile bireylerinin sayfa düzenindeki konumu, bireylerin bütününün ya da uzuvlarının varlığı ya da büyüklüğü, bireyler arasındaki mesafeler, engeller veya resimde yer alan fantastik semboller çocuğun aile algısı, aile içindeki güven ortamı, çocuğun bağlanma stilleri ya da sosyal becerileri hakkında fikir verir. Çizimlerde aile bireylerinin sayfadaki konumları da ayrıca rastgele olmayıp simgesel bilgileri içermektedir [23].



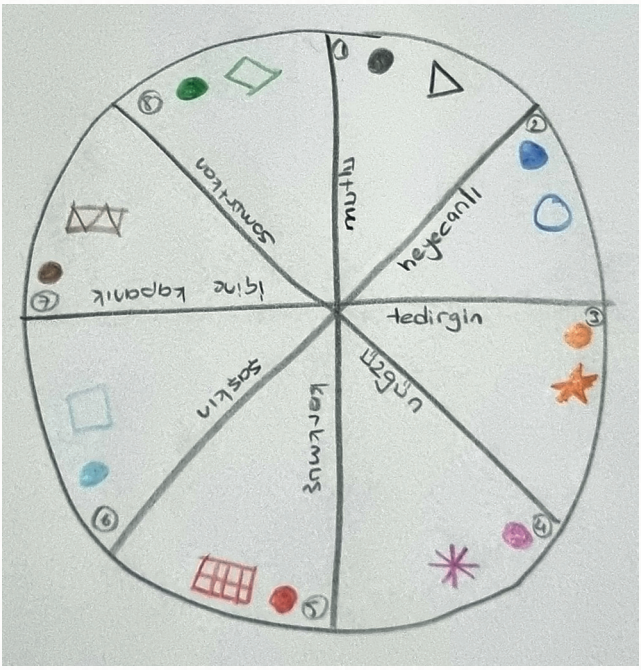
Şekil 9.4. 71 aylık bir kız çocuğu tarafından çizilmiş bir aile resmi örneği

Şekil 9.4'te, bir aile çiz testi kapsamında 6 yaşındaki bir kız çocuğunun çizdiği aile resmi görülmektedir. Çizimlerde çocukların aile bireylerinin tanımlayıcı özelliklerine dikkat çektiği, aile fertlerine özgü bireysel özelliklere daha çok yer verdiği ve figürlerde kimin kim olduğuna dair açık çizimler ile resimde bireyselleşme yaptıkları görülmektedir [23]. Uzmanlar tarafından bireyler arası mesafe, konumlandırmalar ve ifadeler üzerinden birçok anlamlı çıkarım bir aile çiz testi aracılığı ile yapılabilmektedir.

9.2.4. Duygu ve Renk Çemberi Testi

Dört yaşından itibaren çocukların resimlerinde kullandıkları renk ve sembollerin belirli anlamlar içerdiği uzmanlar tarafından kabul görmüş bir gerçektir. Yapılmış araştırmalar veya uzman deneyimleri sonucunda bazı çalışmalarda belli renklere belirli anlamlar verilse de renklerin anlamlandırılmasında her kişiye özgü, kişisel deneyimlerin, toplumsal ve kültürel etkilerin izi olduğu görülebilmektedir [3]. Ancak bu renk ve sembollerle ilgili genel kanı ifade ettiği anlamın evrensel olduğu yönünde olmasına rağmen çocuğun resimlerinde kullandığı renkler ve semboller kendi iç benliğine aittir, kendi dünyasının bir yansımasıdır ve çocuğun kendine özgü

anlamlar içerir. Her çocuğun kullandığı veya kullanmayı tercih ettiği renk ve sembollerin anlamları hem kendi içselliğinde saklıdır hem de içinde bulunduğu zamana göre değişmektedir [22]. Şekil 9.5'de klinik görüşme sırasında çocuk tarafından oluşturulan duygu ve renk çemberi örneği gösterilmiştir.



Şekil 9.5. Örnek Duygu ve Renk Çemberi [22]

Şekil 9.5'de ayrıntıları verilen, Bir Ağaç Çiz testi, Goodenough-Harris Bir Adam Çiz testi, Bir Aile Çiz testi ve Duygu ve Renk Çemberi testi üzerine yapılan analiz ve yorumlamalar, uzman bilgi ve gözetiminin çocuk psikolojisi belirlemede çok önemli olduğunu, bu süreçte yardımcı araç ve uygulamaların sürecin daha hızlı ve etkili yürütülmesinde fayda sağladığını göstermektedir. Uzmanlar tarafından etiketlenmiş verilerden anlamlı ilişkiler çıkarabilen, kısa sürede anlamlı ve etkili geri dönüşler verebilen bir yapay zekâ sisteminin de bu alana önemli katkılar sağlayacağı aşikârdır.

9.3. ÇOCUK ÇİZİMLERİ ANALİZİNDE KULLANILAN YAPAY ZEKÂ YÖNTEMLERİ

Çocuk çizimlerinin, günümüzde her okulda bulunması zorunlu olan rehber öğretmenler tarafından da analiz edilebilmesi ve böylece sorunlu çocukların kısa sürede belirlenebilmesi çok önemlidir. Ancak rehber öğretmenlerinin çocuk çizimlerinin analizi konusunda yeterince bilgi sahibi olmaması veya kalabalık mevcut dolayısıyla söz konusu analizi yapmak için zaman bulamaması durumunda çocukların ihtiyaç duyduğu destek sağlanamamaktadır. Dolayısıyla okullardaki rehber öğretmenlere yardımcı olacak, projektif testler üzerinden yapılan çizimlerin analizi sürecini kısaltarak sıkıntıları olan çocuklara ulaşmayı kolaylaştıracak bir sisteme ihtiyaç duyulmaktadır [8].

Çocuk resimleri analizi için kurulacak sisteme öncelikle uzmanlar tarafından analiz edilmiş ve etiketlenmiş çocuk resimleri eklenerek yapay zekânın eğitilmesi ve analiz örneklerinin tanıtılması gerekmektedir. Bu aşamada, özellikle etiketleme işlemlerinde çocukların ve ilgili çevresinin mahremiyetini ifşa edebilecek her türlü olası çıkarım [24] araştırılmalı, uygun mahremiyet korumalı yöntemler [25] ile verilerin etiketlenmesi ve gerekli ise paylaşılması sağlanmalıdır. Kullanılan çizimler yardımı ile oluşturulacak olan sistemin doğruluğunun test edilmesi için de eğitim verilerine benzer özellikte yeni verilerin kullanılması gerekmektedir. Belirli bir doğruluk oranı yakalandıktan sonra sistem eğitimci ya da psikologların kullanımına sunulabilir. Sisteme sunulan verilerin kalitesi çok önemlidir. Söz konusu çocuk resimleri olduğu için bu çizimlerin dijital ortamda oluşturulmuş olması psikoloji alanındaki uzmanlarca pek uygun görülmemektedir. Çünkü çocukların çizim sırasında kalemi tutuş şekillerinden çizim sırasında uyguladıkları baskıya kadar her detayın analiz açısından büyük önemi bulunmaktadır ve dijital ortamda bu bilgilere ulaşmak mümkün olmayacaktır. Öte yandan normal bir kâğıda çocuklardan yapabildiği en iyi şekilde bir resim çizmesi istendiğinde çocuklar çizim sırasında beğenmediği kısımları silebilir, sildiği noktalara üst üste yeni çizgilere gelebilir, kâğıt hasar görebilir ve sonuç olarak görüntü netliği bozulabilir. Bu gibi durumlarda resim analiz edilmeden önce görüntüyü netleştirmek için görüntü ön işlemeye tabi tutulması gerekebilir. Her durumda çocuk resimleri analizi sırasında verilerin bir kısmının kaybolma ihtimali vardır [12]. Bu aşamada verilerdeki gürültünün giderilmesi amacı ile ön işleme teknikleri kullanılabilmektedir.

9.3.1. Bir Kişi Çiz Testi için Yapay Zekâ Destekli Çözüm Örneği

Rakhmanov ve arkadaşları tarafından, Googenough "Bir Kişi Çiz" testi üzerinde yapılan çalışmanın [12] ayrıntıları aşağıda sunulmuştur. Çalışmaya 1000'den fazla çizim aktarılmış ve 4-11 arasındaki yaş grubu için 8 farklı sınıf oluşturulmuştur. Çalışma sırasında öncelikle etik kurallara uygun olarak velilerden gerekli izinler alındıktan sonra yaşları 4 ile 11 arasında değişen okul öncesi ve ilkokul seviyesindeki çocuklardan öğretmenleri gözetiminde herhangi bir stres ortamı oluşturmadan bir adam resmi çizmesi istenmiştir. Çizilen resimlerden analize uygun olmayanlar çıkartıldıktan sonra elde kalan yaklaşık 1000 resim için doktora seviyesindeki 3 uzmanın resimleri yaş gruplarına göre etiketlemesi istenmiştir. Bu resimlerin bilgisayar üzerinde öncelikle sadece piksel değerlerini kullanarak sınıflandırılması sağlanmıştır. Veri Setini oluşturan çizimlerin %75'i eğitim verisi %25 i test verisi olarak kullanılmıştır. Çalışmayı zenginleştirmek ve en iyi yöntemi keşfetmek için bu veri setinden oluşturulan farklı veri kümeleri şu şekilde gösterilebilir;

- Hepsi 8: Tüm çizimler 4-11 aralığında yaşlara göre sınıflandırılmıştır. Resimler sınıflandırma için kırpılmış ve 120x240 piksele göre yeniden boyutlandırılmıştır.
- Ters 8: Resimler kırpılmıştır ve renkleri ters çevrilmiştir. Makine öğrenmesi işlemi sırasında hesaplama maliyetini düşürmek ve yeniden boyutlandırma sırasında önemli özellikleri kaybetmemek amacıyla resimler genişletilmiştir.
- Çift: Bu veri setinde yaş grupları birleştirerek (4-5, 6-7, 8-9, 10-11) 4 farklı ikili sınıf oluşturulmuştur.
- Çift Ters: Renkleri ters çevirip yaş grupları birleştirilerek 4 farklı ikili sınıf elde edilmiştir.

Araştırma sırasında farklı veri setleri oluşturulup bunlar üzerinde yapay zekâ / makine öğrenmesi yöntemlerinin başarısı ölçülmüş olsa da çalışmada asıl beklenti ilk veri seti üzerinde en yüksek başarıyı elde etmektir. Analiz işlemleri için öncelikle çizimlerin HOG (Histogram of Oriented Gradients) [24] öznitelikleri çıkarılmıştır ve bu değerler üzerinden, yaygın makine öğrenmesi yöntemlerinden SVM (Support Vector Machine) [25] ve ANN (Artificial Neural Network) [26] teknikleri kullanılarak sınıflandırma yapılması sağlanmıştır. HOG özniteliklerinin çıkarılması işleminden sonra ANN'ye

sunulan giriş vektör değeri oldukça büyük olduğu için ANN başlangıçta sınıflandırma için başarılı sonuçlar sunmamıştır. Ayrıca farklı sınıflara ait çocuk çizimlerinde çok fazla ortak nokta bulunması tahminlerdeki doğruluk oranlarının düşmesine sebep olmuştur.

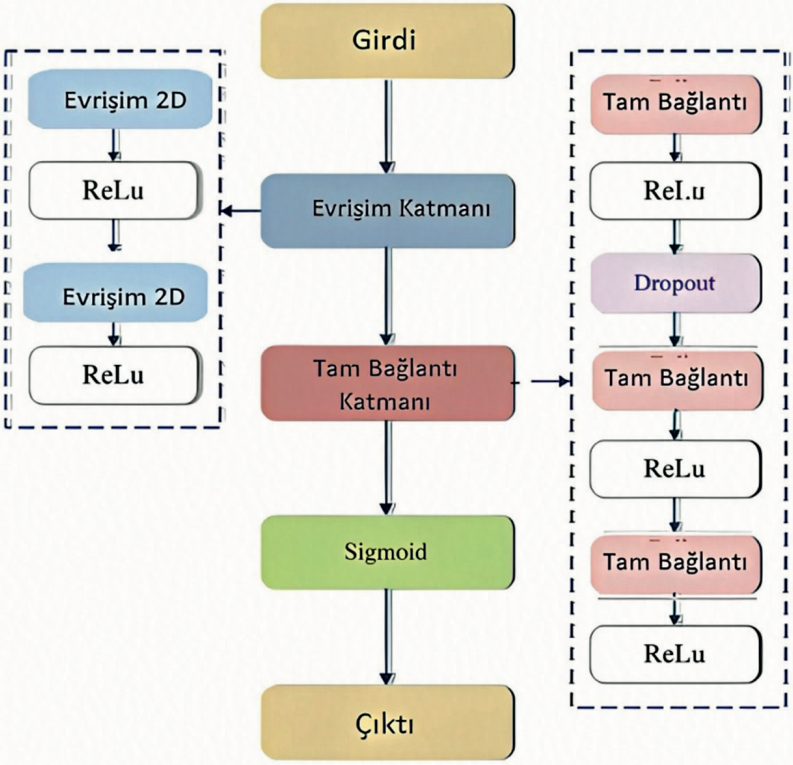
Çalışmanın ikinci kısmında, BoVW (Bag of Visual Words) [27] yöntemi kullanılmıştır. Bu yöntem ile yapılan testlerde giriş vektörlerinin boyutlarının matris biçiminde göstermeye uygun olmaması durumu ile karşılaşılmıştır. Çözüm olarak BoVW metodu iki farklı şekilde uygulanmıştır. Öncelikle manuel olarak uygulanan metotta 400 tanımlayıcı görselden oluşan kelime dağarcığı oluşturulmuş ve bunun 4 farklı veri kümesi için uygun olup olmadığı araştırılmıştır. Ardından standart biçimde BoVW yöntemi uygulandığında, standardizasyon işleminin verileri seri bir şekilde etkilemesi ve böylece sınıflandırıcıyı yanıltması sebebiyle daha düşük doğruluk oranına ulaşılmıştır. Veri seti üzerinde en yüksek başarım BoVW yöntemi kullanılarak elde edildiği tespit edilmiştir.

Çalışmanın son aşamasında CNN (Convolutional Neural Network) [28] tabanlı bir derin öğrenme modeli ile uygulama gerçekleştirilmiştir. Bu çalışmada çocukların adam çizim testi için çok etiketli bir sınıflandırma çerçevesinden oluşan bir CNN modeli denenmiştir. Yapılan uygulamalar arasında en yüksek doğruluk oranı, 8 sınıflı veri seti için %54, 2 sınıflı veri seti için %62 olarak BoVW yönteminin kullanıldığı modelden elde edilmiştir [12].

9.3.2. Bir Ağaç Çiz Testi için Yapay Zekâ Destekli Çözüm Örneği

Yuan ve ark. tarafından [2], Googenough "Bir Ağaç Çiz" testi üzerinde yapılan çalışmanın [2] ayrıntıları aşağıda sunulmuştur. Çalışmada bu testin seçilmesinin amacı ağaç çizimlerinde çocukların kendini daha iyi yansıtması ve duygularını daha iyi ifade etmesi olarak belirtilmiştir. Çocuk çizimleri yüksek soyutlama, daha az ayrıntı, daha fazla boşluk, görüntü dokusu ve renk eksikliği özelliklerine sahipken psikolojik analiz için çeşitli temel özellikleri ve ayrıntıları yakalayabilme özelliğinden dolayı CNN gibi derin öğrenme modellerinin analiz ve sınıflandırma için doğru sonuçlar üreteceği düşünülmüştür. Bu çalışmada çocukların ağaç çizim testi için çok etiketli bir sınıflandırma çerçevesinden oluşan, tam bağlantı katmanındaki Sigmoid

fonksiyonu ile birlikte sığ bir CNN modeli olan SCNN (Shallow Convolutional Neural Network) modeli önerilmiştir. Çalışmada önerilen SCNN modelinin mimarisi Şekil 9.6'da gösterilmektedir.



Şekil 9.6. Yuan ve ark. tarafından bir ağaç çiz testi için önerilen SCNN modeli [2]

Yüksek tanıma oranına sahip bir yapı oluşturabilmek için eğitim verileri çok önemlidir. Görüntüler dikdörtgen kâğıtlara çizildiği için analiz sırasında en boy oranlarının bozulmasını önlemek amacıyla öncelikle dikdörtgen resimler kare resimlere dönüştürülmüştür. Yapılan analizler sonucu hemen hemen tüm parametrelerin test seti doğruluk oranının %67'den fazlasına ulaşabileceği ve en yüksek oran %70,3'e ulaşabileceği görülmüştür. Çalışmada, deneyimsel zihinsel doğrulamaya göre çoklu etiketleme sınıflandırma doğruluğu yaklaşık %70'tir, bu da doktorların tanı kararları vermelerine yardımcı olan kabul edilebilir aralıktadır. Önerilen model aynı zamanda bilgi-

sayar destekli psikanaliz sağlar, çocukların saklayabileceği psikolojik sorunları arar, ebeveynlerin ve doktorların çocuklar için potansiyel psikolojik sorunları önlemelerine yardımcı olur.

9.4. SONUÇ VE DEĞERLENDİRMELER

Bireyler tarafından sergilenen davranışların basit ve tesadüfi olmadığı, bu davranışların analiz edilmesi gerektiği bilişsel gelişim alanında araştırmalar yapan uzmanlar tarafından belirtilmiştir. Özellikle psikolojik analizler ile doğru çıkarımlarda bulunabilmek, alanın uzmanı kişiler tarafından bile zor bir görev olarak kabul edilmektedir. Bu noktada, alanın uzmanı kişilerin işlerini kolaylaştırmak için çeşitli testler, araçlar veya uygulamalar gibi analiz süreçlerine katkı sağlayacak desteklere ihtiyaç vardır. Psikoloji ve pedagoji alanında uzman kişilerin çocukların psikolojik ve sosyolojik durumlarını analiz etmek için aktif olarak kullandıkları bir takım projektif testler mevcuttur. Belirli kısıtlar çerçevesinde çocuklara yaptırılan çizimler üzerinden gerçekleştirilen bu testler alanın uzmanı kişilere çocukların durumu ile ilgili birçok bilgi sunmaktadır. Örneğin; bir ağaç çiz testi ile uzmanlar, çocuklardaki kaygı, güvensizlik, çekingenlik, saldırganlık, baskı ve öğrenme güçlüğü hakkında bilgi sahibi olabilirken, bir adam çiz testi ile çocuklardaki cinsel istismardan motor koordinasyon becerilerine kadar birçok konuda çıkarımlarda bulunabilmektedirler.

Çocuk çizimlerinin bir psikolog veya günümüzde her okulda bulunması zorunlu olan, konu hakkında eğitimli bir rehber öğretmen tarafından analiz edilebilmesi ve bu sayede sorunlu çocukların belirlenebilmesi çok önemli bir ihtiyaçtır. Ancak, her çocuğun, coğrafi veya kültürel sebeplerle bir psikoloğa erişiminin mümkün olmaması, çocukların sorunlarını gizlemeye çalışması, rehber öğretmenlerinin çocuk çizimlerinin analizi konusunda yeterince bilgi sahibi olmaması veya kalabalık mevcut dolayısıyla söz konusu analizi yapmak için zaman bulamaması durumunda çocukların ihtiyaç duyduğu destek sağlanamamaktadır. Bu noktada, günümüz teknolojik gelişmelerinden faydalanılarak hem çocuk çizimlerinin analizi konusunda uzmanların kararlarına destek verebilecek, hem analiz sürecini kısaltabilecek hem de daha fazla çocuğa, daha kısa sürede rahatlıkla ulaşabilmeyi sağlayacak acil çözümlere ihtiyaç duyulmaktadır.

Yapay zekâ destekli çözümler günümüzde özellikle sağlık alanında hem sağlık çalışanlarına hem de hastalara büyük kolaylıklar sunmuştur. Özellikle çocuklar duygularını çizimleri ile daha rahat aktarabildiğinden dolayı, görüntü işleme çalışmalarının yüksek başarımlı derin öğrenme çalışmaları ile birlikte çocuk çizimleri analizinde uzmanların işini kolaylaştırma konusunda oldukça faydalı olacağı açıktır. Çalışmamız kapsamında yapılan araştırmalar neticesinde, çocuk çizimlerinin yapay zekâ ile analizi alanda yapılan sayılı çalışmada bile, elde edilen sonuçların uzmanların karar vermelerinde yardımcı olan kabul edilebilir aralıkta anlamlı çıkarımlarda bulunabildiği sonucuna ulaşılmıştır. Belirli problemlere özgü, beklenen çıkarımlara ek olarak, günümüz açıklanabilir yapay zekâ teknikleri ve yorumlanabilir yapay zekâ uygulamaları yardımı ile ayrıca çocukların saklayabileceği, ortaya çıkarılması daha zorlu potansiyel psikolojik sorunların tespitinde de uzmanlara yardımcı olacak ek bilgilere ulaşılması sağlanabilir.

Yapay zekâ alanında sürekli gelişen teknoloji, yöntem ve araçlar yardımı ile hemen hemen her alanda olduğu gibi çocuk psikolojisi alanında da topluma faydalı uygulama ve araçların geliştirilme potansiyeli oldukça yüksektir. Bu konuda yapılmış sayılı güncel çalışma mevcuttur. Çocuk çizimleri analizinde yapay zekâ tabanlı çalışmaların kısa sürede artması ve faydalı araçların bir an önce uygulamaya konulup topluma yarar sağlayabilmesi için, hem psikoloji / pedagoji alanındaki uzmanların hem de teknoloji geliştiricilerin bu konuda bir arada çalışmalarının teşvik edilmesi, karşılıklı ihtiyaçların belirlenerek gereksinimleri karşılayacak araştırmalara ağırlık verilmesi elzemdir.

KAYNAKLAR

[1] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects?." ACM Transactions on graphics (TOG) 31.4 (2012): 1-10.

[2] Y. Yuan, J. Huang, X. Ma, and K. Yan, “Children’s Drawing Psychological Analysis using Shallow Convolutional Neural Network,” in 2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), 2020.

[3] D. Batı, “4-12 yaş: Çocuk resimleri ve onların iç dünyalarının resimlerine yansıması,” Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü, İzmir, 2012.

[4] C. Malchiodi, “Çocukların Resimlerini Anlamak.,” Tülin Yurtbay Kitabevi, 2005.

[5] B. Ekinci, “Göçe maruz kalmış 9-11 yaş çocukların kendilerini dışa vurma açısından resimlerin yorumlanması,” Sosyal Bilimler Enstitüsü, Malatya, 2008.

[6] P. Canbay, "Sağlıkta Yapay Zekâ: Makine Öğrenmesi Yöntemleri ve Uygulamaları." İçinde: Sağlık Bilimlerinde Yapay Zekâ. Akademisyen Kitabevi (2020).

[7] P. Canbay, "Sağlıkta Yapay Zekâ: Derin Öğrenme Teknikleri ve Uygulamaları." İçinde: Sağlık Bilimlerinde Yapay Zekâ. Akademisyen Kitabevi (2020).

[8] S. Konuş, “Derin Öğrenme Yöntemleri Kullanılarak Çocukların Çizdiği Resimlerdeki Nesnelerin Tanınması,” Trakya Üniversitesi Fen Bilimleri Enstitüsü, Edirne, 2022.

[9] İ. Aydın, E. Güçlü, and E. Akın, “Demiryolu Bağlantı Elemanları için Anahtar Noktalar ve Görsel Kelime Çantası Tabanlı Kusur Tespit Yöntemi,” Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 2021, doi: 10.35234/fumbd.894102.

[10] O. Rakhmanov, "Testing strength of the state-of-art image classification methods for hand drawn sketches." 2019 15th International Conference on Electronics, Computer and Computation (ICECCO). IEEE, 2019.

[11] M. C. Gould and F. L. Goodenough, “Measurement of Intelligence by Drawings,” Am J Psychol, vol. 40, no. 4, 1928, doi: 10.2307/1414371.

[12] O. Rakhmanov, N. N. Agwu, and S. Adeshina, "Experimentation on hand drawn sketches by children to classify Draw-a-Person test images in psychology." The Thirty-Third International Flairs Conference. 2020.

[13] N. Senemoğlu, “Gelişim, öğrenme ve öğretim,” Ankara Anı Yayınevi, 2020.

[14] D. Kaya, “Temel Eğitim Kapsamındaki Çocukların Sanatsal Gelişim Basamaklarının 21. Yüzyılın Koşullarına Göre İncelenmesi,” Anadolu Üniversitesi, Eskişehir, 2022.

[15] G. Uysal, D. Sönmez Düzkaya, G. Bozkurt, and Y. Çöplü, “Hastanede yatan çocukların hemşire algısının resim çizme yöntemi ile incelenmesi: Projektif araştırma,” Mersin Üniversitesi Sağlık Bilimleri Dergisi, 2018, doi: 10.26559/mersinsbd.339045.

[16] Ç. Kağıtçıbaşı and D. Biricik, “‘Bir İnsan Çiz Testi’ Sonuçlarının Nesiller ve Cinsiyetler Arası Karşılaştırılması: 33 Yıl Sonra,” Türk Psikoloji Dergisi, vol. 28, no. 72, 2013.

[17] E. v. Brown, “Developmental Characteristics Of Figure Drawings Made By Boys And Girls Ages Five Through Eleven,” Percept Mot Skills, vol. 70, no. 1, 1990, doi: 10.2466/pms.70.1.279-288.

[18] I. D. Cherney, C. S. Seiwert, T. M. Dickey, and J. D. Flichtbeil, “Children’s drawings: A mirror to their minds,” Educational Psychology, vol. 26, no. 1. 2006. doi: 10.1080/01443410500344167.

[19] R. Lynn, “Sex differences in intelligence and brain size: A paradox resolved,” Pers Individ Dif, vol. 17, no. 2, 1994, doi: 10.1016/0191-8869(94)90030-2.

[20] S. Paksoy, "8-12 Yaş Eğitilebilir Zihinsel Engelli Çocukların Resim-İş Eğitimi," Marmara Universitesi (Turkey) ProQuest Dissertations Publishing, 2003.

[21] P. Eren, "https://egitimheryerde.net/cocuk-resimleri-analizi-bir-insan-ciz-testi/," Mayıs 03, 2020. (Erişim tarihi; Eylül 03, 2022).

[22] E. Afşaroğlu, "Çocuğun Kendi Dili İçinden Resmini Değerlendirmek-Duygu ve Renk Çemberi ile Çocuk Resimlerinin Değerlendirilmesi." Uluslararası Beşeri Bilimler ve Eğitim Dergisi 3.2, 2017.

[23] E. Akgün and A. Ergül, "55-74 aylık çocukların resimlerinde aile algısının değerlendirilmesi," Çukurova Üniversitesi Eğitim Fakültesi Dergisi, vol. 44, no. 2, 2015.

[24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005.

[25] D. A. Pisner and D. M. Schnyer, "Support vector machine," in Machine learning, Elsevier, 2020, pp. 101–121.

[26] Y. Canbay, Ş. Sağıroğlu, "Büyük Veri ve Yapay Zekâ Uygulamalarında Mahremiyet ve Etik." İçinde: Yapay Zekâ ve Büyük Veri: Teknolojiler, Yaklaşımlar ve Uygulamalar (2020).

[27] Y. Canbay, Ş. Sağıroğlu, "Derin Öğrenmede Diferansiyel Mahremiyet." Uluslararası Bilgi Güvenliği Mühendisliği Dergisi 6.1 (2020): 1-16.

[28] K. Mehrotra, C. Mohan, and S. Ranka, Elements of Artificial Neural Networks. 2019. doi: 10.7551/mitpress/2687.001.0001.

[29] X. Peng, L. Wang, X. Wang, and Y. Qiao, "Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice," Computer Vision and Image Understanding, vol. 150, 2016, doi: 10.1016/j.cviu.2016.03.013.

[30] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017.

DİZİN

A

Açıklanabilir Yapay Zekâ, 5, 43, 45, 48, 51, 76
açıklanabilirlik, 41, 50, 51, 54, 55, 56, 60, 61, 63, 64, 65, 67, 70, 71, 72, 73
Aile Çizim Testi, 246
Akıllı sözleşmeler, 168, 169
ALE, 12, 13, 21
Algoritma, 63, 81, 86, 87, 90, 91, 93, 94, 107, 144
Algoritmik Saydamlık, 50, 54, 55, 56
AlphaPortfolio, 12, 19, 22, 40
ANFIS, 62
Anormallik tespiti, 134, 138, 139, 140, 145
Arka kapı saldırıları, 234
artırılmış gerçeklik (AR), 151, 159, 161
Ayna Dünyası, 166
Ayrıştırılabilirlik, 50, 54, 55, 59

B

Bellman-Ford algoritması, 90
Benzetilebilirlik, 50, 54, 55, 56, 57, 59
BFS algoritması, 102, 103
Bir Ağaç Çiz testi, 251
Bireysel Koşullu Beklenti, 12
Birikmiş Yerel Etkiler, 13
Blok Zincir, 168
Büyük Veri, 41, 81, 106, 259

C

CIU, 12, 17, 21
CVE, 213, 237

Ç

Çekişmeli Üretici Ağlar (GAN), 70
Çocuk çizimleri, 241, 244, 245, 246, 252, 254, 256, 257
Çocuk resimleri analizi, 252

D

Derin İnanç Ağları (DBN), 70
Derin Öğrenme, 43, 45, 47, 48, 51, 60, 61, 63, 65, 70, 71, 72, 73, 74, 84, 115, 126, 258, 259
Derin Sinir Ağları (DNN), 63, 70
Derin Yürüyüş (DeepWalk), 93
DFS algoritması, 100, 103
Dijkstra algoritması, 86, 87, 90
DLIME, 12, 17, 21, 40
Doctor XAI, 12, 17, 21, 39
DT, 12, 17, 21, 23

E

En Küçük Yol Ağacı (Minimum Spanning Tree), 93
Evrişimsel Sinir Ağları (CNN), 70

F

FI, 12, 17, 21
Floyd-Warshall algoritması, 91

G

Gaussian, 61, 142, 144, 146
genişletilmiş gerçeklik (XR), 151, 161
GINN, 12, 19, 22
GIRP, 12
GPS, 102, 184, 185, 186, 188, 190, 205
Gradyan Sinir Ağı, 19
Graf algoritmaları, 84
Graf Sinir Ağları, 84
güvenli kodlama, 210, 212, 235

I

ICE, 12, 21
ILIME, 12, 16, 21
IMEI, 194
IML, 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 18, 20, 21, 24, 32, 34, 35, 36, 37

K

Kaçınma saldırıları, 233
Karar Ağaçları, 59, 60, 65
Karar Destek Sistemleri, 46, 76
karma gerçeklik (MR), 151, 161
Kısmi Bağımlılık Grafiği, 12, 17
Kod benzerliği analiz, 228
Konum tespit algoritması, 201
Konum verisi, 194
konuşma tanıma, 2, 84, 110, 153, 178
Kripto Varlıklar, 171
Kruskal algoritması, 96, 97
Kuantum çekişmeli üretici ağlar, 124
Kuantum evrişimli sinir ağı, 121
Kuantum hesaplama, 110, 111, 116
Kuantum makine öğrenimi, 113, 114
Kuantum sinir ağı, 118, 120, 121

L

Lasso, 12, 22
LIME, 11, 13, 14, 16, 21, 65, 66, 67, 74
LORE, 12, 16, 21

M

Makine Öğrenmesi, 2, 3, 5, 41, 43, 44, 45, 47, 48, 49, 50, 51, 53, 54, 55, 56, 60, 65, 70, 74, 215, 258
MAPLE, 12, 16, 21
MATS, 12, 20, 23
METAVERSE, 149, 160, 172
model, 3, 4, 6, 9, 10, 11, 12, 13, 14, 15, 18, 22, 24, 26, 27, 33, 37, 39, 40, 50, 55, 56, 59, 61, 63, 65, 67, 76, 78, 79, 82, 83, 113, 116, 117, 120, 122, 123, 124, 125, 127, 138, 143, 157, 169, 183, 191, 196, 197, 200, 203, 206, 215, 226, 233, 234, 235, 255
Model çıkarma saldırıları, 234

N

NFT, 152, 171, 172, 174, 175, 179, 183
NVD, 213

O

Otomatik Kodlayıcı, 141
Otonom sistemler, 4, 7, 20, 23

P

PDP, 12, 13, 17, 21
Prim algoritması, 94, 95
Psikolojik analiz, 244

R

Renk Çemberi Uygulaması, 246
Robot, 133, 134, 135, 136, 140, 144, 147
Robot İşletim Sistemi, 133, 134, 135, 136, 140
ROC, 18, 142

S

sağlık, 3, 4, 6, 16, 20, 25, 37, 51, 53, 73, 75, 134, 136, 150, 175, 243, 257
sanal evren, 149, 151, 152, 158, 159, 160, 161, 162, 164, 166, 167, 169, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180
sanal gerçeklik (VR), 151, 156
sanal-artırılmış gerçeklik, 149
Sanatsal Gelişim Evreleri, 245
SHAP, 1, 4, 11, 14, 16, 18, 21, 22, 24, 26, 27, 28, 29, 30, 31, 32, 67, 68, 69, 70, 74
SHAP+GIRP, 12, 18, 22
Sınıf Aktivasyon Haritalama, 71
Sollin algoritması, 99
Sondaki adam saldırıları, 216

T

Tersine mühendislik, 194, 214, 235

U

Ulusal Güvenlik Açığı Veritabanı, 213
Uzun-Kısa Dönem Hafıza (LSTM), 70

V

Veri Madenciliği, 155

W

Web 3.0, 150, 152, 168, 171, 172, 179, 182
Welsh-Powell Algoritması, 104

X

XAI, 5, 6, 38, 39, 43, 45, 48, 49, 50, 51, 52, 53, 57, 60, 70, 73, 74, 75, 77, 78, 79

Y

Yapay zekâ, 2, 6, 18, 76, 82, 89, 105, 106, 110, 145, 153, 155, 229, 232, 233, 234, 236, 243, 257
Yaygın Güvenlik Açıkları ve Etkileri listesi, 213
yorumlanabilirlik, 2, 3, 9, 10, 13, 18, 22, 24, 27, 29, 30, 31, 32, 34, 35, 36, 47, 48, 50, 55, 56, 57, 73, 74, 75

Z

Zehirleme saldırıları, 234

YAZARLAR

Prof. Şeref Sağıroğlu

Gazi Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü Öğretim Üyesidir. Gazi Üniversitesi Yapay Zekâ ve Büyük Veri Analitiği ve Güvenliği Merkez Müdürü görevini yürütmektedir. Prof. Sağıroğlu, ülkemizde bilgi güvenliği, siber güvenlik ve büyük veri analitiği, güvenliği ve mahremiyeti konularında çalışmalar yapmaktadır. 20'nin üzerinde yayınlanmış kitabı bulunmaktadır. Biri amerikan patenti olmak üzere alındığı ve müracaat aşamasında olan 10'un üzerinde patenti, 100'ün üzerinde ulusal ve uluslararası indeksli dergilerde yayınlanmış makalesi ile 300'e yakın ulusal ve uluslararası yayımlanmış bildirisi ve 9000'in üzerinde atıfı bulunmaktadır. Bilgi güvenliği alanında iki akademik derginin de baş editörlüğünü yapmaktadır. Uluslararası Bilgi Güvenliği ve Kriptoloji Konferansı (www.iscturkey.org), IEEE Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı (www. ubmk.org), IEEE Uluslararası Makine Öğrenmesi ve Uygulamaları Konferansı (www.icmla-conferences.org), Büyük Veri Analitiği, Güvenliği ve Mahremiyeti Ulusal Kamu Çalıştayı (bigdatacenter.gazi.edu.tr), Ulusal Siber Terör Konferansı (www.siberteror.org), Açık Veri Türkiye Konferansı (www.acikveriturkiye.org), Siber Güvenlik ve Savunma Çalıştayı (www.iscturkey.org), Uluslararası Adli Bilişim ve Güvenlik Sempozyumu (www.isdfs.org) gibi konferansların başkanlığını veya eşbaşkanlığını yürütmüş veya yürütmektedir. Bilgi Güvenliği Derneği (BGD), Türk Bilim Araştırma Vakfı (TÜBAV), Geleceği Önemseyenler Derneği (GÖNDER) Kurucu Üyesidir. İki dönem, BGD Yönetim Kurulu Başkanlığı ve TÜBAV Genel Başkanlığı, Gazi Üniversitesi Fen Bilimleri Enstitüsü Müdürlüğü, Kısa bir süre Gazi Üniversitesi Mühendislik Fakültesi Dekanlığı görevini gibi görevleri yürütmüştür. Gönüllü olarak pek çok sosyal projeyi de yürütmüş olan Sağıroğlu, TÜBİTAK, Cumhurbaşkanlığı DDO, Avrupa Birliği, BAP gibi Bilimsel Araştırma Projeleri tamamlamış veya yürütmeye devam etmektedir. IEEE gibi ulusal ve uluslararası konferanslarda, Bilgi Güvenliği, Büyük Veri, Siber Güvenlik ve Savunma, Akıllı Şebekeler, Yapay Zekâ, Biyometrik Uygulamalar, İnovasyon Kültürü Oluşturma gibi konularda davetli konuşmacı olarak seminer ve konferanslar vermiş ve vermeye devam etmektedir. Son dönemde ise ülkemizde yeni alanlarda Türkçe içerik oluşturma konusunda projeler yürütmekte ve özellikle de Yapay Zekâ ve Büyük Veri Kitap Serisi, Siber Güvenlik ve Savunma Kitap Serisi, Dijital Okur Yazarlık Kitap Serisi gibi projelere Editör olarak katkı vermektedir. Prof. Sağıroğlu; Gazi Üniversitesi ve Erciyes Üniversitesi Bilgisayar Mühendisliği Bölüm Başkanlığı, ISACA Ankara Chapter Akademi Koordinatörlüğü, Yükseköğretim Kurulu Siber

Güvenlik Çalışma Grubu Üyeliği, Bilim Sanayi ve Teknoloji Bakanlığı Yazılım Sektörü Çalışma Grubu Üyeliği, BGD Yönetim Kurulu Üyeliği, IPV6Forum Turkey Başkanlığı, IEEE üyeliği, IEEE Biyometrik Görev Gücü Üyeliği, ACM Üyeliği, Avrupa ETSI Standartları Gözlemci Komisyon Üyeliği, AB Projeleri Danışmanlığı, Sektör Proje Danışmanlığı, Futuretech Genel Müdürlüğü gibi görevleri yürütmektedir. Havelsan, Kişisel Verileri Koruma Kurumu, Bilgi Teknolojileri ve İletişim Kurumu, gibi kurumlarda danışmanlık ta yapmış olan Sağıroğlu, şu anda Gazi Üniversitesi Yapay Zekâ ve Büyük Veri Analitiği ve Güvenliği Merkez Müdürlüğü görevini yürütmektedir. Prof. Dr. Şeref Sağıroğlu'na dair daha detaylı bilgi ve belgelere ulaşmak için https://avesis.gazi.edu.tr/ss adresini ziyaret edebilirsiniz.

M. Umut DEMİREZEN
Dr., ROKETSAN A.Ş., Baş Mühendis (Yönetici)
T.C. Cumhurbaşkanlığı, Dijital Dönüşüm Ofisi,
Yapay Zekâ Birim Müdürü
umut@demirezen.tech, umut.demirezen@cbddo.gov.tr

Avrupa Birliği Yapay Zekâ Birliği/İttifakı üyesi olarak, Roketsan A.Ş.de Yönetici Baş Mühendis olarak Yapay Zekâ alanında araştırma ve geliştirme faaliyetleri yürütmektedir. Daha önce Savunma Teknolojileri ve Mühendislik Anonim Şirketi (STM A.Ş.) de Yapay Zekâ Teknolojileri Grup yöneticisi olarak görev yapmakta olan Dr. M. Umut Demiezen, yine STM A.Ş.de, Siber Güvenlik ve Büyük Veri ARGE Grup Yöneticiliği görevini yürütmüştür. Lisans, Yüksek Lisans ve Doktora çalışmalarını sırasıyla 2000, 2003 ve 2015 yıllarında tamamlamıştır. 2000 yılında Türk Deniz Kuvvetlerine Mühendis Teğmen olarak katılmış ve daha sonra sırasıyla Genelkurmay MEBS Başkanlığı, Genelkurmay İstihbarat Başkanlığı ve Başbakanlıkta görev yaparak, 2013 yılında STM A.Ş.de SSB adına Kıdemli Danışman olarak çalışmalar yürütmüştür.

Siber Güvenlik, Büyük Veri, Veri Bilimi ve Derin Öğrenme alanlarında birçok projede araştırmacı ve yönetici olarak görev alan Demirezen, alandan bağımsız olarak millî büyük veri ve siber güvenlik teknolojilerinin geliştirilmesi göreviyle birçok projede aktif olarak görev almış ve STM Siber Füzyon Merkezinin kurulumu ve faaliyete geçirilmesi çalışmalarında bulunmuştur. Ayrıca yapay zekâ teknolojilerinin askerî ve sivil alanlarda kullanımı üzerine aktif olarak araştırmalarını sürdürmekte olup, veri bilimi teknikleri ile uygulama ve kullanım alanlarından bağımsız olarak veri analitiği ve akıllı optimizasyon yöntemleri konularında çalışmalarına devam etmektedir.

2018 yılından itibaren ROKETSAN A.Ş.de Yapay Zekâ teknolojilerinin farklı alanlarda geliştirilmesinden sorumlu olarak görev yapmakta olup, Sürü Zekâsı (Swarm Intelligence) ve Otonom Sistemler için Yapay Zekâ alanında araştırma ve geliştirme çalışmalarını yürütmektedir. Sürü Zekâsı alanında gerçekleştirilen AR-GE çalışmalarını teknik olarak yönetmiş ve Sürü Zekâsı Tabanlı Otonom Drone Teknolojilerinin ülkemize kazandırılmasına öncülük etmiştir. Hâlen İstanbul Teknik Üniversitesi Aerospace Research Center (ARC)'de Aerospace alanında yapay zekâ teknolojilerinin uygulanması konularında araştırmalar yürütmekte olup, İTÜ Havacılık ve Uzay Mühendisliği, Gazi Üniversitesi Bilgisayar Mühendisliği bölümlerinde Yapay Zekâ, Veri Bilimi ve Büyük Veri Konularında Lisans ve Lisansüstü dersler vermektedir. Aynı zamanda T.C. Cumhurbaşkanlığı, Dijital Dönüşüm Ofisi, Yapay Zekâ ve Büyük Veri Teknolojileri Birim Müdürlüğü görevini de eş zamanlı olarak yürütmektedir.

Dr. Feyza Yıldırım Okay
Gazi Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü
feyzaokay@gazi.edu.tr

Dr. Okay, Yüksek Lisans ve Doktora derecelerini Gazi Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Ana Bilim Dalından sırasıyla 2013 ve 2019 yıllarında almıştır. Araştırma alanları arasında Nesnelerin İnterneti, Sis Hesaplama, Yazılım Tanımlı Ağlar ve Ağ Güvenliği yer almaktadır.

Dr. Mustafa Yıldırım
Türkiye Cumhuriyet Merkez Bankası Veri Yönetişimi ve İstatistik Genel Müdürlüğü
yildirim.mustafa@tcmb.gov.tr

Dr. Yıldırım, Yüksek Lisans ve Doktora derecelerini Gazi Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Ana Bilim Dalından sırasıyla 2014 ve 2021 yıllarında almıştır. Türkiye Cumhuriyet Merkez Bankası'nda uzman olarak çalışmaktadır. Güncel araştırma alanları arasında Büyük Veri Analitiği, Makine Öğrenimi ve Çizge Teorisi bulunmaktadır.

Dr. Utku Köse

Lisans derecesini Gazi Üniversitesi Bilgisayar Eğitimi Bölümü'nden (2008; fakülte birincisi olarak), Master derecesini Afyon Kocatepe Üniversitesi Bilgisayar Anabilim Dalı'ndan (2010; Yapay Zekâ uzmanlığında); Doktora derecesini de Selçuk Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı'ndan (2017; Yapay Zekâ – zeki optimizasyon konularında) almıştır. 2009-2012 yılları arasında Afyon Kocatepe Üniversitesi, 2012-2017 yılları arasında ise Uşak Üniversitesi'nde görev yapmış olan ve gerek ulusal gerekse uluslararası platformda çok sayıda akademik makale, kitap, kitap bölümü ve bildiri çalışmaları bulunan KÖSE; 2017-2019 yılları arasında Süleyman Demirel Üniversitesi'nde Dr. Öğr. Üyesi olarak görev yapmış ve halen aynı kurumda Doç. Dr. olarak görevini sürdürmektedir. 200'den fazla makale, kitap, bildiri ve rapor çalışmaları olan ve aynı zamanda çeşitli akademik dergilerin editör kurulları ile birlikte CRC Press'in Biomedical and Robotics Healthcare kitap serisinin editörleri arasında yer alan Dr. KÖSE'nin genel uzmanlık ve araştırma konuları arasında; başta Yapay Zekâ olmak üzere, Makine Etiği, Yapay Zekâ Güvenliği, Biyomedikal Uygulamalar, Optimizasyon, Kaos Teorisi, Uzaktan Eğitim, E-Öğrenme ve ilgili teknolojiler,

Bilgisayar Eğitimi ve Bilgisayar Bilimleri bulunmaktadır. Edebiyat alanında da çalışmalara imza atan Dr. KÖSE'ye ilişkin daha detaylı bilgiye http://www.utkukose.com adresindeki kişisel Web sayfası üzerinden ulaşabilirsiniz.

Arş. Gör. Elif Eryaşar
eeryasar@sinop.edu.tr, eliferyasarr@gmail.com

Arş. Gör. Elif Eryaşar, lisans eğitimini Gazi Üniversitesi Fen Fakültesi Matematik Bölümünde 2020 yılında tamamlamıştır. Gazi Üniversitesi Fen Bilimleri Matematik Anabilim dalında yüksek lisans yapmakta ve Sinop Üniversitesi Fen Edebiyat Fakültesi Matematik Bölümünde Cebir ve Sayılar Teorisi alanında Araştırma Görevlisi olarak çalışmaktadır.

Prof. Dr. Şerife Büyükköse
sbuyukkose@gazi.edu.tr, serifebuyukkose@gmail.com

Prof.Dr. Şerife Büyükköse, lisans eğitimini Selçuk Üniversitesi Fen Fakültesi Matematik Bölümünde, yüksek lisans ve doktora eğitimini Selçuk Üniversitesi Fen Bilimleri Enstitüsünde tamamladı. Halen Gazi Üniversitesi Fen Fakültesi Matematik Bölümü Cebir ve Sayılar Teorisi Anabilim dalında öğretim üyesi görevini sürdürmektedir. Kitap, makale vb yayın çalışmaları, yürüttüğü ulusal projeler ve uluslararası dergi editör kurulu üyelikleri bulunmaktadır.

Halen Gazi Üniversitesi Fen Fakültesi Matematik Bölümünde Öğretim Üyesi ve GAZİSEM Sürekli Eğitim ve Araştırma Merkezi Müdürü olarak görev yapmaktadır. Çalışma alanları graf teori, spektral graf teori, matris teori ve uygulamalarıdır.

Hande Çavşi Zaim
handecavsi43@gmail.com

Lisans ve yüksek lisans eğitimlerini Kütahya Dumlupınar Üniversitesi Bilgisayar Mühendisliği bölümünde tamamlamıştır. Yüksek lisans tezinde popüler işletim sistemleri ve web uygulamalarında penetrasyon testlerinin incelenmesi konusunu çalışmıştır. Gürallar Cam Ambalaj fabrikasında bilgi teknolojileri mühendisi ve Ericsson şirketinde yazılım geliştirme mühendisi olarak görev almıştır. Eskişehir Osmangazi Üniversitesi Bilgisayar Mühendisliği ana bilim dalı, robotik bilim dalında 100/2000 doktora öğrencisi olarak eğitimine devam etmektedir.,

Metin Yılmaz
metin.yilmaz@tei.com.tr

TEI-TUSAŞ Motor Sanayii A.Ş. Bilişim Teknolojileri Kıdemli Yazılım Geliştirme Mühendisi

2009 yılından beri yazılım alanında uzmanlık-mühendislik olarak görev yapmaktadır. Yüksek lisans öğrenimini ESOGU Bilgisayar Mühendisliğinde yapay zekâ hata tespit sistemleri üzerinde tamamlayıp doktora öğrenimine aynı bölümde robotik alanında devam etmektedir. YÖK 100/2000 doktora bursiyeri ve 2232 TÜBİTAK projelerinde görev almaktadır. 2015 yılından bu yana TEI-TUSAŞ Motor Sanayii A.Ş. Bilişim Teknolojileri Kıdemli Yazılım Geliştirme Mühendisi olarak görev yapmaktadır.

Esra N. Yolaçan
Dr. Öğr. Üyesi, yolacan@ogu.edu.tr

Eskişehir Osmangazi Üniversitesi Mühendislik Mimarlık Fakültesi Bilgisayar Mühendisliği Öğretim Üyesi olarak görev yapmaktadır. Araştırma alanları arasında, bilgi güvenliği, veri bilimi ve yapay zekâ konuları bulunmaktadır. Yüksek lisans çalışmalarını Gazi Üniversitesi Bilgisayar Mühendisliğinde tamamladıktan sonra, Northeastern Üniversitesi, Boston, MA'de Elektrik Elektronik Mühendisliği Bölümü Bilgisayar Mühendisliği alanında doktora derecesini almıştır. Doktora sürecinde, bilgi güvenliği, görüntü işleme ve makine öğrenmesi üzerine çalışmalar yapmıştır. ESOGÜ Akıllı Sistemler Uygulama Merkezi, Akıllı Sistemler Güvenlik (ISS) Laboratuvarı bünyesinde akıllı sistemlerin güvenliği, mahremiyeti ve sistem anormalliklerini tespit etme üzerine araştırma ve çalışmalarda bulunmaktadır.

Betül Ersöz
betul.ersoz@gazi.edu.tr

Betül ERSÖZ, Lisans eğitimini Marmara Üniversitesi Atatürk Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi bölümünde, Yüksek lisans eğitimini Gazi Üniversitesi Bilişim Enstitüsü, Yönetim Bilişim Sistemleri bölümünde tamamlamıştır. Doktora eğitimine Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Bilgisayar ve Öğretim Teknolojileri Eğitimi bölümünde devam etmektedir. İlgi alanları arasında web teknolojileri, eğitimde yapay zekâ, sanal ve artırılmış gerçeklik yer almaktadır.

Prof. Dr. Halil İbrahim Bülbül
bhalil@gazi.edu.tr

Prof. Dr. Halil İbrahim BÜLBÜL, Lisans eğitimini Gazi Üniversitesi, Mesleki Eğitim Fakültesi Teknoloji Eğitimi Bölümünde, Yüksek Lisans eğitimini ABD California University of Pennsylvania, Technology Education Program, Technology Education ve Doktora eğitimini Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü'nde tamamlamıştır. Gazi Üniversitesi, Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü'nde öğretim üyesi olarak görev yapmaktadır. Araştırma ve çalışma alanları arasında Bilişim sistemleri, Eğitim teknolojileri, e-öğrenme, Web tabanlı eğitim, Uzaktan eğitim, Eğitim yazılımı tasarımı, Makine öğrenmesi, Veri madenciliği, Bilgi güvenliği, Siber güvenlik, Yenilenebilir enerji sistemleri, ISO 9001 Kalite Yönetim Sistemi, Meslek standartları ve mesleki yeterlikler vb. konular bulunmaktadır. Kitap, makale vb. çeşitli yayınları ve ICMLA, ICRERA, SMART GRIDS vb. düzenlediği Ulusal ve Uluslararası düzeyde Konferanslarda düzenleme kurulu üyesi görevleri bulunmaktadır. Devlet ve özel sektör tarafından desteklenen ulusal ve uluslararası düzeyde çeşitli projeleri yürütmüştür. Tübitak, Sanayi Bakanlığı, KOSGEP, Ulusal Ajans v.b. kuruluşlarda hakem, panelist ve Komisyon Üyesi vb. görevleri yürütmektedir.

Halen Gazi Üniversitesi Gazi Eğitim Fakültesinde Öğretim Üyesi, Gazi Üniversitesi Rektör Danışmanlığı, Yüksek Öğretim Başkanlığı (YÖK) Uzaktan Eğitim Komisyonu Üyesi ve Hoca Ahmet Yesevi Üniversitesi Uzaktan Eğitim Programları (TÜRTEP) Başkanı olarak görev yapmaktadır.

Murat Akın

Gazi Üniversitesi TUSAŞ Kazan Meslek Yüksek Okulu Öğretim Görevlisi.

Lisans eğitimini Kocaeli Üniversitesi Bilgisayar Mühendisliğinde, yüksek lisans ve doktora eğitimini Gazi Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği ABD tamamlamıştır. İlgi alanları Yapay Zekâ, Büyük Veri, Siber Güvenlik ve Akıllı Ulaşım Sistemleri olmak üzere bu alanlarda çalışmaları mevcuttur.

Emrah Yılmaz

Başarsoft Bilgi Teknolojileri A.Ş. Arge Müdürü

Lisans ve yüksek lisans öğrenimlerini Çankaya Üniversitesi Bilgisayar Mühendisliği bölümünde tamamlamıştır. Doktorasına Eskişehir Teknik Üniversitesinde Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Bölümünde devam etmektedir. Öncelikli ilgi alanları, navigasyon, coğrafi bilgi sistemleri ve engelli bilgisayar etkileşimidir. İlgilendiği başlıklarda ürünleşme sürecini başarılı şekilde tamamlamış bir çok arge çalışması bulunmaktadır.

Tuncay Küçükpehlivan

Başarsoft Bilgi Teknolojileri A.Ş. Genel Müdürü

Lisans öğrenimini Orta Doğu Teknik Üniversitesi Elektrik Elektronik, yüksek lisans öğrenimini Gazi üniversitesi Elektrik Elektronik bölümünde tamamlamıştır. İlgi alanları, Coğrafi Bilgi Sistemleri, akıllı şebekeleridir. Bu başlıklarda üzere birçok ulusal ve uluslararası projede rol almıştır. Ayrıca 96 yılından bu yana misafir öğretim görevlisi olarak üniversitelerde CBS ve Kent Bilgi Sistemleri dersleri vermektedir.

Dr. Çağrı Şahin

Gazi Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü

Lisans eğitimini Selçuk Üniversitesinde tamamlayan Dr. Şahin, yüksek lisans ve doktora derecelerini University of Delaware'den almıştır. Yazılım mühendisliği alanında çalışmalar yapmaktadır.

S. Mine Okumuş

smine.okumus@gmail.com

Sıdıka Mine OKUMUŞ, Lisans eğitimini Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği bölümünde tamamlamıştır. Yüksek lisans eğitimine Kahramanmaraş Sütçü İmam Üniversitesi, Bilişim Sistemleri Anabilim Dalı'nda devam etmektedir. İlgi alanları arasında, yapay zekâ, görüntü işleme, çocuk çizimleri analizi ve bilişsel analizler yer almaktadır.

Pelin Canbay

pelincanbay@ksu.edu.tr

Pelin Canbay, Yüksek lisans ve Doktora eğitimini Hacettepe Üniversitesi Bilgisayar Mühendisliği bölümünde tamamlamıştır. Kahramanmaraş Sütçü İmam Üniversitesi, Bilgisayar Mühendisliği bölümünde Dr. Öğr. Üyesi olarak görev yapmaktadır. İlgi alanları arasında yapay zekâ, metin analizi, nefret söylemi tespiti, açıklanabilir yapay zekâ ve yapay zekâ etiği yer almaktadır.

YORUMLANABİLİR VE AÇIKLANABİLİR YAPAY ZEKÂ VE GÜNCEL KONULAR

Yapay Zekâ ve Büyük Veri Kitap Serisi; ülkemizde büyük veri ve yapay zekâ konusunda açık kaynak veri, kod, kitap, proje, doküman gibi kaynakların artırılmasına yönelik olarak başlatılmış olan bir açık kaynak kitap hazırlama projesidir. Bu kitap serisi ile üniversiteler başta olmak üzere bu alanlarda bilgi birikimi, deneyimi ve uzmanlığı olanların bir araya getirilmesi ve yerli ve millî teknolojiler geliştirilmesine katkılar sağlanması amacıyla ülke için açık kaynak içerik oluşturması ve verilerden değer üretilmesi hedeflenmektedir.

Açık kaynak yaklaşımların; ülkelerin gelişimine büyük katkı sağladığı, yapılan işlerin ve çalışmaların gerek nitelik gerekse niceliğini artırdığı, dijital dönüşümün faydaya dönüştürülmesinde çarpan etkisine sahip olduğu ve ülkelerin bilimsel gelişmelerini hızlandırdığı, özellikle de açık kaynak yaklaşımların bilimsel çalışmaların artmasına katkısının %65'lere kadar yükseldiği bilinmektedir. Dolayısıyla açık kaynak ortam ve sayılarının arttırılması, açık bilim çalışmalarının yaygınlaştırılması, veri paylaşım kültürünün oluşması, ortak çalışma ortamlarının açılması, yeni başarılı modellerin, ürünlerin ve çalışmaların yapılmasının önünü açacağı, ülkemizde de Açık Bilim Felsefe'nin yaygınlaşmasına katkı sağlayacağı değerlendirilmektedir.

Bu kitap serimizin dördüncü cildinde YORUMLANABİLİR VE AÇIKLANABİLİR YAPAY ZEKÂ VE GÜNCEL KONULAR konusuna odaklanılmış olup konu 8 farklı bölüm altında sunulmuştur. Açıklanabilir veya yorumlanabilir yapay zekânın farklı açılardan irdelendiği bu ciltte; yapılan çalışmalar, uygulamalar, projeler, teorik ve pratik çalışmalar, karşılaşılabilecek yeni riskler ve problemler ve sonuçta alınması gereken önlemler ve yapılması gerekenler farklı alanlar için verilmiştir. Yapay zekâ için açıklanabilirlik yöntemleri irdelenerek önemli algoritma ve yöntemler hakkında araştırma sonuçları paylaşılmıştır. Her bir bölüm; bu alana katkı sağlayan, bu alanda eğitim almış, tez hazırlamış, çalışmalar yapmış değerli akademisyenler tarafından birbirinden bağımsız olarak hazırlansa da konu bütünlüğü ve devamlılığının sağlanmasına mümkün olduğunca dikkat edilmiştir.

Yapay Zekâ ve Büyük Veri Kitap Serisi'nin; ülkemizde yapay zekâ ve büyük veri biliminin gelişmesine ve özellikle bu alanda daha çok Türkçe içerik üretilmesine katkılar sağlaması beklenmektedir.

Her şey, veriden değer üreten Türkiye için...

YAPAY ZEKÂ VE BÜYÜK VERİ KİTAP SERİSİ 4

Editörler: Prof. Dr. Şeref Sağıroğlu
Dr. Mustafa Umut Demirezen

nobel
AKADEMİK YAYINCILIK

www.nobelyayin.com E-KİTAP kiralamak için

ISBN 978-625-427-476-3

www.nobelyayin.com
nobelyayin | nobelkitap | nobelcocuk | nobelyasam | nobelcocukyayin

atlaskitap.com
www.nobelkitap.com