T.C.

HİTİT ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ADLİ BİLİMLER ANABİLİM DALI

SUÇ SORUŞTURMALARINDA VERİ MADENCİLİĞİ, YAPAY ZEKA, UYGULAMALARI VE GELECEĞİ

PYTHON VE OPENCV İLE GERÇEK ZAMANLI YÜZ TANIMA UYGULAMASI

Yüksek Lisans Tezi

Murat YÜN

Çorum - 2023

SUÇ SORUŞTURMALARINDA VERİ MADENCİLİĞİ, YAPAY ZEKA, UYGULAMALARI VE GELECEĞİ

PYTHON VE OPENCV İLE GERÇEK ZAMANLI YÜZ TANIMA UYGULAMASI

Murat YÜN

Lisansüstü Eğitim Enstitüsü

Adli Bilimler Anabilim Dalı

Yüksek Lisans Tezi

DANIŞMANI

Doç. Dr. Sevil ÖZKINALI

Çorum 2023

Murat YÜN tarafından hazırlanan "Suç Soruşturmalarında Veri Madenciliği, Yapay Zeka Uygulamaları ve Geleceği-Python ve OpenCV ile Gerçek Zamanlı Yüz Tanıma Uygulaması" adlı tez çalışması …./…./…. tarihinde aşağıdaki jüri üyeleri tarafından oy birliği ile Hitit Üniversitesi Lisansüstü Eğitim Enstitüsü Adli Bilimler Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Doç. Dr. Selda MERCAN ……

Doç. Dr. Sevil ÖZKINALI ……

Dr. Öğr. Üyesi Ömer Faruk AKMEŞE ……

Hitit Üniversitesi Lisansüstü Eğitim Enstitüsü Yönetim Kurulunun …/…/…tarih ve …….. sayılı kararı ile Murat YÜN'ün Adli Bilimler Anabilim Dalında Yüksek Lisans derecesi alması onanmıştır.

Lisansüstü Eğitim Enstitüsü Müdürü

* Jüri Başkanının adı yazılmalıdır.

** Tez danışmanının adı yazılmalıdır.

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını beyan ederim.

Murat YÜN

SUÇ SORUŞTURMALARINDA VERİ MADENCİLİĞİ, YAPAY ZEKA, UYGULAMALARI VE GELECEĞİ PYTHON VE OPENCV İLE GERÇEK ZAMANLI YÜZ TANIMA UYGULAMASI

Murat YÜN

ORCID: 0000-0001-9390-3169

HİTİT ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

Yüksek Lisans Tezi

NİSAN 2023

ÖZET

Tarih boyunca, sosyal hayatta, ekonomide, bilimde ve teknolojide yaşanan gelişmeler insanlık hayatını birçok alanda etkilemiş, bu etki insanlığı iyi yönde geliştirdiği gibi aksi yönde de insanlığın gelişimini sağlamış, ilk dönemlerinden bu yana basit ve çeşit olarak az olan suçun günümüze kadar birçok alana yayılmasına ve karmaşık bir hal almasına sebep olmuştur. Bunun yanında insan nüfusunun artması da eklenince suç soruşturmalarında çeşitli ihtiyaçlar ortaya çıkmıştır. Bunların başında suçluların kimliklerinin tespit edilmesi gelmektedir.

Genellikle suç soruşturmalarında kimlik tespitinde yavaş olan eski geleneksel yöntemler kullanılmakta ya da veri madenciliği ve yapay zeka uygulamaları olan biyometrik sistemler (parmak izi sistemi, retina tarama vb.) kullanılmaktadır. Bu uygulamalarda da kişinin kendine ihtiyaç duyulduğu görülmekte bu da bizim karşımıza farklı bir problem olarak çıkmaktadır.

Suç soruşturmalarında kimlik tespitinde hem hızlı hem de kişinin kendisine ihtiyaç duyulmayan ve aynı zamanda her zaman güncel kişi verileri olan bir sistem şarttır. Bu çalışma da yüz tanıma sistemi ile gerçek zamanlı örnek bir uygulamayla kişinin kendisine ihtiyaç duymadan hızlı bir şekilde kimlik tespiti yapılacağı örnek bir uygulama yapılmıştır. Bu uygulama ile bir fotoğraf, video veya bilgisayara bağlı bir kameradan kişilerin yüzlerini

algılayan ve bilgisayarımızdaki veri tabanına kaydeden, sonrasında ise bu görüntü ile veri tabanındaki yüz görüntüleri karşılaştırılarak kişi kimlik tespiti yapılmaktadır. Uygulama da programa 10 adet fotoğraf verisi öğretilerek, her bir fotoğraf için 30 adet yüz verisi alınmış ve toplamda 300 fotoğraf bulunan bir veri kaynağı elde edilmiştir. Program test edildiğinde (12) adet yüz verisinden (1) tanesini yanlış tanıyıp (1) tanesini de tanıması gerekirken tanımamış, %83,33'lük doğru tahmin ve %36 hata payıyla da başarılı sayılabilecek şekilde sonuç vermiştir. Böylece, klasik olarak yapılan kimlik tespit yöntemlerine alternatif olarak çevrimiçi çalışan, kontrolü yapılabilen, otomatik olarak kimlik tespiti yapabilen ve yüz görüntülerinden kimlik oluşturabilen kişi tanıma uygulaması geliştirilmiştir.

Anahtar Kavramlar: Suç, Veri Madenciliği, Yapay Zeka, Python, OpenCV, Yüz Tanıma, Kimlik Tespiti

Bilim Kodu: 92432

DATA MINING, ARTIFICIAL INTELLIGENCE APPLICATIONS AND ITS FUTURE IN CRIMINAL İNVESTIGATIONS REAL TIME FACE RECOGNITION APPLICATION WITH PYTHON AND OPENCV

Murat YÜN

ORCID: 0000-0001-9390-3169

HITIT UNIVERSITY

GRADUATE SCHOOL

Master of Science Thesis

APRİL 2023

ABSTRACT

Throughout history, developments in social life, economy, science and technology have affected human life in many areas, this effect has not only developed humanity in a good way, but also provided the development of humanity in the opposite direction. In addition, when the increase in the human population is added, various needs have emerged in crime investigations. The first of these is the identification of the criminals.

Generally, old traditional methods that are slow in identification are used in crime investigations or biometric systems (fingerprint system, retina scanning, etc.) with data mining and artificial intelligence applications are used. In these applications, it is seen that the person needs himself, which is a different problem for us.

In criminal investigations, a system that is both fast and does not require the person himself, and at the same time always has up-to-date personal data is essential. In this study, a sample application has been made with a real-time sample application with the face recognition system, in which the identity will be determined quickly without the need for the person himself. With this application, person identification is made by detecting the faces of people from a photo, video or a camera connected to the computer and recording them in the database on our computer, and then comparing this image with the face images in the database. In the

application, 10 photographic data were taught to the program, 30 face data were taken for each photograph, and a data source with a total of 300 photographs was obtained. When the program was tested, (1) of (12) face data was incorrectly recognized and (1) was not recognized when it should have recognized the other, and it yielded results that could be considered successful with a correct prediction of 83.33% and an error margin of 36%. Thus, as an alternative to the traditional identification methods, a person recognition application that works online, can be controlled, can automatically identify and create identity from face images has been developed.

Key Terms: Crime, Data Mining, Artificial Intelligence, Machine Learning, Face Recognition, Policing, Identification

Science Code: 92432

TEŞEKKÜR

Tez çalışmam sırasında kıymetli bilgi, birikim ve tecrübeleri ile bana yol gösterici ve destek olan değerli danışman hocam Doç. Dr. Sevil ÖZKINALI'ya, ilgisini ve önerilerini göstermekten kaçınmayan Kimya Bölümü ve Adli Bilimler Anabilim Dalı Başkanı Prof. Dr. Faruk GÖKMEŞE ve Dr. Öğretim Üyesi Ömer Faruk AKMEŞE''ye, yüksek lisans eğitimim boyunca yardım, bilgi ve tecrübeleri ile bana sürekli destek olan Adli Bilimler bölümündeki tüm hocalarıma, çalışmalarım boyunca destekleriyle beni hiçbir zaman yalnız bırakmayan sevgili eşim Begüm ÜNAL YÜN'e ve yine destekleriyle beni hiçbir zaman yalnız bırakmayan değerli kardeşlerim Melih YÜN ve Mete YÜN'e teşekkürü bir borç bilirim.

Ayrıca bu yüksek lisans çalışmasını sevgili annem ve babam anısına ithaf ediyorum.

Murat YÜN

İÇİNDEKİLER

Sayfa

ÖZET ..... **iv**

ABSTRACT ..... **vi**

TEŞEKKÜR ..... **viii**

İÇİNDEKİLER ..... **ix**

ŞEKİLLER DİZİNİ ..... **xii**

RESİMLER DİZİNİ ..... **xiv**

SİMGELER VE KISALTMALAR ..... **xvii**

GİRİŞ **1**

1. BÖLÜM

LİTERATÜR VE KAYNAK ARAŞTIRMASI

1.1 Suç Soruşturması 3

1.2. Suç ve Suç Analizi 3

1.3. Büyük Veri 6

1.4. Veri Madenciliği 7

1.5. Bilgi Keşfi ve Veri Madenciliği 9

1.6. Veri Madenciliği Yöntemleri ve Modelleri 10

1.7. Veri Madenciliği ve Yapay Zeka 12

1.8. Veri Madenciliği ve Makine Öğrenimi 14

1.9. Makine Öğrenimi Süreci 17

1.10. Öğrenme Modelleri 19

1.10.1. Denetimli Öğrenme 19

1.10.2. Denetimsiz Öğrenme 20

1.10.3. Yarı Denetimli Öğrenme 20

1.10.4. Pekiştirmeli Öğrenme 21
1.11. Derin Öğrenme 21
1.12. Yapay Sinir Ağları 23
1.13. Suç Soruşturmaları ve Güvenlik Alanlarında Veri Madenciliği, Makine Öğrenimi ve Yapay Zeka Kullanımı 27
1.13.1. Dolandırıcılık ve Para Yolsuzlukları 28
1.13.2. Telefon Verileri Kullanma 29
1.13.3. Olay Yerinde DNA gibi Bulguları 30
1.13.4. Suç Kalıplarının Çıkarılması ve Gelecekteki Suç Eğiliminin Tahmini 31
1.13.5. Suç Verilerin İnternet ve Sosyal Medyada Analizi 37
1.13.6. Biyometrik Özelliklerin Güvenlik Alanında Kullanılması 38
1.13.6.1. Yüz Tanıma Sistemi 41
1.13.6.2. İris Tanıma 41
1.13.7. Siber Güvenlik 44
1.14. Suç Soruşturmaları ve Güvenlik Alanlarında Veri Madenciliği, Makine Öğrenimi ve Yapay Zeka'nın Kullanımının Çekinceleri 46
1.15. Suç Soruşturmalarında Yapay Zeka Uygulamalarının Geleceği 47

2. BÖLÜM

MATERYAL VE YÖNTEM

2.1. Yüz Tanımada Temel Konsept 50
2.2. Python ve OpenCV ile Gerçek Zamanlı Kimlik Tespiti Projesi 51
2.3. Kullanılan Yazılım Yapısı 52
2.4. Yüz Tanıma ile Kişi Tespit Programı Veri Kümesi Ara Yüzü 53
2.5. Yüz Tanıma ile Kişi Tespit Programı Eğitim Ara Yüzü 61
2.6. Yüz Tanıma ile Kişi Tespit Programı Veri Kümesi Ara Yüzü 63

3. BÖLÜM

BULGULAR VE TARTIŞMA

3.1. Bulgular ve Tartışma 70

SONUÇ VE ÖNERİLER 76

KAYNAKLAR 77

EKLER 89

ÖZGEÇMİŞ 97

ŞEKİLLER DİZİNİ

Şekil **Sayfa**

Şekil 1.1. Suç Anatomik Görünüm **4**

Şekil 1.2. Veri, Bilgi, Bilgiyi Anlamak, Bilgelik Piramit Gösterimi **7**

Şekil 1.3. Bilgi Keşif Süreci **9**

Şekil 1.4. Veri Analitik Yöntemleri **11**

Şekil 1.5. Turing Testinin Temel İş Akışı **13**

Şekil 1.6. Örnek Makine Öğrenmesi **15**

Şekil 1.7. Makine Öğreniminin Farklı Kullanımları **16**

Şekil 1.8. Kümeleme Grafiği **16**

Şekil 1.9. Veri Türlerinin Görsel Temsili **18**

Şekil 1.10. CRISP-DM Süreci **18**

Şekil 1.11. ML Yöntemlerinin Türleri **19**

Şekil 1.12. Yangın Robot Çalışma Mekanizması **21**

Şekil 1.13. Yapay Zeka Dünyasının Ana Bileşenlerine Üst Düzey Bir Bakış **21**

Şekil 1.14. AI Gelişiminin Zaman Çizelgesi **22**

Şekil 1.15. Makine Öğrenmesi ve Derin Öğrenme Yöntemi **23**

Şekil 1.16. Doğal Nöronlar **24**

Şekil 1.17. Yapay Nöronlar **24**

Şekil 1.18. Biyolojik Nörondan Yapay Nörona **25**

Şekil 1.19. Bir YSA'nın İşleyişi **25**

Şekil 1.20. Kara Kutunun Temsili **26**

Şekil 1.21. HTS Kaydı Örneği **30**

Şekil 1.22. Tahmine Dayalı Polislik İş Süreci **33**

Şekil 1.23. Yüz Tanıma Sistemi İçin Bir Çerçeve **39**

Şekil 1.24. Gözün Önden Görünümü **41**

Şekil 1.25. İris Tanıma Aşamalarının Blok Şeması **42**

Şekil 1.26. Görüntünün Uygun Şekilde Yakalanmasını Gösteren Grafik **42**

Şekil 1.27. İris Lokalizasyonu **42**

Şekil 1.28. İris Özelliklerinin Kodlar Halinde Çıkarılması **43**

Şekil 1.29. İris Görüntüsünün Üst ve Alt Arama Bölgeleri **43**

Şekil 1.30. Siber Tehdit Türleri **45**

Şekil 2.1. Yüz Tanıma Algoritması **50**

Şekil 2.2. Yüz Tanıma Şeması **51**

RESİMLER DİZİNİ

Resim **Sayfa**

Resim 1.1. Dolandırıcılık Tespitinde Bir Programın Ekran Görüntüsü **29**

Resim 1.2. AFIS (Otomatik Parmak İzi Tanımlama Sistemi) Sorgulama Ekran Görüntüsü..... **31**

Resim 1.3. Google Haritalar Aracılığıyla Bir Web Arayüzünde Kırmızı Kutular Olarak Görüntülenen Tahminler. **34**

Resim 1.4. Memurların PredPol Kutularında Geçirdikleri Süre..... **34**

Resim 1.5. PredPol'ün Raporlama Modülü **35**

Resim 1.6. PredPol'ün Raporlama Modülü Tahmine Dayalı Polislik İş Süreci **36**

Resim 1.7. Mala Karşı Suçlar..... **36**

Resim 1.8. Çin Yüz Tanıma Kamera Görüntüsü **40**

Resim 1.9. Çin Yüz Tanıma Sistemine Sahip Akıllı Gözlük Takan Bir Polis Memuru **40**

Resim 2.1. PyCharm Kütüphaneler..... **52**

Resim 2.2. PyCharm Komut Giriş Satırı..... **53**

Resim 2.3. PyCharm Komut Giriş Satırı-1 ..... **54**

Resim 2.4. Veri Klasörü..... **54**

Resim 2.5. Orjinal Resim Brad_PİTT..... **55**

Resim 2.6. Gri Renge Çevrilmiş Resim Brad_PİTT..... **55**

Resim 2.7. Orjinal Resim Angelina_JOLİE..... **55**

Resim 2.8. Gri Renge Çevrilmiş Resim Angelina_JOLİE **56**

Resim 2.9. Orjinal Resim Anthony Edward STARK..... **56**

Resim 2.10. Gri Renge Çevrilmiş Resim Anthony Edward STARK..... **56**

Resim 2.11. Orjinal Resim Zeki ALASYA..... **57**

Resim 2.12. Gri Renge Çevrilmiş Resim Zeki ALASYA **57**

Resim 2.13. Orjinal Resim Rihanna **58**

Resim 2.14. Gr Gri Renge Çevrilmiş Resim Rihanna..... **58**

Resim 2.15. Orjinal Resim Mark SALLİNG **58**
Resim 2.16. Gri Renge Çevrilmiş Resim Mark SALLİNG **58**
Resim 2.17. Orjinal Resim Paul John Vasquez **59**
Resim 2.18. Gri Renge Çevrilmiş Resim Paul John Vasquez **59**
Resim 2.19. Orjinal Resim Ezzatolah Entezami **59**
Resim 2.20. Gri Renge Çevrilmiş Resim Ezzatolah Entezami **60**
Resim 2.21. Orjinal Resim Margot KİDDER **60**
Resim 2.22. Gri Renge Çevrilmiş Resim Margot KİDDER **60**
Resim 2.23. Orjinal Resim Stephen HAWKİNG **61**
Resim 2.24. Gri Renge Çevrilmiş Resim Stephen HAWKİNG **61**
Resim 2.25. PyCharm Komut Giriş Satırı-2 ..... **62**
Resim 2.26. PyCharm Program Çıktısı **62**
Resim 2.27. PyCharm Program Tahmin Hesaplama **63**
Resim 2.28. PyCharm Komut Giriş Satırı-3 ..... **64**
Resim 2.29. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 1 ..... **64**
Resim 2.30. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 2 ..... **65**
Resim 2.31. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 3 ..... **66**
Resim 2.32. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 4 ..... **66**
Resim 2.33. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 5 ..... **67**
Resim 2.34. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 6 ..... **67**
Resim 2.35. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 7 ..... **68**
Resim 2.36. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 6 ..... **68**
Resim 2.37. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 7 ..... **69**
Resim 3.1. Gerçek Zamanlı Yüz Tanıma Programı Çıktı-1 ..... **70**
Resim 3.2. Gerçek Zamanlı Yüz Tanıma Programı Çıktı-2 ..... **71**
Resim 3.3. Gerçek Zamanlı Yüz Tanıma Programı Çıktı-3 ..... **71**
Resim 3.4. Gerçek Zamanlı Yüz Tanıma Programı Çıktı-4 ..... **72**

Resim 3.5. Gerçek Zamanlı Yüz Tanıma Programı Çıktı-5 ..... **72**

Resim 3.6. Gerçek Zamanlı Yüz Tanıma Programı Çıktı-6 ..... **73**

Resim 3.7. Gerçek Zamanlı Yüz Tanıma Programı Çıktı-7 ..... **73**

Resim 3.8. Gerçek Zamanlı Yüz Tanıma Programı Test Veri Çıktıları **74**

SİMGELER VE KISALTMALAR

Simgeler

m	Metre
cm	Santimetre
%	Yüzde
V	Hacim, Hız, Çeşitlilik
SD	Nokta Mesafe

Kısaltmalar

KDD	Veri Tabanlarında Bilgi Keşfi
DM	Veri Madenciliği
ML	Makine Öğrenimi
AI	Yapay Zeka
ID	Kullanıcı Numarası
B. SC.	Fen Fakültesi Mezunu
PHD	Felsefe Doktoru
M. SC.	Yüksek Lisans Derecesi
CRISP-DM	Veri Madenciliği İçin Sektörler Arası Standart Süreç
DL	Derin Öğrenme
YSA	Yapay Sinir Ağları
HTS	Geçmiş Arama Trafiği
DNA	Deoksiribo Nükleik Asit
AFIS	Otomatik Parmak İzi Tanımlama Sistemi
GPU	Grafik İşlemci Birimi
SNS	Sosyal Ağ Hizmetleri
CCTV	Kapalı Devre Televizyon
NIR	Yakın Kızılötesi Yansıtma
BBC	Britanya Yayın Şirketi

NLP	Doğal Dil İşleme
UEBA	Varlık Davranışı Analitiği
UBE	Kullanıcı Davranışı Analitiği
CV	Computer Vision-Bilgisayarla Görme

GİRİŞ

Tarih boyunca, sosyal hayatta, ekonomide, bilimde ve teknolojide yaşanan gelişmeler insanlık hayatını birçok alanda etkilemiş, bu etki insanlığı iyi yönde geliştirdiği gibi aksi yönde de insanlığın gelişimini sağlamış, ilk dönemlerinden bu yana basit ve çeşit olarak az olan suçun günümüze kadar birçok alana yayılmasına ve karmaşık bir hal almasına sebep olmuştur. Suçta ki bu farklı alanlara yayılım ve karmaşıklık ile birlikte üretilen veri miktarı da günden güne artmaktadır. Teknolojinin gelişmesi bu verilerin saklanmasında kolaylık sağlasa da, önemli olan bu verilerin suç çözümünde ve suçun önlenmesinde kuruluşlara fayda sağlayan bilgiler haline dönüştürülmesidir. Amaç hep daha hızlı, daha güvenli, daha akıllı ve daha kullanışlı sistemler oluşturmaktır. Teknolojinin gelişmesiyle bundan elli yıl önce yapılmayan bazı uygulamalar günümüzde kolaylıkla yapılabilmektedir. Teknolojide ki bu hızlı gelişim suçları ve suçluları da hızla geliştirmiş, buna nüfus ve imkanların da artması eklenince suç soruşturmalarında çeşitli ihtiyaçlar ortaya çıkmıştır. Bunların başında suçluların kimliklerinin hızlı bir şekilde tespit edilerek biran önce yakalanması gelmektedir.

Genellikle suç soruşturmalarında kimlik tespitinde eski geleneksel yöntemler kullanılmakta olduğu, kişilerin kolluk tarafından daha önceki suça karışma durumlarından tanındığından kimlik tespitinin hızlı yapıldığı, kişinin suç konusunda illiyet bağı varsa yani tanıdığı kişiler tarafından suç mağduru olduysa kimlik tesptinin hızlı olduğu ya da elde bulunan kamera kayıtlarından yola çıkılarak kişinin kimliğinin tespit edilmeye çalışıldığı görülmektedir. Kişi mağdur tarafından tanınmıyorsa ya da daha önceden kolluk tarafından bilinmiyorsa kimlik tespiti zaman almakta hatta bazen kişi kimlik bilgileri tespit edilmemektedir. Günümüz teknolojisinin ne kadar geliştiği göz önünde bulundurulursa kişi kimlik tespitinde teknolojiden yararlanmamak hem zaman açısından hem de bu alanda ilerleme kaydetme açısından büyük eksiklik olacaktır.

Biyometrik sistemlerden yararlanmak bu eksiklikleri büyük ölçüde ortadan kaldırmaktadır. Parmak izi, yüz yapısı, avuç içi bilgisi, retina, iris, ses, yürüyüş vb. biyometrik özelliklerden yararlanarak kişi kimlik tespitinde büyük avantaj sağlanabilir. Biyomertik sistemlerin en büyük faydası biyometrik özelliklerin unutulmayan ve kaybedilmeyen yapılar olmasıdır. Biyometrik sistemlerin bu özelliğinden dolayı suç soruşturöalarında bu sistemlerin kullanılması bizim için hem daha net sonuçlar verecek olup aynı zamanda yararlanılan sistemin hızlı olmasından dolayı zamandan kazanmamızı sağlayacaktır.

Yüz tanıma; günümüzde neredeyse her yerde kullanılmaya başlanmıştır. Güvenlik sistemleri, giriş çıkış denetlenmesi gibi alanlar bunların başında gelmektedir. Bu sistemlerde yapay zeka tanıma ve doğrulama görevi üstlenmektedir. İnsanlar bildikleri yüzleri tanıma kabiliyetleri oldukça iyidir. Bu tanıma kabiliyetine eşit otomatik bir sistem geliştirmek oldukça zordur. Yüz tanıma uygulamarı ddiğer parmak izi, retina tanıma gibi biyometri uygulamalarıyla karşılaştırıldığında kişinin kendine ihtiyaç duymadığından bu konuda daha yararlı bir uygulamadır. Yüz tanıma sistemleri bir fotoğraf, video veya bilgisayara bağlı bir kameradan

kişilerin yüzlerini algılayan ve öncesinde oluşturulan veri tabanında ki yüz biyometrik verisi ile karşılaştırarak kişi kimlik tespitini yapan uygulamalardır.

Bu tez kapsamında suç soruşturmalarında kullanılan yapay zeka uygulamalarının neler olduğu ele alınmış, suç soruşturmalarında kullanılabilmek üzere literatürde suç soruşturmalarında kullanılmayarak yer almayan ve tarafımızdan ilk kez kullanılan görüntü işleme kütüphanesi olan OpenCV ve Python programı ile Gerçek Zamanlı Yüz Algılama ve Tanıma programıyla kimlik tespiti uygulaması yapılmıştır. Bu uygulama da öncelikle bir fotoğraf, video veya bilgisayara bağlı bir kameradan kişilerin yüzlerini algılanarak bir veri tabanı oluşturulmuştur. İkinci aşamada oluşturulan veri tabanında ki yüzlerin sistemimize eğitimi yapılmıştır. Uygulamanın son aşamasında elde ki yüz verileri sisteme kamera ile gösterilerek kişi kimlik tespiti yapılmış, elde edilen bulgular belirtilmiştir. Çalışmamızın son bölümünde tez kapsamında yapılan çalışma özetlenmiş ve değerlendirme yapılmıştır.

Tezin amacı; suç soruşturmalarında suç olayının çözülmesi ve suçluların yakalanması için kimlik tespiti çok önemlidir. Günümüzde suç soruşturmalarında kimlik tespitinde eski geleneksel yöntemlerin halen kullanılmakta olduğu görülmektedir. Kimlik tespitinde kullanılan biyometrik sistemlere bakıldığında kişinin kendine ihtiyaç duyulmaktadır. Suç soruşturmalarında kimlik tespitinde bir diğer sorun hızdır. Elde olan sistemler kişinin kendisine ihtiyaç duyulması ve geleneksek yöntemler olduğundan yavaştır. Bunlar suçluların kimliklerinin tespitinde büyük sorunlar yaratmaktır.

Suç soruşturmalarında kimlik tespitinde hem hızlı hem de kişinin kendisine ihtiyaç duyulmayan ve aynı zamanda her zaman güncel kişi verileri olan bir sistem şarttır. Yüz tanıma sistemi ile gerçek zamanlı örnek bir uygulamayla kişinin kendisine ihtiyaç duymadan hızlı bir şekilde kimlik tespiti yapılacağı örnek bir uygulamayla kanıtlanmıştır.

1. BÖLÜM

LİTERATÜR VE KAYNAK ARAŞTIRMASI

1.1. Suç Soruşturması

Araştırma, bir amaca ulaşmak için bilgi toplama sürecini ifade eder (Kabir, 2016). Örneğin, iyi bir araba satın almak istiyorsak arabayı satın almadan önce aracın özellikleri hakkında bilgi toplamamız gerekir. Aynı şekilde oluşan bir suçu çözmek veya suçun oluşmaması için gereklilikleri yapmakta bir araştırma yapmayı yani bir bilgi toplama sürecini gerektirir. Bu süreç suç soruşturmalarına uygulandığında, bir suç soruşturma: bir suçun işlenip işlenmediğini belirlemek, faili tespit etmek, faili yakalamak ve suçla ilgili kanıtların toplanması olarak karşımıza çıkar (Britannica, 2017). Yani suç soruşturması; suçların incelendiği ve suçluların yakalandığı yöntemlerdir.

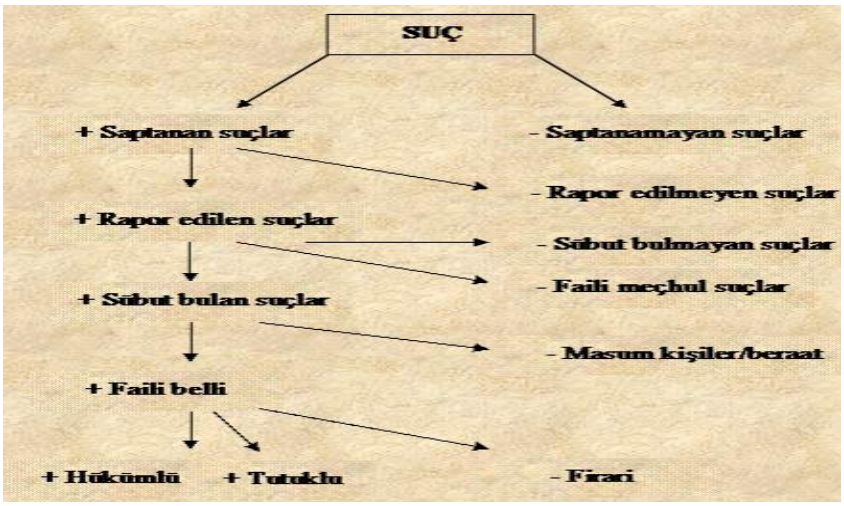
Suç soruşturmalarında ilk aşama keşiftir (Encyclopedia, 2022) Suç soruşturması bir suçun işlendiğinin keşfedilmesi, mağdurun ya da tanığın suçun oluştuğunu anlaması ve mağdurun ya da tanığın suçu bildirmesi soruşturmanın ilk aşamasını, soruşturma sonlanana kadar ki diğer tüm iş ve işlemler ise soruşturmanın ikinci aşamanı oluşur (Encyclopedia, 2022). Bu iş ve işlemler suçla ilgili tanıkları bulmak ve onlarla görüşmek, olay görüntülerini ve diğer kanıtları toplamak, suçluyu (eğer biliniyorsa ve mevcutsa) tutuklamaktır (Brandl ve Frank, 1994). Suç soruşturmasının ilk aşamasında da görüldüğü gibi soruşturmasının başlaması için mağdurun ya da tanığın suçun oluştuğunu anlaması ve yine aynı şekilde mağdurun ya da tanığın suçu bildirmesi kritik eşiktir. Burada insanların nelerin suç olduğunu ve suçun ne zaman oluştuğunu bilmesinin önemi ortaya çıkmaktadır.

İkinci önemli bir konu ise suç soruşturmasının ikinci aşamasında yapılan bilgiler(kanıtların) toplanması ve faydasının belirlenmesidir. Soruşturma aşamasında çok fazla bilgi elde edilebilir. Elde edilen bu bilgilerin konuyla ne kadar ilgili olduğu, doğruluğu, eksiksiz oluşu ve bize suçlunun kimliğini belirlemede ne kadar yardımcı olacağının değerlendirilmesi gerekir. Bazı kanıt türleri bize daha somut ve faydalı bilgiler verecektir. Görüldüğü üzere suçun önlenmesi ve çözülmesinde iki önemli detay ortaya çıkmaktadır. Bunlardan biri suçun oluştuğunun bildirilmesi, ikincisi ise kanıtların toplanması ve faydasının belirlenmesidir.

1.2. Suç ve Suç Analizi

Suç insanlarla var olan ilk zamanlardan bu yana süre gelen bir kavramdır (Benton, 1971). İlk dönemlerde suçlar basit yollarla işlenirken, teknolojinin ve medeniyetin değişim ve gelişimi insan yaşamını her anlamda değiştirmenin yanı sıra suç olgusunda da değiştirmiştir (Ward, 2011). Değişen zamanla suçlar daha karmaşık bir hal alırken, çeşit olarakta artmış, suç olarak ele alınan kavramlar değişmiş hatta ortadan kalkmıştır. Örneğin, kürtaj bir zamanlar olağandışı

durumlar dışında dünyanın bazı bölgelerinde ciddi bir suç olarak yasaklananırken, şimdi birçok ülkede yasaldır (Bernard, 2020). Bu bize suç olarak ele alınan kavramın zamanla değişebileceğini gösterdiği gibi, bir ülkede yasal olan davranışın başka bir ülkede suç teşkil edeceğini de gösterir. Suçun zamanla değişmesi, çeşidinin artması, suç olarak ele alınan kavramın zamanla ortadan kalkması ve ülkeler, toplumlar tarafından suçun farklı değerlendirilmesinden dolayı suçun standart bir tanımı yapılamaz. Bugüne kadar birçok tanım yapılarak suç kavramı tanımlanmaya çalışılsada bunlardan en çok kabul edileni "toplumun yasalarına bağlı, bir yasaklanmış eylem veya eylemler bütünü veya bir yükümlülüğün ihlalidir." tanımıdır (Ateş, 2021). Diğer tanımları; "Genel olarak suç, bir devlet tarafından korunan değerleri ihlal eden, sosyal olarak zararlı bir eylem veya ihmaldir." (Marchuk, 2014); "Suç, savunma veya mazeret olmaksızın işlenen, devlet tarafından suç veya kabahat olarak cezalandırılan, ceza hukukunu ihlal eden kasıtlı bir eylemdir." (Tappan, 1947) olarak kaşımıza çıkar. Yapılan eylemin suç teşkil etmesi için tanımdanda anlaşılacağı üzere, devlet tarafından suç veya kabahat sayılan eylemi yasal olarak kabul edilebilir savunma yada gerekçe olmadan kişinin yapmaya niyet etmiş olması gerekir. Yapılan eylemin devlet tarafından suç veya kabahat sayılması suçun "kanuni unsurunu", yasal olarak kabul edilebilir savunma yada gerekçe olmadan yapılması "hukuka aykırılık unsurunu", eylemin yapılmış olması "maddi unsuru(hareket/fiil), kişinin eylemi yapmaya niyet etmiş olması ise "manevi unsuru" oluşturur (Thotakura, 2011). Maddi unsurda fiilden kasıt, insanın kendi iradesiyle eylemi gerçekleştirmesidir. Manevi unsurda ise önemli olan kişinin eylemi kasıtlı yapmasıdır. Kasıt yoksa eylem taksire girer. Hukuka aykırılık unsurundaysa meşru savunma, kamu hukukuna dayanan bir yetkinin kullanılması gibi bir takım istisnalar mevcuttur. Devletin suç ve kabahat saydığı eylemleri insanlar bilmek zorundadır. Çünkü yapılan eylemin suç olduğunun bilinmemesi bazı durumlar hariç cezai sorumluluğu düşürmeyeceği gibi, suç olan eylemin bazı durumlar hariç bildirilmemeside bir suçtur. İstisnalar; suç olan eylem farklı ülkede suç olmaması ve kan-kayın altsoy-üstsoy hısımlığıdır. Eylemin suç olduğunun bilinmesi ve bildirilmesi suç soruşturmalarının ilk aşamasını oluşturur (Thotakura, 2011).



Şekil 1.1. Suç Anatomik Görünüm (Yücel, 2007)

Suçun anatomik görünümü Şekil 1.1.'de olduğu gibidir. Suç ile ilgili bilinmesi gereken bir diğer önemli olgu ise kanıtlardır. Kanıtların neler olduğunu bilmek ve değerlendirerek kanıtlardan faydalı bilgiler elde etmek suçun çözümünde önemli bir rol oynar. Kanıtlardan **f**iziksel kanıt, bize doğrudan suçla ilgili somut nitelikte bilgi veren kanıtlardır. Fiziksel delil; parmak izleri, kan, lifler ve suç aletleri (bıçak, tabanca, levye vb.) gibi öğeleri içerir (Sharma, 2020). Fiziksel kanıtlar, soruşturma veya yargı sürecinde en az iki önemli işleve hizmet edebilir (Peterson ve ark., 1987). İlk olarak, fiziksel kanıtlar bir suçun unsurlarının belirlenmesine yardımcı olabilir. Örneğin, bir aracın camında meydana gelen kırık (fiziksel kanıt), bir hırsızlığın meydana geldiğini belirlemeye yardımcı olabilir. İkincisi, fiziksel kanıtlar mağdurlar ile suç mahalli, suçlular ile suç mahalli, mağdurlar ile mağdurları, aletler ile suç mahalli, suçlular ile suç aletleri vb. ile ilişkilendirebilir veya bağlayabilir (Peterson ve ark., 1987). Örneğin, bir cinayet vakasında, boş bir arazide boynunda elektrik kablosu düğümlü (fiziksel kanıt) bulunan yaşlı bir kadın cesedi bulunduğunu varsayalım. Ölüm nedenin elektrik kablosuyla boğulma olduğu belirlensin. Çevrede kanıt ararken terk edilmiş bir çiftlik evi bulunsun ve arandığında kadının boynunda ki kabloya benzer bir elektrik kablosu parçası bulunsun. Bu kanıttan yola çıkılarak, gerçekte olayın meydana geldiği yerin çiftlik evinin olabileceği değerlendirilir ve ikamet incelendiğinde daha fazla delile ulaşılabilir. Bir suç soruşturmasında fiziksel delillerin yanı sıra bir diğer önemli bilgi kaynağı da kişiler, yani tanıklar ve şüphelilerdir (Encyclopedia, 2022). Tanıklar birincil veya ikincil olarak sınıflandırılabilir (Davies, 2020). Birincil tanıklar, olayı duydukları veya gözlemledikleri için suç hakkında doğrudan bilgisi olan kişilerdir. Bu sınıflandırma, suçu gözlemleyen veya başka bir şekilde suça karışan suç mağdurlarını içerecektir. Burada görgü tanıkları da yer alacak. İkincil tanıklar, suçtan önce veya sonra ilgili olaylar hakkında bilgi sahibidir. Suçu gözlemlemeyen muhbirler (veya sokak kaynakları) ve mağdurlar en iyi şekilde ikincil tanıklar olarak sınıflandırılır (Davies, 2020).

Fiziksel kanıtlar, tanıklar ve şüphelilerin yanı sıra, bir soruşturmada bize faydalı bilgiler sağlayabilecek diğer bir kaynak ise suç analizidir. Suç analizi suçların tamamen rastgele, izole olmadığı varsayımına dayanarak, suç olayları ve potansiyel suç faaliyetlerinin; nasıl, ne zaman, nerede ve kimin tarafında olduğu gibi özelliklerini inceleyerek, problem çözmeye yardımcı olmak için ortaya çıkmıştır (Ekblom, 1988). Suç analizini basitçe tanımlamak gerekirse, suç analizi, suç olaylarındaki kalıpları veya eğilimleri belirleme sürecidir (Lawrence, 2020). Suçlunun suç ile olan ilişkisini, suçun nasıl oluştuğunu, suçun oluştuğu zamanı ve yeri, farklı zamanlarda işlenen suçlar ve bunların gelecekteki ilişkisinin tahminini, aralarındaki gizli örüntünün tespitini, oranını ve konumuna göre bu tahminlerin sınıflandırılmasını, inceleyerek suçu önleme ve suçu çözmede kolluk kuvvetlerine katkı ve kolaylıklar sağlar. Analizde hangi teknik ve algoritmaların kullanıldığı ya da kullanılacağı güvenilirlik ve etkin sonuçlar açısından önemlidir.

Suç analizi suçlara, suçların rastgele ve izole olmadığı varsayımıyla yaklaşır ve suçları ortak özellikleri bakımından inceleyerek, farklı model kümelerle birleştirir (Ekblom, 1988). Birçok suçlunun suç işlerken izlediği suç dizisi ve örüntüleri birbirine benzerdir ve suçlular suç

mahallini terk ederken ipucu olarak kullanılabilecek bazı izler bırakırlar (Manonmaniam, 2017). Bu izler ile daha önceki suç kalıplarından elde edilen bilgilerle birlikte değerlendirilerek suçlu analizi edilebilir, suçun nedeni, zamanı hakkında bilgiler tahmin edilebilir. Bu bilgiler ışığında tahmin, suçların önceden tespit edilmesi veya çözülmesinde fayda sağlayabilir. Bütün yönleriyle suç olaylarını inceleyen bilim ise kriminolojidir (Brazil, 2019). Kriminoloji, suç ve suç özelliklerini belirlemek için kullanılan süreç olmakla birlikte, kriminoloji teknikleri yardımıyla suçlular ve suçun oluşma olasılığı değerlendirilebilir, suç mahallinden toplanabilecek suç izlerine, özelliklerine ve yöntemlerine dayanarak bir suçlunun gerçek özelliklerini belirlenebilir (Manonmaniam, 2017).

Günümüz şartlarında suçun fazlalığı ve çeşitliliği gibi nedenlerden dolayı suçla ilgili gerçeklere dayalı teoriler üzerinde çalışmak çok fazla verinin depolanmasını gerektirir. Bu çalışmalarda verimlilik elde etmek için ise depolanan verilirin doğru bir şekilde analiz edilerek hızlıca sonuca bağlanmasını gerektirir. Bilgisayarlı sistemlerin kullanımının artmasıyla birlikte, bilgisayar veri analistleri, kolluk kuvvetlerine suçları çözme sürecini hızlandırma konusunda yardım etmeye başlamış; 1900'ların ortalarında büyük veri kümelerinden faydalı bilgiler çıkarmak ve verinin nitelikleri arasındaki ilişkiyi bulmak için güçlü bir araç olarak veri madenciliği, ortaya çıkmıştır (Manonmaniam, 2017).

1.3. Büyük Veri

"Büyük veri" kavramının tanımlar üzerinden inceleyecek olursak;

• Büyük veri kavramı her sosyal alanda büyüme potansiyeli sunan benzeri görülmemiş bir hızla büyüyen, soyut disiplinler arası bir kavramdır (Chen ve diğerleri, 2014),

• Standart bilgiler aracılığıyla ulaşılması zor olan veri yakalama, işleme, toplama, görüntüleme ve büyük veri kümeleri üzerinde analiz yöntemleri olarak tanımlanmaktadır (Vaidhyanathan&Bulock, 2014).

Büyük veri kavramını daha fazla araştırmak ve anlamak için literatürde büyük veriyi tanımlayan 'beş V' (5V-Hacim, Hız, Çeşitlilik, Değer, Doğruluk) kavramlarına bakmak büyük veriyi daha iyi anlamamızı sağlar (Abdullah et al. 2015; Ateş ve diğerleri, 2020; Wamba ve diğerleri, 2015; Beyaz, 2012).

Hacim: Bu terim, sürekli artan miktarda veriyi tanımlamak için kullanılır. Verilerin artmasıyla birlikte bir bilgi yığını da oluşmaya başlamıştır. Böyle bir hacim artışı nitelikli bilgiye ulaşmayı daha da zorlaştırır (Ateş, 2021),

Çeşitlilik: Üretilen verilerin çoğu yapısal olmayan bilgisayarlar, sosyal ağlar, cep telefonları ve tabletler vb. gibi ortamlarda üretilir. Veri kaynaklarına gün geçtikçe yeni kaynaklar eklenmektedir. Bu veri kaynağı sayısını artırarak, verilerin farklı formatlarda

olmasına ve veri toplamadan sonra değişkenlerin ortaya çıkmamasına, analizörün işinin daha da zorlaşmasına neden olacaktır (Ateş, 2021).

Hız: Verilerin büyüme oranını ifade eder.

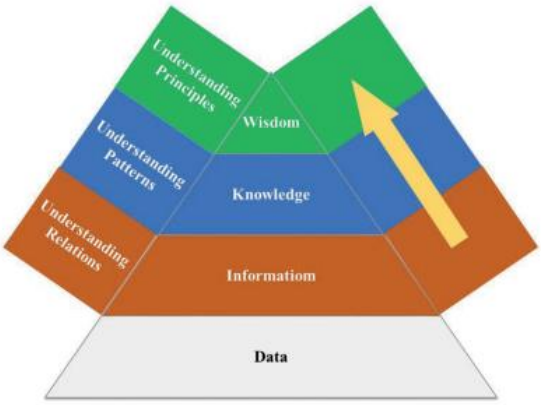
Doğruluk: Verilerin işlem sırasında güvenliğini garanti eden önemli bir değişkendir.

Üst düzey güvenlik önlemleri altında korunması önemlidir.

Değer: Karar destek birimlerinin ham verileri dönüştürerek anlamlı bilgilerle büyük verinin değerlendirilmesini ve bunun için sağladığı yardımı ifade eder (Ateş, 2021).

1.4. Veri Madenciliği

Veri madenciliği kavramını incelemek için öncelikle veri, bilgi terimi, bilgi ve bilgelik kavramı, dikkatle incelenmelidir (Ateş, 2021). Bu bağlamda, "veri" terimi, yorumlanmış veya analiz edilmiş (Zins, 2007), gizlenmiş ancak henüz ortaya çıkmamış, basit gözlemler ve sembollere dayalı, ham/organize edilmemiş gerçekler (Ahmed, 2020) ve belirli soruları yanıtlayan bir metin (Shao ve diğerleri, 2017) gibi tanımlarla ifade edilmektedir. "Bilgi" terimi, insan zihnindeki bilgi, deneyim dahil bağlamsal bilgiler, değerler, tahminler, tutarlı bir şekilde düzenlenmiş bir veri koleksiyonu, neden-nasıl sorularına cevap veren bir çıktı olarak tanımlanabilir (Ahmed, 2020; Cooper, 2017). "Bilgi" kelimesi kim, ne zaman, ne veya nerede gibi soruları yanıtlayan bir metin, bir kişinin bilgi düzeyini ve amacını, alıcısının algısını değiştiren algı mesajını ifade eder (Ateş, 2021). Bilgelik de tahmin edileceği gibi mevcut bilgiden yola çıkarak gelecek için öngörüde bulunmayı ifade eder. Şekil 1.2'de bir piramit olarak bu kavramların gösterimi verilmektedir.



Şekil 1.2. Veri, Bilgi, Bilgiyi Anlamak, Bilgelik Pramit Gösterimi (Hey, 2004; Cooper, 2017),

Veri kavramı, analiz edilmemiş genel bir kavramı; bilgi terimi, verilerin zihin tarafından süzülmesini; bilgi kavramı ise anlamlı hale gelen bir veri biçimini ifade eder (Ateş, 2021). Bilgelik kavramı, veri ve bilgi adımlarının birleşiminden meydana gelir ve bilgelik kavramıyla birlikte, işlendikçe anlam kazanan, karar vermemize yardımcı olan büyük verinin ham hali ortaya çıkar (Pauleen ve diğerleri, 2017). ‘‘Büyük veri’’ terimi, ham ve işlenmemiş, her düzeydeki verileri tanımlamak için kullanılır. Büyük verilerin işlenerek anlamlı veri elde edilmesinde istatistiksel yöntemler yetersiz kaldığından, veri madenciliği kavramı ortaya çıkmıştır.

Veri madenciliği kavramının farklı kaynaklarda birçok tanımı vardır. Bunlardan bazıları:

- Daha önceden bilinmeyen veriler hakkında potansiyel olarak yararlı bilgileri ortaya çıkarmak için bir yol (Srinivas ve diğerleri, 2010),
- İstatistiksel kavramlar, araçlar, algoritmalar, makine öğrenimi ve çok büyük veri kümelerinin analizi ile birleştirerek kavrama, anlama ve eyleme dönüştürülebilir bilginin elde edilmesini sağlayan bir yol (Williams, 2009).
- Büyük miktarda veriden bilgi edinme veya madencilik çıkarma işlemi olarak tanımlanabilecek tüm bilgi keşfi sürecindeki adımlar ve özellikle belirli bir alandaki sorunları çözmek için ihtiyaç duyulan bir bilgi keşfi şekli (Beniwal ve Arora, 2012; Olson ve Lauhoff, 2019).
- Kavram olarak, literatürde “bilgi çıkarma”, “bilgi toplama”, “bilgi keşfi”, “veri arkeolojisi”, “veri/kalıp analizi” ve “bilgi madenciliği” olarak karşımıza çıkar (Khare ve Shrivasta, 2018; Koyuncugil ve Özgülbaş, 2009).

Bu tanımlarda ki ilişki dikkate alınarak aşağıdaki tanımı yapmak mümkündür:

Veri madenciliği, istatistik, makine öğrenmesi, yapay zeka gibi birçok farklı disiplin bünyesinde gelişen yöntemleri kullanarak geçmişteki büyük veriden bilgiyi elde etmeyi ve bu bilgiyi geleceğe yönelik karar destek faaliyetlerinde kullanmayı amaçlayan bir süreçtir (Campbell ve Ying, 2011; Hand ve Adams, 2014). Teknolojinin gelişmesiyle birlikte her gün çok daha fazla veri toplanmaktadır. Parmak izleri, retina verileri, çeşitli sensör verileri, jeolojik veriler, tıbbi kayıtlar, araç ve adres bilgileri dijital veri toplamanın ve saklamanın yaygınlığını göstermektedir. Gelişime paralel olarak veri boyutlarının ve türlerinin arttığı bilinmektedir. Bilgisayar ve internet kavramlarının hayatımıza girmesi veri hacmindeki artışı hızlandırmış, verilerin daha karmaşık bir hale gelmesine neden olmuştur. Bu nedenle verilerin doğru bir şekilde analiz edilmesi her geçen gün daha da önem kazanmaktadır. Veri madenciliği, büyük veri içindeki farklı kalıpları ve ilişkileri tanımlamak ve analiz etmek için sistematik bir yaklaşımdır (Blei & Smyth, 2017). Veri madenciliği, genel olarak geçmiş örneklemden yararlanarak gelecekte ortaya çıkması ve karar verme süreçlerini desteklemesi öngörülen olaylar hakkında tahmin modelleri geliştirmeyi amaçlar (Kelleher ve Tierney, 2018). Bu

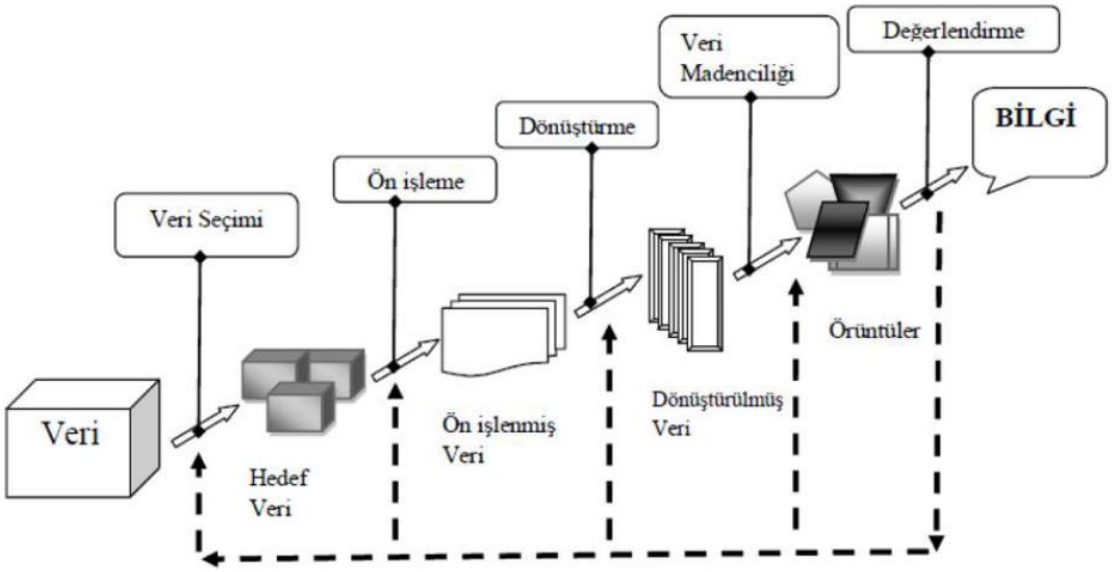
tahmin modellerinin doğruluğunun artması için mevcut verilerin analizinden ziyade çok sayıda verinin toplanması gerekir.

1.5. Bilgi Keşfi ve Veri Madenciliği

Veri madenciliği, tahmine dayalı veya tanımlayıcı bir model olarak tanımlanan, verilen verilerden faydalı model veya bilgi çıkarmak için hesaplama tekniklerinin uygulanmasından oluşan, veri tabanlarında bilgi keşfi olarak bilinen bilgi keşfi (KDD) sürecinde bir adımdır (Alfred, 2005).

Veri tabanlarında bilgi keşfi başlangıçta verilerden örtük, önceden bilinmeyen ve potansiyel olarak faydalı bilgilerin önemsiz olmayan bir şekilde çıkarılmasıdır (Frawley, 1991). 1996 yılında, KDD'nin tanımı revize edilerek şu şekilde açıklanmıştır; verilerdeki geçerli, yeni, potansiyel olarak faydalı ve nihai olarak anlaşılabilir örüntü veya bilgiyi tanımlamanın önemsiz olmayan sürecidir (Fayyad, 1996)

Aşağıdaki Şekil 1.3 veri madenciliğini yinelemeli bir bilgi keşif sürecinde bir adım şeklinde göstermektedir (Bharati, 2010).



Şekil 1.3. Bilgi keşif Süreci (Bharati, 2010)

Yinelemeli süreç adımları (Zaïane, 1999):

• Veri temizleme (gürültülü ve alakasız verilerinin çıkarılması)

• Veri entegrasyonu [(veri bütünleştirme)-bu aşamada, genellikle heterojen olan çoklu veri kaynaklarının ortak bir kaynakta birleştirilir.]

• Veri seçimi (bu adımda, analizle ilgili verilere karar verilir.)

• Veri dönüştürme (veri konsolidasyonu olarak da bilinir; seçilen veriler madencilik prosedürüne uygun formlara dönüştürülür.)

• Veri madenciliği (akıllı tekniklerin uygulandığı kritik adımdır, potansiyel olarak yararlı olan kalıplar çıkarılır.)

• Kalıp değerlendirmesi (bu adımda, ölçümlere göre elde edilmiş bilgiyi temsil eden ilginç kalıplar tanımlanır.)

• Bilgi temsili (bilgi sunumu-keşfedilen bilginin ortaya çıktığı, kullanıcıya görsel olarak sunulduğu son aşamadır).

Keşfedilen bilgi kullanıcıya sunulduktan sonra, madencilik daha da rafine edilebilir, değerlendirme yapılabilir, yeni veriler seçilebilir veya daha fazla dönüştürülebilir, farklı daha uygun sonuçlar elde etmek için yeni veri kaynakları entegre edilebilir.

1.6. Veri Madenciliği Yöntemleri ve Modelleri

Bilginin yapısı ve insanların bilgi ihtiyaçlarının artmasıyla veri madenciliği günümüzde oldukça popüler hale gelmiştir. Doğru bilgiyi elde etmek için kullanılan yöntemlerin de değişebildiği göz önüne alındığında her geçen gün yeni bir yöntem veya algoritma literatüre eklenmektedir. Bu da veri madenciliği yöntem ve algoritmalarının statik değil dinamik bir şekilde hareket ettiğini gösterir (Ateş, 2021). Doğru bilgiyi elde etmek için kullanılan yöntemlerde, veri kaynağı yapıları da verinin durumuna göre değişebilir. Veri madenciliği üç ana bölüme odaklanır ve verilerin düzenlenebilirliğine göre yapılan sınıflandırma aşağıda verilmiştir (Agrahari ve Rao, 2017; Ge vd., 2017).

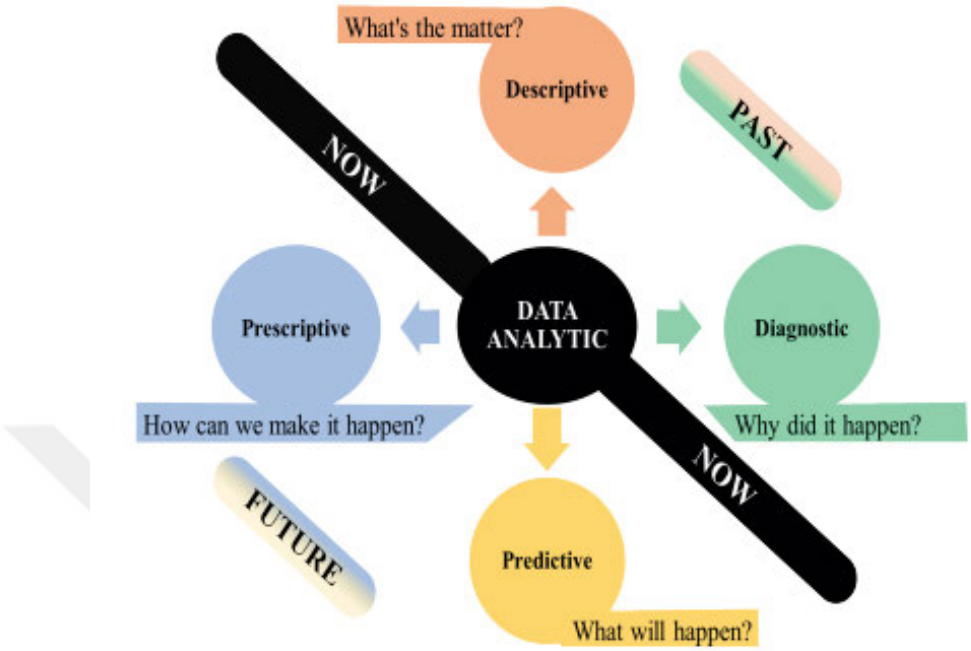
-Yapılandırılmış

- Yapılandırılmamış

- Yarı yapılandırılmış

Açıkça belirlenmiş bir veri kümesini tanımlayan yapılandırılmış veriler, sınıflandırmalarının uygunluğu nedeniyle her zaman veri analizinde önemli bir rol oynamaktadır (Kumar ve Nagpal, 2019). Güvenlik alanındaki tanımlarda kullanılan biyometrik doğrulamaların çoğu yapılandırılmış verilere dayanmaktadır (Ateş, 2021). Tüm verilerin %90'ının yapılandırılmamış veriler olarak kabul edilmesi, analiz edilmesi zor olan daha büyük bir veri kümesi gerçeğini ortaya koymaktadır (Tirgari, 2012). Bu durum özellikle her geçen gün daha sık kullanılan ve yapılandırılmamış verilerde önemli bir rol oynayan sosyal ağ siteleri için geçerlidir. Yapılandırılmamış veriler, veri kümelerinden daha fazla miktarda bilgi elde etme yeteneğini geliştirir; ancak elde edilen verilerin doğruluğu yapılandırılmış veriler kadar büyük olmayabilir (Ateş, 2021). Yarı yapılandırılmış veriler, bilgilerin kısmi olarak sınıflandırılmasına

ve tanımlanmasına yardımcı olur. Yarı yapılandırılmış verilerin en yaygın örnekleri arasında özellikle epostalar, belge oluşturma dilleri yer almaktadır.



Şekil 1.4. Veri Analitik Yöntemleri (Mesgarpour ve Dickinson, 2014)

Yapılandırılmış, yapılandırılmamış ve yarı yapılandırılmış olarak üç grupta incelenen veri madenciliği; öngörücü, tanımlayıcı, tanılayıcı ve kuralcı (Şekil 1.4) genel başlıkları altında dört model yapı içinde değerlendirilmektedir (Mesgarpour ve Dickinson, 2014);

- Tanımlayıcı Yöntemler: Veri yığınında karar vermeyi destekleyecek veriler arasındaki bağlantıları ve ilişkileri belirleyen yöntemleri içerir (Bock ve diğerleri, 2019).
- Teşhis Yöntemleri: "Neden" sorularına ilişkin nedenler ortaya koyan ve bilgi sağlayan yöntemlerdir.
- Öngörü Yöntemleri: Veri tabanındaki bağımsız değişkenleri kullanarak bağımlı değişkeni tahmin etmek için kullanılan yöntemleri içerir. Bunlar, bilinen geçmiş sonuçları kullanarak gelecekteki olayların bir tahminini yapmak için kullanılır.
- Kuralcı Yöntemler: Bu yöntemler, bilinen geçmiş sonuçları kullanarak gelecekteki olayların tahminini yapmak için kullanılır.

Dünya verilerinin büyük çoğunluğunun yapılandırılmamış olduğunu iddia edersek, en sık kullanılan yöntemlerin tanımlayıcı ve tanısal olduğunu rahatlıkla söyleyebiliriz. (Ateş.2021).

Veri madenciliği modelleri işlevlerine göre aşağıdaki dört ana başlık altında şöyledir:

Sınıflandırma: Sınıflandırmada genellikle verilerin sınıflandırılması ön plandadır ve veri yönelimlerine bu şekilde karar verilir (Mukhopadhyay vd., 2013; Rutkowski vd., 2020).

Örneğin, suç verileri üzerine bir sınıflandırma modeli, şehirdeki suç olaylarının yoğunluğuna göre şehrin güvenli mi yoksa tehlikeli mi olduğunu belirtebilir. Özellikle sınıflandırma kapsamındaki denetimli öğrenmede, verilerden örüntü keşfi yapılarak yeni eklenen konuların sınıflandırılmasını tahmin etmek mümkündür (Ateş, 2021).

Regresyon: Veri yönelimlerini tahmin ederek bir model kurabilen analiz yöntemidir ve sınıflandırmadan farklı olarak regresyonda esas olan tahmindir (Mittal vd., 2019). Örneğin, regresyon modeli, şehirdeki suçların mevcut türlerini, zamanlarını ve sıklıklarını analiz ederek suçlar için gelecekteki potansiyel verileri tahmin etmek için kullanılabilir (Ateş, 2020).

Kümeleme: Belirli bir yakınlık kriterine göre kümeler adı verilen gruplara ayrıştırma işlemidir (Berkhin, 2006; Gupta ve Chandra, 2019). Kümeleme işlemi yapıldıktan sonra genel olarak aynı kümedeki verilerin birbirine benzer olması ve farklı kümelerde benzerliğin çok daha küçük olması beklenir (Rutkowski vd., 2020). Ses ve görüntü işleme, konuşma tanıma, sık telefon görüşmeleri, mesajlaşma ve veri kullanımı, müşteri sıralama gibi birçok alanda kullanılır.

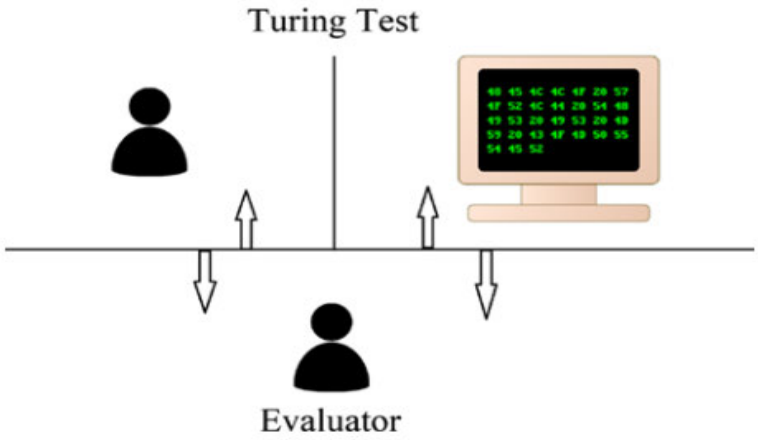
Birliktelik kuralları: İlişkili davranışları belirleyerek ve geçmişten elde edilen verilerden yararlanarak gelecekteki çalışmaları desteklemek için kullanılır (Ngai ve diğerleri, 2009; Mittal ve diğerleri, 2019). Özellikle organize ve sıklıkla işlenen suçlarda karşılaşılan tutumların ortaya çıkarılması, tespitin çok daha hızlı yapılmasını sağlayacaktır.

Veri madenciliği (DM), istatistik, makine öğrenimi, yapay zeka gibi bilimsel disiplinleri içeren disiplinler arası bir alandır (Kamber, 2001). Gerçek olayları analiz etmek için matemetik, bilgisayar bilimi, istatistik, veri tabanı teknolojileri, bilgi bilimi, makina öğrenmesi, yapay zeka ve görselleştirme, programlama gibi bir çok farklı teknik ve teori kullanır.

1.7. Veri Madenciliği ve Yapay Zeka

Yapay zeka, insan zekasını anlamak, insan zihninin problem çözme ve karar verme yeteneklerini taklit etmek için akıllı makineler yapma, kullanma teknolojisidir (Ateş, 2020). Yani yapay zeka insan zekası süreçlerinin makineler, özellikle bilgisayar sistemleri tarafından simülasyonudur diyebiliriz. İnsandan ilham alan ve insanlaştırılmış akıllı davranışı simüle etmek için bilgisayarları ve sistemleri kullanma düşüncesi ilk olarak 1950'de Alan Turing tarafından ele alınmıştır (Yu, 2011). AI'nın babası olarak da bilinen Alan Turing AI'nın ana hedefini belirleyerek; "Makineler düşünebilir mi?" sorusuna cevap aramıştır ve özünde de, AI'nın yanıtlamaya çalıştığı şey budur (Taulli, 2019). Bu soruyu cevaplamak için, "Bilgisayar Makineleri ve Zeka" (Turing, 1950) adlı makalesinde bir makinenin akıllı olup olmadığını belirlemek için kendi adıyla anılan "Turing Testi" adında bir test yapmıştır. Taklit oyunu olarak da bilinen test, temel olarak iki oyuncunun -bir insan ve bir bilgisayar- ve birde hangisinin bilgisayar hangisinin insan olduğunu tahmin etmek için onlara açık uçlu sorular soran

değerlendirici bilgisayardan oluşur (Şekil 1.5). Değerlendirici, aralarında ayrım yapamıyorsa bilgisayar testi geçecektir (Taulli, 2019).



Şekil 1.5. Turing Testinin Temel İş Akışı (Taulli, 2019)

Turing testini geçen bilgisayar için gerçekten düşünmediği, düşünme için Turing testini geçmekten çok fazlasını yapmasını gerektiğini savunan John Searle tarafından bu durum şiddetle eleştirilmiştir. Bunun sebebi Searle'nin sadece bilinçli olan varlıkların zeka sahibi olacağını düşünmesidir (Searle, 2004). Bunu kanıtlamak için Searle "Çin Odası" adlı kendi deneyini kurmuştur. Çin odası deneyinde Oda 1'e ingilizceden başka dil bilmeyen bir kişi Oda 2'ye de Çince bilmeyen bir kişiye tek başına yerleştirilir. Oda 1'de ki kişide üzerinde Çince bazı sorular yazılı bir kağıt parçası, Oda 2'de ki kişiye ise bir İngilizce kurallar kitabı ve üzerinde Çince semboller olan bir dizi kart verilir (Dore, 2012). Oda 1'deki kişi kağıt parçasını duvardaki bir delikten oda 2'ye iter ve oda 1'de kişi bir süre sonra oda 2'deki kişiden yanıtları Çince olarak geri alır. Burada oda 2'deki diğer kişinin dili anladığını varsayılır. Anlaşılması gereken oda'2 de ki kişinin yazıları anlamamasıdır. Kart dizileri arasındaki bağdaştırma kurallar kitabı yardımıyla sağlanmış, kartlardaki sembollerin tanınması, tamamen biçimleriyle sınırlı kalmıştır (Dore, 2012). Odada ki kişi sadece verilen verilerle verilen kuralları takip ederek, ne olup bittiğini anlamadan, kendisinden istenen "hiçbir şey anlamadığı semboller"i elle manipüle etmiştir ve bunu farkında olmadan yapmıştır (Taulli, 2019). Sonuç olarak, Searle'nin argümanı olan "Çin Odası" deneyi ve diğer birçok AI sistemi görevlerini çözüyor gibi görünmektedir; ancak sistem sadece programlar tarafından belirlenen bazı sabit adımları takip ederler (Taulli, 2019).

Bu testlerin asıl amacı, makinenin büyük miktarları veri işleme kapasitesine sahip olup olmadığını göstermektir (Olatz, 2020). Yapay zeka sistem dış kaynaklardan gelen verileri analiz eder, bu verilerden öğrenir ve bu verileri kullanarak çeşitli problemlerin çözümünü sağlar. Burdaki asıl amaç, insan zekasına benzer zekanın makinelerde gelişimini sağlamaktır. Bu insan beyninin nasıl çalıştığını ve öğrendiğini anlayan, insan beynini taklit etmeye çalışan eğitim verileri ve öğrenme algoritmaları ile gerçekleştirilebilir (McClendon & Meghanathan, 2015). Makine öğrenimi algoritmaları veri ile beslenir, bu verilerle kendi mantığını oluşturur

ve yönlendirme olmadan öğrenme amacıyla örnek verilere dayalı matematiksel modelleme işlemleri kullanır. Makine öğrenimi disiplinler arası bir alandır ve geleneksel programlama yöntemlerini kullanarak programlaması zor olan karmaşık problemlerin çözümlerini otomatikleştirmeye yönelik algoritmaları ve teknikleri inceler (Rebala, Ravi ve Churiwala, 2019). Bu amaçla, istatistiksel algoritmalar, büyük eğitim veri kümelerinin analizi yoluyla kendi prosedürlerini geliştirerek insan bilişsel görevlerini taklit eder. Sonuç olarak, ML terimi, açıkça programlanmak zorunda kalmadan öğrenebilen bir sistemi ima eder (Taulli, 2019). Geleneksel programlama yöntemleri birçok gerçek dünya sorununa uygulanamazken, ML yaklaşımları uygulanabilir. Geleneksel bir programlama yöntemi iki aşamadan oluşacaktır (Olatz, 2020);

Başlangıç olarak, programın spesifikasyonuna, yani yaratılış amacının ne olduğuna karar verilmeli ve ardından tasarım ortaya konmalıdır. Tasarım, sorunu çözmek için sabit bir dizi prosedür veya kural içermelidir (birinci aşama). Bunun üzerine nihai tasarım, uygulanması için bilgisayar dilinde bir program olarak yeniden modellenir (ikinci aşama) (Rebala, Ravi ve Churiwala, 2019).

Veri madenciliğinde, sinir ağları tekniği sınıflandırma amacıyla kullanılır. Kullanılan bu teknik yapay zekanın bir ürünüdür. Sinir ağları tekniği ilerleyen bölümlerde incelenecektir.

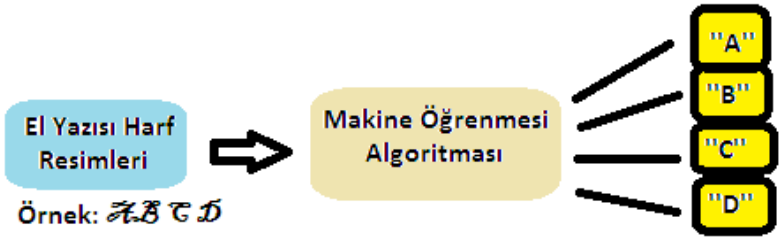
1.8. Veri Madenciliği ve Makine Öğrenimi

Makine öğrenimi kavramının hayatımıza girmesi ilk olarak Alan Turing tarafından ileri sürülen "Makineler düşünebilir mi?" sorusuyla başlamıştır (Turing, 1950). Makine öğrenmesi, bir bilgisayarın eldeki mevcut verileri ve deneyimleri kullanarak kendi kendine öğrenme yeteneğidir (Ateş, 2021). Makine öğrenmesinde eğitim verilerinin etiketlenmesine göre farklı öğrenme yöntemleri bulunmaktadır (Aggarwal, 2018). Bilgisayarın, çeşitli makine öğrenme algoritmaları yardımıyla eğitim verilerini öğrenmeye çalıştığı yönteme denetimsiz öğrenme yöntemi denilir. Denetimsiz öğrenme yönteminde bilgilerin ne kadar başarılı öğrenildiğinin tespiti test verilerinin değerlendirilmesi ile olmaktadır. Test verileri ile eğitim verileri arasındaki en önemli fark, verilerin makine tarafından tahmin edilip edilmediği veya sınıflandırılıp sınıflandırılmadığıdır (Ateş, 2021).

Denetimsiz öğrenmede eğitim verileri yoktur. Tüm öğrenme veriler arasındaki benzerlik ve farklılıklara göre yapılır. Yarı denetimli öğrenmede ise hem eğitim verileri hem de etiketlenmemiş veriler bulunur. Karma bir yapı ile gerçekleştirilen makine öğrenmesidir.

Makine öğrenmesi kavramı veri analizi için olduğu kadar veri madenciliği için de kullanılmaktadır ve ayrıca yapay zeka kullanarak örüntüleri tanımlayan, tanıyan, analizin tanımını ve yapısını belirleyen yapay zeka türüdür (Mcclendon & Meghanathan, 2015). Bilgisayarlar ve insanlar arasındaki en önemli fark, insanların deneyim dedikleri geçmişte

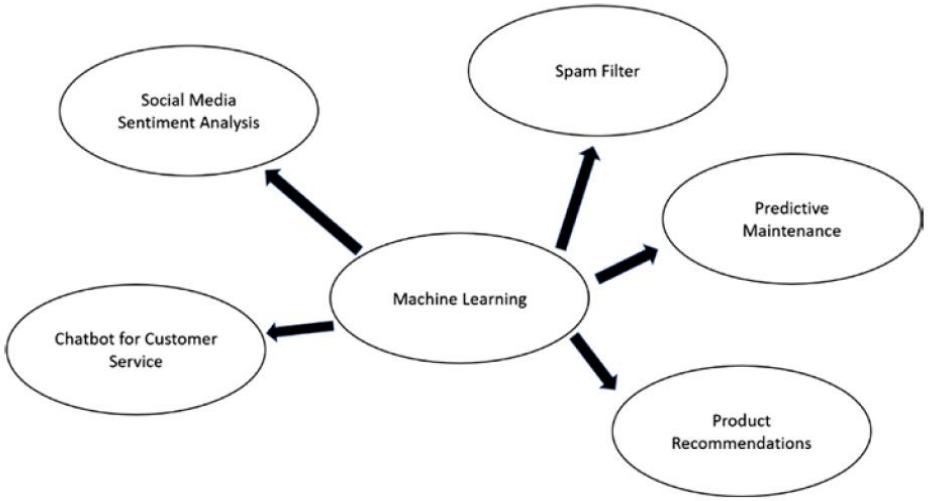
yaşadıkları olaylardan ders çıkarması, makine öğrenmesinin ise bir karar mekanizması ile bu noktaya ulaşmayı hedeflemesidir (Ateş, 2021). Makine öğrenmesinde, oluşturulan sistemin öğrenme sürecini kendi kendine gerçekleştirmesi temel ilkedir. Yapay zekanın bir alt dalı olan makine öğrenmesi ile veriler öğrenilerek tahminler yapılabilir. Bilgiye erişimde otomasyonu artıran eğitim verileri aracılığıyla öğrenmenin sağlanması, çok az insan gücüyle gerçekleştirilecek verimli otomatikleştirilmiş teknikleri mümkün kılar (Jackson, 2002). Özünde makine öğrenmesi, bir bilgisayar sisteminin örnekleme yoluyla öğrenme yöntemidir ve çeşitli problemlerde veya veri türlerinde birçok farklı makine öğrenmesi algoritması vardır (Şekil 1.6.).



Şekil 1.6. Örnek Makine Öğrenmesi

Veri madenciliği ve makine öğrenmesi kavramları temel düzeyde incelendiğinde, yapay zekanın kullanılması ile veri sistemlerinden müdahale olmadan ham verilerden bilgi elde edildiği elde ettiği sonucuna götürür (Kelleher ve Tierney, 2018). Ek olarak, veri madenciliği öncelikle veri yığınlarından anlamlı veri alımı sağlar. Ayrıca, makine öğrenimi kavramı aracılığıyla insan emeğinin en aza indirilmesine ve sonuçların optimizasyonuna olanak tanır. Bu nedenle doğa bilimleri başta olmak üzere bilimin tüm alanlarında bilgi edinme sürecinde veri madenciliği ve makine öğrenmesinin aktif olarak kullanılabileceğini söylemek mümkündür.

Makine öğrenimi, veri kümeleri içindeki yapılar ve kalıplar hakkında bilgi sağlarlar. Ve sadece bununla kalmayıp aynı zamanda sonuçlar veya davranışlarla ilgili tahminlerde bulunmak için bu yapı ve kalıplardan öğrenerek modeller oluşturur. Ayrıca ML'de kullanılan algoritmalar, açık ve ayrıntılı bir tasarımla tasarlanmadıkları için birçok zorlu problemi çözebilir (Şekil 1.7.). Yaptıkları, verilerden ayrıntılı tasarımı öğrenmektir (Rebala, Ravi ve Churiwala, 2019).



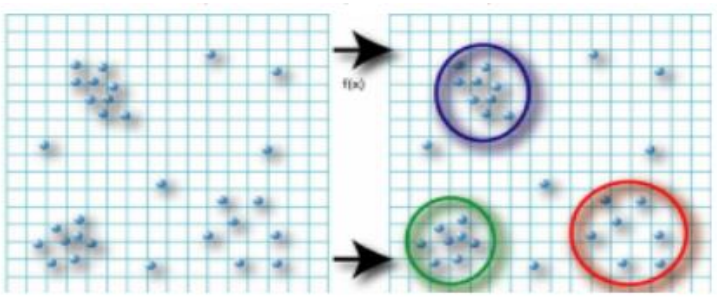
Şekil 1.7. Makine Öğreniminin Farklı Kullanımları (Taulli, 2019).

Makine öğreniminin çözdüğü birkaç sorunu geniş anlamda listelemek gerekirse (Rebala, Ravi ve Churiwala, 2019):

1) Sınıflandırma: verileri kategorilere ayırır (örn. e-postaları spam veya spam değil).

2) Tahminler: geçmiş verilere dayalı bir modele dayalı olarak gelecekteki değerleri tahmin etme (örneğin, bir suçlunun mükerrer cezaya çarptırılma olasılığının tahmin edilmesi, bir suçun işlendiği bölge tahmini)

3) Kümeleme: Verileri alarak ve öğeleri ortak özelliklerine göre kümeler halinde gruplayın (örneğin, müşteri segmentasyonu, suçun yaş dağılımı) (Şekil 1.8.).



Şekil 1.8. Kümeleme Grafiği (Öztürk, 2020).

Bu teknolojinin faydaları, maliyetleri azaltmaktan iş fırsatlarını tespit etmeye veya riskleri izlemeye kadar sayısızdır (Taulli, 2019). Ve bu ML algoritmaları genellikle insanlardan daha doğru sonuçlar verir çünkü veri işleme kapasiteleri daha büyüktür ve ön bilgi nedeniyle modelde yanlılık oluşturmazlar (Rebala, Ravi, & Churiwala, 2019).

1.9. Makine Öğrenimi Süreci

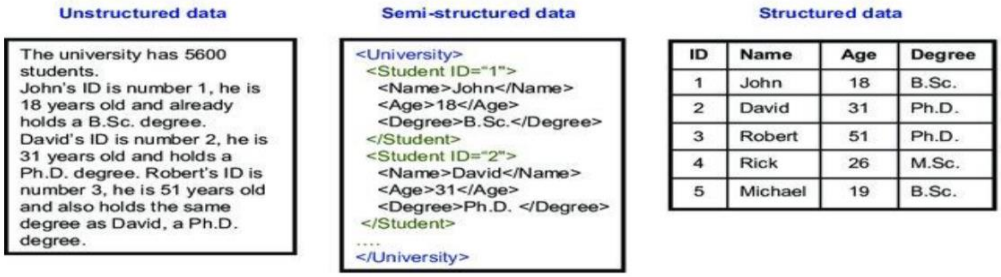
Çözülmesi gereken problem belirlendikten sonra atılması gereken ilk adım, algoritmanın hangi verilerle besleyeceğini seçmek, ikinci adım ise mevcut verilere ve çözülmesi gereken probleme bağlı olarak tahmin yoluyla ne tür bir algoritmaya karar verilmesi gerektiğidir (Ateş,2021). Üçüncü adım ise algoritmanın kalıpları bulmak için eğitim verilerinin kullanılarak eğitim verilerinin ötesinde doğru tahminlerin üretileceği bir model oluşturulacak eğitim aşamasıdır. Son olarak, son adım, parametrelerinin değerlerini ayarlayarak algoritmayı geliştirmek olacaktır (Taulli, 2019).

Bu bağlamda bir algoritma, makineye sağlanan matematiksel talimatların sabit adımları olarak anlaşılmalı; ayrıca bir model, makine algoritmayı çalıştırdığında ortaya çıkan gerçek dünyanın nasıl işlediğine dair bir hipotez olmalıdır (Egido, 2020).

Veriler algoritmalar için bir yakıtı görevi görür. Veriler olmasaydı algoritmaların başlamak için yakıtı olmazdı. Bundan dolayı veriler yapay zekanın hayati bir yönüdür. Veriler, algoritmaların sorunlara çözümler bulmak için kalıp bulmayı öğrenmesini sağlar (Egido, 2020). Veri kümesi ne kadar büyük olursa, algoritmanın doğruluğu da o kadar yüksek olur (Taulli, 2019). Amaç, ML algoritmasının verilen veri kümesi üzerinde kesin tahminler yapabilen bir model oluşturmak için verilerden desen veya kurallar kümesini öğrenmesidir (Egido, 2020). Verileri algoritmaya verilmeden önce rastgele seçmenin önemli olduğunu, aksi takdirde algoritmanın doğru bir model olmayacağı ve tespitleri ile sonuçları bozabileceği umutulmamalıdır.

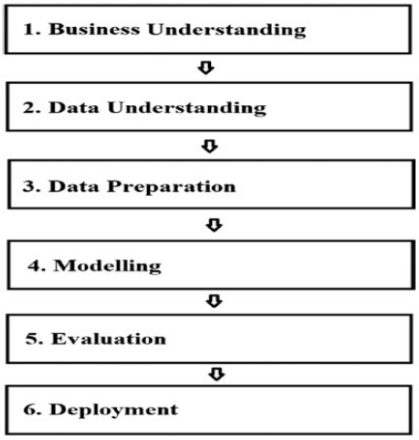
Veriler, sosyal medya, kurumsal veri tabanları ve elektronik tablolar gibi birçok farklı kaynaktan gelebilir (Taulli, 2019). Bu bilgilerin nasıl organize edildiğine bağlı olarak veriler üç ana gruba ayrılabilir; yapılandırılmış veriler, yarı yapılandırılmış veriler ve yapılandırılmamış veriler (Şekil 1.9.) (Egido, 2020):

- Yapılandırılmış veriler: Verilerin %20'sini oluşturur. Bunlar etiketlenmiş, biçimlendirilmiş ve normal olarak kaydedilmiş veri tabanları veya elektronik tablolardır. (ör. telefon numaraları, Sosyal Güvenlik numaraları, adresler, vb.)(Egido,2020).
- Yapılandırılmamış veriler: Verilerin büyük bir kısmını oluşturan biçimi olmayan verilerdir. (örn. resimler, videolar, ses dosyaları, sosyal ağ içeriği vb.). Daha sonra yapılanma gerektirir (Egido,2020).
- Yarı yapılandırılmış veriler: yapılandırılmış ve yapılandırılmamış veriler arasında, sınıflandırmaya yardımcı olan bir tür etikete sahip %5-10'luk bir oranı oluşturan verileridir (örn. JavaScript nesnesi Notasyonu veya Genişletilebilir İşaretleme Dili) (Egido,2020).



Şekil 1.9. Veri türlerinin görsel temsili (Cardoso, 2006).

Tüm verilerin nasıl işleneceği ile ilgili olarak modelde esas olarak olasılık analizi olmak üzere üst düzey istatistikler kullanılır. Bunu yapmanın benzersiz bir yolu olmamasına rağmen, 1900'lerin sonlarında oluşturulan ve yaygın olarak kabul edilen ve CRIP-DM süreci olarak adlandırılan bir yaklaşım vardır (Egido,2020). Şekil 1.10 süreci özetlemektedir.



Şekil 1.10. CRISP-DM Süreci (Taulli, 2019).

1 ve 3 aşamalarının veri sürecine ayrılan zamanın %80'ini oluşturduğu düşünüldüğünde, genel bir bilgi edinmek için sadece bunlar açıklanacaktır (Egido,2020)

1. İş Anlayışı: Bu adımda yapılacak ilk şey bir problem, bir hedef belirlemektir (Ateş, 2021). Problem ve hedef belirlendikten sonra en iyi sonuçları elde etmek için ön yargılardan uzak doğru görev gücü oluşturulmalıdır. Bunlar veri biliminde uzman kişiler ve yapay zeka projesinin belirli alanında uzman kişilerden oluşması gerekir. Son olarak, teknik ihtiyaçlar tanımlanmalıdır.

2. Veri Anlama: Bu adımda proje için veri kaynağını seçlmelidir. Bu veriler kurum içinden, mevcut açık kaynaklardan veya üçüncü şahıslardan toplanabilir. Verilerin nereden geldiğine bakılmaksızın, güvenilir olması gerekir. Verilerin eksiksiz olup olmadığını, verileri kimin değiştirdiğini, herhangi bir kalite sorunu olup olmadığını vb. kontrol etmemiz gerekir. Veriler etiketlenmişse (yapılandırılmış veriler) bu adım daha az zaman alır (Ateş, 2021).

3. Veri Hazırlama: Hangi veri setlerini kullanılacağı önemlidir. Çünkü sadece bir değişkeni seçmek veya hariç tutmak, sonuçlar üzerinde olumsuz sonuçlar doğurabilir (Egido,2020). Ardından, kalitesini artırmak için sorunlu verilerden kurtulmanız gerekir (Yineleme var mı? Alakalı mı? Ve tutarlı mı?).

Günümüzde kullanıma hazır veri hacimleri; internet erişimi, akıllı telefonlar, giyilebilir cihazlar vb. nedeniyle durmaksızın artmakta ve etkileyici büyüme oranlarına ulaştığı görülmekredir. Bu üstel sıçramayla başa çıkmak için, "Büyük Veri" olarak bilinen teknoloji oluşturulmuştur (Egido, 2020). Büyük Veri'nin ne olduğuna dair ortak bir tanım yoktur, ancak üç V olarak bilinen üç ana özellikten oluşur: hacim (verinin ölçeği, genellikle yapılandırılmamış), çeşitlilik (daha önce açıklanan verilerin çeşitliliği) ve hız (hız)(Laney, 2001). Sonuncusu, insanların yavaş verilerle hüsrana uğradığı günümüz dünyasında çok önemlidir, ancak evriminden doğruluk, değer, değişkenlik ve görselleştirme gibi yeni özellikler ortaya çıkmıştır (Taulli, 2019). Büyük Veri tarafından kullanılan tekniklerden biri, karar vermeye yardımcı olmak için kalıpları bulan ve büyük hacimli verileri sentezleyen veri madenciliğidir (Han, Kamber ve Pei, 2011).

1.10. Öğrenme Modelleri

İnsanlar geçmiş deneyimlerden öğrenirken, makineler verilerden öğrenir (Egido, 2020). Hangi tür verilerle çalışılacağına karar verildikten sonra, makinelerin ondan nasıl öğreneceğinin dört yolu vardır (Şekil 1.11.): denetimli, denetimsiz, yarı denetimli ve pekiştirmeli öğrenme (Taulli, 2019) .

Şekil 1.11. ML yöntemlerinin türleri (Taulli, 2019).

1.10.1. Denetimli öğrenme

Bu öğrenme süreci, verileri sınıflandırmak veya sonuçları doğru bir şekilde tahmin etmek için etiketlenmiş veri kümelerinin kullanılmasıyla, algoritmaların eğitmesine ve karşılaştırmaya dayanır (Rebala, Ravi, & Churiwala, 2019). Denetimli öğrenmede eğitim seti kullanılır. Bu eğitim setinin amacı modellere istenen çıktıyı verecek şekilde öğretme sağlamaktır. Bu eğitim veri seti, modelin zaman içinde öğrenmesini sağlayan girdileri ve doğru çıktıları içerir. Verileri

ve o verilerden çıkan sonuçları makineye tekrar baştan vererek bu bilgilerden bir fonksiyon (giriş verileri ile sonuç verileri arasında bir eşleşme) çıkartılmasını sağlamaktadır (Ateş, 2021). Algoritma, hata yeterince minimize edilene kadar ayar yaparak, kayıp fonksiyonu aracılığıyla doğruluğunu ölçer. Amaç veriler arasındaki ilişkiyi öğrenmektir ve bu yekilde öğrenilmiş olur. Denetimli öğrenme modelleri, görüntü ve nesne tanıma, tahmine dayalı analitik, istenmeyen posta algılama, kredi kartı sahtekarlığında kullanılabilir (Rebala, Ravi, & Churiwala, 2019).

Denetimli öğrenmede, algoritma, veriler üzerinde yinelemeli olarak tahminler yapar ve verileri uygun şekilde etiketlemek için önceden insan müdahalesi gerekir (Ateş, 2021). Yapılan tahminlerle doğru yanıtı ayarlayarak eğitim veri kümesinden "öğrenir". Denetimli öğrenmeye örnek olarak yol-zaman tahmini (makine öğrenimi) örnek verilebilir. Varılacak noktaya ne kadar sürede gidileceğimizi tahmin etmemizi sağlayacak bir makine eğitebiliriz. Ancak şartlar konusunda örneğin günün bayram olması, tatil olması, iş çıkışı olması gibi durumlarda bunu makineye öğretmemiz gerekir.

1.10.2. Denetimsiz öğrenme

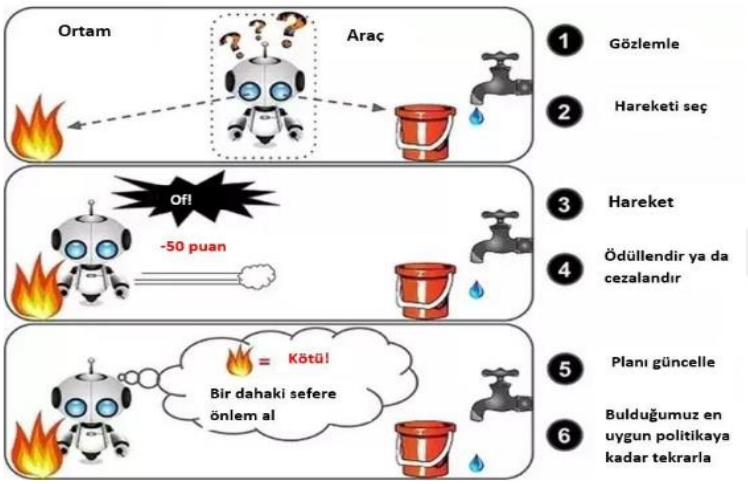
Mevcut verilerin büyük bir kısmı etiketlenmemiştir, bu öğrenme etiketlenmemiş verileri ele alacaktır. Buradaki kalıplar, “Derin Öğrenme” algoritmaları kullanılarak keşfedilir. İnsan bilgisinin çoğu öğrenme etiketleri ile değil gözlem yoluyla elde edildiğinden, bu tür öğrenme gelecekte çok önemli hale gelecektir. Keşfederek ve benimseyerek kendi başına öğrenilmiş, girdi modeline dayalı olarak adlandırılır (Ateş,2021). Bu öğrenmede ortak noktaları bulmak için veriler farklı kümelere ayrılır. Veriler yorumlanarak kümeleştirme işlemi ile anlamlı veri elde edilir. Bundan dolayı öğrenmeye kümeleme algoritması denir. Kümeleme için kullanılan genel yaklaşımdır (Taulli, 2019). Kümeleme; benzer ya da ilişkili verilerin yakınlık, uzaklık ve birbirlerine benzerliğine göre yorumlayarak sınıflara ayrılmasına denir (Ateş,2021). Bu algoritmalar insan müdahalesine ihtiyaç duymazlar. Verilerdeki gizli kalıpları keşfederler, dolayısıyla "denetimsizdirler”. Denetimli öğrenme ile arasındaki fark etiketlenmiş verilerdir (Rebala, Ravi, & Churiwala, 2019). Denetimsiz öğrenmeye örnek olarak; kitap sitesinden alınan bir kitabın yanına başka bir kitabın tavsiye olarak belirlenmesi gösterilebilir.

1.10.3.Yarı denetimli öğrenme

Etiketlenmiş verilerin çoğunluğunun ve etiketlenmemiş verilerin küçük bir bölümünün bulunduğu önceki öğrenmelerin bir kombinasyonudur (Ateş, 2021). Algoritmaların uygulanabilmesi için denetimsiz veriler derin öğrenme sistemleri kullanılarak denetimli hale dönüştürülür (Rebala, Ravi ve Churiwala, 2019). Örnek olarak ağ saldırıları tahmin edilemezdirler ve bu öğrenme için kullunılabilir.

1.10.4.Pekiştirmeli öğrenme

Takviyeli Öğrenme, bir deneme yanılma sürecinden oluşur (Ateş, 2021). Olumlu ve olumsuz pekiştirmelere dayalı olarak sonuçlar iyileştirilir (Taulli, 2019). Takviyeli öğrenme, bazı uzun vadeli ödül kavramlarında en yüksek ödül düzeye çıkarmak için hangi eylemlerin bir ortamda nasıl harekete geçmesi, yapılması gerektiğiyle ilgilenen ve buna ilişkin çıktıya dayanan bir makine öğrenmesi yaklaşımıdır (Nahid, 2021). Takviyeli öğrenme, denetimli öğrenme probleminden, doğru girdi/çıktı çiftlerinin hiçbir zaman sunulmaması veya alt-optimal eylemlerin açıkça düzeltilmemesi bakımından farklıdır (Ateş, 2021). Örnek verilecek olursa; oyun performans iyileştirmeleri için, oyunculara ödül puan ve bonus karşılığı bir oyunun belirli seviyelerini tamamlatılması ve geri bildirim sağlanması gösterilebilir (Şekil 1.12.). Takviyeli öğreme, eğitim robotlarında, kendi kendine çalışan arabalarda vb. kullanılır.



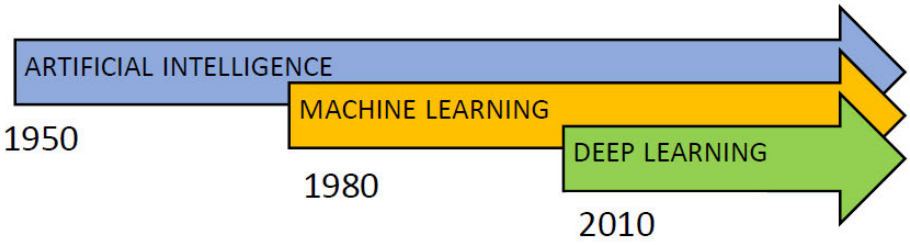
Şekil 1.12. Yangın Robot Çalışma Mekanizması (Öztürk, 2020).

1.11. Derin Öğrenme

Derin öğrenme (DL), günümüzün en popüler konularından biri olan makine öğrenmesinin alt kategorisi olup, bir veya daha fazla gizli katmana sahip yapay sinir ağlarını ve benzeri makine öğrenmesi algoritmalarını içeren bir çalışma alanıdır (Ateş, 2021). Kavramı açıklığa kavuşturmak için Şekil-9, DL'yi AI temel bileşenleriyle ilişki içinde konumlandırılmıştır (Taulli, 2019).

Şekil 1.13. Yapay zeka dünyasının ana bileşenlerine üst düzey bir bakış (Taulli, 2019).

Gördüğümüz gibi AI, Makine Öğrenimi (ML) ve aynı zamanda DL'yi içerir (Şekil 1.13.). Aynı zamanda derin öğrenme AI'nın bir alt grubu olan ML'nin bir alt kümesidir. Bu hiyerarşi, AI'nın önce, ML'nin sonra ve ML'nin bir parçası olarakda Deep Learning (DL) ortaya çıkması gerçeğinden kaynaklanmaktadır (Şekil 1.14.) (Ateş, 2021).



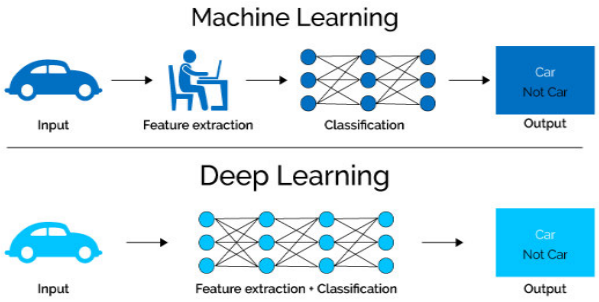
Şekil 1.14. AI gelişiminin zaman çizelgesi (Taulli, 2019).

DL, son on yılda başlayan otomasyon hedefine dayanan işleme kapasitesinde ML'nin bir evriminden oluşmuştur (LeCun, Bengio, & Hinton, 2015). DL algoritmaları, bir insan beyninin işleme sisteminin yaklaşık değerleridir (Egido, 2020). Beynimizin sınıflandırma için kullanılacak kalıpları algılamasıyla, DL algoritmalarının da benzer şekilde insan sinir ağlarını taklit etmek için aynı şeyi yaptığı düşünülmektedir. Önce bilgi alınır (giriş katmanı) ve daha sonra bilinen bir öğeyle (gizli katman) karşılaştırılır (varsa anlamlandırılır), çıktı daha sonra gelir (çıktı katmanı) (Şekil 1.15.) (Egido, 2020).

Derin öğrenme, karmaşık yapıları çok katmanlı sinir ağları ile öğrenebilen, klasik makine öğrenmesi algoritmalarının kolayca öğrenemediği modelleri gizli katmanlar aracılığıyla daha kolay ve başarılı bir şekilde öğrenebilen yapılardan oluşmaktadır (Dey, 2016). Gizli katmanlar, istenen çıktıyı elde etmek için matematiksel fonksiyonların ve hesaplamaların yapıldığı bir yapının parçasıdır; katmanlarının sayısı değişebilir (Ateş, 2021).

Derin öğrenme yeni bir yaklaşım olmasına rağmen anlambilim, transfer öğrenme, doğal dil işleme, görselleştirme, suç soruşturması gibi birçok alanda aktif olarak kullanılmaktadır (Dong, 2021). Derin öğrenmede eğitim sırasında klasik algoritmalara göre çok daha fazla veri gerektirir (Robarts, 2018). Teknolojinin gelişmesiyle oluşan veri bolluğu nedeniyle bu günümüzde bir sorun teşkil etmemektedir. Ancak analiz edilecek veri miktarı az ise klasik algoritmalar birçok yönden daha başarılı sonuçlar üretebilir. Derin öğrenmede katman sayısı

arttıkça öğrenme süreci için gerekli donanım (öncelikle grafik işlem birimi (GPU) gereksinimi) ve sürenin (parametre sayısının fazla olması nedeniyle) artacağı bilinmelidir (Robarts, 2018).



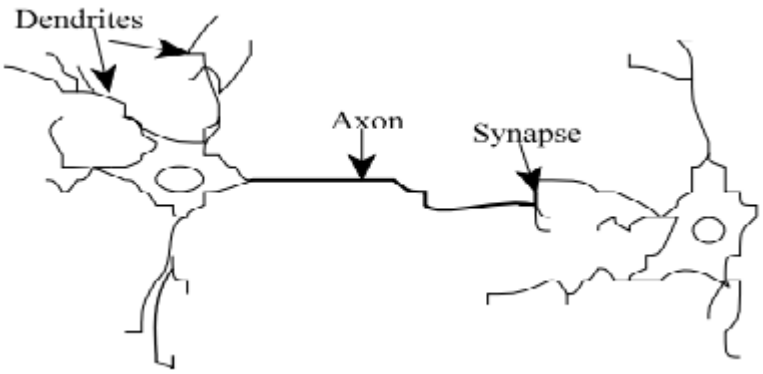
Şekil 1.15. Makine Öğrenmesi ve Derin Öğrenme Yöntem (Gulal, 2018).

Klasik makine öğrenmesi algoritmaları, örüntüyü tanımak için istatistiksel analize dayalı yöntemleri kullanırken, derin öğrenmede insan beyninin nöronlarına benzer bir modelleme vardır. Bu nedenle kayıp fonksiyonunu en aza indiren hiperparametre(tuning) optimizasyonu birçok farklı şekilde yapılabilmektedir (Ateş, 2021). Derin öğrenme ile özellikle öznitelik çıkarma yeteneği sayesinde insan müdahalesi olmadan metin, ses ve görsel veriler için iyi bir sınıflandırma performansı elde edilir. DL yöntemlerinin konuşma tanıma, bilgisayarla görme, örüntü tanıma, öneri sistemleri ve doğal dil işlemeye uygulanmasında dikkate değer bir başarı ile büyük veri çalışmasına uygun olduğu bulunmuştur (Zhou, 2017). Günümüzde, görüntü tanımlama, nesne algılama, görüntü sınıflandırma ve yüz tanımlama görevlerinde DL'nin yeniliği büyük başarıya sahiptir (Zhou ve ark., 2017).

1.12. Yapay Sinir Ağları (YSA)

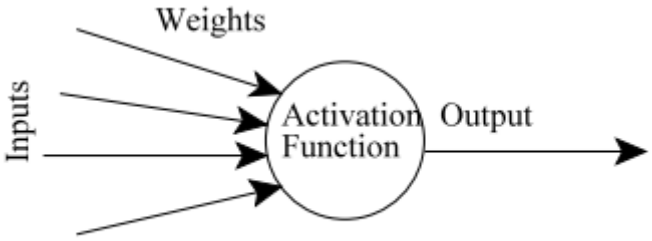
Karmaşık problemleri çözmenin ve anlamanın etkili bir yolu, problemi bölerek incelemektir. Yani karmaşık bir sistemi anlamak için sistem mümkün olduğu kadar basit elemanlara ayrıştırılabilir. Ayrıca karmaşık bir sistem oluşturmak için de basit elemanlar bir araya getirilebilir (Yam, 1997). Ağlar bunu yapmak için var olan bir yaklaşımdır. Çok sayıda farklı ağ türüü vardır ve hepsi bir dizi düğüm ve düğümler arasındaki bağlantılar ile karakterize edilebilir (Gershenson, 2003). Düğümler bir çıktı elde etmek için girdileri alırlar işlerler. Bundan dolayı düğümler hesaplama birimleri olarak görülebilir. Bu işlem girdilerin toplanması şeklinde basit olabileceği gibi bir düğümün başka bir ağ içerebileceği şeklinde karmaşık olabilir (Gershenson, 2003). Düğümler arasında ki bilgi akışını bağlantılar sağlar. Bilginin yalnızca bir anlamda aktığı durumlarda tek yönlü olabileceği gibi, her iki anlamda da aktığında çift yönlü olabilir. Ağlar, fizikte, bilgisayar bilimlerinde, biyokimya, etoloji (hayvan bilimi), matematik, sosyoloji, ekonomi, telekomünikasyon ve birçok alanda çok çeşitli fenomenleri (bir nesne, olay ya da süreç) modellemek için kullanılır. Bunun nedeni, birçok sistemin bir ağ olarak görülebilmesidir: proteinler, bilgisayarlar, topluluklar, vb.

DL'de kullanılan popüler bir yöntem insan beyninden esinlenerek geliştirilen ‘‘Yapay Sinir Ağlarıdır (YSA)’’. Yapay bir nöron, doğal nöronlardan (Şekil 1.16.) esinlenilen bir hesaplama modelidir. Doğal nöronlar, sinyalleri dendritlerde bulunan sinapslar veya nöronun zarı aracılığıyla alır (Gershenson, 2003). Alınan sinyaller yeterince güçlü olduğunda yani belirli bir eşik seviyesini geçtiğinde, nöron aktive edilir ve aksondan bir sinyal yayar. Bu sinyal bir sinapsa gönderilir ve diğer nöronları aktive edebilir.



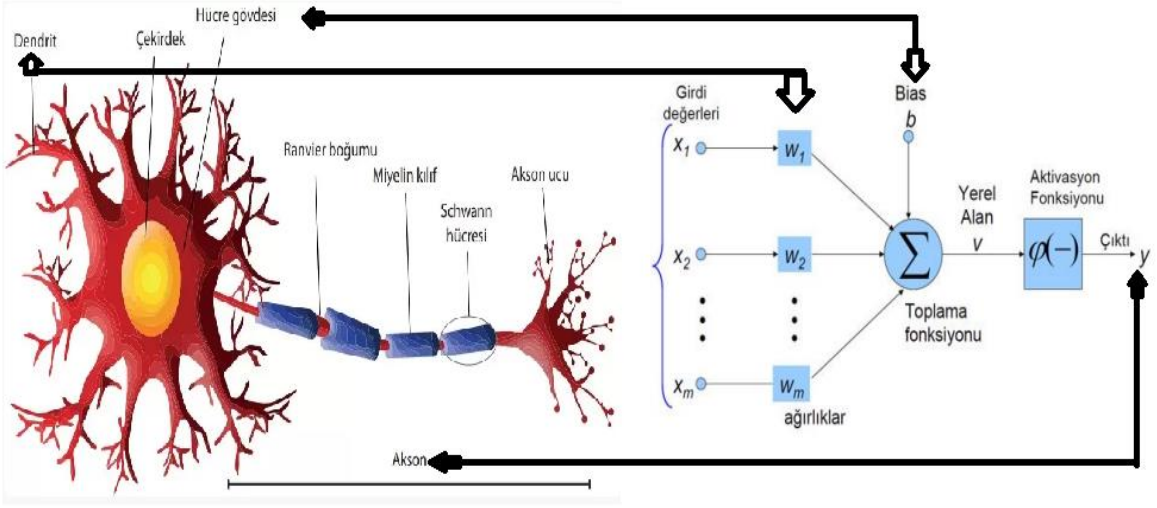
Şekil 1.16. Doğal nöronlar (Gershenson, 2003).

Gerçek nöronların karmaşıklığı, yapay modelleme yapılırken oldukça soyutlanır. Nöronlar temel olarak ağırlıklarla çarpılan (ilgili sinyallerin gücü) girdilerden (sinapslar gibi), oluşur ve daha sonra matematiksel bir fonksiyonla hesaplanır. Bu nöronun aktivasyonunu belirler (Gershenson, 2003).



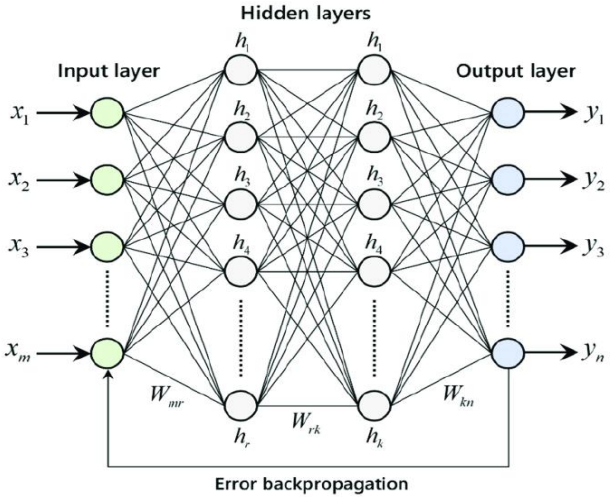
Şekil 1.17. Yapay Nöronlar (Gershenson, 2003).

YSA'ları bilgiyi işlemek için yapay nöronları (Şekil 1.17.) birleştirir. Bir yapay nöronun ağırlığı ne kadar yüksekse, girdi o kadar güçlü olacaktır. Ağırlıklara bağlı olarak, nöronun hesaplanması yapılır. Yani yapay bir nöronun ağırlıkları ayarlanarak, belirli bir girdi sağlayarak bir çıktı elde elde edilir. Ancak yüzlerce veya binlerce YSA olduğunda nöronlar için, gerekli tüm ağırlıkları elle bulmak veya elle belirlemek oldukça karmaşık olacaktır. İstenilen değeri elde etmek için YSA'nın ağırlıklarını ayarlayabilen algoritmalar olmalıdır. Ağırlıkları ayarlama işlemine öğrenme veya eğitim denir (Gershenson, 2003).



Şekil 1.18. Biyolojik Nörondan Yapay Nörona (Gorgoda, 2020 & Yayık, 2013).

DL'de kullanılan popüler bir yöntem insan beyninden esinlenerek geliştirilen ‘‘Yapay Sinir Ağlarıdır (YSA)’’. Günümüzde kodlamada çok sık kullanılır. Şekil 1.18'de gösterildiği gibi, aynı dikey çizgideki tüm düğümler, aynı katmana karşılık gelen girdi birimleridir ve giriş katmanı dışındaki tüm katmanlarda, düğümler, içinde aktivasyon işlevi olan nöronları -ya da yapay nöronları- temsil eder (Ateş, 2021). En sağdaki mavi katman, çıktı birimleri tarafından oluşturulur ve aktivasyonlarının sonucundan sonra adlandırılan çıktı katmanını içerir, ortadaki beyaz düğümler, gizli katmanlara uyan gizli birimlerdir (Rebala, Ravi ve Churiwala, 2019).



Şekil 1.19. Bir YSA'nın İşleyişi (Fernández-Cabán, Masters, & Phillips, 2018).

Aldığı (giriş birimleri) ve bundan ne çıktığı (çıkış birimleri), arada ne olduğu (orta süreç, kararlar, davranışlar) bilinmez, sinir ağının (Şekil 1.19.) ne olduğunun farkında olunmasına rağmen gizli olarak adlandırılır ve kara kutu (Şekil 1.20.) adını alır (Ateş,2021). Sonuç olarak,

bir hatanın kaynağını veya sonucu etkileyen baskın faktörlerin hangileri olduğunu bulmak çok karmaşıktır.

Şekil 1.20. Kara Kutunun Temsili (Taulli, 2019)

DL, Yapay Sinir Ağları (ANN) (Bhatt, Bhatt ve Prajapati, 2017) gibi alt alanları ile birlikte insan beyninin biyolojik yapısının ve yeteneklerinin adaptasyonunu ve kopyalanmasını araştırır (Martín del Brío & Sanz, 2006). Öğrenmek için daha fazla kendi kendine yeterliliği nedeniyle ML'de farklı bir alan olarak kabul edilir (Ateş, 2021). Örneğin, DL, özelliklerini not alarak verilerin bir sınıflandırmasını yapabilir (yapılandırılmamış veriye dönüştürülür), ML'de ise bir insanın bu özellikleri bulmasını (yapılandırmasını) ve daha önce makineye tanıtmasını gerektirir (Egido, 2020). Denetimsiz Öğrenmede DL algoritmalarının kullanılmasının nedeni budur.

Bu karmaşık insan görevini yerine getirmek için, bir başka eşitsizlik; derin öğrenme, doğru sonuçlar üretmek için makine öğrenimi eğitim verisine göre çok daha fazla miktarda eğitim verisine ihtiyaç duyması ve bu teknolojiyi desteklemek için gereken makinelerin üst düzey performansa sahip olması gerektiğidir (Egido, 2020). Bu nedenle, günlük hayatımızda DL modelleriyle etkileşimlerin meydana gelmesi ML ile olduğu kadar sık değildir. Bununla birlikte, kanser tedavisi önerileri ve kalp hastalığının erken tespiti, dikkat çekici kullanımları arasındadır (Patrick, 2020). Nitekim Google, 2015 yılına kadar arama motorunu geliştirmek için DL kullanmıştır (Taulli, 2019).

Uygulama alanlarındaki YSA modellerine bakıldığında; Bilgisayarla görme, konuşma ve örüntü tanıma, yüz hizalama ve algılamadaki zorlukların üstesinden gelmek için yapay sinir ağı tekniği benimsenmiştir (Oludare, 2018). Bilgisayarlı görme; bilgisayarların videolar ve görüntüler gibi görsel verileri doğru bir şekilde anlamasını ve işlemesini sağlamayı amaçlar (Choon, 2004). Bilgisayarlı görmenin temel amacı, bilgisayarlara insan beyninin işlevselliğini sağlayacak türden yetenekler sağlamaktır. Teorik olarak, bilgisayarla görme, verilerin yapay çerçevelerdeki görüntülerden nasıl ayrılacağını inceleyen mantıksal kontrole atıfta bulunur. Bilgisayar görüşünün alt alanları arasında nesne algılama ve nesne tanıma, nesne tahmini, nesne konumu, olay algılama, sahne yeniden yapılandırma, görüntü restorasyonu, görüntü düzenleme, video iyileştirme ve istatistiksel öğrenme yer alır (Almazrooie, 2018). Bu nedenle bilgisayarlı görmede YSA modelleri çok kullanışlıdır. Yüz hizalama, çeşitli görsel uygulamalarda önemli bir rol oynar. Son zamanlarda ANN'ler yüz hizalamada ve yüz tanımada başarılı şekilde kullanılmaktadır (Toufiq, 2014).

1.13. Suç Soruşturmaları ve Güvenlik Alanlarında Veri Madenciliği, Makine Öğrenimi ve Yapay Zeka'nın Kullanımı

Veri madenciliği ve makine öğrenimi kavramları, paket yazılımlar aracılığıyla kullanımları daha kolay olduğu için son zamanlarda daha popüler hale geldiği bilinmektedir. Bu popülerlik veri madenciliğinin kullanım alanlarını da arttırmaktadır. İncelenecek veri yığını büyüdükçe veri madenciliği, yapay zeka, makine öğrenmesi gibi kavramların önemi bir kez daha kendini göstermiştir. İnsanlığın nüfusu arttıkça ve dünyada kurallar var oldukça, diğer veri yığınları gibi suç verileri de artmaya devam edecek.

Suç kavramı, toplumlarda normlara ve sapmalara göre yasaklanmış faaliyetlerin uygulanmasını ifade eder (Ateş, 2021). Toplumdan topluma değişmekle birlikte suç kavramı genel olarak mevcut yasalara aykırı hareket etmek olarak tanımlanmaktadır. Cinsel suçlardan asayiş suçlarına, insan kaçakçılığından uyuşturucuya dayalı suçlara, çocuk suçluluğundan savaş suçlarına kadar suç türlerini kategorilere ayırmak mümkündür ve bunlar insanın hayal gücünün sınırlarını zorlayacaktır. Bu durumda suçla mücadele, var olan toplumsal düzenin devamı için toplumlar için daha önemli hale gelmektedir.

Suçla mücadelede kolluk uygulamaları genel olarak önleyici ve tepkisel (yani yargısal) olmak üzere iki başlık altında incelenmektedir (Clarke, 2006). Tepkisel suç soruşturmaları, suç işlendikten sonra faili meçhul veya suç mağduru gibi unsurları tespit etmeyi amaçlar. Bu araştırmalar olay yeri inceleme desteği ile yürütülmektedir. Önleyici suç soruşturmaları, suç işlenmeden önce etkili suçla mücadele operasyonları ile suçluları caydırarak suç mağdurlarının farkındalığını artırmayı içerir. Ayrıca önleyici tedbirlerde suç analizi kavramı daha önemlidir. Zira suç analizi, var olan suç ve suç eğilimlerinin tespit edildiği ve alınması muhtemel tedbirlerin de dikkate alındığı bir kavramdır.

Makine öğrenimi, tanımlama, sınıflandırma ve örüntüler olarak kümeleme yoluyla veri madenciliğinde suçun kullanılması, biyoloji, kimya, fizik, psikoloji, tıp ve kanıtların incelenmesi için bilgi gibi çeşitli mühendislik ve bilimsel disiplinlerde de önemli konulardan biridir ve nedeni, suç soruşturmalarının çok disiplinli olmasıdır (Hassani ve diğerleri, 2016).

Suç unsurlarının kullanıldığı çeşitli bilim dallarında kullanım alanlarından bazı örnekler şu şekildedir (Ateş, 2021):

- El yazısı ve imzaların karşılaştırmalı örneklerle incelenmesi,
- Adli biyoloji incelemeleri (kan, kemik, saç gibi mevcut biyolojik örneklerin mevcut veri tabanları ile karşılaştırılması işlemleri),
- Trafik kazalarında, kazaların sebeplerinin bulunması amacıyla kazanın yeniden yapılandırılması,
- Adli kimya incelemelerinde (maddenin, özellikle uyuşturucu gibi suçla ilgili türevlerinin tespiti),

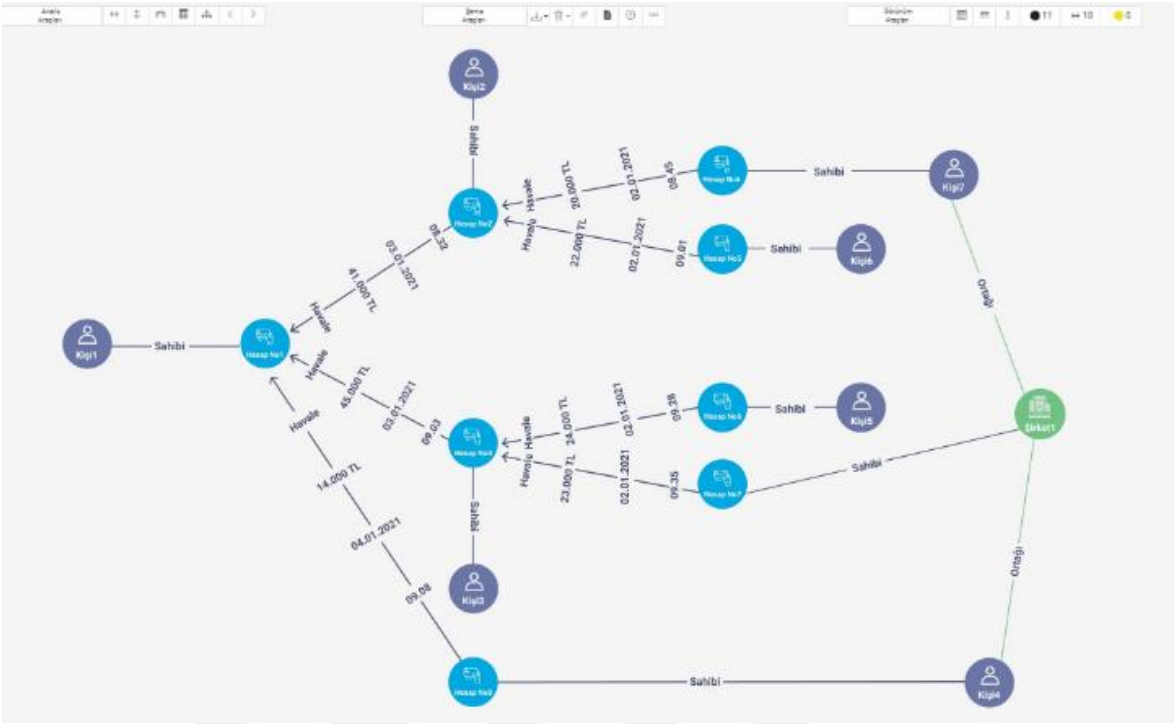
- Adli ses ve görüntü incelemeleri kapsamında geçmiş ses ve görüntü kayıtları ile karşılaştırmaya konu örneklerin incelenmesi (Quick & Choo, 2016),
- Adli muhasebe incelemelerinde (olağan dışı para transferleri veya suça konu tutara uygun para transferleri),
- Bilgisayar, internet, sosyal medya, cep telefonu vb. ortamlarda yapılan işlemlerin adli bilişim kapsamında işlenen suçlarla bağlantısının araştırılması (Quick & Choo, 2016),
- Suç analizi, mağdur ve şüpheli profili çıkarma gibi kriminoloji kapsamında yapılacak suç önleme faaliyetlerinde, suçun yoğun olduğu bölgeler vb. (Chan & Moses, 2017).

Yukarıdaki kullanım alanları örneklerinden de anlaşılacağı üzere büyük veri ve veri madenciliği kavramları multidisipliner bir alanın parçasıdır. Kullanımları her geçen gün artmakla birlikte istenilen düzeyde değildir. Bu nedenle aşağıdaki ana başlıklar altında veri madenciliği ve makine öğrenmesinin kullanım alanlarından bazı örnekler verilmiştir.

1.13.1. Dolandırıcılık ve para yolsuzlukları

Yolsuzluk, kalpazanlık ve dolandırıcılık kavramı, kolluk kuvvetleri tarafından önlenmesi zor olan suç kavramlarından biridir (Ateş, 2021). Ortaya çıkan her kalpazanlık türü insanları yeni tekniklerle aldatma eğiliminde olduğundan, suçu başlangıç aşamasında tespit etmek çok zordur. Ayrıca, teknolojideki gelişmelerle birlikte yolsuzluk ve dolandırıcılık kavramları da giderek daha karmaşık hale gelmektedir (Chau, Pandit ve Faloutsos, 2006). Haksız eylemlerle birlikte toplumda devlete ve özellikle devlete bağlı güvenlik birimlerine olan güven sarsılmakta ve insanlar olumsuz etkilenmektedir.

Teknolojinin gelişmesiyle birlikte bankaların kullanım kolaylığı nedeniyle maddi paranın sanal paraya dönüştüğü ve hatta sanal işlemler kullanılarak gerçek tutarlarda para tablolarının oluşturulduğu günümüzde bilinen bir gerçektir. Söz konusu para transferi verilerinde dünya genelinde çoğu ülkede sorunlar ancak büyük miktar para söz konusu olduğunda dikkate alınmakta, bankalar gerekli bilgileri ilgili adli makamlarla paylaşmakta, yani aslında veri madenciliği kullanılmaktadır (Odia ve Akpata, 2020). Ancak küçük miktarlarda dahi olsa sıklıkla tekrarlanan şüpheli para transferlerinin adli makamlara bildirilmesi, yasa dışı işlemlerin en baştan tespit edilmesini kolaylaştıracaktır (Agu vd., 2019).



Resim 1.1. Dolandırıcılık Tespitinde Bir Programın Ekran Görüntüsü (Data, 2021)

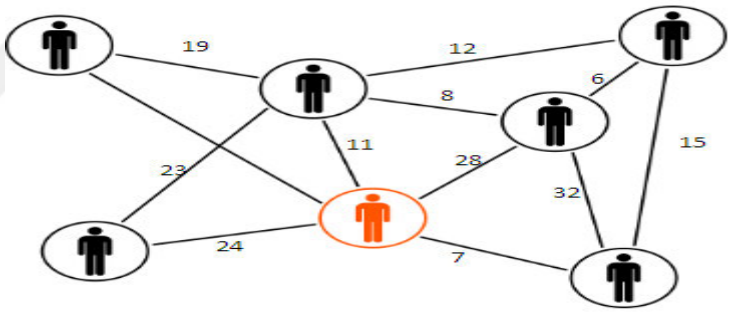
Büyük veri yığınlarında suç teşkil edebilecek unsurları belirlemek kolay olmasa da uygun algoritmik analizler ile şüpheli olarak adlandırılabilecek değerleri belirlemek mümkündür. Bu nedenle veri madenciliği kapsamında yapılan makine öğrenmesi çalışmaları ile birden fazla hesaptan bir hesaba veya bir hesaptan birden fazla paravan hesaba yapılan düzenli para transferleri tespit edilebilmekte ve bu işlemlerin şüpheli işlem olarak incelenmesi suçun daha hızlı tespit edilmesini sağlamaktadır (Resim 2.1.). Şüpheli işlemlerde suç tespit edilirse adli kolluk birimleri devreye girecek, suç tespit edilmezse suçun önlenmesinde önemli olan önleyici kolluk faaliyeti yapılacaktır. Ayrıca, bankadaki müşterilerin istatistiksel verilerinin müşterilerin olağan davranış kalıpları ile karşılaştırılması, yeni şüpheli hareketlerin tespit edilmesine yardımcı olacaktır. Bu bağlamda literatürde özellikle kredi kartı ve banka hesaplarında başarılı makine öğrenmesi ve derin öğrenme modelleri mevcuttur (Adewumi ve Akinyelu, 2017; Perols, 2011; Roy vd., 2018).

1.13.2. Telefon verileri kullanma

Cep telefonu kullanımının dünya genelinde, özellikle gelişmiş ülkelerde nüfusun %68'i olmasıyla birlikte, oranın oldukça yüksek olduğu aşikardır (Wearesocials & Hootsuide, 2018). Cep telefonlarının hayatımızın bir parçası olduğunu rahatlıkla söyleyebiliriz. Avrupa'da ortalama cep telefonu kullanım süresinin ayda 251 dakika olduğu düşünüldüğünde cep telefonu kullanım oranının yüksek olduğu rahatlıkla söylenebilir (Ajans Press, 2017).

Cep telefonu kullanımımız günlük hayatta kiminle ve ne sıklıkla haberleştiğimiz, telefonun hangi coğrafi koordinatta ilgili GSM (Global System for Mobile Communications) firmasının baz istasyonunu aldığı gibi birçok bilgiyi sağlamaktadır. Ayrıca gün içinde sık sık arama yapmak ve her farklı zaman aralıklarında veri yığınları oluştururlar. Bu tür veri yığınlarına HTS (Tarihsel Trafik Arama) denir (Ateş,2021). Bazı kaynaklarda "Çağrı Veri Kaydı (CDR)" olarak geçmektedir (Steenbruggen vd., 2015). Arayan, aranan, arama süresi, aranma süresi, arama konumu ve arama yeri gibi bilgiler dahil olmak üzere kişilerin telefonlarında yaptıkları aramalar hakkında bilgi sağlarlar (Boyd & Crawford, 2012). Böylece bu anlamsız veri yığınlarından yola çıkarak kişinin suç anındaki yaklaşık konumunu bulmak ve şüpheli ile olay yeri arasındaki ilişkiyi kurmak mümkün hale gelmektedir.

Ceza soruşturmaları incelendiğinde, şüpheliler genellikle mağdurun çevresindeki ve mağdurla ilgili kişilerdir (Tilley ve Sidebottom, 2017). Suç sebepsiz işlenmediği için genellikle şüpheli ile mağdur arasında bir ilişki vardır. Bu ilişkiyi ortaya koyan bulgulardan biri de geçmiş görüşmelerin bilgisidir ve özellikle mağdurun zarar gördüğü durumlarda olaydan önceki son arama trafiği çoğu zaman olaya ışık tutabilir (Ateş, 2021). Bu bağlamda literatürde özellikle cep telefonu kullanımına ilişkin başarılı modeller bulunmaktadır (He vd., 2020; Traunmueller vd., 2014).



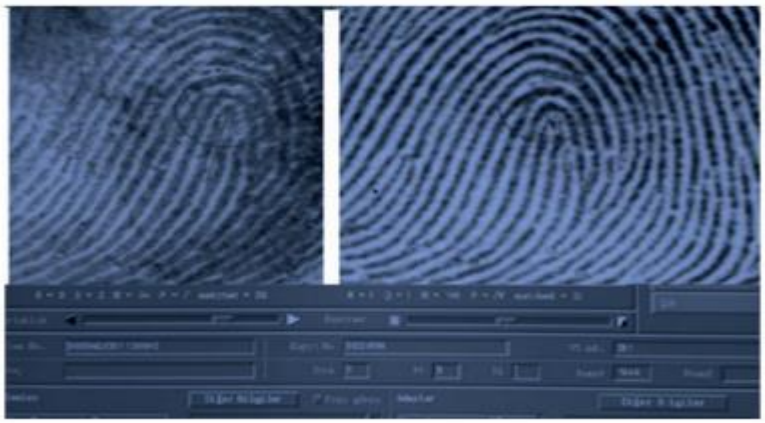
Şekil 1.21. HTS Kaydı Örneği

Şekil 2.1'de de belirtildiği gibi HTS analizi, özellikle organize suç örgütlerinde ilişki boyutunun ortaya çıkarılması açısından oldukça önemlidir.

1.13.3. Olay yerinde DNA bulguları

Olay yerinde tespit edilen bazı bulgular, belirli bir kimlik tespiti için oldukça önemlidir ve parmak izi, DNA örnekleri, şüpheye yer bırakmayacak şekilde şüüpheliyi tespit etmeyi sağlamaları nedeniyle ayrı bir öneme sahiptir (Bostancı, 2015; Wilson ve ark., 2010). İnsanlarda henüz fetüs olduklarında parmak izleri oluşmaya başlar (Ateş,2021). Herkesin parmak izleri benzersizdir bundan dolayı insanları tanımlamak için kullanılabilir. Başka bir deyişle, hiçbir iki kişi tam olarak aynı parmak izlerine sahip değildir. Bu nedenle kimlik belirleme amacıyla kullanılırlar.

Ana papiller çizgiler aynı kalırken, yaralanmalar parmak ucunda iz bırakır, ancak ana karakteristik yapı her zaman aynıdır (Bhuyan ve diğerleri, 2010). Söz konusu tüm parmak izi verileri veri yığınını oluşturur. Adli vakalarda mümkünse olay yerinden alınan parmak izi bulgularından yararlanılarak şüphelilerin kimlikleri tespit edilir.



Resim 1.2. AFIS (Otomatik Parmak İzi Tanımlama Sistemi) sorgulama ekran görüntüsü

(Ateş, 2021)

Resim 2.2'de, AFIS (Otomatik Parmak İzi Tanımlama Sistemi) veri tabanındaki parmak izlerinin bir karşılaştırmasını göstermektedir (Commission, 2017). Yukarıda bahsedilen sistem, belleğine yüklenen yeni parmak izini tarayarak kendi veri tabanına aktarır ve burada makine öğrenmesi algoritmaları ile veri havuzundan en benzer parmak izi örneklerini uzmanlara sunar (Win vd., 2020). Bu bağlamda literatürde özellikle parmak izi tanıma ve sınıflandırma konusunda başarılı makine öğrenmesi ve derin öğrenme modelleri bulunmaktadır (Uliyan vd., 2020; Wani vd., 2020; Pandya vd., 2018).

Saç, kıl, tükürük, ter, idrar, kemik, sperm gibi biyolojik materyelden elde edilen DNA ile veri tabanı karşılaştırılarak parmak izi örneğinde olduğu gibi daha önce alınmış DNA örneklerini içeren veri havuzunda benzer veriler varsa inceleme için uzmanlara gönderilir (Xu, 2020). DNA'nın tek dezavantajı, DNA örneklerinin monozigotik ikizlerde özdeş olmasıdır. Literatürde özellikle DNA konusunda başarılı makine öğrenmesi ve derin öğrenme modelleri bulunmaktadır (Aledhari vd., 2018; Lau & Fung, 2020).

1.13.4. Suç kalıplarının çıkarılması ve gelecekteki suç eğiliminin tahmini

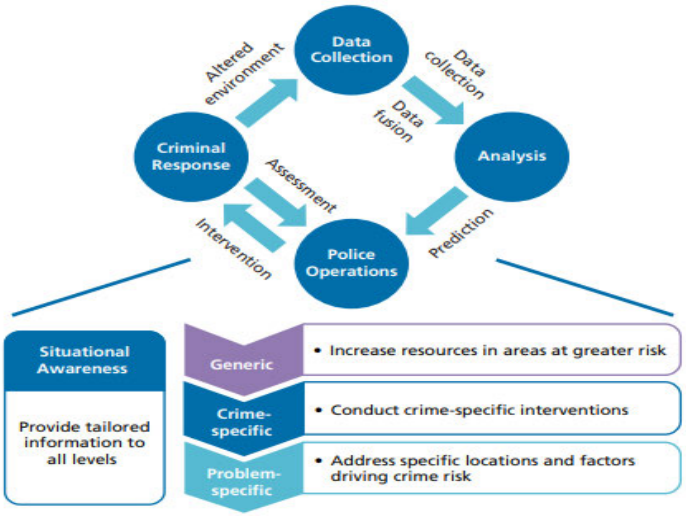
Suç istatistik bilgilerinin analizi, mevcut suçların tespit edilerek suç eğilimlerinin ortaya çıkarılması ve bu tür suçlara karşı gerekli önlemlerin alınmasını içeren bir uygulamadır (Lei, 2019). Çünkü suç genel olarak öğrenilmiş bir davranıştır ve ayrıca gelecekteki suç tahmini için geçmiş geleceğin habercisidir (Wang ve diğerleri, 2016).

Suçla mücadelede suç davranışının tespiti kriminolojinin en önemli konusudur. Suç davranışı genellikle farklı suç motivasyonlarının bir sonucu olarak ortaya çıkar (Ateş, 2021). Duyular, bireyleri sosyal dünyaya bağlayan temas noktalarıdır. Sosyal dünya ile birlikte bencil davranış ve şiddet gibi birçok psikolojik davranış ortaya çıkar. Bu davranışların tanımlanması ve sınıflandırılması, şüphelinin kendisinin ve içinde bulunduğu toplumun duygusal durumunu anlamayı ima eder (Umair, vd., 2015). Bu nedenle suç istatistikleri önemlidir ve suçu anlamak için varsayımsal ve deneysel suç modelleri ile suç davranışını tahmin eden bir dizi dinamik mikro veya makro alanı içeren modeller geliştirilmiştir (Snaphaan ve Hardyns, 2019; Bulgakova ve ark., 2019). Ancak suç verilerindeki temel sorun, analiz öncesi karmaşık ve büyük bir veri yığını şeklinde olmasıdır (Ateş, 2021). Sıradan istatistiksel analiz yöntemleriyle, belirli bir konumda bulunan ve başkalarıyla etkileşime giren bağımlı ve bağımsız değişkenlerin eklenmesiyle de bu kaos içinde bilgi edinmek her zaman kolay değildir. Bu nedenle söz konusu veri yığınından anlamlı bilgilere ulaşmak, verilerin analizi ile mümkündür. Veri madenciliği tekniklerinin kullanılması, suç faaliyetlerine ve potansiyel güvenlik risklerine karşı proaktif olarak harekete geçme yeteneği verir (Feng ve diğerleri, 2019).

Kriminoloji ve sosyoloji alanındaki birçok çalışma, tanımladıkları spesifik analizin boyutuna bakılmaksızın, mikro ve makro coğrafi konum, suç yapısı vb. düzeylerde suç yoğunluğu hakkında önemli miktarda bilgi sağlar (Ateş, 2021). Söz konusu suç verilerinin analizi ile işlenen suç türleri başta olmak üzere suçla ilgili birçok konuda bilgi sahibi olunabilir:

- Suçların ağırlıklı olarak işlendiği saatler,
- Suç noktalarının bulunduğu yerler,
- Şüpheli profilleme,
- Kurban profili oluşturma,
- Suç türü ile diğer değişkenler arasında bir ilişki olup olmadığı,
- Gelecekteki suç oranlarının oranlarının ve sıklığının tahmini.

Bu tür makul şüphe sınıflandırması, ancak veri madenciliği yoluyla elde edilebilen türde tahmin ve kural koyma faaliyeti ile mümkündür (Ateş, 2021). Bu tür bir faaliyet, suçu açıklayan verilerin ortaya çıkarılmasını sağlayacaktır. Her suçta, verilerin ortaya koyacağı nesnel eylem biçimleri ile suç profilinin belirlenmesi suçla mücadelede en önemli kavram olacaktır (Yoo, 2019). Başarılı makine öğrenmesi ve derin öğrenme ile ilgili çalışmalar, özellikle suç tahmin modelleme konusunda literatürde mevcuttur (Berk, 2017; Mittal vd., 2018; Wheeler & Steenbeek,2020).



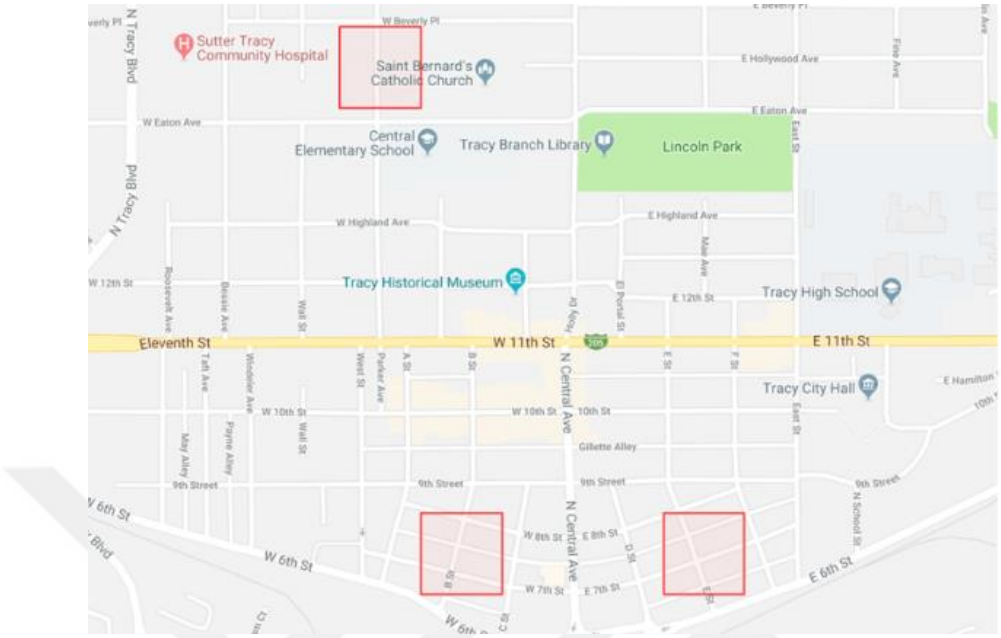
Şekil 1.22. Tahmine Dayalı Polislik İş Süreci (Perry, 2013)

Şekil 2.2'de gösterilen işlemin merkezinde dört adımlı bir döngü vardır (Şeklin üst kısmı). İlk iki adım, suç, olay ve suçlu verilerini toplamak ve analiz etmek, tahminler üretmektir. Toplanan verilerle daha iyi analiz yapılıp daha iyi karar verilebilmesi için farklı kaynaklardan gelen veriler biçimlendirilerek birleştirlir (Veri füzyonu) (Perry,2013). Üçüncü adım, öngörülen olaylara müdahale eden polis operasyonlarını yürütmektir. Müdahalenin türü duruma göre değişir. Bunlar, en basitten en karmaşığa doğru, genel müdahale, suça özel müdahale ve soruna özel müdahaledir (Perry,2013). Müdahalenin türü ne olursa olsun, müdahaleyi başarıyla yürütmek için müdahaleyi yapanların bilgiye ihtiyacı vardır. İhtiyacı karşılayan bilgilerin sağlanması memurlar ve personel arasında durumsal farkındalık için herhangi bir müdahalenin kritik bir parçasıdır. Müdahaleler, suçu ideal olarak azaltan veya çözen bir cezai müdahaleye (tutuklama vb.) yol açar. Buda dördüncü adımı oluşturur (Perry, 2013).

Tahmine dayalı polislik, potansiyel suç faaliyetleri hakkında istatistiksel tahminler yapmak için kolluk kuvvetleri tarafından analitik tekniklerin kullanılması anlamına gelir (Rand, 2013). Tahmine dayalı polislik ürünleri her geçen gün daha da artmakta ve farklı algoritmalarla mağduriyet oluşmadan suçların çözülmesini amaçlamaktadır. Tahmine dayalı polislik ürünleri, karmaşıklıkları bakımından farklılık gösterir (Shapiro, 2017). Bu ürünlere örnek verecek olursak; PredPol, antropolog Jeffrey Brantingham ve matematikçi Andrea Bertozzi tarafından deprem artçı şoklarını tahmin etmek için üretilen algoritmalardan uyarlanarak geliştirilmiş, tahminleri hesaplamak için üç değişkene dayanan (suç türü, tarih- saat ve yer) bir makine öğrenimi algoritması kullanan, şuçu thmin etmeye çalışan bir uygulamadır (Crunchbase, 2013).

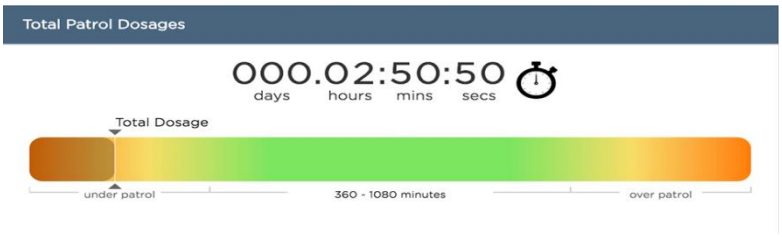
PredPol'ün kolluk kuvvetlerine sunduğu değer önerisi üç katlıdır. İlk olarak şirket, üç veri noktasına dayalı makine öğrenimi algoritması aracılığıyla tahmine dayalı polislik sunar: suç türü, suç yeri ve suç tarihi/saati (Crunchbase, 2013). Algoritmayı eğitmek için tarihsel olay veri kümelerini kullanılır. Tahminler, Google Haritalar aracılığıyla bir web arayüzünde kırmızı

kutular olarak görüntülenir. Kutular, her gün ve ilgili vardiya için en yüksek riskli alanları temsil eder (Resim 2.3.).

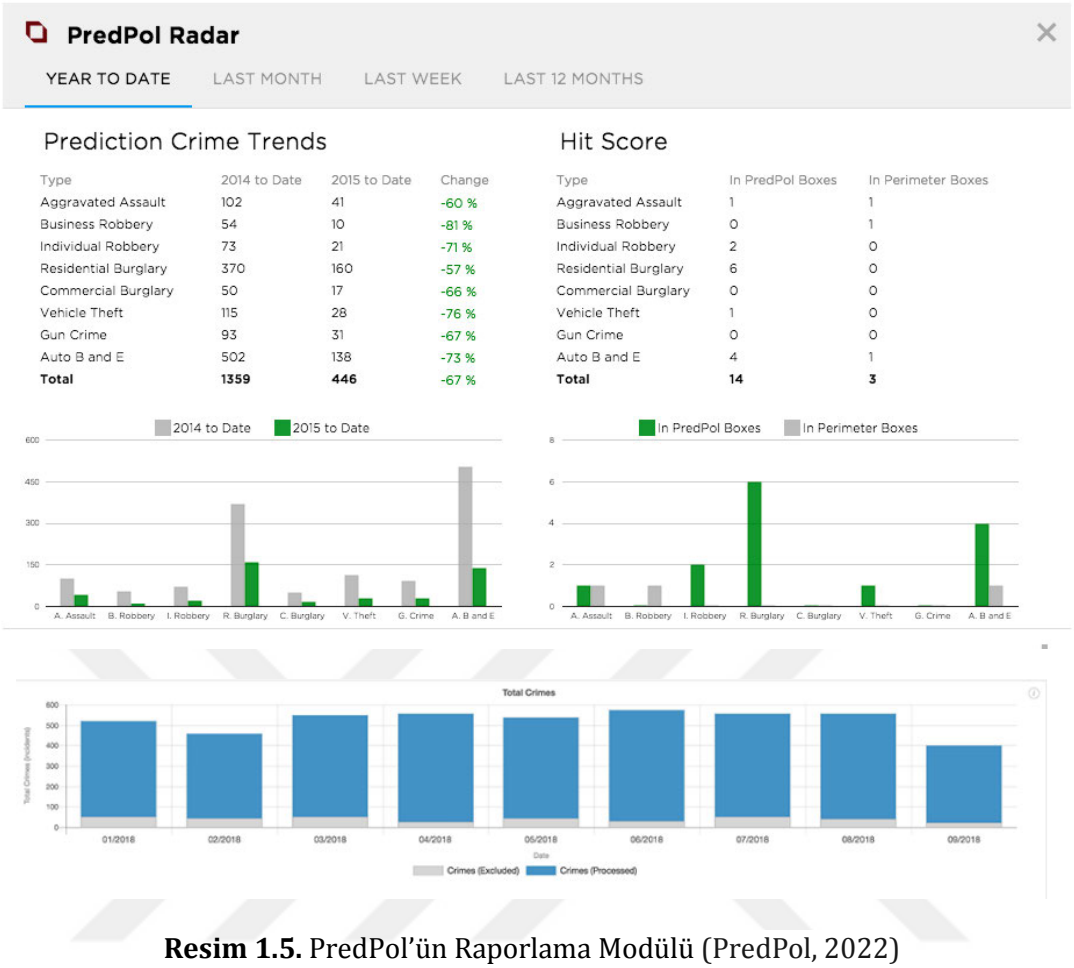


Resim 1.3. Google Haritalar aracılığıyla bir web arayüzünde kırmızı kutular olarak görüntülenen tahminler (PredPol, 2022)

İkincisi, görevlilerin devriye sürelerini ve PredPol kutularında harcadıkları süreyi görmeyi sağlar. Kısaca algoritma sonuçlarına göre görev planlama ve konum yönetimi hizmetleri sağlar. Son olarak PredPol, belirli bir tarih aralığında herhangi bir suç türü, görev, bölge vb. kombinasyonuna göre özel raporlara izin veren bir analitik raporlama modülü sunar (Resim 2.4-2.5.) (PredPol, 2018).



Resim 1.4. Memurların PredPol kutularında geçirdikleri süre (PredPol, 2022)



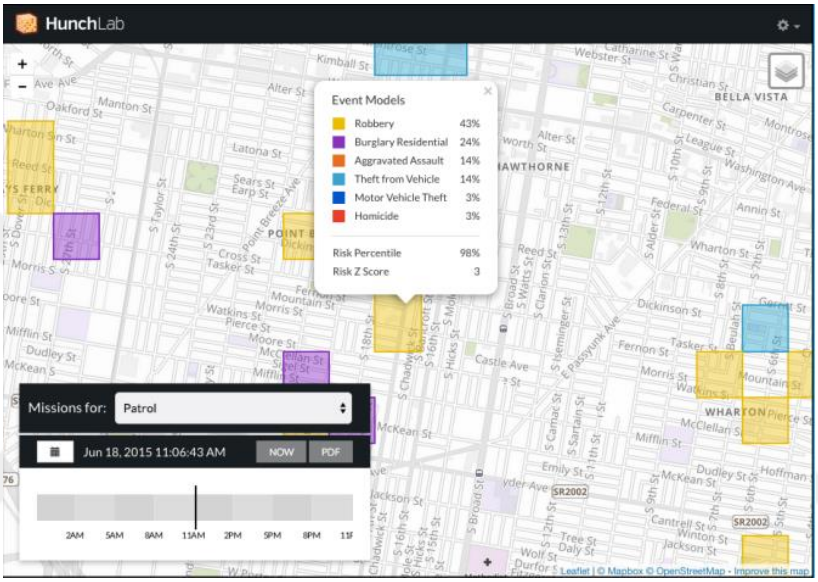
Resim 1.5. PredPol'ün Raporlama Modülü (PredPol, 2022)

Suçu tahmin etmek için makine öğrenimi ve yapay zeka algoritma uygulamalarını kullanan bir diğer şirket ise HunchLabtır. Diğer sistemlere göre daha iyi kararlar vermek için daha fazla veri içerir ve bu veriler, kamuya açık suç raporları ve polis yardımı taleplerinin yanı sıra hava durumu, eğlence mekanları ve ulaşım merkezleri gibi coğrafi özellikleri, önemli etkinliklerin veya okul döngülerinin verilerinden oluşur (Shapiro, 2017). HunchLab, tahminlerde bulunmak için makine öğrenimini kullanır. Bu, HunchLab'ın bir bilgisayara belirli bir suç türünün (saldırı, motorlu araç hırsızlığı vb.) belirli bir zaman diliminde çeşitli konumlarda ne kadar olası olduğunu belirlemeyi öğrettiği anlamına gelir. Bunu yapmak için, HunchLab kolluk tarafından sağlanan geçmiş verilerden, suç raporlarından ve kolluk yardım çağrılarından bir eğitim seti oluşturur (Shapiro, 2017). Bu verilere dayalı modeller oluşturmak için çalışır. Her model bazı hatalar yapacaktır, ancak Hunchlab mümkün olan en iyi modeli yapmak için geçmiş hataları yinelemeli bir şekilde iyileştirmeye çalışmaktadır. HunchLab'ın çalışma sistemi şu şekildedir (HunchLab, 2014).

HunchLab tahmin etmeye eğitim örneklerinden rastgele bir seçim yaparak başlar. Dahasonra, değişkenlere göre suçların nerede işlendiği veya işlenmediği örneklerini ayıracak bir karar ağacı oluşturur. (Örneğin, ağaçtaki ilk "karar", "geçen yıl içinde bu konumda bir olay meydana gelmediyse, muhtemelen bugün de olmayacak" şeklinde yorumlanabilir.) Her belirli veri

kümesinde, bu süreç, suçların neden işlendiği yerde işlendiğini açıklayan bir ağaç oluşturmaya devam edecektir. Hunchlab bu işlemi tekrar edecek ama bu kez eğitim setinin başka bir rastgele bölümünü seçecektir. Ancak suçların işlendiği ve işlenmediği yerleri ayırmak için bu kez bir ağaç oluşturmak yerine, bu ağaç ilk ağaç modelinin bu yeni veri kümesine uygulanmasından kaynaklanan hataları tahmin etmek için oluşturulmuştur.

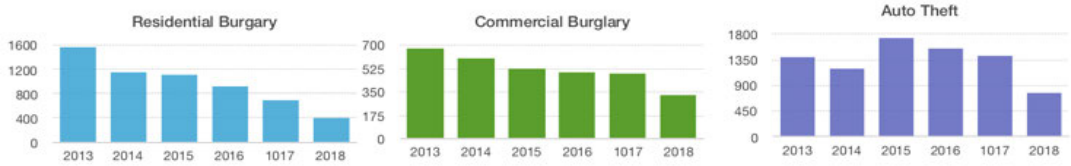
HunchLab tüm bu süreci birkaç kez tekrarlar. Her yeniden başladıklarında, ilk örnek verilerin bir kısmı geri tutulur. Nihai sonuç, olasılık olarak ifade edilen, bir suçun olup olmayacağına dair bir evet/hayır tahminidir. Tahminler, belirli bir süre için bir haritada hücreler halinde gösterilir (Resim 2.6.).



Resim 1.6. Tahmine Dayalı Polislik İş Süreci (Hunchlab, 2014)

Amerika Birleşik Devletler' inin Kaliforniya eyaletinde tahmine dayalı yapay zeka polislik ürünü olan PredPol uygulaması devreye girdikten sonra suçlarda ciddi azalmalar görülmüştür (Resim 2.7.) (PredPol, 2018).

DESCRIPTION	2013	2014	2015	2016	1017	2018
Auto Theft	1389	1195	1735	1553	1412	778
Res Burglary	1573	1149	1112	925	696	390
Comm Burglary	678	605	527	496	492	327



Resim 1.7. Mala Karşı Suçlar (PredPol, 2018)

1.13.5 Suç Verilerinin İnternet ve Sosyal Medyada Analizi

Bilgi çağında internetin ortaya çıkması ve gelişmesiyle birlikte zaman ve mekan kavramları en aza indirgenmiş ve başta iletişim alanı olmak üzere birçok alanda büyük değişimler meydana gelmiştir (Ateş, 2021). Bu değişimlerle birlikte insanların boş zaman aktiviteleri ve tutumları değişmiş, yeni iletişim yöntemleri keşfedilmiş ve başta gazetecilik, eğitim, bankacılık ve ticaret hizmetleri olmak üzere birçok işletme sanal ortamda faaliyet göstermeye başlamıştır. Son zamanlarda yapılan çalışmaların çoğu, internet kullanımının günlük bir rutin haline geldiğini ve bireylerin birçok günlük aktiviteyi sürdürmek için internette çok fazla zaman harcadıklarını göstermektedir (Ateş, 2021). Bireyler sahip oldukları teknolojik cihazların sayısının artmasıyla birlikte özellikle taşınabilir dizüstü bilgisayar, tablet, cep telefonu gibi cihazlarla ister özel alanda ister kamusal alanda internete her an erişebilecekleri noktaya gelmişlerdir. İnternet ve sosyal ağlar, bireylerin iletişim kurma ve sosyalleşme biçimini değiştirmiştir. Bu nedenle sanal ortamlar iletişim için popüler platformlar haline gelmiştir. İnsanların interneti ve sosyal ağları giderek daha fazla kullandığı dünyada, kullanılan veri miktarı her geçen gün artmaktadır (Agrahari ve Rao, 2017; Ateş vd., 2020).

İnternetin ve sosyal paylaşım sitelerinin sosyal bir platform olarak ortaya çıkmasıyla birlikte internet ve sosyal medya aracılığıyla işlenen suçlarda artış gözlemlenmiştir (Ateş, 2021). Sosyal medya özellikle son yıllarda çocuk pornografisi, siber zorbalık, taciz, hakaret, internet dolandırıcılığı ve siber terörizm gibi suçlarda artış göstermiştir (Heickerö, 2014). Arap baharında görüldüğü gibi, bazen mevcut hükümete karşı insan eylemlerini organize etmek için sosyal ağlar kullanılabilir.

İnternet ve sosyal medya üzerinden yapılan paylaşımların da suçların aydınlatılmasında etkisi vardır. Amerika Birleşik Devletleri'nde yapılan bir araştırmada, kolluk kuvvetlerinin %80'inden fazlası sosyal ağları aktif olarak kullanmaktadır (Guenther, 2012). Bu konu ile ilgili net bir istatistiksel çalışma olmamasına rağmen internet ve özellikle internetteki sosyal ağlar dünya çapında çoğu ülkede her türlü adli suç ve güvenlik soruşturması için aktif olarak kullanılmaktadır (Ateş, 2021). Örneğin intihar sayılan bir adli vakada kişinin paylaşımlarının normal olması bunun intihar süsü verilmiş bir cinayet olabileceği şüphesini doğururken, terör suçlarında kişinin terör örgütünü öven paylaşımları olabilir. Örgüt liderinin düşünceleri ve resimleri de bu kişinin örgütle ilgili olabileceği şüphesini artıracaktır. Bunların dışında internetteki çeşitli uygulama ve programlar, özellikle temas tabanlı arama motoru taramaları ve kişi paylaşım uygulamaları (getcontact ve truecaller gibi) suça ilişkin verileri destekleyebilmektedir.

İnternet ve sosyal ağlardaki (tipik olarak "Sosyal Ağ Hizmetleri" (SNS) olarak adlandırılan) veriler kişi bazında kullanılmasının yanı sıra, genel çerçevede önleyici ve bilgilendirici görevlerde de aktif olarak kullanılmaktadır (Arshad et al., 2019). Dünyadaki istihbaratın %60-80'i açık kaynaktan elde edilmektedir ve bu nedenle açık kaynak verilerinin analizi giderek daha önemli hale gelmektedir (Power, 2016; Chen ve diğerleri, 2014). Çeşitli istihbarat

birimleri, teröristlerin faaliyetleri başta olmak üzere çeşitli konularda aktif olarak bilgi toplamakta ve analiz etmektedir (Power, 2016).

Özellikle anahtar kelime aramayı sağlayan sosyal ağ araçları aracılığıyla, ilgili güncel konuların genel paylaşımları da incelenebilir (Ayre ve Craner, 2017). Ayrıca literatürde özellikle internet ve sosyal medya çalışmalarında başarılı makine öğrenmesi ve derin öğrenme modelleri mevcuttur (Ch vd., 2020; Williams vd., 2020; Ristea vd., 2020; Muneer & Fati, 2020).

1.13.6. Biyometrik Özelliklerin Güvenlik Alanında Kullanılması

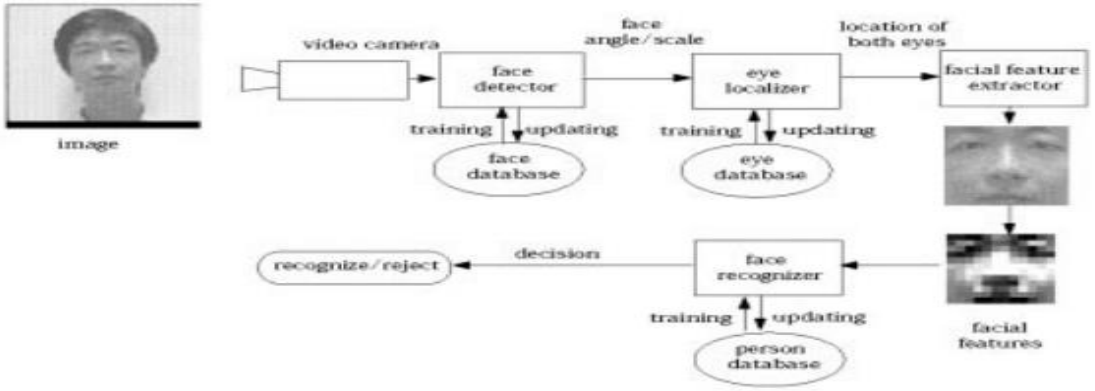
Biyometrik kimlik doğrulama tipik olarak insanların fiziksel ve davranışsal özelliklerini ölçmek için kullanılır (Ateş, 2021). Söz konusu ölçülebilir değerlerin otomasyon sistemleri aracılığıyla ayırt edilmesini sağlayarak kimlik doğrulama başta olmak üzere çeşitli güvenlik alanlarında kullanılır (Stewart, 2019). Biyometri kavramının en önemli özelliği, tanımlamalarda herhangi bir nesne ya da veri kullanılmadan, yalnızca kendileri tarafından kullanılmasıdır.

Biyometrik sistemler temel olarak fiziksel (pasif) ve davranışsal (aktif) olmak üzere iki grup altında incelenir (Tiwari, et al., 2015). Fiziksel biyometride ses, yüz, el geometrisi, parmak izi, iris ve retina kullanılırken, davranışsal biyometride yazı stili, imza, yürüme şekli ve konuşma sırasındaki dudak hareketleri gibi özellikler kullanılır. Temel olarak fiziksel biyometri, bir bireyin diğer insanlardan ayırt edilmesini sağlayan sabit fiziksel özelliklerine dayanmaktadır (Martinovic, vd., 2017). Davranışsal biyometri, belirli bir amaç doğrultusunda ve belirli bir zamanda birbirinden farklı kişiler tarafından gerçekleştirilen davranışlara dayanır. Bu biyometriklerin güvenilirliği, diğer doğrulama yöntemlerinde olduğu gibi aktarılabilir veri olmadığı için güvenlik açısından çok yüksektir (Ateş, 2021).

Biyometrik sistemlerin günlük yaşam pasaport veya görüntüleme sistemlerinde, sınır kontrolünde, video gözetlemede, suç tespitinde, giriş kontrolünde, bilgisayar girişlerinde, akıllı telefonlarda kullanıcı doğrulamada, kalabalık taramada, e-ticarette, elektronik bankacılıkta, adli bilişim ve kimlik kartlarında kullanılır.

Biyometrik doğrulama çok güvenilir bir yöntemdir çünkü denetimli kalıp öğrenmeye dayalı veri madenciliği kullanır. Ayrıca bu tür sistemlerin tasarımı ve simülasyonu da yapay sinir sistemleri ve sinyal işleme teknikleri kullanılarak çok daha basit hale gelmektedir. Daha fazla kimlik doğrulama ve güvenlik gibi olumlu yönleri göz önüne alındığında, yakın gelecekte biyometri kullanım sıklığının artması beklenmektedir.

1.13.6.1. Yüz Tanıma Sistemi



Şekil 1.23. Yüz tanıma sistemi için bir çerçeve (Lin, 2000)

Şekil 2.3'de yüz tanıma sisteminin bir örneğini gösterilmektedir. Yüz tanımada insanların yüzleri bir kamera ile çekilerek analiz edilir. Analiz gözler, burun, çene, ağız dahil olmak üzere bütün yüz yapısını ölçer. Bu ölçümler veritabanlarında saklanır ve kullanıcı kamera karşısına geldiğinde sistem karşılaştırarak kişiyi tanır. Yukardaki sistemde yüz dedektörü (yalnızca bir yüzün yaklaşık konumu bulmak için), göz konumlandırıcı (yüzün tam konumunu bulmak için her iki gözün konumlarını bulan) ve yüz tanıyıcı (yüz hatlarından veri tabanından yüzü bulan) olmak üzere üç fonksiyonel modül vardır (Lin, 2000).

Teknolojinin gelişmesiyle yüz tanıma sistemleri her geçen gün daha da gelişmektedir. Ülkeler özellikle yüz tanıma teknolojisini kullanarak vatandaşları hakkında her geçen gün daha fazla bilgi elde etmekte ve elde ettikleri verileri daha çok alanda kullanmaktadır. Bu teknolojiyi en çok kullanan ülkelerden biride Çin'dir. Çin bu teknolojiyi geliştirmeye 2003 yılında ''Akıllı Şehirler'' projesiyle başladı ve hava kalitesi, trafik akışı, tıkanıklık, atık su bertarafı dahil olmak üzere şehir yaşamının her yönünden verileri ölçmeyi, izlemeyi ve analiz etmeyi amaçlayan yüzlerce akıllı şehirler yaratmayı hedefledi (Gershgorn, 2021). Bu kapsamda Çin, son derece gelişmiş yüz tanıma teknolojisine ulaşmıştır (Resim 2.8.). Teknoloji, ulusal bir gözetim ve veri paylaşım platformu oluşturmak için büyük veri ve yapay zeka gibi diğerleriyle entegre edilmiştir. Akıllı sistem, trafik kurallarını ihlal eden karşıdan karşıya geçenleri ve motorsuz araçları kaydetmek için yüz tanıma teknolojisi ile donatılmıştır (Zhang, 2021).

Resim 1.8. Çin yüz tanıma kamera görüntüsü (Comet Labs Research,2018)

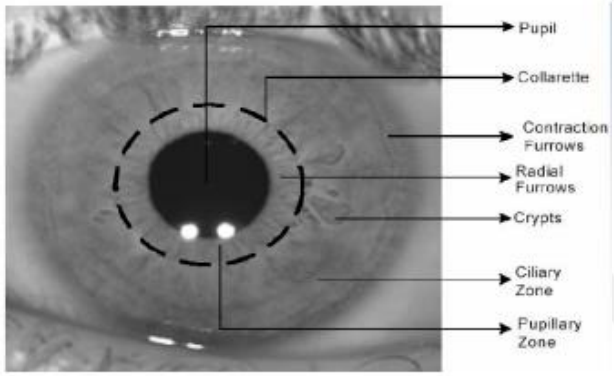
Aynı şekilde yüz tanıma sistemine sahip ekipman kullanan polisler (Resim 2.9.) ve programlanmış robotlar halkın yoğun olarak kullandığı istasyonlar, alışveriş merkezleri ve tarihi alanlar gibihalka açık yerlere yerleştirilmiştir (Halder,2017). 2020'de, ülke genelindeki Kamu Güvenliği Büroları, sağlık hizmetleri, sosyal medya faaliyetleri ve internet tarama faaliyetlerinden veri toplamak ve faaliyetleri izlemek ve tahmin etmek için "Polis Bulutu" sistemlerini uygulamaya başladı. Polis bulut sistemi, ikamet adresleri, aile ilişkileri, doğum kontrol bilgileri, dini bağlantılar, otel, uçuş kayıtları, tren kayıtları, biyometri, CCTV görüntüleri ve bilgiler dahil ancak bunlarla sınırlı olmamak üzere tüm mevcut kaynaklardan gelen bilgileri kamu güvenliği bürolarına entegre etmeyi amaçlar ve bunları diğer devlet daireleriyle paylaşılır.

Resim 1.9. Çin yüz tanıma sistemine sahip akıllı gözlük takan bir polis memuru (Chan,2018)

Bir diğer proje olan ‘‘Keskin Gözler’’ 2016'da yayınlanan 13. Beş Yıllık Plan'a göre, 2020 yılına kadar Çin'deki kamusal alanın yüzde yüzünü yüksek çözünürlüklü kameralarla gözetlemeyi amaçlayan bir projedir. Yapay zeka ve insanlar tarafından evlerinde dahil canlı izlenebilen görüntüler sayesinde Çin suç oranını ve kolluk güçü ihtiyacını azaltmayı hedeflemektedir.

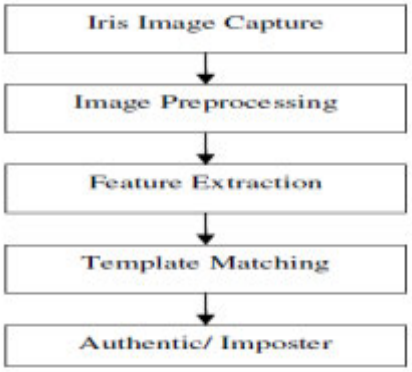
1.13.6.2. İris Tanıma

İnsanların fiziksel özellikleri zamanla değişebilir, kaybolabilir veya bir kişiden diğerine geçebilir. Bundan dolayı fiziksel özelliklerden yararlanılarak kimlik tespiti güvenli bir yöntem değildir. Teknolojinin gelişmesi ile kimlik doğrulama konusunda daha güvenilir bir yol olan biyometrik kimlik doğrulama ortaya çıkmıştır. Biyometrik kimlik doğrulamanın çok sayıda yöntemi vardır. Bu yöntemler retina tanıma (Şekil 2.4.), göz veya yüz çevresi, parmak izi, avuç içi, ses, imza tarzı tanımadan oluşmaktadır. Yöntemler kullanılan alana göre değişebilir. Örneğin bankalarda yüz tanıma sistemi kullanıldığı görülürken, sağlık kuruluşlarında avuç içi tanıma sistemi kullanıldığı görülmektedir. Biyometrik yöntemlerin de kendi aralarında güvenlik seviyeleri vardır. Bunlardan güvenilirlik ve değiştirilebilirlik açısından en güvenilirlerden biride iris tanıma yöntemidir. "İris" terimi, insan gözünün dışarıdan görülebilen oldukça dokulu halka şeklindeki kısmını ifade eder (Nguyen, 2017).



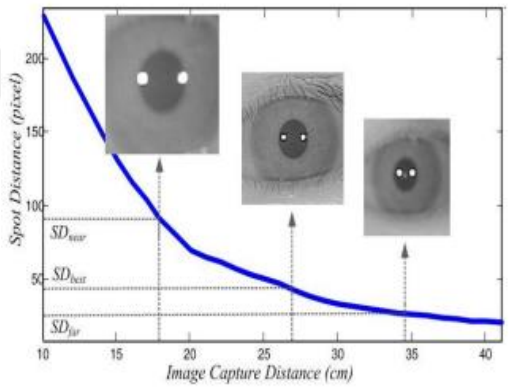
Şekil 1.24. Gözün Önden Görünümü (Ross, 2010)

Bir iris tanıma sistemi, bireyleri ayırt etmek için bu dokusal modellerin zenginliğinden yararlanır ve bu iris deseninin zengin dokusal bilgisini yakalamak için kızılötesi (NIR) sensörler kullanılır. İris Tanıma kavramı ilk olarak Dr. Frank Burch tarafından 1939'da önerilmiş, 1990 yılında Dr. John Daugman tarafından örüntü tanıma ve bazı matematiksel hesaplamalar kullanılan algoritmalar geliştirilmiştir (Şekil 2.5.) (Verma, 2012). İris, göz içinde bulunan ışık miktarını kontrol ederek gözün boyutunu ayarlayan, gözün renkli kısımını oluşturan bir kastır. İrisin cerrahi yöntemle değiştirilememesi, doğal izolasyonu ve dış ortamdan korunması, yaşla stabil olması güvenilirlik açısından kimlik tespitinde çok büyük önem oluşturmaktadır (Verma, 2012). Ancak irisin görüntüsünün alınma mesafesi son birkaç yılda 1m ila 60m arasında değişse de daha uzun mesafelerde iris tanıma sistemleri tasarlamak ve uygulamak çok zordur (Nguyen, 2017)



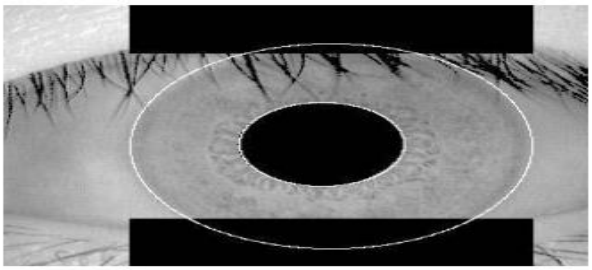
Şekil 1.25. İris Tanıma Aşamalarının Blok Şeması (Verma, 2012)

İris tanıma sisteminin ilk aşaması irisin yüksek kaliteli görüntüsünün yakalanması, görüntü elde edilmesidir (Verna,2012). Bu aşamada önemli olan görüntü elde etme donanımlarının yeterli çözünürlük, keskinlik, iyi kontrast ayrımı, uygun mesafeden algılama gibi özelliklerinin yeterli olmasıdır. Aşağıdaki şekilde görüntü yakalama mesafesi arttıkça irisin görüntüsünün uygun şekilde yakalanması gösterilmiştir (Şekil 2.6.). Mesafe arttıkça iris görüntüsünün çözünürlüğü düşmüştür.



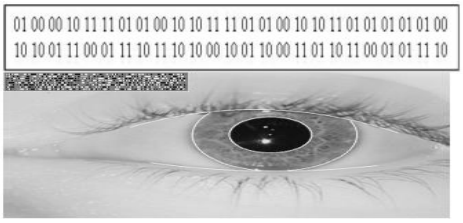
Şekil 1.26. Görüntünün uygun şekilde yakalanmasını gösteren grafik (Ross, 2010)

İris lokalizasyonu (Şekil 2.7.), elde edilen görüntüden iris bölgesini izole etme işlemidir (Verma,2012). Yani sklera (sert tabaka, beyaz kısım) ve damarsal ağ tabakadan(koroid) irisin ayırt edilmesidir.



Şekil 1.27. İris lokalizasyonu (Wang, 2017)

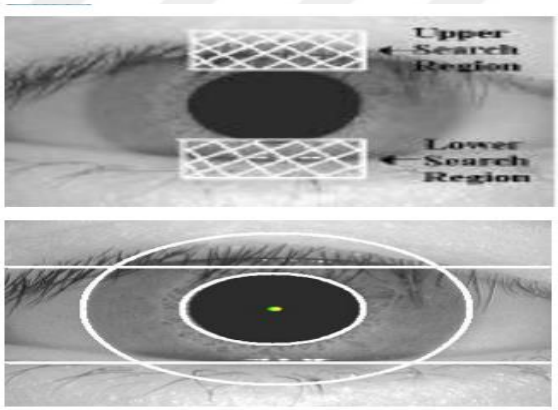
Özellik Çıkarma; iris özelliklerinin kodlar halinde çıkarılmasıdır (Şekil 2.8.).



Şekil 1.28. İris özelliklerinin kodlar halinde çıkarılması (Abikoye, 2014)

Şablon eşleştirme; yüksek kaliteli iris taramasından elde edilen şablonların eşleştirilmesi sağlanır.

İris dış sınır lokalizasyonuna benzer şekilde üst ve alt göz kapağı algılama yöntemide bir diğer kimlik tespit ve doğrulama yöntemidir (Şekil 2.9.) (Verma, 2012). Güvenilirlik açısından değerlendirildiğinde tek başına yeterli olmasada iris tanıma yöntemi ile birlikte kullanıldığında güvenilirliği arttırıcı bir yöntemdir.



Şekil 1.29. (a) İris görüntüsünün üst ve alt arama bölgeleri.

(b) Üst ve alt göz kapaklarının tespiti. (Verma, 2012)

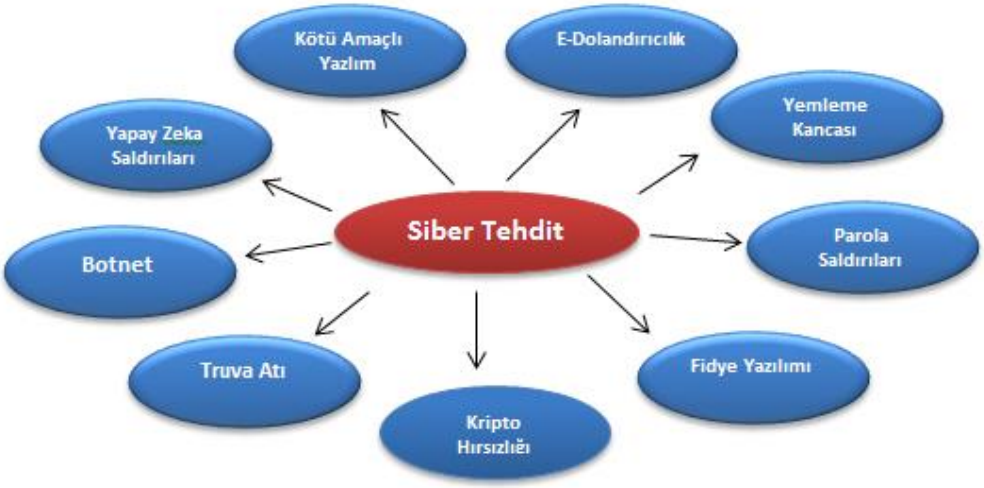
Diğer biyometrik kimlik doğrulama sistemleride (avuç içi, retina tanıma, kulak, el damarı, vücut kokusu, davranışsal ses, yürüyüş, imza, klavye tuşlarına basma şekli vb.) yüz ve iris tanıma sistemine benzer şekilde çalışan kimlik doğrulama ve tespit etme yöntemleridir. Kullanılan alan ve kurum sayısı gün geçtikçe artan bu sistemlerin suç soruşturmalarında da kullanımı teknik olarak giderek yaygınlaşmaktadır. İlk çağlardan günümüze kadar olan süreçte bu yöntemlerin birçoğu kullanılsa da veri madenciliği, makine öğrenimi ve yapay zekanın gelişmesi bu yöntemlerden elde edilen sonuçların daha hızlı ve güvenilir olamasını sağlamış ve kullanım alanını arttırmıştır. Örneğin Çin Hümeti CCTV kamera ağını ve yüz tanıma teknolojisinde ne kadar geliştiklerini göstermek için Aralık 2017'de BBC muhabirini sisteme ekleyerek yaptığı bir testte BBC muhabirini bulup yakalaması yedi dakika sürmüştür (Russell, 2017).

1.13.7. Siber Güvenlik

Siber güvenlik bilişim sekteründeki gelişmelere bağlı olarak günümüzün büyük ve zor sorunlarından biri haline gelmiştir. Siber güvenlik teriminin gerçekte ne anlama geldiği ve çeşitli bağlamlarda nasıl yer aldığı konusunda yetersizlik vardır (Diakun, 2014). Bundan dolayı siber güvenliğin birden fazla tanımı yapılmıştır:

- Siber güvenlik, büyük ölçüde olası davetsiz misafirleri tespit etmek ve engellemek için kullanılan savunma yöntemleridir (Kemmerer, 2003),
- Siber güvenlik, bilgisayar ağlarının ve içerdikleri bilgilerin sızma ve kötü niyetli hasar veya kesintilerden koruyan uygulamadır (Lewis, 2006),
- Siber Güvenlik, ağlara, yazılımlara ve bilgisayarlara yönelik kötü niyetli saldırı risklerini azaltmayı içerir. Siber güvenlik kötü niyetli erişim ve izinsiz girişleri tespit ederek engellemek, virüslerden korumak ve durdurmak, kimlik doğrulama sağlamak, şifreli iletişimin devamını ve etkileşimini sağlamak için kullanılan araçları içerir (Amoroso, 2006),
- Siber güvenlik, siber ortamı siber saldırılara karşı koruma veya savunma yeteneğidir (CNSS, 2010),
- Gizlilik, bütünlük ve kullanılabilirliği sağlamak için ağları, bilgisayarları, programları ve verileri saldırı, hasar veya yetkisiz erişimden korumak için tasarlanmış teknolojiler, süreçler, uygulamalar, müdahale ve azaltma önlemleri bütünüdür (Kozlenkova, 2014)
- Bir ulusun bilgi toplumunun varlığını ve devamlılığını sağlama, siber uzayda bilgi, varlık ve kritik altyapısını güvence altına alma ve koruma sanatıdır (Canongia & Mandarino, 2014),
- Elektronik verilerin suça veya yetkisiz kullanımına karşı korunma durumu veya bunun için alınan tedbirlerdir (Oxford, 2014),
- Bilgi ve iletişim sistemlerinin ve içerdiği bilgilerin hasar, yetkisiz kullanım veya değişiklik veya istismardan korunduğu ve/veya bunlara karşı korunduğu faaliyet, süreç, yetenek veya durumdur (DHS, 2014).

Bu tanımlardan bazıları teknik olmayan faaliyetlere ve insan etkileşimlerine göndermeler içermesine rağmen, literatürde teknik bakış açısının baskın olduğu görülmektedir (Diakun, 2014). Siber güvenlik alanında ki zorluklar; geniş saldırı alanı, yüzlerce saldırı vektörü, nitelikli profesyonel eksikliği, çok fazla veri yığınlarıdır (Diakun, 2014). Bu zorlukların bir çoğunu öğrenen ve analiz eden, yapay zeka tabanlı siber güvenlik yönetim sistemi, bu zorlukların çoğunun üstesinden gelebilir. Makine öğrenimi ve yapay zeka sayesinde tehlike algılama otomatikleştirilebilir ve geleneksel sistemlerden, yazılımlardan daha iyi çözümler bulunabilir (Diakun, 2014).



Şekil 1.30. Siber Tehdit Türleri

Siber güvenlik tehdit türlerine (Şekil 2.10.) bakıldığında günümüzde her insanın karşılaşabileceği yöntemlerden biri E-dolandırıcılık yöntemidir (Cosmin, 2013). Bilinen ve yaygın olarak kullanılan kaynaklardan gelen e-postalara gibi sahte e-postalar gönderme uygulamasıdır. Burda amaç, kişisel veriler, kartı numaraları, şifre verileri gibi giriş bilgilerini çalmaktır. Kötü amaçlı e-postaları filtreleyen yapay zeka uygulamaları ile koruma sağlanabilir. Bir diğer siber güvenlik tehdidi fidye yazılımıdır. Kötü amaçlı bir yazılım olan fidye yazılımı istenen ödeme olana kadar kişisel verilere, dosyalara, bilgisayar sistemlerine ya da programlara erişimi engelleyerek parayı, değerli olanı gasp etmek için tasarlanmıştır (Popoola, 2017). İsteneni yapmak, kişisel verilerin geri getirilmesini, dosyaların kurtarılmasını veya sisteme tekrar uluşılacağını sağlamaz, garanti etmez. Sosyal mühendislik bir diğer siber güvenlik tehdidir ve hassas bilgilerin ifşa edilmesi için insanları kandırmada kullanılan bir taktiktir. Para veya gizli veriler istenebilir. Günümüzde bu şekilde birçok siber güvenlik tehdidi mevcuttur. Ve bu eylemler (bir sisteme izinsiz girme, verileri ele geçirme, sistem işleyişini bozma, veri değiştirme) siber suçlar kapsamında değerlendirilir.

Siber güvenlik çözümlerinde kullanılan yapay zeka, derin öğrenme ve veri madenciliği uygulamaları zaralı yazılım tespiti, saldırı tespiti ve engellenmesi, ağ trafiği analizi, doğal dil işleme, kullanıcı davranış analitiğidir (Alberto, 2018). Ağ trafiği analizi, güvenlik tehditlerini tespit etmek ve bunlara yanıt vermek için ağ trafiği iletişim modellerini yakalama, kaydetme ve analiz etme sürecidir (Theyazn, 2020). Belirli bir zamanda bir ağda hareket eden veri incelenir. Doğal dil işleme (NLP), bir bilgisayar programının insan dilini konuşulduğu ve yazıldığı şekliyle anlama yeteneğidir (Reshamwala, 2013). Yapay zekanın bir bileşenidir ve kullanılan örnekleri istenmeyen e-posta algılama, sosyal medya duyarlılık analizidir. Kullanıcı davranışı analitiği (UBA) olarak da bilinen kullanıcı ve varlık davranışı analitiği (UEBA), kullanıcıların her gün oluşturduğu ağ olaylarına ilişkin içgörü toplama sürecidir (Callara,2018). Ağ olaylarına ilişkin veriler oplanıp analiz edildikten sonra, güvenliği ihlalleri, alınan kimlik bilgileri, hareketin ve diğer kötü niyetli davranışların kullanımı tespit edilebilir

veya tespit etmek için kullanılabilir. UBA, makine öğrenimi ve analitiği kullanarak, kurumsal ortamlarda dolaşırken tehdit aktörlerinin davranışlarını tanımlar ve takip eder, kullanıcı normlarından sapan eylemleri tespit etmek için verileri bir dizi algoritma aracılığıyla çalıştırır (Callara, 2018). UBA, kimlik bilgileri hırsızlığı, dolandırıcılık ve diğer kötü niyetli faaliyetleri gösterebilecek şüpheli kalıpları tespit etmek için değerli bir araçtır.

1.14. Suç Soruşturmaları ve Güvenlik Alanlarında Veri Madenciliği, Makine Öğrenimi ve Yapay Zeka'nın Kullanımının Çekinceleri

Suç soruşturmalarında kullanılacak veri madenciliği, makine öğrenimi ve yapay zeka uygulamaları üzerinde çalışılırken günümüz uygulamalarında yaşanan sorunları da göz önünde bulundurmak gerekir. Suç soruşturmalarında kullanılan günümüz veri madenciliği, makine öğrenimi ve yapay zeka uygulamalarına bakıldığında bazı etik ve hukuki sorunların olduğu görülmektedir. Bunlardan ilki gizlilik ve gözetimdir (Petković, 2007). Bilgi teknolojisinde mahremiyet ve gözetim hakkında, esas olarak özel verilere ve kişisel olarak tanımlanabilir verilere erişimle ilgili genel bir tartışma vardır (Macnish 2017; Roessler 2017). Mahremiyetin kendi başına bırakılma hakkı, bilgi gizliliği, kişiliğin bir yönü olarak mahremiyet, kendi hakkındaki bilgiler üzerinde kontrol ve gizlilik hakkı gibi bilinen birkaç yönü vardır (Bennett ve Raab 2006). Gizlilik çalışmaları, tarihsel olarak gizli servisler tarafından devlet gözetimine odaklanmıştır, ancak şimdi diğer devlet görevlileri, işletmeler ve hatta bireyler tarafından yapılan gözetimi de içermektedir (Müller,2020). Bununla birlikte günümüzde artık yaşantılarımız giderek dijitalleşmekte ve neredeyse tüm veri toplama ve depolama artık dijital olmuştur. Dijital dünyada kimin hangi verileri topladığını ve kimin erişimi olduğunu kontrol etmek oldukça zordur. Birçok yeni yapay zeka teknolojisi ise bu sorunları artırır. Örneğin, fotoğraflarda ve videolarda yüz tanıma, bireylerin profillerinin çıkarılmasına ve sistem tarafından tanımlanmasına olanak tanır (Whittaker ve diğerleri. 2018: 15ff). Bu, cihaz parmak izi gibi tanımlama için kullanılan bilginin elde edilmesine olanak sağlar. Sonuç olarak kişisel bir bilgi izinsiz olarak elde edilmiş olur.

Bir diğer sorun manipülasyondur. Yapay zekanın gözetimdeki etik sorunlar, yalnızca veri birikiminin ve yönünün ötesinde, özerk rasyonel seçimi baltalayacak şekilde, çevrimiçi ve çevrimdışı davranışları manipüle etmek için bilginin kullanımını içerir (Müller, 2020). Kullanıcılar dürtmelere, manipülasyona ve aldatmaya karşı savunmasızdırlar. Yeterli ön veriyle, algoritmalar, yalnızca belirli bireyleri veya küçük grupları etkilemesi muhtemel girdi türüyle hedeflemek için kullanılabilir (Thaler ve Sunstein 2008). Geliştirilmiş yapay zeka teknolojileri, bir zamanlar güvenilir olan kanıtları güvenilmez kanıtlara dönüştürür. Örneğin; dijital fotoğraflara, ses kayıtlarına ve videolara uygulanarak verileri bozabilir. Yapay zekada ki makine öğrenimi teknikleri çok büyük miktarda veriyle eğitime dayanır (Müller, 2020). Eğitim verisinde manipilasyon sonuçları etkileyerek doğru sonuçlar elde edilmesinin önüne geçecektir.

Yapay zeka sistemlerinde opaklık (donuk, şeffaf olmayan) ve yanlılık günümüzde bazen “veri etiği” veya “büyük veri etiği” olarak adlandırılan temel konulardır (Floridi ve Taddeo 2016; Mittelstadt ve Floridi, 2016). Otomatik karar desteği ve öngörüye dayalı analitik için yapay zeka sistemleri; gerekli süreç, hesap verebilirlik, topluluk katılımı ve denetim eksikliği hakkında önemli endişeler uyandırmaktadır (Whittaker, 2018). Bunlar, insan katılımı için fırsatları kısıtlayan ve sınırlayan karar verme süreci yaratan bir güç yapısının parçasıdır (Danaher, 2016). Aynı zamanda, etkilenen kişinin sistemin bu çıktıya nasıl geldiğini bilmesi çoğu zaman imkansız olacaktır, yani sistem o kişiye “opaktır” (Müller, 2020). Sistem makine öğrenimi içeriyorsa, belirli bir kalıbın nasıl tanımlandığını bilmeyen uzman için bile genellikle anlaşılmaz olacaktır. Karar sistemlerindeki ve veri kümelerindeki yanlılık, bu opaklık tarafından şiddetlenir. Veriler bir önyargı içeriyorsa (örneğin, şüphelilerin ten rengiyle ilgili polis verileri), program bu önyargıyı tekrarlı üretecektir. Buolamwini ve Gebru'nun 2018 araştırmasına göre, bazı yüz analizi algoritmalarının siyah kadınları uygulandığı zamanın yaklaşık %35'inde yanlış sınıflandırdığı ve beyaz erkekler için neredeyse her zaman doğru yaptığı sonucuna varmıştır (Kade, 2020).

Otomatikleştirilmiş AI karar destek sistemleri tahmini analitik veriler üzerinde çalışır ve "çıktı" olarak bir karar üretir. Tahmine dayalı polislik ile ilgili çoğu kişinin korkuları kamu özgürlüklerinin aşınmasına yol açabileceğinden dolayıdır. (Ferguson, 2017). Çünkü davranışları tahmin edilen insanlardan gücü alınabilir. Örneğin bir bölgedeki kolluk kuvvetini artırarak ve o bölgede daha fazla suç oluşmasını önlemek aynı zamanda orda ki insanlar üzerinde de bir kontrol baskısı oluşturabilir. Gerçek "tahmini polislik" veya "istihbarat güdümlü polislik" teknikleri, esas olarak polis kuvvetlerine en çok nerede ve ne zaman ihtiyaç duyulacağı sorusuyla ilgilenir. Önyargı tipik olarak adil olmayan kararlar verildiğinde ortaya çıkar, çünkü yargıda bulunan kişi, eldeki konuyla gerçekten alakasız bir özellikten etkilenir; tipik olarak bir grubun üyeleri hakkında ayrımcı bir önyargıda bulunur (Müller, 2020). Yapay zeka sistemlerinin böyle bir bilişsel önyargıya sahip olup olmayacağı veya olması gerektiği konusunda sorunlar vardır.

Makine öğrenimi ve yapay zeka da yaşanan ve yaşanabilecek diğer sorunlardan yapay zekanın yarattığı şeylerden (mülkiyet aideti) ve hukuk ihlalinden kimin sorumlu olacağı; yapay zeka için kullanılan enerjinin çevresel etkisi ve yapay zekanın kontrolü ele alabileceği gibi temel etik sorunlar yapay zeka ve makine öğrenimi konusunda çalışırken göz önünde bulunulmalıdır.

1.15. Suç Soruşturmalarında Yapay Zeka Uygulamalarının Geleceği

Suç olaylarında en önemli verilerden biri olayın meydana geldiği anın kamera görüntüleridir. Soruşturma aşamasında kolluk kuvvetleri olayla ilgili kamera kayıtlarını toplar ve adli makamlara sunmak üzere çözümlemesini yapar. Çözüm aşamasında karşılaşılan zorluklardan bir tanesi olayın meydana geldiği saatin belirlenmesidir. Örneğin ateşli silahla meydana gelen

mala zarar verme olayında olayın ne zaman gerçekleştiği kestirilemeyebilir. Bu durum saatlerce kamera izlenmesine ve kamera kaydı toplanmasına neden olur. Makine öğrenmesi ve yapay zeka kullanılarak sesli kayıt yapan kameralar için silah sesinden ya da namludan çıkan alevden olay anının tespiti mümkün olabilir.

Kamera konusuna değinmişken; teknolojinin gelişmesiyle kameralara ulaşım oldukça artmış, yaşam alanlarımızın neredeyse her yerine girmiştir. İkametlerin dışında, içinde, araçlarımızda kameralar mevcuttur ve herşeyi kaydederler. Bunun nedeni insanların güvenlik kaygılarıdır. Güvenlik endişesi içinden bulunan insanlar birçok güvenlik tedbirine başvurabilir. Teknolojinin gelişmesiyle gelişen kontakt lens teknolojisi sayesinde görüntü alabilen ve aktaran kontakt lensler yapılabilir.

Olayların çözümü için bazen elde hiçbir veri olmayıp, olay ile ilgili olay yerinde bir veya birden fazla şüpheli olarak değerlendirilenler şahıslar olabilir. Bu tarz durumlarda şüpheli ya da şüpheliler arasında ki çelişkileri bulmak, davranış analizi yapma çok önemlidir. Suç soruşturmalarıda çelişkilerden yararlanarak olayla ilgili veri elde edilse de bunu teknik olarak yapmak delil niteliği kazanması yönünden önemlidir. Yapay zeka ve makine öğrenmesinden yararlanılarak elde bulunan şüphelilerin sorgu öncesi, sırası ve sonrasında hareket analizi yapılabilir, aralarındaki çelişkiler bulunabilir.

Çin'in uygulamış olduğu yüz tanıma sistemi ve bazı ülkelerin uyguladığı elektronik bileklik sisteminden yola çıkılarak; gelişen çip teknolojisi kullanılarak şiddet uygulayan ya da her insana çip takılabilir. Bu çip sayesinde konum bilgilerine ulaşılarak şüphelinin mağdura yaklaşması önlenebilir, elektronik kelepçenin çıkarılarak suç işlenmesinin önüne geçilebilir. Ayrıca fizyolojik bilgilerden yola çıkılarak suç işleyen kişinin vücudunda yaşanan değişimlerden suçlu tespit edilebilir.

Tüm ülkelerde akıllı ve anında göresel analizler yapabilen CCTV izleme sistemleri yaygınlaştırılarak eş zamanlı analiz kabiliyetleri sayesinde şüpheli insan davranışları tespit edilebilir.

Yüksek çözünürlüklü kamera sistemlerinin sayıları arttırılarak makine öğrenimi ile suç işleyen ya da işleyecek insanların biyolojik, fizyolojik özellikleri (vucut ısısı, göz bebeği büyümesi) tespitler edilebilir.

Dolandırıcılık ve siber suçlarla ilgili daha önce meydana gelen olayların analizi yapılarak, makine öğrenimi ve yapay zeka sayesinde şüpheli veriler ve hareketler tespit edilebilir, erken uyarı sistemi geliştirilebilir. Örneğin; telefonlara yüklenen bir program ile konuşma içerisinden anahtar kelimeler seçilerek kelimeler yapay zeka tarafından değerlendirilir ve hedef şahıs uyarılabilir.

Yapay zeka sayesinde bankacılık sistem rutinleri öğrenilip farklı bir durumda erken uyarı sistemi geliştirilebilir.

Olay yeri çalışması robotlar tarafından yapılarak, toplanan verilerle olayın simülasyonu hazırlanıp anlık olay yerindeki kolluk kuvvetine aktarılabilir ve olayın hızlı birşekilde değerlendirilmesi ve çözülmesi sağlanabilir.

Yapay zekaya sahip asistanlar geliştirilerek suç soruşturmalarında insan doğasından kaynaklı hatalar (duygusal düşünme, yorgunluk, stres vb.) minumuma indirilebilir.

Geçmiş veriler ve gelişen durum analiz edilerek olaylar meydana gelmeden similasyon çalışmalarıyla öngörülüp, olaylar olmadan önlenmeye çalışılabilir.

Yapay zeka aracılığıyla, yasa dışı faaliyetler, suç örgütlerinin faaliyetleri, hedeflerindeki, iş stratejilerindeki ve hiyerarşilerindeki değişiklikleri bulmak ve ortaya çıkarmak için makine öğrenimi, AI teknolojisinden yararlanılabilir.

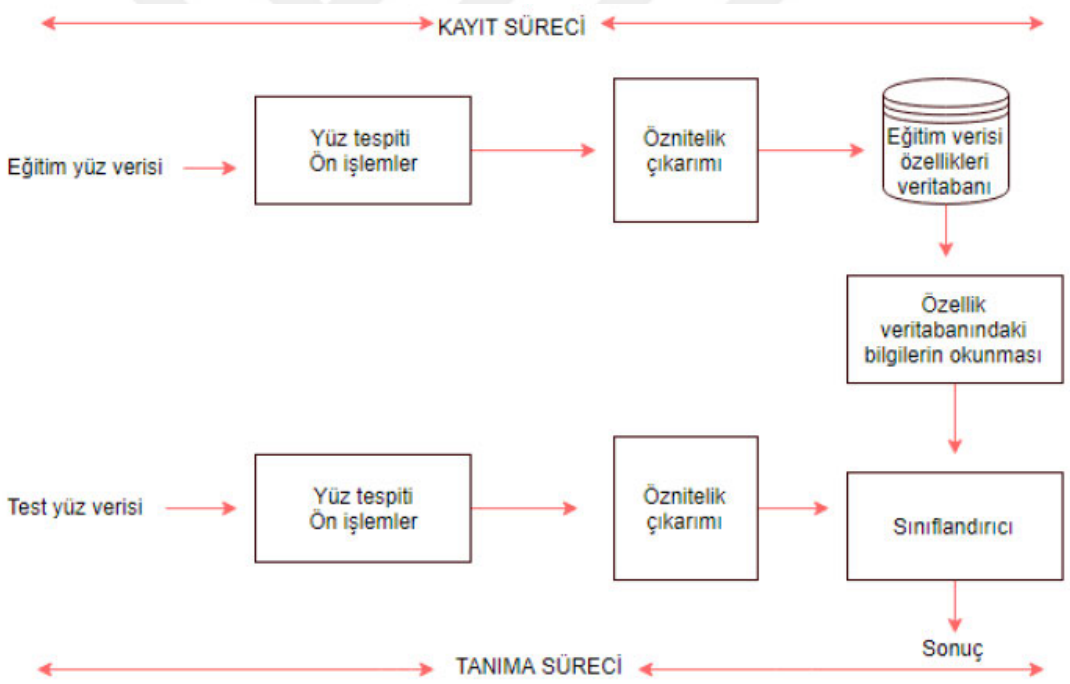
2. BÖLÜM

MATERYAL VE YÖNTEM

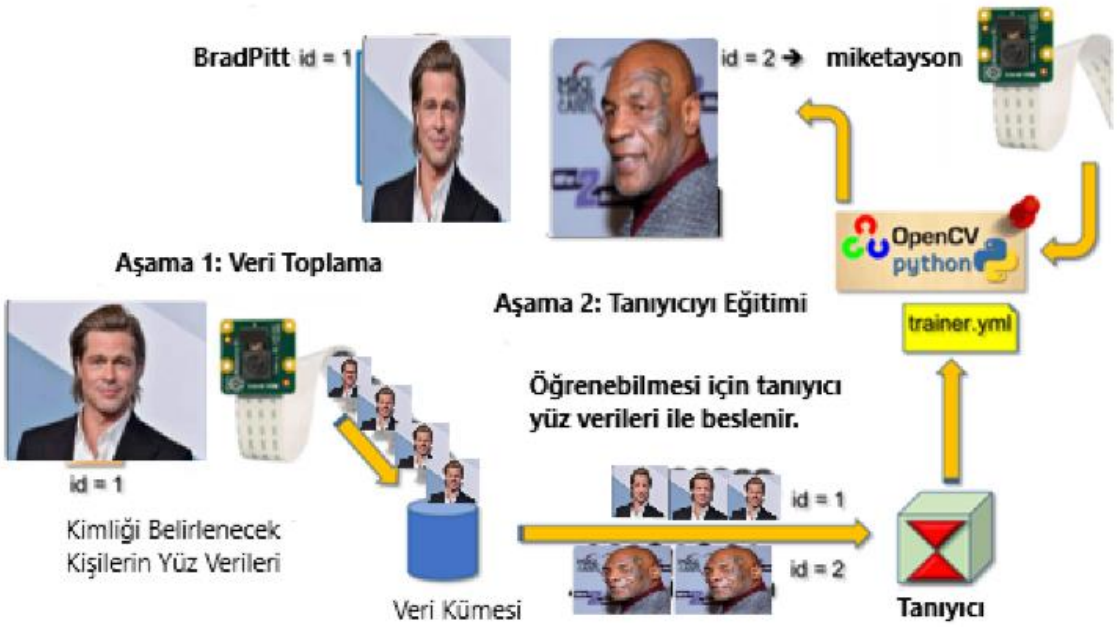
Bu bölümde; yüz tanıma çalışmamızın genel mimarisi, tez kapsamında geliştirilen algoritması detaylı bir biçimde anlatılmıştır.

2.1. Yüz Tanımada Temel Konsept

Bir yüz tanıma sistemi; kayıt, eğitim ve tanıma aşaması olmak üzere üç bölümden oluşur. Kayıt sürecinde; sistem kamera karşısında gerçek zamanlı yüz görüntüsünü alır ve veri tabanı oluşturur. Oluşturulan veri tabanında ki görüntü ile sistem eğitilir, eğitilen veri ile kamera karşısında ki gerçek zamanlı veri tabanında saklanan yüz özellikleri ile karşılaştırılarak tanıma işlemi gerçekleştirilir. Şekil 2.1'de yüz tanıma algoritması işleyiş adımları ve Şekil 2.2'de yüz tanıma şeması gösterilmiştir.



Şekil 2.1 Yüz Tanıma Algoritması



Şekil 2.2 Yüz Tanıma Şeması

2.2 Python ve OpenCV ile Gerçek zamanlı Kimlik Tespiti Projesi

Bu çalışmada görüntü işleme kütüphanesi olan OpenCV ve Python programı İle bir makine öğrenimi olan Gerçek Zamanlı Yüz Algılama ve Tanıma programıyla kimlik tespiti uygulaması oluşturulmuştur.

Bu program ile bir fotoğraf, video veya bilgisayara bağlı bir kameradan kişilerin yüzlerini algılayan ve bilgisayarımızdaki veri tabanına kaydeden, sonrasında ise bu görüntü ile veri tabanındaki yüz görüntüleri karşılaştırılarak kişi kimlik tespiti yapılmaktadır. Böylece, klasik olarak yapılan kimlik tespit yöntemlerine alternatif olarak çevrimiçi çalışan, kontrolü yapılabilen, otomatik olarak kimlik tespiti yapabilen ve yüz görüntülerinden kimlik oluşturabilen kişi tanıma uygulaması geliştirilmiştir.

Bu projede, kimlik tespiti yapay zeka tarafından yapıldığından zamandan tasarruf sağlanacaktır. Ayrıca, belirli bir süre de insan yardımıyla yapılan kimlik tespitinden aynı süre de sayıca daha fazla kimlik tespiti yapılacak ve bu da bize yine zamandan tasarruf sağlayacaktır. Sürekli izleniyor olmak ve sistem tarafından otomatik olarak tanınmak suçluları kontrol altında tutmada ve suç işlemeden önce bir kez daha düşünmelerine neden olacaktır.

Biyometrik tanımlama yoluyla yüz görüntülerinden oluşturulan kimlik bilgisi sisteme eklendiğinde ve bir merkezi yönetim ile kişi girişleri yapıldığında kayıt dışılığın önüne geçilecektir. Yine bir merkezi yönetim ile veri tabanın da kimlik bilgisi olmayan veya eski verileri olan kişiler bu program ile tespit edilerek sisteme eklenmeleri, yüz görüntü verileri güncellenmesi sağlanacaktır.

2.3. Kullanılan Yazılımın Yapısı

Bu tez çalışmasında; python programlama dili kullanılarak, suç soruşturmalarında kullanılmak üzere sistemde tanımlanan kişi kimlik bilgileri ile web kameradan gelen yüz görüntüsü bilgilerini yöneten bir sistem geliştirilmiştir. Burada sisteme kişiyi yine bizim tanımlamamız ve bir veri kümesi hazırlamamız gerekecektir. Görüntü işleme kütüphanesi olarak Opencv kullanılmıştır. Kodlarımıı Çapraz platform bir python geliştirme ortamı olan PyCharm'da yazılacaktır.

Yüz algılama için hesaplama açısından ucuz, hızlı ve yüksek doğruluk oranına sahip olan popüler algoritmalardan biri "haarcascade"dir (Geek, 2021). Haar karanlık bölgelerin şiddetlerinin toplamı ile aydınlık bölgelerin şiddetlerinin toplamının farkını alarak ve bütünsel görüntüdeki belirli bir piksellerin karşılaştırmasını yaparak nesneyi tanımlar. Bu XML dosyaları, cv2 modülünün cascadeClassifier yöntemiyle yüklenebilir. Burada yüzleri tespit etmek için haarcascade_frontalface_default.xml kullanılması tercih edilmiştir.

Package	Version	Latest version
Pillow	8.4.0	▲ 9.5.0
Pillow-PIL	0.1.dev0	
click	8.0.4	▲ 8.1.3
colorama	0.4.5	▲ 0.4.6
dlib	19.24.0	▲ 19.24.2
face-recognition	1.3.0	1.3.0
face-recognition-models	0.3.0	
importlib-metadata	4.8.3	▲ 6.6.0
imutils	0.5.4	0.5.4
numpy	1.19.5	▲ 1.24.3
opencv-contrib-python	4.6.0.66	▲ 4.7.0.72
opencv-python	4.6.0.66	▲ 4.7.0.72
pip	21.3.1	▲ 23.1.2
setuptools	59.6.0	▲ 67.8.0
typing-extensions	4.1.1	▲ 4.6.3
wheel	0.37.1	▲ 0.40.0
zipp	3.6.0	▲ 3.15.0

Resim 2.1 PyCharm Kütüphaneler

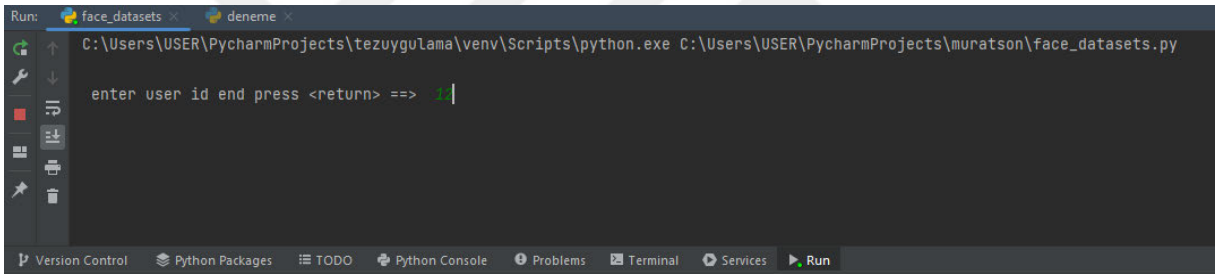
Python ve PyCharm kurulduktan sonra PyCharm'da Resim 2.1'de görünen ve programın sağlıklı çalışmasını sağlayacak kütüphaneler yüklenildi ve programın ilk kısmı olan veri kümemizi oluşturma kısmına geçildi.

2.4. Yüz Tanıma ile Kişi Tespit Programı Veri Kümesi Ara yüzü

Yapılan çalışmada bir fotoğraf, video veya bilgisayara bağlı bir kameradan kişilerin yüzlerini algılanarak bilgisayarda ki veri tabanına kaydedildiği ara yüzdür. Bu ara yüzde öncelikle kamera tanımlanarak aktif hale getirildi. Yüz tespiti için open-cv kütüphanesiyle cascade sınıflandırıcısı bir değişkene atadık. Tanımlanan her bir yüz için farklı rakamlarla ilişkilendirdi. Arayüzü her çalıştırıldığında yüz için atanacak sayıyının girilmesi istenecektir. Sayı girildikten sonra sayıya karşılık veri klasörü içine kişi yüz verisini kaydedecektir. Burada dikkat edilmesi gereken kısımlar; "cv2.waitKey(20" kod satırında '20' olarak belirlenen sayı elde edilen fotoğraf kalitesidir. Bu sayı azaltılarak elde edilecek görüntü kalitesi iyileştirilebilir.

Bir diğer dikkat edilmesi gereken nokta "elif count> 30" kod satırıdır. Burada veri kümesine kaç adet yüz verisi alınacağı gösterilir. Uygulanan progrmda 30 adet yüz verisi alınmış ve program sonlandırılarak her bir kişi için veri kümesi elde edilmiştir.

Programımızı çalıştırdığımızda bizden veri tabanına kaydedeceğimiz kullanıcı kimliğini girmemizi isteyecek ve biz her kullanıcı kimliği için bir numara tanımlanacaktır. Örneğin sisteme Resim 2.2 'de ki gibi ilk kullanıcıyı tanımlammak için '1' komutu girilmiştir.



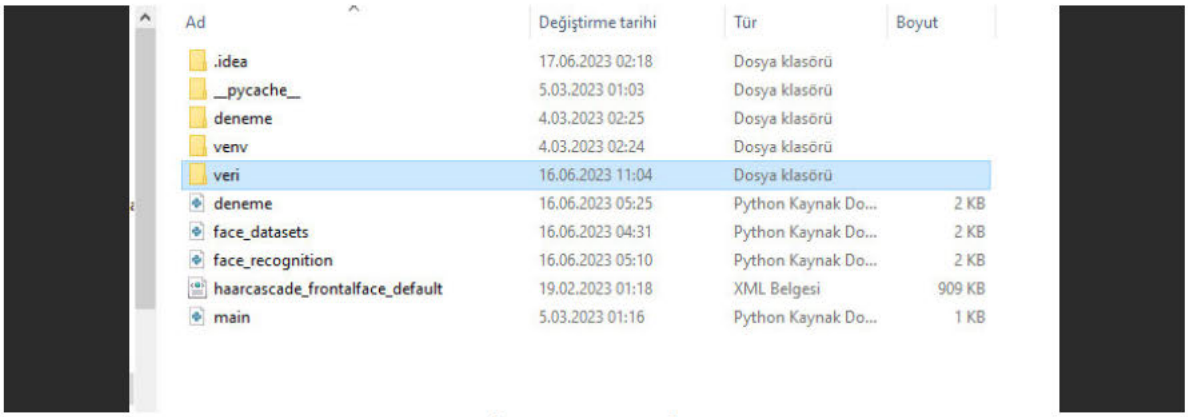
Resim 2.2. PyCharm Komut Giriş Satırı

Program kamerayı açarak karşısındaki yakalanan görüntüdeki yüzü bulunarak yeşil kenarlı dikdörtgen içine almakta ve veri kümesini oluşturmaktadır. Program belirlenen kadar yüz verisi aldıktan sonra kamerayı kapatacaktır. Bu işlemleri yapan kodlar Resim 2.3'de verilmiştir.

```
import cv2
import os

vid_cam = cv2.VideoCapture(0)
# Kamera bir değişkene tanımlanır ve aktif edilir

face_detector = cv2.CascadeClassifier('haarcascade_frontalface_default.xml')
# Yüz tespiti için open-cv kütüphanesiyle cascade sınıflandırıcısı bir değişkene atanır
# Tanımlanan yüzler ayırmak için farklı rakamlar atandı.
face_id = input('\n enter user id end press <return> ==>  ')

print("\n [INFO] Initializing face capture. Look the camera and wait ...")
# yüz sayımı başlatılır
count = 0
```

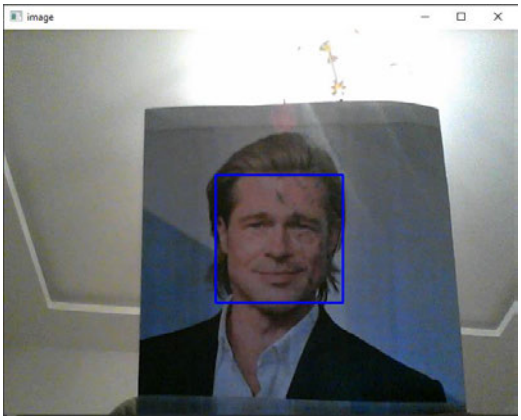
```
while(True): # While döngüsü oluşturuldu

    # kamera okutuldu
    ret, img = vid_cam.read()

    # Başlangıçta görüntü üç katmanlı bir görüntüdür. Bu nedenle tek katmanlı bir görüntüye
(yani gri tonlamalı) dönüştürülür.
    gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)

    # resim için alt ve üst sınırlar belirler
    faces = face_detector.detectMultiScale(gray, 1.3, 5)



    for (x,y,w,h) in faces: #for döngüsünde çerçeve ebatları için değişkenler belirlenir
# çerçeve rengi ve kalınlığı belirlendi
        cv2.rectangle(img, (x,y), (x+w,y+h), (255,0,0), 2)
        count += 1
# fotoğraf adet artışı tanımlandı.
        # resimler veri klasörüne yazdırıldı
        cv2.imwrite("veri/User." + str(face_id) + '.' + str(count) + ".jpg", gray[y:y+h,x:x+w])

        cv2.imshow('image', img) #kameraya göster komutu atandı
# fotoğraf kalitesi belirlendi ve çıkış tuşu atandı
    if cv2.waitKey(20) & 0xff == ord('q'):
            break
    elif count> 30:  # kameranın çekeceği fotoğraf sayısı sınırlandırıldı
            break


print("\n [INFO] Exiting Program and cleanup stuff")
vid_cam.release() # kamera durduruldu
cv2.destroyAllWindows() # tüm pencereler kapatıldı
```

Resim 2.3. PyCharm Komut Giriş Satırı-1



Resim 2.4. Veri Klasörü

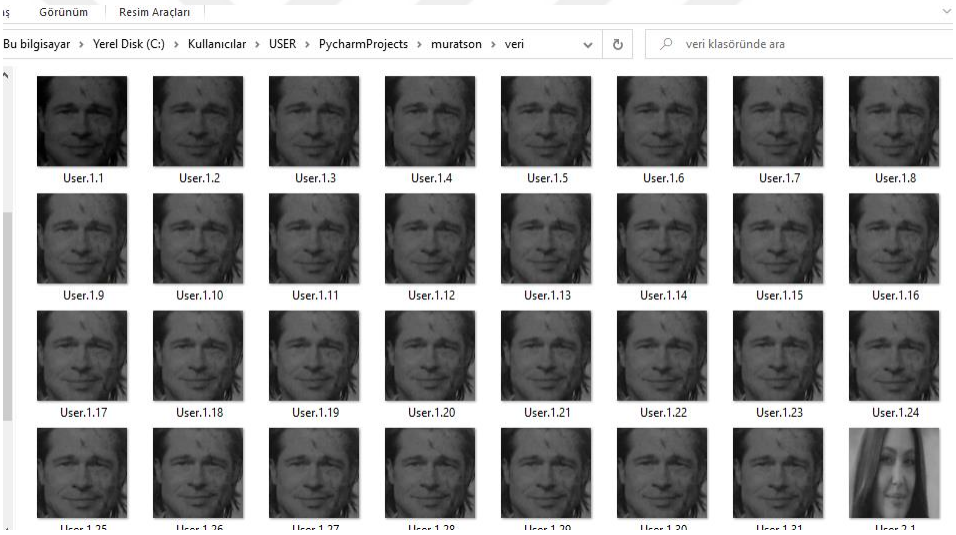
Kamera karşısından elde edilen veriler Resim 2.4'de gösterilen klasöre kaydedilir.



Resim 2.5. Orjinal Resim Brad_PİTT

Programımız çalıştırılarak kamera karşısındaki yüzü tespit ettiği görüntü Resim 2.5'de verilmiş, diğer yüzlere ait tespit görüntüleri de aşağıda sırayla verilecektir.

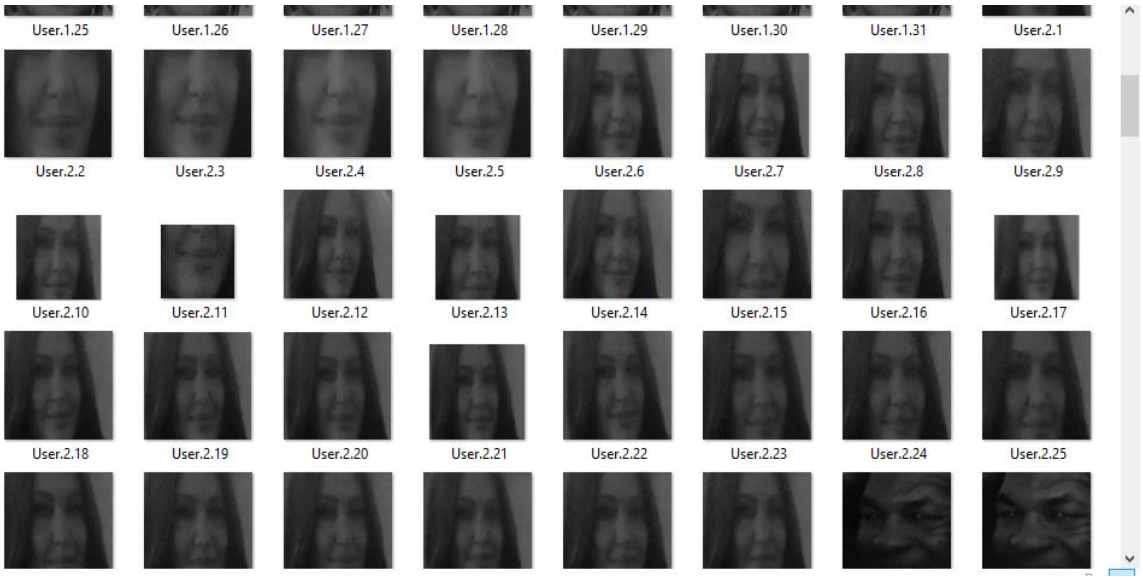
Yukarda belirttiğimiz ve Resim 2.6'da göründüğü gibi her bir yüz için otuz adet fotoğraf alarak veri kümesi oluşturuldu.



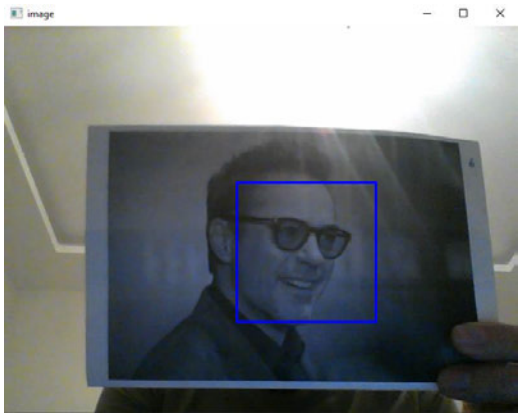
Resim 2.6. Gri Renge Çevrilmiş Resim Brad_PİTT

Her bir kişi için bu uygulamayı tekrarlayarak veri kümemizi oluşturuldu. Yapılan çalışmada için kişi sayısı 10 ile sınırlı tutulmuştur.. Ve her kişi için aynı işlemler yapıldı.

Resim 2.7. Orjinal Resim Angelina_JOLİE

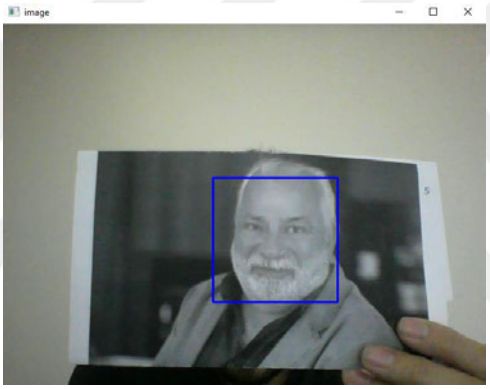


Resim 2.8. Gri Renge Çevrilmiş Resim Angelina_JOLİE

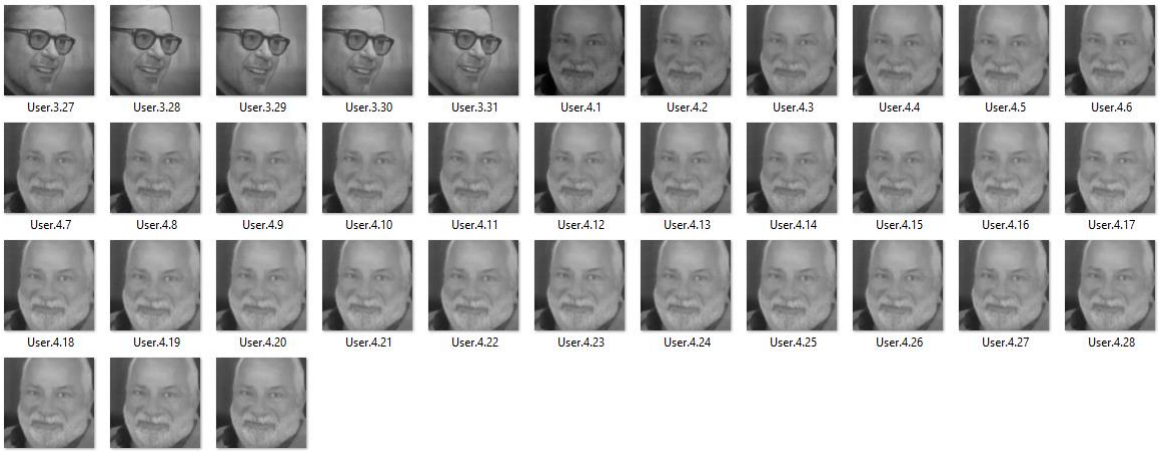


Resim 2.9. Orjinal Resim Anthony Edward STARK

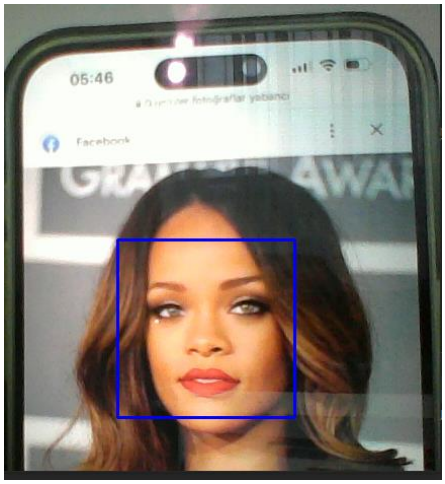
Resim 2.10. Gri Renge Çevrilmiş Resim Anthony Edward STARK



Resim 2.11. Orjinal Resim Zeki ALASYA

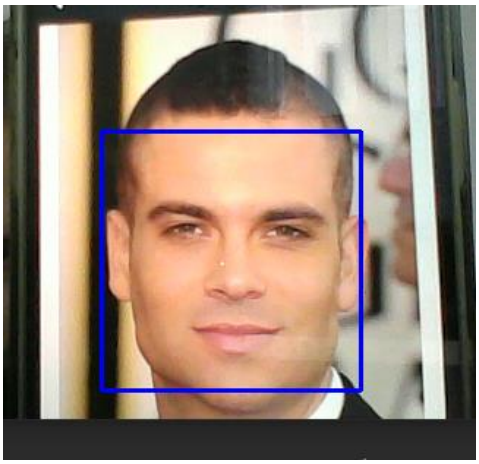


Resim 2.12. Gri Renge Çevrilmiş Resim Zeki ALASYA

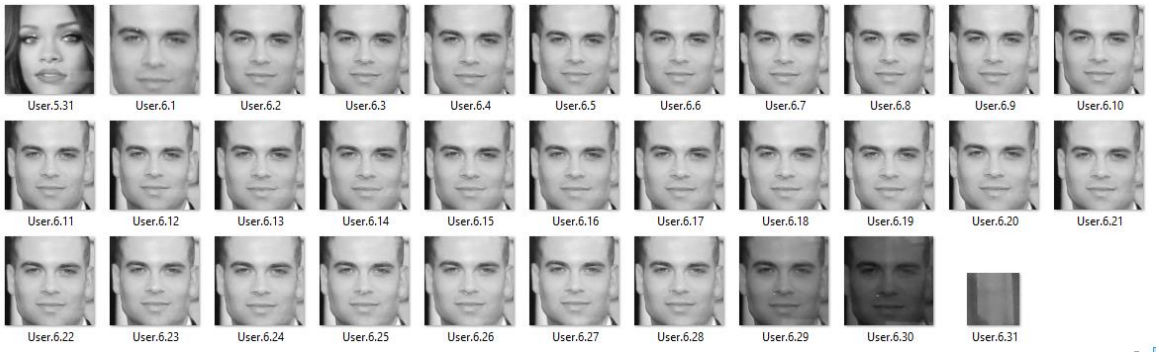


Resim 2.13. Orjinal Resim Rihanna

Resim 2.14. Gri Renge Çevrilmiş Resim Rihanna

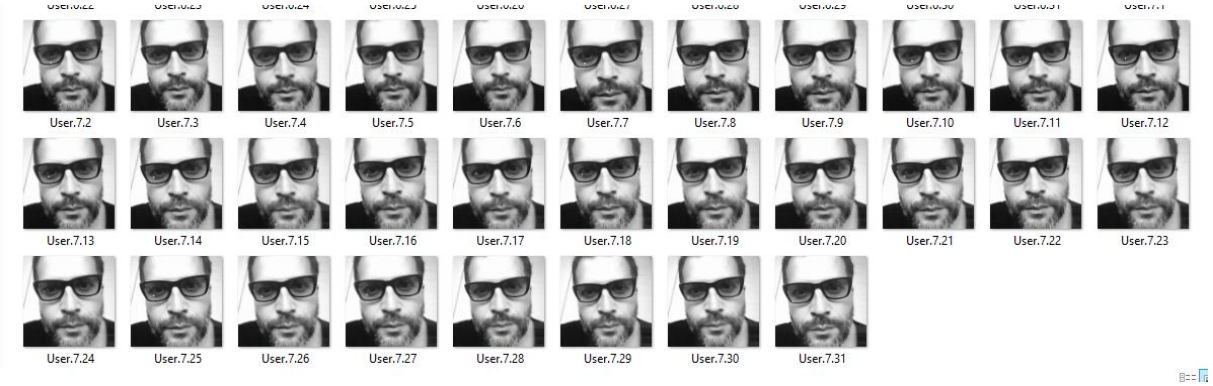


Resim 2.15. Orjinal Resim Mark SALLİNG

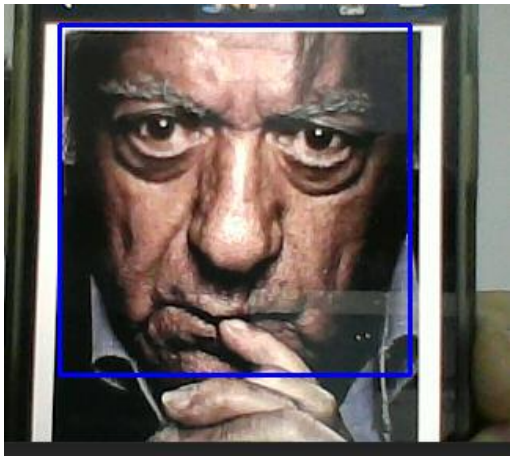


Resim 2.16. Gri Renge Çevrilmiş Resim Mark SALLİNG

Resim 2.17. Orjinal Resim Paul John Vasquez

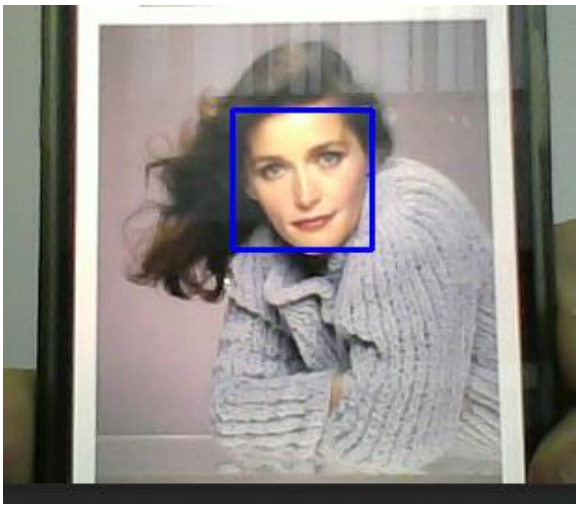


Resim 2.18. Gri Renge Çevrilmiş Resim Paul John Vasquez

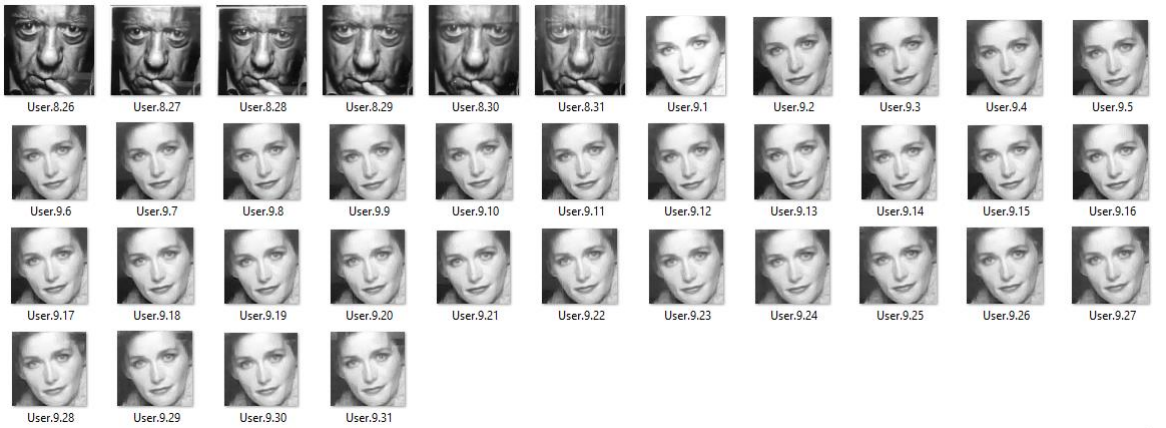


Resim 2.19. Orjinal Resim Ezzatolah Entezami

Resim 2.20. Gri Renge Çevrilmiş Resim Ezzatolah Entezami

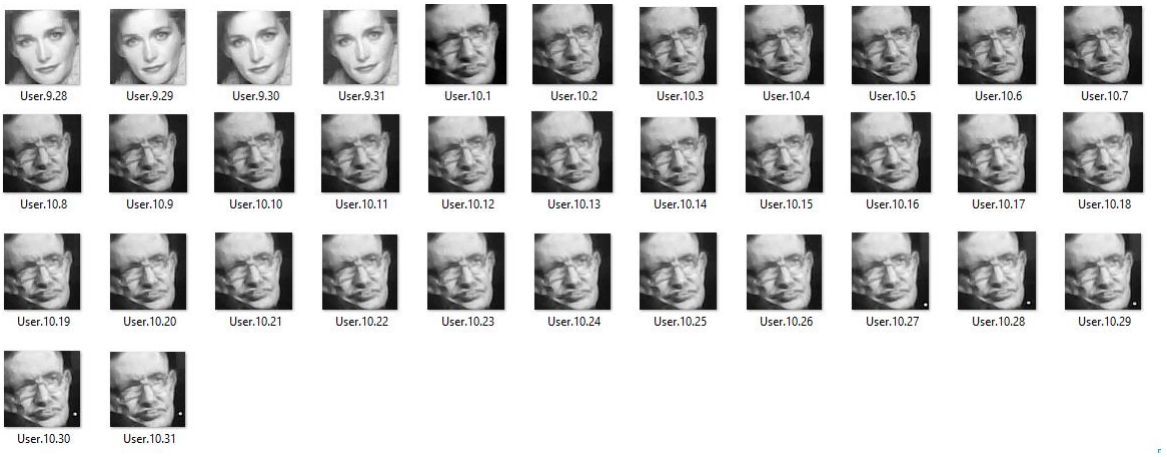


Resim 2.21. Orjinal Resim Margot KİDDER



Resim 2.22. Gri Renge Çevrilmiş Resim Margot KİDDER

Resim 2.23. Orjinal Resim Stephen HAWKİNG



Resim 2.24. Gri Renge Çevrilmiş Resim Stephen HAWKİNG

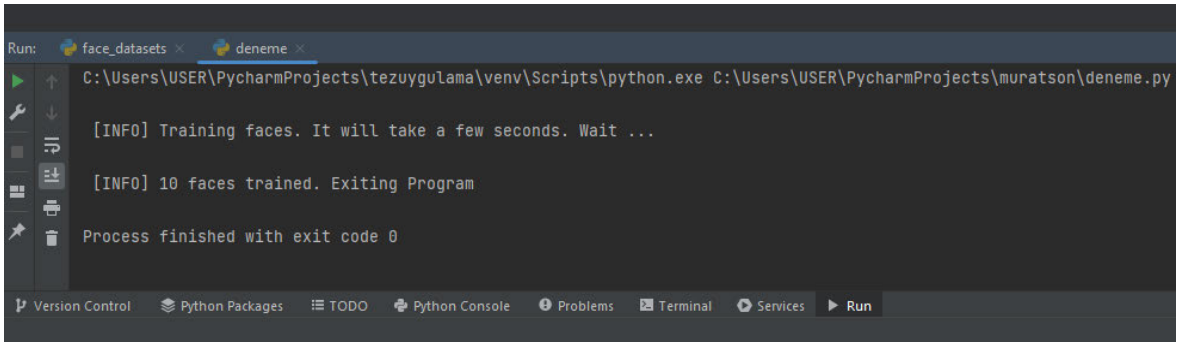
2.5. Yüz Tanıma ile Kişi tespit Programı Eğitim Ara yüzü

Yüz tanımanın algılaması yani öğrenmesi için deneme arayüzü olarak Resim 2.25'de gösterilen kod python geliştirme ortamı olan PyCharm'da yazıldı. Burada yine ilk başta kütüphane aktif hale getirilmiştir. Görüntüler ve tanımlar için yol atadık ve resimler için bir döngü oluşturuldu.

İsmin yazılacağı yer, yüz verisi için değişkenler vb. tanımlanarak yüzler eğitildi ve kaç tane yüz eğitildiği Resim 2.26'da gösterildiği gibi çıktı olarak verildi.

```
import cv2
import numpy as np
from PIL import Image
import os

#
path = 'veri'

recognizer = cv2.face.LBPHFaceRecognizer create() #yüz tanıyıcı
oluşturuldu
detector = cv2.CascadeClassifier("haarcascade_frontalface_default.xml");
#sınıflandırıcı xml dosyası eklendi

# görüntüler ve tanımlar için yol atandı
def getImagesAndLabels(path):
# resim yolunu bulmak için döngü oluşturuldu
    imagePaths = [os.path.join(path,f) for f in os.listdir(path)]
    faceSamples=[] #değişken tanımlandı
    ids = [] #değişken tanımlandı

    for imagePath in imagePaths: # for döngüsü oluşturuldu
# resmin okuyacağı dosyayı açmak için pıl img den yararlanıldı
        PIL_img = Image.open(imagePath).convert('L') # convert it to
grayscale
        img numpy = np.array(PIL img,'uint8') # pozitif tam sayı olduğu
belirtildi

        id = int(os.path.split(imagePath)[-1].split(".")[1]) # ismin
yazılacağı yer belirtildi
        faces = detector.detectMultiScale(img numpy) # yüz ölçeklerinin
belirleneceği kod satırı atandı

        for (x,y,w,h) in faces: # yüz çerçevesi için değişkenler for
döngüsüne eklendi
            faceSamples.append(img numpy[y:y+h,x:x+w])
            ids.append(id)

    return faceSamples,ids

print ("\n [INFO] Training faces. It will take a few seconds. Wait ...")
faces,ids = getImagesAndLabels(path)
recognizer.train(faces, np.array(ids))

# model trainer/trainer.yml dosyasına kaydedildi
recognizer.save('deneme/deneme.yml')

# eğitilen yüzlerin sayısı yazdırılarak program sonlandırıldı
print("\n [INFO] {0} faces trained. Exiting
Program".format(len(np.unique(ids))))
```

Resim 2.25. PyCharm Komut Giriş Satırı-2



Resim 2.26. PyCharm Program Çıktısı

Şekilde kaç adet yüz öğrenidiği bize çıktı olarak on adet yüzün eğitildiği bir çıktı alındı.

2.6. Yüz Tanıma ile Kişi Tespit Programı Yüz Tanıma Ara yüzü

Yüz tanıma kod satırlarında daha önceden olduğu gibi gerekli kütüphanelerin eklemesi yapıldı. Recognizer.read('deneme/deneme.yml') kod satırı ile okunacak dosya belirtildi. Çünkü bir önceki kod satırlarında makine eğitilmiştir. Kamera ve yazı, çerçeve boyutu, rengi, kişi ismi, çıkış vb. tanımları yapılmış ve programımızı yüz tanımaya hazır hale getirilmiştir.

```
if (confidence < 100):
    id = names[id]
    confidence = "  {0}%".format(round(100 - confidence)) #hata payı
hesaplandı
else:
    id = "unknown"
    confidence = "  {0}%".format(round(100 - confidence))
```

Resim 2.27. PyCharm Program Tahmin Hesaplama

Bu ara yüzde önemli olan kısım yüz tanıma hata payını hesaplamaktı. Bu da Resim 2.27'de ki şekilde görünen kod satırı ile hesaplandı. Yüz tanıma işlemini yapan kodlar Resim 2.28'de verilmiştir.

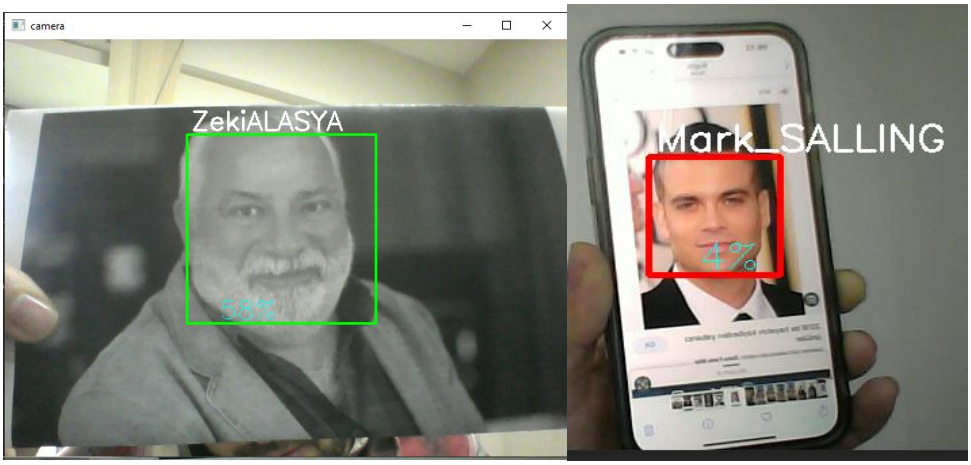
```
import cv2
import numpy as np
import os

recognizer = cv2.face.LBPHFaceRecognizer create() #yüz tanıyıcı
oluşturuldu
# okunacak dosya belirtildi
recognizer.read('deneme/deneme.yml')
cascadePath = "haarcascade frontalface default.xml" #sınıflandırıcı xml
dosyası eklendi
faceCascade = cv2.CascadeClassifier(cascadePath); #kullanılacak yol
atandı
font = cv2.FONT_HERSHEY_SIMPLEX #yazı tipi belirlendi
# kimlik sayacı başladı
```

```
id = 0
# kimlik isimleri verildi
names =
["YOK",'BradPITT','Angelina','Anthony Edward STARK','ZekiALASYA','Rihanna
','Mark SALLING','Paul John Vasquez','Ezzatolah Entezami','Margot KİDDER'
,'Stephen HAWKİNG']
# gerçek zamanlı kamera başlatıldı
cam = cv2.VideoCapture(0)
cam.set(3, 640)  # görüntü genişliği ayarla
cam.set(4, 480)  # görüntü yüksekliği ayarla
# minimum pencere boyutunu ayarla
minW = 0.1 * cam.get(3)
minH = 0.1 * cam.get(4)
while True:
    ret, img = cam.read() # kamera okutuldu
    img = cv2.flip(img, 1)
    gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)

    faces = faceCascade.detectMultiScale(
        gray,
        scaleFactor=1.2,
        minNeighbors=5,
        minSize=(int(minW), int(minH)),
    )
    for (x, y, w, h) in faces:
        cv2.rectangle(img, (x, y), (x + w, y + h), (0, 0, 255), 3) #
çerçeve ebatları ayarlandı
        id, confidence = recognizer.predict(gray[y:y + h, x:x + w])
        # eşleşme yüzdesi tanımlandı
        if (confidence < 100):
            id = names[id]
            confidence = "  {0}%".format(round(100 - confidence)) #hata
payı hesaplandı
        else:
            id = "unknown"
            confidence = "  {0}%".format(round(100 - confidence))

        cv2.putText(img, str(id), (x + 5, y - 5), font, 1, (255, 255,
255), 2)
        cv2.putText(img, str(confidence), (x + 5, y + h - 5), font, 1,
(255, 255, 0), 1)

    cv2.imshow('camera', img)
    # çıkış komutu tanımlandı
    if cv2.waitKey(10) & 0xFF == ord('q'):
        break

print("\n [INFO] Exiting Program and cleanup stuff")
cam.release()
cv2.destroyAllWindows()
```

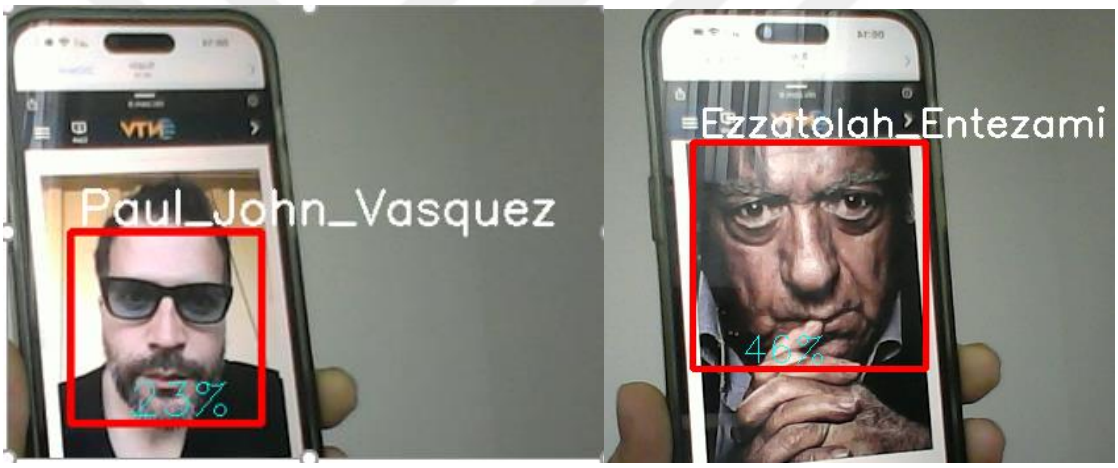
Resim 2.28. PyCharm Komut Giriş Satırı-3

Programımı çalıştırdığımızda programımızı bize kameradan kişilerin yüzlerini algılanarak veri tabanında ki yüzlerle karşılatırdıktan sonra kişiyi tanıdıysa kişinin bilgilerini verecek tanımadıysa tanımadığı çıktısı alınacak. Bu kimlik bilgileri bizim sisteme ne kadar ver verdiğimizle ilgili olarak değişecektir. Burda kişi isim soy isim bilgilerini tanımlanacak ve sadece isim soy isim bilgileri olarak çıktı alınacaktır. Bunun haricinde sistem sağlıklı bir veri sistemine entegre edilirse örneğin bir devlet kurumu veri tabanına bize yüz tespiti yapılan kişi

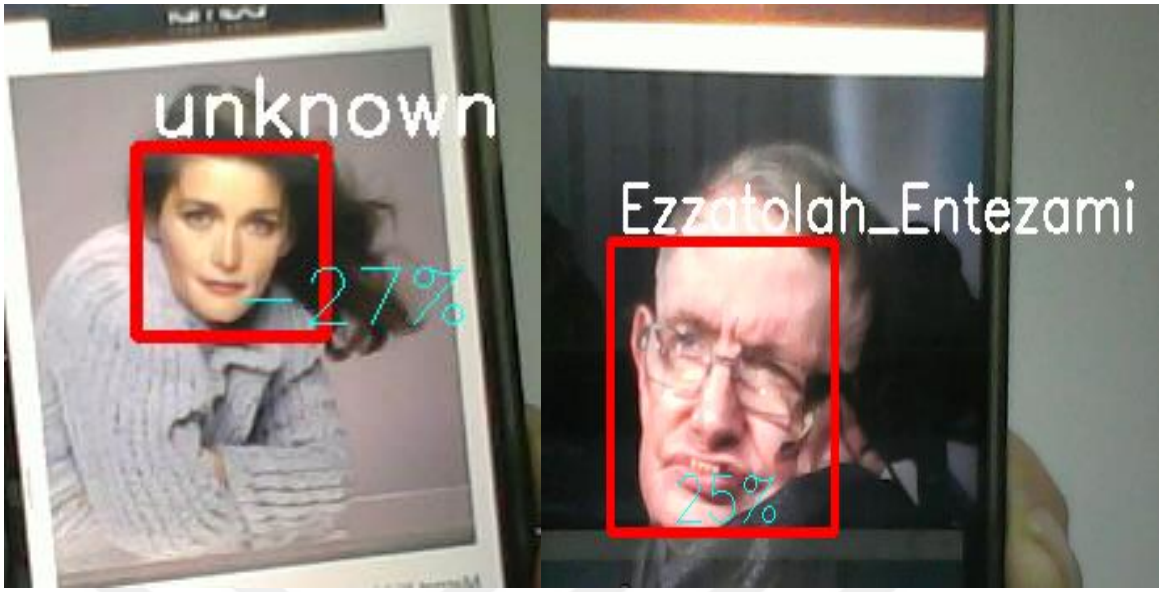
hakkında daha fazla bilgi verebilir. Programı çalıştırarak yüz tespit ve tanıma yapıldığında aşağıdaki resimlerde ki çıktıları elde edildi.



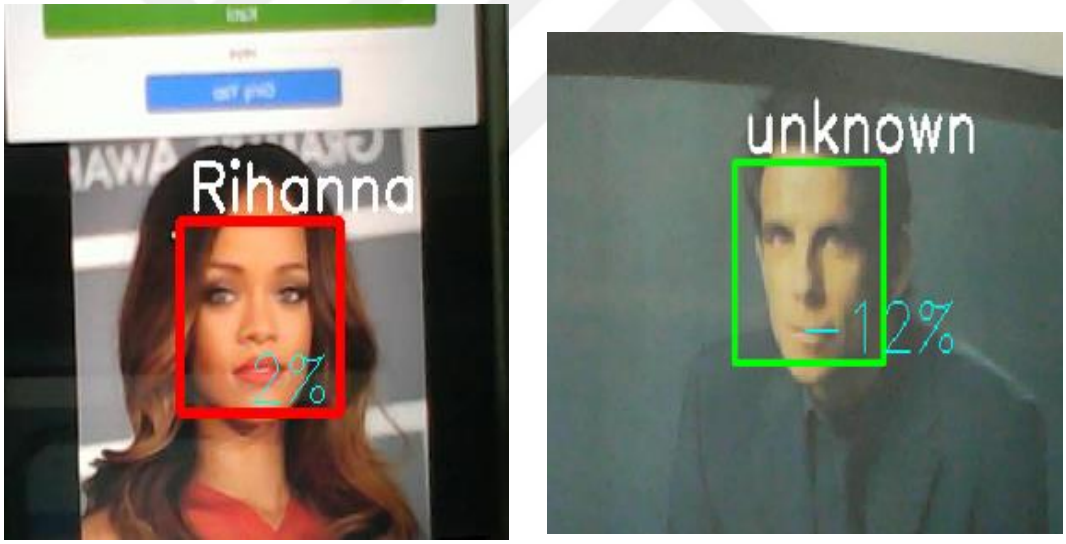
Resim 2.29. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 1



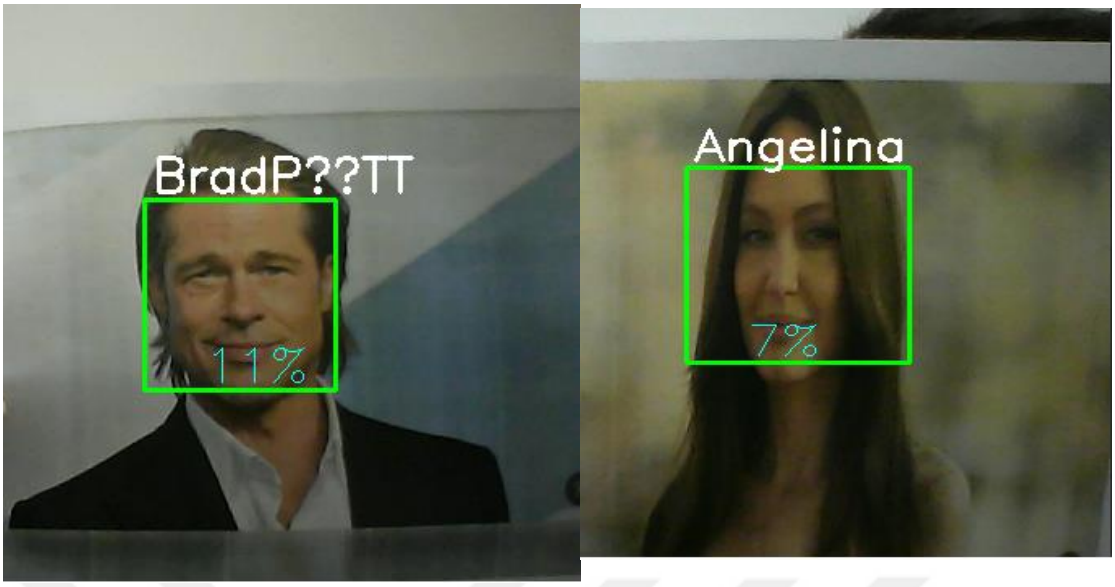
Resim 2.30. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 2



Resim 2.31. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 3

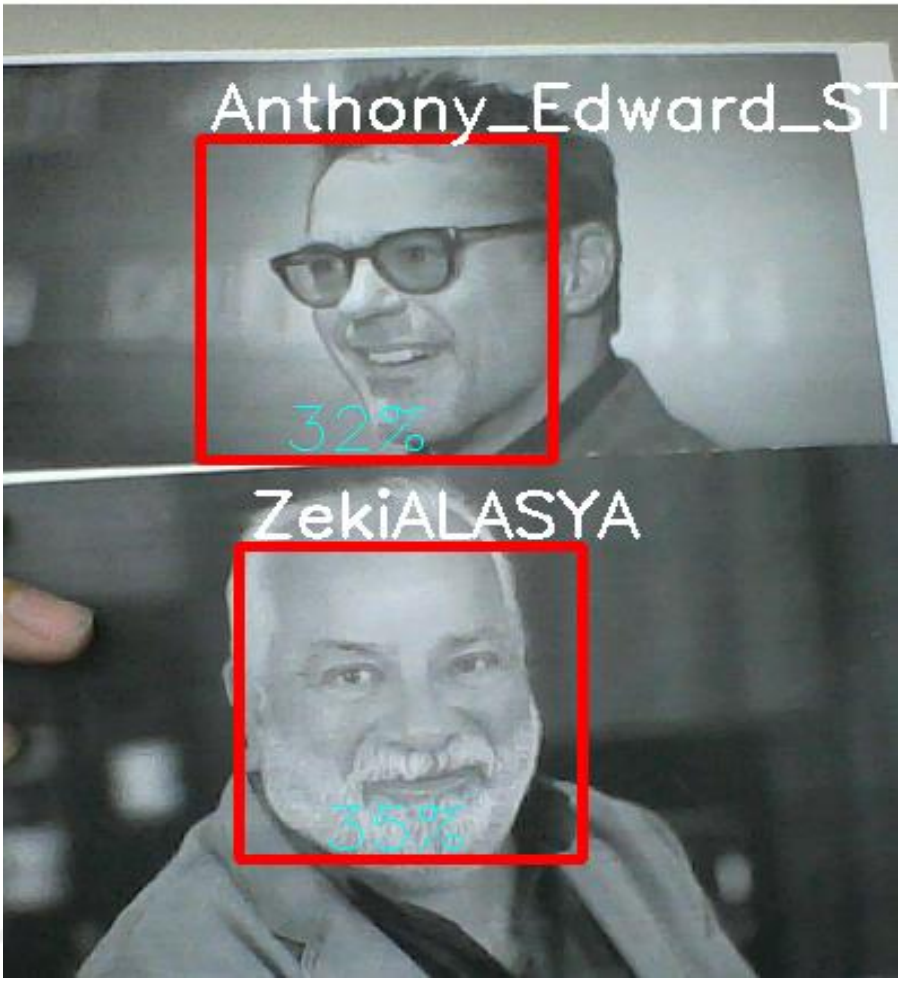


Resim 2.32. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 4

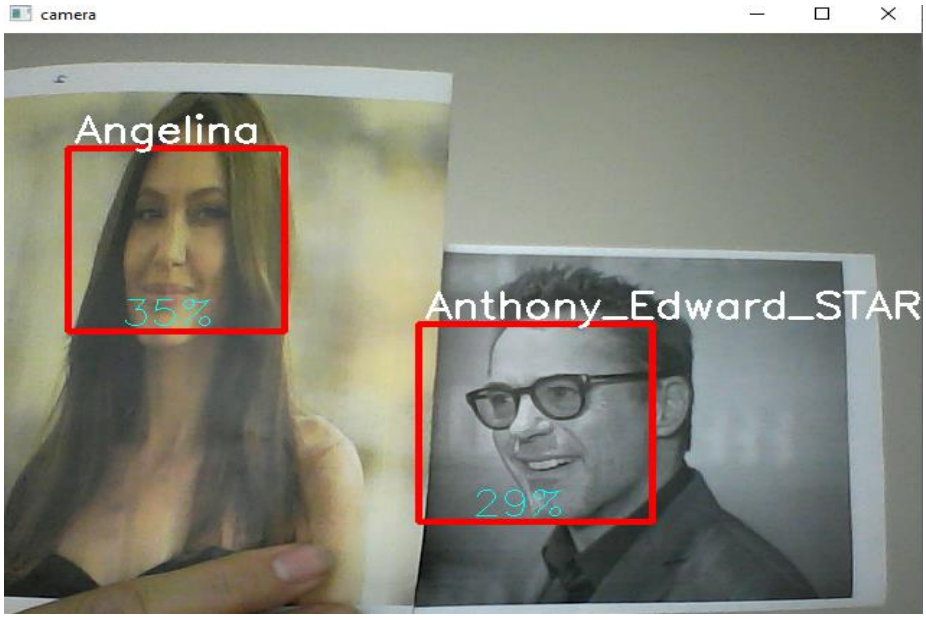


Resim 2.33. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 5

Resim 2.34. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 6



Resim 2.35. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 7



Resim 2.36. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 8

Resim 2.37. Gerçek Zamanlı Yüz Tanıma Programı Ekran Görüntüsü- 9

3. BÖLÜM

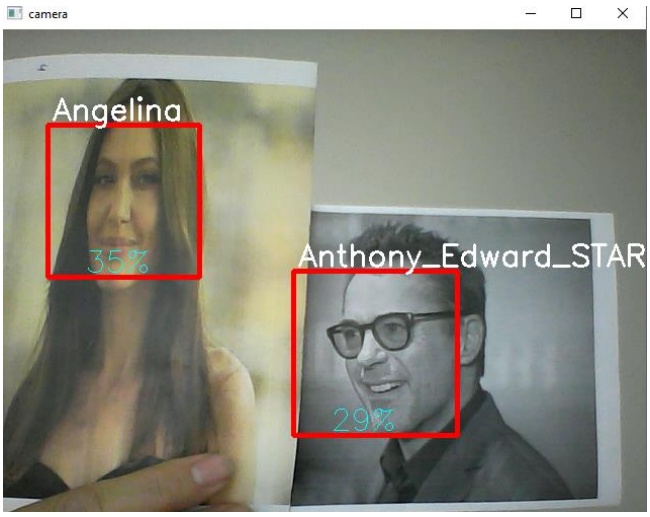
BULGULAR VE TARTIŞMA

Bu uygulama Intel Core i3-4005U, 1.70GHz CPU ve 4 GB RAM özelliklerine sahip standart bir dizüstü bilgisayarda geliştirilmiştir. Yapmış olduğumuz yüz tanıma uygulaması başarılı olarak çalışmış kamera karşısında ki yüzü alglayıp uygun veri kaynağını oluşturmuş sonrasında kamera karşısındaki yüzleri tanımayı başarmıştır. Bu bölümde yapay zeka makine öğrenme tekniği kullanılarak eğitilen yüz tanıma algoritmasının başarı oranları sonuçları incelenecektir.

Hazırlamış olduğumuz programımıza 10 adet fotoğraf verisi öğreterek programımızı test ettik. Programımız her bir fotoğraf için 30 adet yüz verisi aldı ve toplamda 300 fotoğraf bulunan bir veri kaynağı elde ettik. Programımız kamera karşısına geçirilen yüzlerden bir tanesini tanıyamamış, bir tanesini yanlış tanımış, 8 tanesini tanımıştır. Aynı şekilde veri tabanına yüklenmeyen iki yüz verisi de kameraya gösterilmiş, program beklenildiği gibi tanımaması gereken yüleri de tanımamıştır.

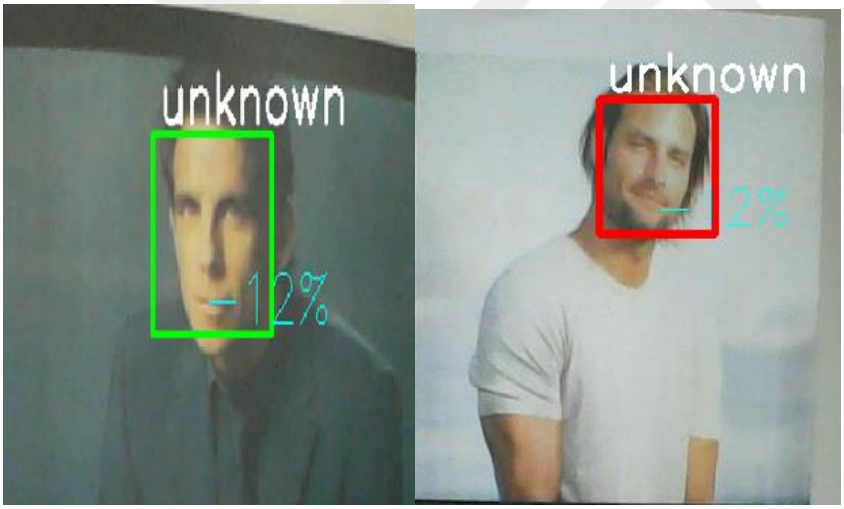
Yapmış olduğumuz program bize sadece kişileri tanıma veya tanımama çıktısını değil aynı zamanda yüzde olarak ne kadar hata payıyla tanımakta yada tanımamakta olduğunu da bildirmektedir. Resim 3.1'de örnekte soldaki resimdeki Ezzatoloh'u programımız % 46 hata payıyla tespit etmiş, sağda ki Paul'u % 23 hata yanılma payıyla tahmin etmiştir.

Resim 3.1. Gerçek Zamanlı Yüz Tanıma Programı Çıktı- 1



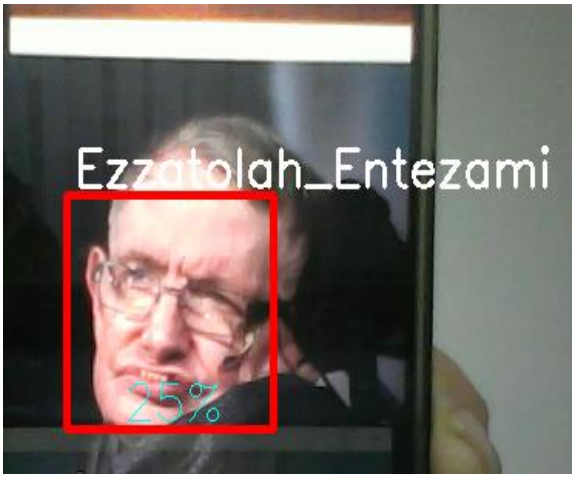
Resim 3.2. Gerçek Zamanlı Yüz Tanıma Programı Çıktı- 2

Resim 3.2'de programımız % 35 pay Angelina'le, Edward'I ise % 29 gibi bir hata payıyla tanımıştır.



Resim 3.3. Gerçek Zamanlı Yüz Tanıma Programı Çıktı- 3

Resim 3.3'de programımıza daha önceden tanıtmadığımızdan program bekleneni yaparak bu iki resimde ki olan şahışları %12 hata payıyla tanımamıştır.

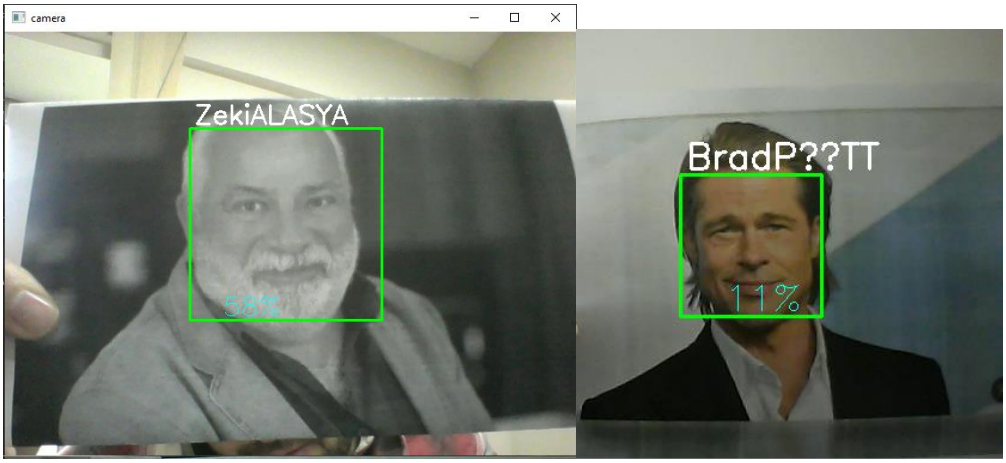


Resim 3.4. Gerçek Zamanlı Yüz Tanıma Programı Çıktı- 4

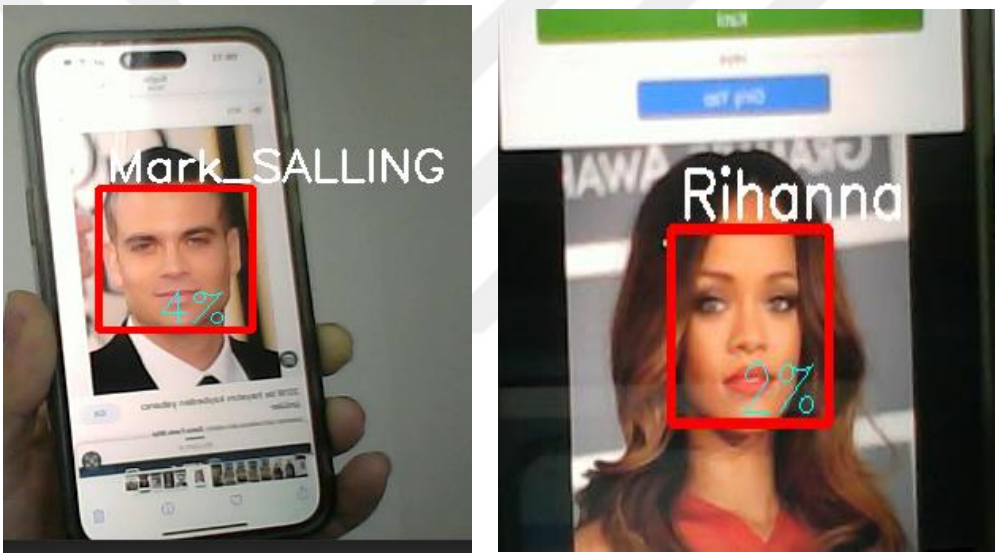
Resim 3.4'de programa Stephen HAWKİNG'I tanıtmamıza ragmen program yanlış çıktı vererek Ezzatolah ENTEZAMİ'yi % 25 hatayla çıktı olarak vermiştir.

Resim 3.5. Gerçek Zamanlı Yüz Tanıma Programı Çıktı- 5

Resim 3.5'de bu kez de programa tanıttığımızı Margot KİDDER'I program tanımamıştır.



Resim 3.6. Gerçek Zamanlı Yüz Tanıma Programı Çıktı- 6



Resim 3.7. Gerçek Zamanlı Yüz Tanıma Programı Çıktı- 7

Program Zeki ALASYA'YI % 58 hata payıyla, Brad PİTT'I %11 hata payıyla, Mark SALLING'I %4 hatayla, Rihanna'yı da yüzde iki hatayla tespit etmiştir.

Program çıktı değerleri Resim 3.6'da ki toplada gösterilmiştir. Kırmızı rankle yazılan tahmin programın yanlış tahminini göstermektedir. Tahmin değeri pozitif olan değer kamera karşısında ki yüzü tanıdığını, tahmin değeri negative olan değer kamera karşısında ki yüzün tanınmadığını göstermektedir. Tahmin yüzdesi ise programımızın hata payını göstermektedir. Yani tahmin yüzdesi (11) ise programımız kamera karşısında ki yüzü %11 hata payıyla tanımaktadır şeklinde açıklayabiliriz.

Eğitim ve Test Verileri	Tahmin	Tahmin Değeri	Tahmin Yüzdesi
Brad PITT	**Brad PITT**	+	11
Angelina	**Angelina**	+	35
Anthony Edward STARK	**Anthony Edward STARK**	+	29
Zeki ALASYA	**Zeki ALASYA**	+	58
Rihanna	**Rihanna**	+	2
Mark SALLING	**Mark SALLING**	+	4
Paul John VASQUEZ	**Paul John VASQUEZ**	+	23
Ezzatolah ENTEZAMİ	**Ezzatolah ENTEZAMİ**	+	46
Margot KİDDER	**Tanınmıyor**	-	27
Stephen HAWKİNG	**Ezzatolah ENTEZAMİ**	+	25
Bilinmiyor	**Tanınmıyor**	-	12
Bilinmiyor	**Tanınmıyor**	-	12

Resim 3.8. Gerçek Zamanlı Yüz Tanıma Programı Çıktıları

Programımızın başarısını kamera karşısında ki yüzü tanıyıp tanımadığına göre ve tanıma ya da tanımama gerçekleştiğinde yüzdesine bakarak yüzde olarak hesaplayacak olursak;

- Programımız (12) adet yüz verisinden (1) tanesini yanlış, (1) tanesini de tanıması gerektiği halde tanımamıştır.

$$\text{Hata Yüzdesi} = \frac{\text{I Deney Sonucu} - \text{Gerçek Sonuç I}}{\text{Gerçek Sonuç}} x100$$

$$\text{Hata Yüzdesi} = \frac{\text{I } 10 - 12 \text{ I}}{12} x100 \quad = 16{,}67$$

Programımızın yüz tanıma başarısı %83,33'tür.

- Program yukarda bahsedildiği gibi (12) adet yüz verisinden (1) tanesini yanlış tanıyıp (1) tanesini de tanıması gerekirken tanımamıştır. Tahmin yüzdesinde yanlış tanıdığı tahminlerde tahmin yüzdesine (-100) alarak tahmin yüzdesindeki hata yüzdesini hesaplamak uygun olacaktır. Çünkü tamamen hatalı bir çıktı olmuştur.

$$\text{Hata Yüzdesi} = \frac{\text{Hata Payları toplamı}}{\text{Toplam Veri Sayısı}}$$

$$\text{Hata Yüzdesi} = \frac{11+35+29+58+2+4+23+46+100+100+12+12}{12} = 36$$

Programımızın yüz tanıma tahmin yüzdesi başarısı ortalama %36'dir.

Her ne kadar programın hatasız bir şekilde çalışmasını istesek de bazı sebeplerden dolayı sistemimizin doğruluk performansı yüzde yüz olmamıştır. Ancak program çalışarak %83,33'lük doğru tahmin ve %36 hata payıyla başarılı sayılabilecek şekilde sonuç vermiştir. Daha keskin sonuçlar için yüksek performanslı cihazlarla gerçek zamanlı tarama yapmak, daha fazla yüz verisi kullanmak, daha donanımlı işletim sistemleri, bilgisayar ve kamera sistemleri kullanmak gerekir.

Yapılan program her ne kadar yüksek doğruluk oranıyla ve kişinin kendisine ihtiyaç duymadan gerçek zamanlı çalışıyor olsa da kullanılan verilerin önemi açısından ve bir suç soruşturmasında kişinin adı geçeceğinden dolayı sistemimizin daha az hata payıyla çalışıyor olması gerekirdi. Öyle ki bu sistem zaten kamera ile çalışıyor olmasından dolayı maske, makyaj vb. yöntemlerle aldatmaya çok müsait olan bir sistem olduğundan ve kullanımında daha önceden bahsedilen çekincelerden dolayı da kullanılması konusunda büyük yankı oluşturacak bir sistem olduğundan hata payı olmaması ya da çok az olması beklenirdi. Hazırlanılan program gibi yüz tanıma programlarının kullanıldığında gizliliği ihlal edebilecek olması, manipüle edilebilir olması, etik ve hukuki sorunlar ortaya çıkabilir olması, insanlar üzerinde kontrol baskısı oluşturabilir olması ve hukuk ihlalinden kimin sorumlu olacağı gibi tartışma yaratacak sorunlar ortaya çıkarabileceği aşikardır.

4. BÖLÜM

SONUÇ VE ÖNERİLER

Bu yüksek lisans tez çalışmasında suçu önlemek, suçu tahmin etmek ve suçlu tespitinde yani suç soruşturmalarında kullanılan yapay zeka ve veri madenciliği teknik ve yöntemleri araştırılmış, biyometrik güvenlik sistemlerinden yüz tanımanın aşamaları ve kullanılan uygulamanın detaylı açıklaması ile mimarisi ve çalışma prensibinden bahsedilerek bir yüz tanıma uygulaması yapılmıştır. Görüntü işleme kütüphanesi olan OpenCV ve Python programı ile bir makine öğrenimi olan gerçek zamanlı yüz algılama ve tanıma programıyla kimlik tespiti programı oluşturulmuştur. Bir fotoğraf veya videodan bilgisayara bağlı bir kamera yardımıyla kişilerin yüzlerini algılayan ve bilgisayardaki veri tabanına kaydeden, sonrasında ise bu görüntü ile veri tabanındaki yüz görüntülerini karşılaştırılarak kişi kimlik tespiti yapıldı. 10 kişilik veri tabanı oluşturuldu program 1 kişiyi yanlış tespit, 1 kişiyi ise tanımayarak hata yaptı. Veri tabanına kaydedilmeyen 2 kişiyi de veri tabanına kaydedilmediği için tanımayarak sistem dahilinde olmayan kişilere karşı da programın çalıştığı görülmüş oldu. Programın en yüksek performans oranı %98'lik benzerlik tespiti oldu. Programımız her ne kadar başarılı olsa da, etik ve hukuki sorunlara yol açmamak için daha az hata payıyla hatta hatasız çalışıyor olması gerekirdi.

Bu programda veri tabanımızı geliştirmek, yüksek performanslı kameralar ve sistemler kullanmak daha keskin sonuçlar almamızı sağlayabilir. Biyometrik verileri kullanan diğer uygulamalarda olduğu gibi yüz tanımada da gizliliğin ihlal edebilecek olması, etik ve hukuki sorunlar ortaya çıkabilecek olması, insanlar üzerinde kontrol baskısı oluşturabilecek olması ve hukuk ihlalinden kimin sorumlu olacağı gibi sorunlardan dolayı kullanımına çok dikkat edilmesi, şu aşamada suçlu tespitinde yada suç soruşturmasında direk kullanılmasındansa bir yol gösterici olarak değerlendirelebilir bir unsur olaması daha uygun olacaktır.

KAYNAKÇA

Britannica, T. (2017). Forensic investigation Encyclopedia Britannica. Date of access:12 Nisan 2022.https://www.britannica.com/topic/criminal-investigation

Encyclopedia, (2022). "Police: Criminal Investigations." Encyclopedia of Crime and Justice. 28 Mart 2022'de Encyclopedia.com'dan alındı. Erişim tarihi: 12 Nisan 2022. https://www.encyclopedia.com/law/legal-and-policy-magazines/police-criminal-investigations

Brandl SG, Frank J. (1994). The relationship between evidence, detective effort, and the disposition of burglary and robbery investigations. American Journal of Police, XIII, 3, 149-168.

Marchuk, I. (2014). Concept of Crime in International Criminal Law. In: The Basic Concept of Crime in International Criminal Law. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-28246-1_4

Benton, W. (1971). " Encyclopedia Britannica". Encyclopedia Britannica Inc, Vol. 1,1971.

Ward, C. H., Yu, M. W., Morabito, M., & Ding, W. (2011). Crime forecasting using data mining techniques. In 2011 IEEE 11th international conference on data mining workshops (pp. 779-786). IEEE.

Bernard, TJ, Kenar,. Ian David, Thomas, . David A., Clarke, . Donald C. ve Allott, . Antony Nicolas (2020). Crime Encyclopedia Britannica. https://www.britannica.com/topic/crime-law

Tappan, P. W. (1947). Who is the Criminal? American Sociological Review, 12(1), 96–102. https://doi.org/10.2307/2086496

Olatz, Cibrian Egido (2020). ''ARTIFICIAL INTELLIGENCE IN CRIMINAL JUSTICE SETTINGS: Where should be the limits of Artificial Intelligence in legal decision-making? Should an AI device make a decision about human justice?'' http://hdl.handle.net/10810/48983

McClendon & Meghanathan, (2015). Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.1, March 2015

Peterson, J., Ryan, J., Holden, P, & Mihajlovic, S. (1987). The uses and effects of forensic science in the adjudication of felony cases. Journal of Forensic Sciences, 32, 1730-1753.

Baker, R.S].d. and Yacef, K. (2010). The state of educational data mining in 2009; A review and future visions. Journal of Educational Data Mining, 1(1), 3-17.Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0- 387-31073-2.

Ateş, E. C. (2021). Big Data, Data Mining, Machine Learning, and Deep Learning Concepts in Crime Data. Journal of Penal Law & Criminology, 293-319. https://doi.org/10.26650/JPLC2020-813328

Ekblom, P., & Tilley, N. (2000). GOING EQUIPPED: Criminology, Situational Crime Prevention and the Resourceful Offender. The British Journal of Criminology, 40(3), 376–398. http://www.jstor.org/stable/23638938

Rebala, G., Ravi, A., & Churiwala, S. (2019). An Introduction to Machine Learning. Cham, CH: Springer

Aggarwal, C. C. (2018). Machine learning for text. Cham: Springer International Publishing.

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. Journal of the American society for information science and technology, 58(4), 479-493.

Ahmed, A. (2020). "From Data to Wisdom" Using Machine Learning Capabilities in Accounting and Finance Professionals. Talent Development & Excellence, 12.

Manonmaniam S.U, (2017). India, Fredrick David, H. B., & Suruliandi, A. (2017). SURVEY ON CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES. ICTACT Journal on Soft Computing, 7(3), 1459-1466. https://doi.org/10.21917/ijsc.2017.0202

Perry, W. L. (2013). Predictive policing: The role of crime forecasting in law enforcement operations. RAND.

Bhatt, N., Bhatt, N., & Prajapati, P. (2017). Deep Learning: A New Perspective. International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS), 6(6), 136-140.

Cardoso, J. (2006). Developing Dynamic Packaging Applications using Semantic Web based Integration. In A. F. Salam, & J. R. Stevens, Semantic Web Technologies and eBusiness: Toward the Integrated Virtual Organization and Business Process Automation (pp. 1-39). Hershey, PA: Idea Group.

Dore, F. (2012). The Chinese Room Argument against the Strong Artificial Intellingence. Afyon Kocatepe Üniversitesi, SBD,. 53-71.

Martín del Brío, B., & Sanz, A. (2006). Redes neuronales y sistemas borrosos. Zaragoza: Ra-Ma. Patrick, B. (2020). What is Artificial Intelligence? Journal of Accountancy, 229(2), 69-73.

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Waltham, MA: Elsevier. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436-444.

Dey, A. (2016). Machine learning algorithms: a review. International Journal of Computer Science and Information Technologies, 7(3), 1174-1179.

Read, M. (2021). CLEAR - Contact lens technologies of the future. Contact Lens and Anterior Eye, 44(2), 398-430. https://doi.org/10.1016/j.clae.2021.02.007

Clarke, C. (2006). Proactive policing: Standing on the shoulders of community-based policing. Police Practice and Research, 7(1), 3-17.

Hassani, H., Huang, X., Silva, E. S., & Ghodsi, M. (2016). A review of data mining applications in crime. Statistical Analysis and Data Mining: The ASA Data Science Journal, 9(3), 139-154

Fernández-Cabán, P. L., Masters, F. J., & Phillips, B. M. (2018). Predicting Roof Pressures on a Low-Rise Structure From Freestream Turbulence Using Artificial Neural Networks. Frontiers in Built Environment, 4(68), 1-16.

Thotakura, Dr. (2011). Crime: A Conceptual Understanding. Indian Journal of Applied Research. 4. 196-198. 10.15373/2249555X/MAR2014/58.

Quick, D., & Choo, K. K. R. (2016). Big forensic data reduction: digital forensic images and electronic evidence. Cluster Computing, 19(2), 723-740.

Chan, J., & Bennett Moses, L. (2017). Making sense of big data for security. The British journal of criminology, 57(2), 299-319.

Dong, Shi & Wang, Ping & Abbas, Khushnood. (2021). A survey on deep learning and its applications. Computer Science Review. 40. 100379. 10.1016/j.cosrev.2021.100379.

Jackson, J. (2002). Data mining; a conceptual overview. Communications of the Association for Information Systems, 8(1), 19.

Wang, L., Zhao, G., Cheng, L., & Pietikäinen, M. (Ed.). (2011). Machine Learning for Vision-Based Motion Analysis. Springer London. https://doi.org/10.1007/978-0-85729-057-1

Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press

Shao, L., Duan, Y., Sun, X., Gao, H., Zhu, D., & Miao, W. (2017, July). Answering Who/When, What, How, Why through Constructing Data Graph, Information Graph, Knowledge Graph and Wisdom Graph. In SEKE (pp. 1-6). Hey, J. (2004). The data, information, knowledge, wisdom chain: the metaphorical link. Intergovernmental Oceanographic Commission, 26, 1-18.

Pauleen, D. J., Rooney, D., & Intezari, A. (2017). Big data, little wisdom: trouble brewing? Ethical implications for the information systems discipline. Social Epistemology, 31(4), 400-416.

Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering (IJCSE), 2(02), 250-255.

Williams, G. J. (2009). Rattle: a data mining GUI for R. The R Journal, 1(2), 45-55.

McClendon, L., & Meghanathan, N. (2015). Using machine learning algorithms to analy-ze crime data. Machine Learning and Applications: An International Journal, 2(1), 1–12. https://doi.org/10.5121/mlaij.2015.2101

Beniwal, S., & Arora, J. (2012). Classification and feature selection techniques in data mining. International journal of engineering research & technology (IJERT), 1(6), 1-6

Ogunleye J.O. (2020): Review of Data Mining Techniques in Software Effort Estimation.

Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press

Blei, D. M., & Smyth, P. (2017). Science and data science. Proceedings of the National Academy of Sciences, 114(33), 8689-8692.

Bird, Carolyne & Found, Bryan. (2016). The modular forensic handwriting method. Journal of Forensic Document Examination. 26. 7-83.

Campbell, C., & Ying, Y. (2011). Learning with support vector machines. Synthesis lectures on artificial intelligence and machine learning, 5(1), 1-95.

Khare, A. R., & Shrivasta, P. (2018). Data mining for the internet of things. In Exploring the Convergence of Big Data and the Internet of Things (pp. 181-191). IGI Global.

Wang, Qi & Liu, Zhipeng & Tong, Shu & Yang, Yuqi & Zhang, Xiangde. (2017). Efficient Iris Localization via Optimization Model. Mathematical Problems in Engineering. 2017. 1-9. 10.1155/2017/7952152.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile networks and applications, 19(2), 171-209.

Lei, C. (2019). Legal Control over Big Data Criminal Investigation. Social Sciences in China, 40(3), 189-204

Agrahari, A., & Rao, D. (2017). A review paper on Big Data: technologies, tools and trends. Int Res J Eng Technol, 4(10), 640-649. Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. Journal of Experimental Criminology, 13(2), 193-216.

Mittal, M., Goyal, L. M., Hemanth, D. J., & Sethi, J. K. (2019). Clustering approaches for high-dimensional databases: A review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3), e1300.

Wheeler, A. P., & Steenbeek, W. (2020). Mapping the risk terrain for crime using machine learning. Journal of Quantitative Criminology, 1-36.

Yoo, J. S. (2019, December). Crime data warehousing and crime pattern discovery. In Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems (pp. 1-6).

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. Santa Monica, CA: RAND.

Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y., & Liu, Q. (2019). Big data analytics and mining for effective visualization and trends forecasting of crime data. IEEE Access, 7, 106111-106123.

Bulgakova, E., Bulgakov, V., Trushchenkov, I., Vasilev, D., & Kravets, E. (2019). Big data in investigating and preventing crimes. In Big Data-driven World: Legislation Issues and Control Technologies (pp. 61-69). Springer, Cham.

Snaphaan, T., & Hardyns, W. (2019). Environmental criminology in the big data era. European Journal of Criminology, 1477370819877753.

Sharma, Mukesh. (2020). Physical Evidence Interpretation and Analysis by Practical Study.

Umair, S., Muhammad, S., Amna, U., Aniqa, M., Abdul, B.S., Sheikh, K.R., (2015). Application of Machine learning Algorithms in Crime Classification and Classification Rule Mining. Res. J. Recent Sci. (pp. 106–114).

Wang, H., Kifer, D., Graif, C., & Li, Z. (2016, August). Crime rate inference with big data. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 635-644).

Kumar, R., & Nagpal, B. (2019). Analysis and prediction of crime patterns using big data. International Journal of Information Technology, 11(4), 799-805.

Tirgari, V. (2012). Information technology policies and procedures against unstructured data: A phenomenological study of information technology professionals. Journal of Management Information and Decision Sciences, 15(2), 87.

Toufiq R., (2014) M.R. Islam Face recognition system using PCA-ANN technique with feature fusion method International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), IEEE (2014), pp. 1-5

Choon L.S. (2004). A. Samsudin, R. Budiarto Lightweight and cost-effective MPEG video encryption Information and Communication Technologies: from Theory to Applications, 2004. Proceedings. 2004 International Conference on (pp. 525–526), IEEE (April 2004)

Almazrooie M., A. Samsudin, A.A.A. Gutub, M.S. Salleh, M.A. Omar, S.A. Hassan (2018) Computer and Information Sciences

Nahid P. F (2021), Deep reinforcement learning in transportation research: A review, Transportation Research Interdisciplinary Perspectives, https://doi.org/10.1016/j.trip.2021.100425.

Callara, Matias & Wira, Patrice. (2018). User Behavior Analysis with Machine Learning Techniques in Cloud Computing Architectures. 1-6. 10.1109/ICASS.2018.8651961.

Yücel, M.T., (2007). ‘‘Türk Ceza Siyaseti ve Kriminolojisi’’, T.B.B. Yayınları, 4. Baskı, Ankara, 2007, s.127

Theyazn A, (2020). A review of network traffic analysis and prediction techniques.

Alberto P. V., (2018). Applications of AI to Network Security. 10.13140/RG.2.2.29373.56803.

Popoola, Segun & Iyekekpolo, Ujioghosa & Ojewande, Samuel & Sweetwilliams, Faith & John, Samuel & Atayero, Prof. Aderemi. (2017). Ransomware: Current Trend, Challenges, and Research Directions.

Kemmerer, R. A. 2003. Cybersecurity. Proceedings of the 25th IEEE International Conference on Software Engineering: 705-715. http://dx.doi.org/10.1109/ICSE.2003.1201257

Cosmin, Deaconescu & Polkowski, Zdzislaw & Gruber, Jacek. (2013). E-Fraud.

Oxford University Press. 2014. Oxford Online Dictionary. Oxford: Oxford University Press. October 1, 2014: http://www.oxforddictionaries.com/definition/english/Cybersecurity

DHS. (2014). A Glossary of Common Cybersecurity Terminology. National Initiative for Cybersecurity Careers and Studies: Department of Homeland Security. October 1, 2014: http://niccs.us-cert.gov/glossary#letter_c

Kozlenkova, I. V., Samaha, S. A., & Palmatier, R. W. (2014). ResourceBased Theory in Marketing. Journal of Academic Marketing Science, 42(1): 1-21. http://dx.doi.org/10.1007/s11747-013-0336-7

ITU. (2009). Overview of Cybersecurity. Recommendation ITU-T X.1205. Geneva: International Telecommunication Union (ITU). http://www.itu.int/rec/T-REC-X.1205-200804-I/en

CNSS. (2010). National Information Assurance Glossary. Committee on National Security Systems (CNSS) Instruction No. 4009: http://www.ncix.gov/publications/policy/docs/CNSSI_4009.pdf

Reshamwala, Alpa & Mishra, Dhirendra & Pawar, Prajakta. (2013). REVIEW ON NATURAL LANGUAGE PROCESSING. IRACST – Engineering Science and Technology: An International Journal (ESTIJ). 3. 113-116.

Canongia, C., & Mandarino, R. (2014). Cybersecurity: The New Challenge of the Information Society. In Crisis Management: Concepts, Methodologies, Tools and Applications: 60-80. Hershey, PA: IGI Global. http://dx.doi.org/10.4018/978-1-4666-4707-7.ch003

Diakun T. N., (2014). Defining Cybersecurity. Technology Innovation Management Review. 2014.

Lewis, J. A. (2006). Cybersecurity and Critical Infrastructure Protection. Washington, DC: Center for Strategic and International Studies. http://csis.org/publication/cybersecurity-and-criticalinfrastructure-protection

Amoroso, E. (2006). Cyber Security. New Jersey: Silicon Press

Oludare I, A, (2018), Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, Humaira Arshad, State-of-the-art in artificial neural network applications: https://doi.org/10.1016/j.heliyon.2018.e00938.(https://www.sciencedirect.com/science/article/pii/S2405844018332067)

Vaidhyanathan, S., & Bulock, C. (2014). Knowledge and dignity in the era of “big data”. The Serials Librarian, 66(1-4), 49-64.

Mesgarpour, M., & Dickinson, I. (2014). Enhancing the value of commercial vehicle telematics data through analytics and optimisation techniques. Archives of Transport System Telematics, 7.

Abdullah, N., Ismail, S. A., Sophiayati, S., & Sam, S. M. (2015). Data quality in big data: a review. International Journal of Advances in Soft Computing & Its Applications, 7(3).

Abikoye, Oluwakemi & S., Sadiku & S., Adewole & Gbenga, Jimoh. (2014). Iris Feature Extraction for Personal Identification using Fast Wavelet Transform (FWT). International Journal of Applied Information Systems (IJAIS). 6. 1-6.

Bock, F. E., Aydin, R. C., Cyron, C. J., Huber, N., Kalidindi, S. R., & Klusemann, B. (2019). A review of the application of machine learning and data mining approaches in continuum materials mechanics. Frontiers in Materials, 6, 110.

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Coello, C. A. C. (2013). A survey of multiobjective evolutionary algorithms for data mining: Part I. IEEE Transactions on Evolutionary Computation, 18(1), 4-19.

Mittal, M., Goyal, L. M., Hemanth, D. J., & Sethi, J. K. (2019). Clustering approaches for high-dimensional databases: A review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3), e1300.

Berkhin, P. (2006). A survey of clustering data mining techniques. In Grouping multidimensional data (pp. 25-71). Springer, Berlin, Heidelberg.

Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert systems with applications, 36(2), 2592-2602.

Chau, D. H., Pandit, S., & Faloutsos, C. (2006, September). Detecting fraudulent personalities in networks of online auctioneers. In European Conference on Principles of Data Mining and Knowledge Discovery (pp. 103-114). Springer, Berlin, Heidelberg.

PredPol. (2018). The Three Pillars of Predictive Policing. Available at: http://www.predpol.com/law-enforcement/ [Erişim tarihi 12 Kasım 2018].

HunchLab, (2014) ‘‘Predictive Policing Systems’’, https://teamupturn.gitbooks.io/predictive-policing/content/systems/hunchlab.html, (Erişim Tarihi: 29.04.2022).

Odia, J. O., & Akpata, O. T. (2020). Role of Data Science and Data Analytics in Forensic Accounting and Fraud Detection. In Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics (pp. 203-227). IGI Global.

RAND Şirketi (2013). Tahmine Dayalı Polislik: Kolluk için Suç Tahmini. Santa Monica. https://www.rand.org/pubs/research_briefs/RB9735.html [Erişim tarihi 28 Nisan 2022].

Kabir, S. M. (2016). INTRODUCTION TO RESEARCH.

Robarts, K. (2018), "Makine Öğrenimi ve Derin Öğrenme Arasındaki Farklar." Benzer Terimler ve Nesneler Arasındaki Fark, 23 Temmuz 2018, http://www.differencebetw een.net/technology/ differences-between-machine-learning-and-deep-learning/.

Read more: Differences Between Machine Learning and Deep Learning | Difference Between http://www.differencebetween.net/technology/differences-between-machine-learning-and-deep-learning/#ixzz7SnZTp6jg

Shapiro, A. (2017). Reformed policing. Doğa 541, 458-460 https://doi.org/10.1038/541458a

Crunchbase. (2018). PredPol'e Genel Bakış.https://www.crunchbase.com/organization /predpol#section-overview [Erişim tarihi 28. Nisan 2018].

Agu, S. C., Ajah, I., & Ibe, W. E. (2019). Impact of Human Character and Information System on Corruption Risk in Nigeria. International Journal of Scientific Research and Engineering Development, 2(4), 481-485

HunchLab, (2014). ‘‘Predictive Policing Systems’’,https://teamupturn. gitbooks.io/predictive-policing/content/systems/hunchlab.html, (Erişim Tarihi: 29.04.2022).

Martinovic, I., Rasmussen, K., Roeschlin, M., & Tsudik, G. (2017). Authentication using pulse-response biometrics. Communications of the ACM, 60(2), 108-115.

Müller, V. C. (2021). " Artificial Intelligence and Robotics Ethics", Stanford Felsefe Ansiklopedisi (Yaz 2021 Baskısı), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/ sum2021/entries/ethics-ai/>.Per Nis 30, 2020

Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8(2), 937-953.

Lin, S.-H. (2000). An Introduction to Face Recognition Technology. Informing Science: The International Journal of an Emerging Transdiscipline, 3, 001-007. https://doi.org/10.28945/569

Gershgorn, Dave (2021). "Çin'in 'Keskin Gözler' Programı Kamusal Alanın %100'ünü Gözetlemeyi Amaçlıyor". orta. 2021-08-02

Nguyen, M. T.(2021). Linh H. Truong, and Trang TH Le. "Video surveillance processing algorithms utilizing artificial intelligent (AI) for unmanned autonomous vehicles (UAVs)." MethodsX 8 (2021): 101472.

Agrahari, A., & Rao, D. (2017). A review paper on Big Data: technologies, tools and trends. Int Res J Eng Technol, 4(10), 640-649.

Kade C. (2020). ‘‘How is Face Recognition Surveillance Technology Racist?’’. https://www.aclu.org/news/privacy-technology/how-is-face-recognition-surveillance-technology-racist. 27.04.2022

Guenther, A.J. (2012). Role of Social Media in Law Enforcement Significant and Growing [Online]. Available:http://www.lexisnexis.com/en-us/about-us/media/press-release.page?id =1342623085481181

Fayyad, U.M.(1996). Haussler, D. and Stolorz, Z. 1996. KDD for Science Data Analysis; Issues and Examples. Proc. 2nd Int. Conj. on Knowledge Discovery and Data Mining (KDD-96), Menlo Park, CA: AAAI Press.

Yu, C., Ward, M. W., Morabito, M., Ding, W. (2021). “Crime Forecasting Using Data Mining Techniques”, 11th IEEE International Conference on Data Mining Workshops, 779-786,(2011) Zhang, Yixuan; Shang, Dan.Kod çözme yüz tanıma teknolojisi: "Yüzünüzü kaydırın" hayatınıza giriyor[Günlük hayatımızda yüz tanıma teknolojisi hakkında bilgi sahibi olmak]. www.xinhuanet.com (Çince (Çin)). 2018-11-21

Arshad, H., Jantan, A., & Omolara, E. (2019). Evidence collection and forensics on social networks: Research challenges and directions. Digital Investigation, 28, 126-138.

Wearesocials & Hootsuide, (2018). Digital in 2018: World’s ınternet users pass the 4 billion mark. URL https:// wearesocial.com/blog/2018/01/global-digital-report-2018 (accessed 16.02.20).

Ayre, L. B., & Craner, J. (2017). Open data: What it is and why you should care. Public Library Quarterly, 36(2), 173-184

Alfred Rayner, (2005). Knowledge Discovery: Enhancing Data Mining and Decision Support Integration, [www-users.cs.york.ac.uk]

Bharati M. Ramageri, (2010). DATA MINING TECHNIQUES AND APPLICATIONS, Modern Institute of Information Technology and Research, Department of Computer Application, Yamunanagar, Nigdi Pune, Maharashtra, India-411044. Zaïane Osmar R., (1999): “Principles of Knowledge Discovery in Databases - Chapter 8: Data Clustering”. http:// www. cs. ualberta. ca/~ zaiane/ courses /cmput690 /slides/ Chapter8 /index.html.

Ch, R., Gadekallu, T. R., Abidi, M. H., & Al-Ahmari, A. (2020). Computational System to Classify Cyber Crime Offenses Using Machine Learning. Sustainability, 12(10), 4087.

Petković, M. (2007). Security, privacy and trust in modern data management: With 13 tables. Springer.

Davies, Pamela & Cook, Ian. (2020). Victims, Witnesses and the Criminal Justice System. Power, D. J. (2016). “Big Brother” can watch us. Journal of Decision systems, 25(sup1), 578-588

Halder, B. (2017). "CHINA TURNED INTO ROBOTIC POLICE". Ozy . Archived from the original on 2019-11-28. 2019-11-28

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the machine: anti-Black and antiMuslim social media posts as predictors of offline racially and religiously aggravated crime. The British Journal of Criminology, 60(1), 93-117.

Ristea, A., Al Boni, M., Resch, B., Gerber, M. S., & Leitner, M. (2020). Spatial crime distribution and prediction for sporting events using social media. International Journal of Geographical Information Science, 1-32.

Muneer, A., & Fati, S. M. (2020). A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter. Future Internet, 12(11), 187.

Stewart, L. (2019). Big Data Discrimination: Maintaining Protection of Individual Privacy without Disincentivizing Businesses’ Use of Biometric Data to Enhance Security. BCL Rev., 60, 349.

Tiwari, S., Chourasia, J. N., & Chourasia, V. S. (2015). A review of advancements in biometric systems. International Journal of Innovative Research in Advanced Engineering, 2(1), 187-204.

Martinovic, I., Rasmussen, K., Roeschlin, M., & Tsudik, G. (2017). Authentication using pulse-response biometrics. Communications of the ACM, 60(2), 108-115.

Ajans Press, 2017. Steenbruggen, J., Tranos, E., & Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities?. Telecommunications Policy, 39(3-4), 335-346

Lawrence J. F., (2020). İn Handbook of Loss Prevention and Crime Prevention (Sixth Edition) Pages 201-205,https://doi.org/10.1016/B978-0-12-817273-5.00018-1., (https://www.sciencedirect.com/science/article/pii/B9780128172735000181)

Russell J., (2021-03-02). "China’s CCTV surveillance network took just 7 minutes to capture BBC reporter". https://techcrunch.com/2017/12/13/china-cctv-bbc-reporter 2022-04-26

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5), 662-679.

Xu, H. (2020). Big data challenges in genomics. In Handbook of Statistics (Vol. 43, pp. 337-348). Elsevier

Aledhari, M., Di Pierro, M., Hefeida, M., & Saeed, F. (2018). A deep learning-based data minimization algorithm for fast and secure transfer of big genomic datasets. IEEE Transactions on Big Data.

Comet L. R. (2018). ‘‘Surveillance Redefined: Advances in Vision AI Drive China’s Leading Security Market’’ Erişim tarihi: 24 Nisan 2022.https://blog. cometlabs. io/surveillance-redefined-advances-in-vision-ai-drive-chinas-leading-security-market-4cf757a48275

Tilley, N., & Sidebottom, A. (2017). Handbook of crime prevention and community safety. Routledge.

Brazil, Kristopher & Whittingham, Lisa. (2019). Criminology. 10.1007/978-3-319-16999-6_3193-1.

Yao, S., Wei, M., Yan, L., Wang, C., Dong, X., Liu, F., & Xiong, Y. (2020). Prediction of Crime Hotspots based on Spatial Factors of Random Forest. In 2020 15th International Conference on Computer Science & Education (ICCSE) (pp. 811-815). IEEE.

Song, G., Bernasco, W., Liu, L., Xiao, L., Zhou, S., & Liao, W. (2019). Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices. Journal of Quantitative Criminology, 35(4), 831-854.

Mistek, E., Fikiet, M. A., Khandasammy, S. R., & Lednev, I. K. (2018). Toward locard's exchange principle: recent developments in forensic trace evidence analysis. Analytical chemistry, 91(1), 637-654.

Data, (2021). ''Dolandiricilik-tespiti-ve-onlenmesinde-finans-sektorunun-elindeki-guc-datactive'' Erişim Tarihi: 14.05.2022. https://datateam.com.tr/dolandiricilik-tespiti-ve-onlenmesinde-finans-sektorunun-elindeki-guc-datactive/

Thotakura, Dr. (2011). Crime: A Conceptual Understanding. Indian Journal of Applied Research. 4. 196-198. 10.15373/2249555X/MAR2014/58.

Bode, J. (2019). Every Contact Leaves a Trace: A Literary Reality of Locard's Exchange Principle. In Outside the Box: A Multi-Lingual Forum (p. 18).

Öztürk, E.E., (2020) ''Makineler Nasıl Öğrenir.'' Erişim Tarihi: 14.05.2022. https://www.veribilimiokulu.com/makineler-nasil-ogrenir/

He, L., Páez, A., Jiao, J., An, P., Lu, C., Mao, W., & Long, D. (2020). Ambient Population and Larceny-Theft: A Spatial Analysis Using Mobile Phone Data. ISPRS International Journal of Geo-Information, 9(6), 342

Roessler, B. (2017). Privacy as a human right. Proceedings of the Aristotelian Society, 2 (CXVII).

Macnish, K. (2017). The ethics of surveillance: An introduction. London: Routledge,

Müller, V. C. (2020). Can machines think? Fundamental problems of artificial intelligence. New York: Oxford University Press.

Traunmueller, M., Quattrone, G., & Capra, L. (2014, November). Mining mobile phone data to investigate urban crime theories at scale. In International Conference on Social Informatics (pp. 396-411). Springer, Cham

Wilson, D. B., McClure, D., & Weisburd, D. (2010). Does forensic DNA help to solve crime? The benefit of sophisticated answers to naive questions. Journal of Contemporary Criminal Justice, 26(4), 458-469.

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V. Schultz, J. (2018). AI Now Report 2018. From https://ainowinstitute.org/AI_Now_2018_Report.html

Floridi, L., & Taddeo, M. (2016). What is Data Ethics? Phil. Trans. R. Soc. A, 374 (2083).

Ferguson, A. G. (2017). The rise of big data policing: Surveillance, race, and the future of law enforcement. New York: NYU Press.

Danaher, J. (2015) How might algorithms rule our lives? Mapping the logical space of algocracy, 15 June. Available at: http://philosophicaldisquisitions.blogspot.co.za/2015/06/how-might-algorithms-rule-our-lives.html (accessed 9 April 2017).

Gershenson, C. (2003). Artificial Neural Networks for Beginners.

Yam, Y. (1997). Dynamics of Complex Systems. Addison-Wesley.

Bostanci, E. (2015). 3D reconstruction of crime scenes and design considerations for an interactive investigation tool. arXiv preprint arşiv:1512.03156.

Wachter, S., Mittelstadt, B. D., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of La & Technology, 31 (2).

Chan, T. F. (2018). ‘‘Beijing police are using facial-recognition glasses to identify car passengers and number plates’’ Erişim tarihi: 24 Nisan 2022.https://www.businessinsider.com/china-police-using-smart-glasses-facial-recognition-2018-3

Gulal, O., (2018). ‘‘Life/artificial-intelligence-deep-learning-and-machine-learning-differences/’’ Erişim Tarihi:14.05.2022. http://ozhangulal.com/hayat/yapay-zeka-derin-ogrenme-ve-makine-ogrenimi-arasindaki-farkliliklar/

Geek,(2021).”face-detection-using-cascade-classifier-using-opencv-python” https://www.geeksforgeeks.org/face-detection-using-cascade-classifier-using-opencv-python/?ref=rp

Gorgoda, (2020). ‘‘Nerve Cell-Neuron’’ Erişim Tarihi: 14.05.2022. https://www.gorgoda.com/nerve-cell-neuron.html

Win, K. N., Li, K., Chen, J., Viger, P. F., & Li, K. (2020). Fingerprint classification and identification algorithms for criminal investigation: A survey. Future Generation Computer Systems, 110, 758-771.

Taulli, T. (2019). Artificial Intelligence Basics: A Non-Technical Introduction. Monrovia, CA: Apress. Turing, A. M. (1950). Computing Machinery and Intelligence. Mind 49, 433-660.

Bhuyan, M. H., Saharia, S., & Bhattacharyya, D. K. (2012). An effective method for fingerprint classification. arXiv preprint arXiv: 1211.4658.

Yayık, Apdullah & Kutlu, Yakup. (2013). Yapay Sinir Ağları ile Kriptoloji Uygulamaları. 10.13140/RG.2.2.29938.61125.

Commission, (2017). Kriminalistik. Gendarmerie and Coast Guard Academy, Ankara.

Verma P. , Dubey M., Verma P., Basu S., Shankaracharya S., (2012), DAUGHMAN"S ALGORITHM METHOD FOR IRIS RECOGNITION-A BIOMETRIC APPROACH. International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 6, June 2012)

Uliyan, D. M., Sadeghi, S., & Jalab, H. A. (2020). Anti-spoofing method for fingerprint recognition using patch based deep learning machine. Engineering Science and Technology, an International Journal, 23(2), 264-273

Song, G., Bernasco, W., Liu, L., Xiao, L., Zhou, S., & Liao, W. (2019). Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices. Journal of Quantitative Criminology, 35(4), 831-854.

Ross A., (2010), "Iris Recognition: The Path Forward", Computer , cilt. 43, hayır. 2, s. 30-35, Şubat 2010, doi: 10.1109/MC.2010.44.

Zhou X., W. Gong, W. Fu, F. Du (2017), Application of deep learning in object detection IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), IEEE (2017), pp. 631-634

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417-457.

Nguyen K., Fookes C., Jillela R., Sridharan S., Ross A., (2017). Long range iris recognition: A survey, Pattern Recognition. University of York Department of Computer Science, Deramore Lane, Heslington, YO10 5DD, York, United Kingdom https://doi.org/10.1016/j.patcog.2017.05.021. (https://www.sciencedirect.com/science/article/pii/S0031320317302182)

EKLER

EK A: Kamera Karşısı Kişi Tespit Program Kodları

EK B: Kişi Tespit Eğitim Ara Yüzü Program kodları

EK C: Gerçek Zamanlı Yüz Tanıma Program kodları

EK A:

```
import cv2
import os

vid_cam = cv2.VideoCapture(0)
# Kamera bir değişkene tanımlanır ve aktif edilir

face_detector = cv2.CascadeClassifier('haarcascade_frontalface_default.xml')
# Yüz tespiti için open-cv kütüphanesiyle cascade sınıflandırıcısı bir değişkene atanır
# Tanımlanan yüzler ayırmak için farklı rakamlar atandı.
face_id = input('\n enter user id end press <return> ==>  ')

print("\n [INFO] Initializing face capture. Look the camera and wait ...")
# yüz sayımı başlatılır
count = 0

while(True): # While döngüsü oluşturuldu

    # kamera okutuldu
    ret, img = vid_cam.read()

    # Başlangıçta görüntü üç katmanlı bir görüntüdür. Bu nedenle tek katmanlı bir görüntüye (yani gri tonlamalı) dönüştürülür.
    gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)

    # resim için alt ve üst sınırlar belirler
    faces = face_detector.detectMultiScale(gray, 1.3, 5)
```

```
   for (x,y,w,h) in faces: #for döngüsünde çerçeve ebatları için değişkenler belirlenir
# çerçeve rengi ve kalınlığı belirlendi
      cv2.rectangle(img, (x,y), (x+w,y+h), (255,0,0), 2)
      count += 1
# fotoğraf adet artışı tanımlandı.
      # resimler veri klasörüne yazdırıldı
      cv2.imwrite("veri/User." + str(face_id) + '.' + str(count) + ".jpg", gray[y:y+h,x:x+w])


      cv2.imshow('image', img) #kameraya göster komutu atandı
# fotoğraf kalitesi belirlendi ve çıkış tuşu atandı
   if cv2.waitKey(20) & 0xff == ord('q'):
         break
   elif count> 30:  # kameranın çekeceği fotoğraf sayısı sınırlandırıldı
         break



print("\n [INFO] Exiting Program and cleanup stuff")
vid_cam.release() # kamera durduruldu
cv2.destroyAllWindows() # tüm pencereler kapatıldı
```

EK B:

```
import cv2
import numpy as np
from PIL import Image
import os

#
path = 'veri'

recognizer = cv2.face.LBPHFaceRecognizer_create() #yüz tanıyıcı oluşturuldu
detector = cv2.CascadeClassifier("haarcascade_frontalface_default.xml"); #sınıflandırıcı xml dosyası eklendi

# görüntüler ve tanımlar için yol atandı
def getImagesAndLabels(path):
# resim yolunu bulmak için döngü oluşturuldu
   imagePaths = [os.path.join(path,f) for f in os.listdir(path)]
   faceSamples=[] #değişken tanımlandı
   ids = [] #değişken tanımlandı

   for imagePath in imagePaths: # for döngüsü oluşturuldu
# resmin okuyacağı dosyayı açmak için pıl_img den yararlanıldı
      PIL_img = Image.open(imagePath).convert('L') # convert it to grayscale
      img_numpy = np.array(PIL_img,'uint8') # pozitif tam sayı olduğu belirtildi

      id = int(os.path.split(imagePath)[-1].split(".")[1]) # ismin yazılacağı yer belirtildi
      faces = detector.detectMultiScale(img_numpy) # yüz ölçeklerinin belirleneceği kod satırı atandı
```

```
        for (x,y,w,h) in faces: # yüz çerçevesi için değişkenler for döngüsüne eklendi
            faceSamples.append(img_numpy[y:y+h,x:x+w])
            ids.append(id)

    return faceSamples,ids

print ("\n [INFO] Training faces. It will take a few seconds. Wait ...")
faces,ids = getImagesAndLabels(path)
recognizer.train(faces, np.array(ids))

# model trainer/trainer.yml dosyasına kaydedildi
recognizer.save('deneme/deneme.yml')

# eğitilen yüzlerin sayısı yazdırılarak program sonlandırıldı
print("\n [INFO] {0} faces trained. Exiting Program".format(len(np.unique(ids))))
```

EK C:

```
import cv2
import numpy as np
import os

recognizer = cv2.face.LBPHFaceRecognizer_create() #yüz tanıyıcı oluşturuldu
# okunacak dosya belirtildi
recognizer.read('deneme/deneme.yml')
cascadePath = "haarcascade_frontalface_default.xml" #sınıflandırıcı xml dosyası eklendi
faceCascade = cv2.CascadeClassifier(cascadePath); #kullanılacak yol atandı
font = cv2.FONT_HERSHEY_SIMPLEX #yazı tipi belirlendi
# kimlik sayacı başladı
id = 0
# kimlik isimleri verildi
names = ["YOK",'BradPITT','Angelina','Anthony_Edward_STARK','ZekiALASYA','Rihanna','Mark_SALLING','Paul_John_Vasquez','Ezzatolah_Entezami','Margot_KİDDER','Stephen HAWKİNG']
# gerçek zamanlı kamera başlatıldı
cam = cv2.VideoCapture(0)
cam.set(3, 640)  # görüntü genişliği ayarla
cam.set(4, 480)  # görüntü yüksekliği ayarla
# minimum pencere boyutunu ayarla
minW = 0.1 * cam.get(3)
minH = 0.1 * cam.get(4)
while True:
    ret, img = cam.read() # kamera okutuldu
    img = cv2.flip(img, 1)
```

```
gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)

faces = faceCascade.detectMultiScale(
    gray,
    scaleFactor=1.2,
    minNeighbors=5,
    minSize=(int(minW), int(minH)),
)
for (x, y, w, h) in faces:
    cv2.rectangle(img, (x, y), (x + w, y + h), (0, 0, 255), 3) # çerçeve ebatları ayarlandı
    id, confidence = recognizer.predict(gray[y:y + h, x:x + w])
    # eşleşme yüzdesi tanımlandı
    if (confidence < 100):
        id = names[id]
        confidence = "  {0}%".format(round(100 - confidence)) #hata payı hesaplandı
    else:
        id = "unknown"
        confidence = "  {0}%".format(round(100 - confidence))

    cv2.putText(img, str(id), (x + 5, y - 5), font, 1, (255, 255, 255), 2)
    cv2.putText(img, str(confidence), (x + 5, y + h - 5), font, 1, (255, 255, 0), 1)

cv2.imshow('camera', img)
# çıkış komutu tanımlandı
if cv2.waitKey(10) & 0xFF == ord('q'):
    break
```

```
print("\n [INFO] Exiting Program and cleanup stuff")
cam.release()
cv2.destroyAllWindows()
```

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, Adı	:	YÜN, Murat
Uyruğu	:	T.C.
Doğum tarihi ve yeri	:	[redacted]
Medeni hali	:	Evli
Telefon	:	[redacted]
[redacted]		[redacted].com

Eğitim

Derece	**Eğitim Birimi**	**Mezuniyet tarihi**
Lise	Zile Anadolu Lisesi	2006
Lisans	Harran Üniversitesi Çevre Mühendisliği	2008
Lisans	Ondokuz Mayıs Üniversitesi Mühendislik Fakültesi Elektrik Elektronik Mühendisliği	2013
	Polis Akademisi	2017
	Yüksek Lisans Polis Akademisi Güvenlik Yönetimi (Tezsiz)	2017

İş Deneyimi

Yıl	**Yer**	**Görev**
2017-	[redacted]	[redacted]

Yabancı Dil

İngilizce