

T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
İSTATİSTİK ANA BİLİM DALI



**MAKİNE ÖĞRENMESİ KULLANARAK İNSAN
HAREKETLERİNİN SINIFLANDIRILMASI: AKILLI
TELEFON ALGILAYICILARI ÖRNEĞİ**

Yüksek Lisans Tezi

Dina Duraid Haqı AL-MOMAYEZ

Danışman

Prof. Dr. Kamil ALAKUŞ

SAMSUN
2023

TEZ KABUL VE ONAYI

Dina ALMOMAYEZ tarafından, Prof. Dr. Kamil ALAKUŞ danışmanlığında hazırlanan “MAKİNE ÖĞRENMESİ KULLANARAK İNSAN HAREKETLERİNİN SINIFLANDIRILMASI: AKILLI TELEFON ALGILAYICILARI ÖRNEĞİ ” başlıklı bu çalışma, jürimiz tarafından 20.2.2023 tarihinde yapılan sınav sonucunda oy birliği ile başarılı bulunarak Yüksek Lisans Tezi olarak kabul edilmiştir.

	Unvanı Adı Soyadı Üniversitesi Ana Bilim/Ana Sanat Dalı	Sonuç
Başkan	Prof. Dr. Kamil ALAKUŞ Ondokuz Mayıs Üniversitesi İstatistik Ana Bilim Dalı	☒ Kabul ☐ Ret
Üye	Doç. Dr. Hasan BULUT Ondokuz Mayıs Üniversitesi İstatistik Ana Bilim Dalı	☒ Kabul ☐ Ret
Üye	Doç. Dr. Tolga ZAMAN Çankırı Karatekin Üniversitesi İstatistik Ana Bilim Dalı	☒ Kabul ☐ Ret

Bu tez, Enstitü Yönetim Kurulunca belirlenen ve yukarıda adları yazılı jüri üyeleri tarafından uygun görülmüştür.

Prof. Dr. Ahmet TABAK
Enstitü Müdürü

BİLİMSEL ETİĞE UYGUNLUK BEYANI

Hazırladığım Yüksek Lisans tezinin bütün aşamalarında bilimsel etiğe ve akademik kurallara riayet ettiğimi, çalışmada doğrudan veya dolaylı olarak kullandığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin Kaynaklar'da gösterilenlerden oluştuğunu, her unsurun enstitü yazım kılavuzuna uygun yazıldığını ve TÜBİTAK Araştırma ve Yayın Etiği Kurulu Yönetmeliği'nin 3. bölüm 9. maddesinde belirtilen durumlara aykırı davranılmadığını taahhüt ve beyan ederim.

Etik Kurul Gerekli mi?

Evet ☐ (Gerekli ise ekler kısmına ekleyiniz)

Hayır ☒

xx / xx / 2023

Dina ALMOMAYEZ

TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI

Tez Başlığı: MAKİNE ÖĞRENMESİ KULLANARAK İNSAN HAREKETLERİNİN SINIFLANDIRILMASI: AKILLI TELEFON ALGILAYICILARI ÖRNEĞİ

Yukarıda başlığı belirtilen tez çalışması için şahsım tarafından 06.01.2023 tarihinde intihal tespit programından alınmış olan özgünlük raporu sonucunda;

Benzerlik oranı : % 14

Tek kaynak oranı : % 3 çıkmıştır.

06 / 01 / 2023

Prof. Dr. Kamil ALAKUŞ

ÖZET

MAKİNE ÖĞRENMESİ KULLANARAK İNSAN HAREKETLERİNİN SINIFLANDIRILMASI: AKILLI TELEFON ALGILAYICILARI ÖRNEĞİ

Dina ALMOMAYEZ

Ondokuz Mayıs Üniversitesi

Lisansüstü Eğitim Enstitüsü

İstatistik Ana Bilim Dalı

Yüksek Lisans, Ocak/2023

Danışman: Prof. Dr. Kamil ALAKUŞ

Giyilebilir teknolojideki gelişmelerle birlikte ortaya çıkan cihazlar hızla gündelik hayatın bir parçası haline gelmiştir. Özellikle sahip oldukları algılayıcılar, bu cihazların kullanışlılığını artırmaktadır. İnsan Aktivitesi Tanıma (İAT), insan hareketlerinden etkilenen duyarlı sensörler kullanarak bir kişinin aktivitesini sınıflandırır. İnsan Aktivitesi Tanıma özelliğinin akıllı telefonlarda olması nedeniyle hem kullanıcıların deneyimleri hem de akıllı telefonların yeteneklerini (sensörleri) artırmaktadır ve kullanıcılar bu nedenle akıllı telefonlarını yanlarında taşırlar. Bu gerçekler İAT'ı daha önemli ve popüler hale getirmektedir. İnsan Aktivitesi Tanıma (İAT), akıllı telefonlar ve video kameralar gibi yakalama cihazlarının küresel kullanımı ve bunların insan aktivitesi verilerini yakalama yetenekleri nedeniyle birden fazla uygulamaya sahiptir. Elektronik cihazların ve uygulamalarının sayısı artmaya devam ederken, yapay zekadaki (AI) ilerlemeler, doğru tanıma ve yorumlama için derinden gizlenmiş bilgileri çıkarma becerisinde devrim yarattı. İnsan Aktivitesi Tanıma (İAT), son yirmi yılda uzaktan sağlık izleme, güvenlik ve gözetim ve akıllı ortamlar gibi uygulamalarla büyük ilgi gördü. Bu çalışma, farklı makine öğrenimi sınıflandırma yaklaşımlarını kullanan akıllı telefon sensörlerini kullanarak insan etkinliğinin tanınmasına odaklanmaktadır. Akıllı telefonlarda bulunan ivmeölçer ve jiroskop sensörlerinden alınan veriler, insan aktivitesinin tanınmasında kullanılır. Bu çalışma amacı akıllı telefonun sahip olduğu sensörlerden yararlanarak insan hareketlerinin tespit edilmesidir. Bu nedenle 18 farklı insan hareketi için algılayıcı verileri akıllı telefon aracılığı ile toplanmıştır ve oluşturulan veriler Lojistik Regresyon, Karar Ağacı, Destek Vektör Sınıflandırma ve Rasgele Orman makine öğrenmesi yöntemleri ile test edilmiş ve performansları karşılaştırılmıştır. Elde edilen sonuçlar, Karar Ağacı ve Lojistik Regresyon alt örneklemi yaklaşık %99 doğrulukla sınıflandırabildiğini göstermektedir. Bu veri seti, akıllı telefonların ve diğer akıllı cihazların yeni etkinlikleri belirlemesini sağlayacak ve araştırmacıların pratik HAR verilerine dayalı daha ince modeller geliştirmelerine yardımcı olacaktır.

Anahtar Sözcükler: Makine öğrenme, Akıllı telefon, Etkinlik tanıma, Sınıflandırma

ABSTRACT

CLASSIFICATION OF HUMAN MOVEMENTS USING MACHINE LEARNING: THE EXAMPLE OF SMARTPHONE SENSORS

Dina ALMOMAYEZ
Ondokuz Mayıs University
Institute of Graduate Studies
Department of Statistic
Master, January/2023
Supervisor: Prof. Dr. Kamil ALAKUŞ

Devices that emerged with the developments in wearable technology have quickly become a part of everyday life. Especially the sensors they have increase the usefulness of these devices. Human Activity Recognition (HAR) classifies a person's activity using sensitive sensors that are affected by human movements. Since the Human Activity Recognition feature is on smartphones, it increases both the users' experience and the capabilities (sensors) of the smartphones, and users carry their smartphones with them for this reason. These facts make IAT more important and popular. Human Activity Recognition (HAR) has multiple applications due to the global use of capture devices such as smartphones and video cameras and their ability to capture human activity data. As the number of electronic devices and their applications continues to grow, advances in artificial intelligence (AI) have revolutionized the ability to extract deeply hidden information for accurate recognition and interpretation. Human Activity Recognition (HAR) has received a lot of attention over the past two decades with applications such as remote health monitoring, security and surveillance, and smart environments. This study focuses on the recognition of human activity using smartphone sensors using different machine learning classification approaches. Data from accelerometer and gyroscope sensors in smartphones are used to recognize human activity. The aim of this study is to detect human movements by using the sensors of the smart phone. For this reason, sensor data for 18 different human movements were collected via smart phone and the generated data were tested with Logistic Regression, Decision Tree, Support Vector Classification and Random Forest machine learning methods and their performances were compared. The results show that Decision Tree and Logistic Regression can classify sub-samples with approximately 99% accuracy. This dataset will enable smartphones and other smart devices to identify new activities and help researchers develop finer models based on practical HAR data.

Keywords: Machine learning, Smartphone, Activity recognition, Classification

ÖN SÖZ VE TEŞEKKÜR

Akademik eğitim sürecimin bir üst noktası olan yüksek lisans tez çalışmalarım boyunca yardım ve desteğini benden esirgemeyen danışman hocam Sayın Prof. Dr. Kamil Alakuş'a teşekkürü bir borç bilirim.

Tez çalışmam boyunca benden yardımlarını ve desteklerini esirgemeyen İstatistik Bölümü akademisyenlerine ve arkadaşlarıma teşekkür ederim.

Beni bugünlere getirmek için hiçbir fedakarlıktan kaçınmayan anneme, başarılarımı uzaklardan izlediğine inandığım rahmetli babama sonsuz teşekkür ederim.

Dina ALMOMAYEZ

İÇİNDEKİLER

TEZ KABUL VE ONAYI.....	i
BİLİMSEL ETİĞE UYGUNLUK BEYANI.....	ii
TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI.....	ii
ÖZET.....	iii
ABSTRACT	iv
ÖN SÖZ VE TEŞEKKÜR	v
İÇİNDEKİLER.....	vi
SİMGELER VE KISALTMALAR.....	vii
ŞEKİLLER DİZİNİ	ix
TABLOLAR DİZİNİ	x
1. GİRİŞ.....	1
1.1. Literatür Araştırması.....	2
2. MATERYAL VE METOT.....	8
2.1. Makine Öğrenimi	8
2.1.1. Makine Öğrenimi Tanımı.....	8
2.1.2. Makine Öğrenimi İhtiyacı	8
2.1.3. Makine Öğreniminin Nedenleri ve Zamanı	8
2.1.4. Makine Öğreniminin Modelleri	9
2.1.5. Makine Öğrenimi Uygulamaları.....	11
2.2. Kullanılan Makine Öğrenme Yöntemleri	11
2.2.1. Denetimli Makine Öğrenme Yöntemleri.....	11
2.2.1.1. Lojistik Regresyon	12
2.2.1.2. Destek Vektörü Sınıflandırıcı (DVS)	13
2.2.1.3. Karar Ağaçları (KA)	14
2.2.1.4. Rastgele Orman Algoritması	15
2.2.2. Değerlendirme Teknikleri	16
3. UYGULAMA VE BULGULAR.....	18
3.1. Python Tarihçesi ile Makine Öğrenmesinin Kullanımı	18
3.2. Veri Kümesi	19
3.3. Uygulama Yöntemi	20
3.4. Sınıflandırma Sonuçları	26
4. SONUÇ VE ÖNERİ	29
KAYNAKLAR	30
ÖZ GEÇMİŞ	33

SİMGELER VE KISALTMALAR

İAT (HAR)	: İnsan Aktivitesi Tanıma (Human Activity Recognition)
DVS	: Destek Vektör Sınıflandırıcısı
MÖ	: Makine Öğrenimi
AR	: Activity Recognition
CNN	: Evrişimsel Sinir Ağları (Convolutional Neural Network)
VGG-19	: Görsel Geometri Grubu, (19, katman sayısını belirtmektedir)
FSR	: Güce Duyarlı Direnç (Force Sensitive Resistor)
EYK (KNN)	: K-En Yakın Komşu (K Nearest Neighbour)
DVM	: Destek Vektör Makinesi
PCA	: Temel Bileşen Analizi (Principal Component Analysis)
LSTM	: Uzun Kısa Süreli Bellek (Long Short-Term Memory)
DNN	: Derin Öğrenme (Deep Neural Network)
YOLO	: Sadece Bir Kez Bak (You Only Look Once)
DVM	: Destek Vektör Makinesi
RBF	: Radyal Tabanlı Fonksiyon (Radial Basis Function)
BiLSTM	: Çift Yönlü Uzun Kısa Süreli Bellek
CNN-GRU	: Evrişimli Sinir Ağları - Geçitli Yineleme Birimi (Convolutional Neural Network- Gated Recurrent Unit)
WISDM	: Kablosuz Sensör Veri Madenciliği (Wireless Sensor Data Mining)
AUTOML	: Otomatik Makine Öğrenmesi (Automated Machine Learning)
HMM	: Saklı Markov Modelleri (Hidden Markov Model)
NB	: Naive Bayes
LR	: Lojistik Regresyon
KA (DT)	: Karar Ağacı (Decision Trees)
RO(RF)	: Rastgele Orman (Random Forest)
MLP	: Çok Katmanlı Algılayıcı (Multilayer Perceptron)
DNN	: Derin Öğrenme (Deep Neural Network)
CNN BiLSTM	: Hybrid Bidirectional LSTM and CNN Architecture
IOT	: Nesnelerin İnterneti (Internet of Things)
XGBOOST	: eXtreme Gradient Boosting
UKSB	: Birleşik Krallık Biyomalzemeler Topluluğu
ConvLSTM	: Evrişimsel Uzun Kısa Süreli Bellek

UCI HAR	: Human Activity Recognition Using Smartphones
P	: Performans
G	: Görev
D	: Deneyim
TP	: Gerçek Olumlu (True Positive)
TN	: Gerçek Olumsuz (True Negative)
FB	: Yanlış Olumlu (False Positive)
FN	: Yanlış Negatif (False Negative)
KU-HAR	: Khulna University- Human Activity Recognition (İnsan Aktivitesi Tanıma)



ŞEKİLLER DİZİNİ

Şekil 2.1. Makine Öğrenimi Modeli	10
Şekil 2.2. Karışıklık matrisini ve bileşenleri.....	16
Şekil 3.1. Lojistik Regresyon yönteminden elde edilen karmaşıklık matrisi.....	20
Şekil 3.2. Doğrusal SVC'nin (Destek Vektör Sınıflandırıcısı) yönteminden elde edilen karmaşıklık matrisi.....	22
Şekil 3.3. Karar Ağacı yönteminden elde edilen karmaşıklık matrisi	24
Şekil 3.4. Rastgele orman yönteminden elde edilen karmaşıklık matrisi	25



TABLÖLAR DİZİNİ

Tablo 3.1. Kesinlik deęerleri	26
Tablo 3.2. Duyarlılık deęerleri	27
Tablo 3.3. F-skoru deęerileri	28



1. GİRİŞ

Günümüzde insanların günlük yaşamını akıllı telefonsuz hayal etmek neredeyse imkânsız hale gelmiştir. Çoğu kişi akıllı telefonları araştırma yapmak, film izlemek, oyun oynamak ve sosyal topluluğa erişmek için kullanır; bununla birlikte akıllı telefonlarla ilgili birçok yararlı çalışma da vardır. Etkinlik tanıma (AR), kalp atışı, kandaki oksijen miktarı gibi sağlık verilerini takip edebilme, kaybolma ve yere düşmenin algılanması, bağlamsal mobil uygulamalara erişim, GPRS takibi yapılabilmesi, ev otomasyonu ve diğerleri gibi birçok akıllı telefon özelliğinin arkasındaki temel teknolojiler bunlardan bazılarıdır. Akıllı telefon tabanlı etkinlik tanıma sistemleri, yeni tür akıllı uygulamalara yol açabileceği için önemli bir çalışma alanıdır.

Bu akıllı cihazların günlük yaşamımızdaki kombinasyonu hızla gelişmektedir. Örneğin, bu tür akıllı cihazlar günlük aktivitelerimizi sorunsuz bir şekilde takip edebileceği için bundan sonra ne yapacağımız konusunda daha iyi kararlar vermemize yardımcı olacaktır. Bu durum, ortam zekasına dayanmaktadır. Bu çalışmada, akıllı teknoloji aktivite uygulamaları ile insan etkinliğini tanımak için bir akıllı telefon uygulaması örneği verilecektir.

Etkinlik tanıma, bir kişi tarafından gerçekleştirilen hareketleri, kendisi ve çevresindeki ortamla ilgili bir dizi gözleme dayalı olarak tanımlamaya çalışır. Örneğin, ivmeölçerler gibi atalet sensörlerinden elde edilen veriler kullanılarak tanıma yapılabilir. Bazı akıllı cihazlarda, bu sensörler varsayılan olarak entegre edilmiştir. Entegre edilen sensörler, donanım için denetimli bir makine öğrenimi algoritması (MÖ) kullanılarak atalet çerçeve göstergelerini kaydedebilmektedir. Kaydedilen bir dizi fiziksel etkinliği (ayakta durma, yatma, yürüme, yukarı ve aşağı gitme gibi) sınıflandırmak için sınırlı kaynaklarla kullanabilmektedir.

Sensörler, akıllı telefonların işlevselliğini ve çevre bilincini artırmada büyük bir rol oynar. Kullanıcıların günlük hayatı ve faaliyetleri hakkında bilgi toplamaya olanak tanıyan akıllı telefonlar, çeşitli yerleşik sensörlerden oluşmaktadır. Bu cihazlar ivmeölçerler ve jineskop sensörleri içerir.

İvmeölçer, neredeyse tüm cep telefonu üretimlerinde kullanılan standart bir parçadır. İvmeölçer, hızın kendisini değil hızdaki değişimi ölçmektedir. İvmeölçerden elde edilen bilgiler, hareketteki ani değişiklikleri algılamak için işlenebilir. Akıllı

telefonlarda standart olarak kullanılan bir diğer sensör ise yerçekimini kullanarak yönü belirleyen jiroskoptur. Jiroskop tarafından toplanan sinyaller, cihazın konumunu ve hizalamasını tespit etmek için kaydedebilir. Bu sensörlerden elde edilen verilerde değişik birçok fonksiyon olduğu belirlenmiştir. Bu nedenle cihazı giyen veya takan kişinin davranışını belirlemek için bu sensör verilerinden farklı opsiyonlar oluşturulabilir.

Bu çalışmada, bir akıllı telefonun ivmeölçer ve jiroskopundan gelen sinyallerden oluşan elde edilen veriler ile dört farklı makine öğrenmesi yöntemi: Lojistik Regresyon, DVS'nin (Destek Vektör Sınıflandırıcısı), Karar Ağacı, Rastgele orman ile sınıflandırılmış ve sonuçları karşılaştırılmıştır.

1.1. Literatür Araştırması

Sağlık sektöründe rehabilitasyon hizmetlerinin teknolojik gelişimine ve yapay zekanın sağlık perspektifinden hastalara ve sağlık personellerinin faydalarına odaklanmışlardır. Araştırma sonucunda, rehabilitasyon sürecinde yapay zeka kullanımının hem yönetim hem de sağlık hizmetlerinin kalite ve verimliliğinin artırılması açısından zamansal, mekânsal ve maddi birçok avantajı beraberinde getirdiğini söylemişlerdir (Akalin ve Demirbaş, 2022).

Spor salonlarının video sınıflandırması için derin öğrenmeye dayalı bir sınıflandırma modeli geliştirilmiştir. CNN uygulamalı transfer öğrenme ile geliştirilen model, VGG-19 ile. 32827 çerçeve üzerinde CNN ve VGG-19 modelleri kullanılarak yapılan deneysel çalışmalar, VGG-19'un doğruluk açısından CNN'den daha başarılı bir sınıflandırmaya sahip olduğunu göstermiştir. Deneysel sonuçlar, geliştirilen modelin literatürde yapılan çalışmalara göre daha başarılı sonuçlara sahip olduğunu göstermiştir (Gençaslan, 2022).

Akıllı telefonlardan toplanan jiroskoplar ve ivmeölçerler kullanarak çeşitli iç ve dış ortamlardaki aktivitelerini izleyerek yaşlılara yardım etmekle ilgilidir. Akıllı telefonlar, engellilerin faaliyetlerini izlemek için düzenli olarak kullanılmaktadır. Oturma, yürüme, merdiven çıkma, ayakta durma ve uzanma gibi rutin aktiviteler veri setinde yer almaktadır. İnsan etkinliğini tespit etmek için k-en yakın komşular, rastgele orman, destek vektör makinesi, yapay sinir ağı ve uzun süreli kısa süreli bellek ağı gibi geleneksel makine öğrenimi ve derin öğrenme algoritmaları kullanılır (Hayat, 2022).

Günlük aktivite olan yürüme, sola ve sağa eğilme, arabayı geri çekme, çömelme, inme, merdiven çıkma, koşma ve düşme dahil olmak üzere on üç yaygın insan hareketini algılayan ayakkabıların iç tabanlarına entegre edilmiş bir ayak tabanı basınç sensör sistemi (P2S2) sunmaktadır. Çeşitli hareketler sırasında basınç değişikliğinin elektriksel imzasını yakalamak için tabanlıklardaki kritik basınç noktalarına altı kuvvete duyarlı direnç (FSR) yerleştirilmiştir. Toplam 34 yetişkin katılımcı P2S2 ile test edilmiştir. Basınç verileri, k-NN, sinir ağları ve destek vektör makinesi (DVM) algoritmaları dahil olmak üzere makine öğrenme algoritmalarını (ML) beslemek için temel bileşen analizi (PCA) kullanılarak toplanmış ve işlenmiştir. MÖ modelleri dörtlü çapraz doğrulama ile eğitilmiştir. Her aşamada, deneklerin verileri diğer aşamalarda bağımsız tutulmuştur. Modelin %86 doğrulukla etkili olduğu tespit edilmiş ve ayakkabılara entegre edilmiş P2S2 kullanarak insan hareketlerini tahmin etmede umut verici sonuçlar göstermiştir (Anderson, 2022).

Sınıflandırma için akıllı sensörler ve çeşitli makine öğrenimi yöntemlerini kullanarak insan aktivitesinin tespitine odaklanmaktadır. Akıllı telefonun ivmeölçeri ve jiroskopu tarafından toplanan veriler, insan etkinliği tanıma olarak sınıflandırılır. Kullanılan yaklaşımların sonuçları verimlilik ve doğruluk açısından karşılaştırılmıştır (Gade, 2022).

Farklı makine öğrenimi algoritmaları ile optimize edilmiş özellikleri kullanarak insan etkinliği algılamayı uygular. Bu amaçla t-dağılımlı stokastik komşuluk gömme tekniği kullanılmış ve sonuçlar elde edilmiştir. Geniş bir öznitelik uzayı nedeniyle varyansı azaltarak yanlış sınıflandırma hata oranını azaltmada öznitelik indirgemenin rolü araştırılmış ve analiz edilmiştir. Önerilen çalışma, sorunu anlamak için veri görselleştirmeye de odaklanmaktadır (Abdulhussein, 2022).

HAR tabanlı akıllı telefon sensör problemlerini çözmek için Ensem-HAR dediğimiz CNN ve LSTM tabanlı dört derin öğrenme modelini tanıtmıştır. Yaygın ve herkese açık üç veri kümesine bakmış; önerilen modelin bu veri kümelerinde iyi performans gösterdiği ve özellikleri iyi bir doğrulukla tahmin etme konusunda övgüye değer bir iş çıkardığını vurgulamıştır. Bazı durumlarda, özellikler arasındaki yüksek korelasyon yanlış sınıflandırmaya yol açsa da çalışmada belirtilen veri kümelerine uygulanan birkaç yeni yöntemden daha iyi performans gösterebilmiştir (Bhattacharya, 2022).

Bu makale, nesneyi gerçek dünya ortamlarında kısıtlama olmaksızın tanıyan gelişmiş bir akıllı telefon tabanlı yürüyüş tanıma sistemi sunmaktadır. Önerilen model, 7-10 günlük bir süre boyunca toplanan 44 kullanıcıdan oluşan denetimsiz yürüyüş veri setine duyulandırmış, kullanıcılardan yalnızca günlük aktivitelerini gerçekleştirmeleri istendi. Katılımcılara belirli faaliyetlerle ilgili hiçbir koşul, kontrol veya beklenti dayatılmamıştır. Deneyin bir parçası olarak dört tip hareket modellenmiştir: her bir kullanıcı için normal yürüme, hızlı yürüme ve aşağı ve yukarı yürüme. Önerilen modelin değerlendirilmesinde normal, hızlı, aşağı, yukarı ve tüm aktiviteler için sırasıyla %11,38, %11,32, %24,52, %27,33 ve %15,08'lik tek tip hata oranı elde edilmiştir. Sonuçlar, yürüyüş tanımanın makul sınırlar içinde gerçek dünyada kullanım için uygun bir teknik olduğunu göstermektedir (Alobaidi, 2022).

Derin sinir ağı (DNN) kullanarak iki farklı fiziksel aktiviteyi, yürüme ve hızlı yürümeyi sınıflandırmışlardır. Birden fazla cep telefonu sensörü kullanılarak her iki fiziksel aktivite için kişisel veriler toplanmıştır. İvmeölçer, jiroskop, manyetometre vb. gibi cep telefonu sensörleri toplandı, incelendi ve DNN modelini eğitmek ve test etmek için kullanılmıştır. Bir cep telefonundaki sayısız sensör arasında, verilen hareketlere/işlemlere en uygun sensörleri belirlemişlerdir. Bu çalışma sonucunda %96,5'lik bir sınıflandırma doğruluğu elde etmişlerdir (Khan, 2022).

Suçlu/şüpheli davranış görüntülerinde şüpheli davranışı tespit etmek için gerçek zamanlı bir nesne algılama aracı olan YOLO kullanılmış ve başarılı sonuçlar elde edilmiştir. YOLO'nun araştırması, temel olarak günlük makine öğreniminde ve görünür nesnelerin tanınmasında kullanılır, bu da YOLO'nun davranış tanımda uygulanabilirliğini destekler. Nesne algılama aracı YOLO, şüpheli etkinliği algılamada fazlasıyla başarılı olmuştur (Çalışkan ve Demir, 2022).

Makine öğrenmesi sınıflandırma yöntemlerinden SVM, KNN ve Naive Bayes olmak üzere üç farklı yöntem kullanılmış ve sonuçlar karşılaştırılmıştır. Sınıflandırma performansını etkilemesi beklenen farklı normalleştirme ve çapraz doğrulama yöntemleri kullanılmıştır. Kullanılan bu yöntemlerin sınıflandırma performansı üzerindeki etkileri karşılaştırılmış ve yorumlanmıştır. Parkinson hastalığının tespitinde yürüyüş verilerinin kullanılmasının önemli ve karakteristik sonuçlar sağladığını göstermiştir (Altuntaş, 2022).

İnsan aktivitesi tespiti için en iyi algoritmayı bulmanın yolunu aramıştır ve veri analizi için lojistik regresyon, SVM ve RBF çekirdeği, CNN, LSTM, çift yönlü LSTM

ve CNN-LSTM algoritmaları kullanılmıştır. Veri analizinde doğruluk ve eğitim süresi ölçülerek karşılaştırılmıştır, CNN-LSTM algoritması ve Çift Yönlü LSTM algoritması, doğruluk özelliğinde en fazla sonucu vermiştir. Eğitim süresi içinde, RBF kernel algoritmali SVM en az eğitim süresine sahipken, CNN-LSTM en fazla eğitim süresine sahiptir. En iyi seçim LSTM'dir (Moradi, 2022).

Akıllı telefonun dokunmatik paneline, yerçekimine, doğrusal ivmeye ve jiroskopik verilere ve ayrıca yazma davranışına bakmıştır. Çalışma, akıllı telefon yazma özelliklerini kullanarak stres tespiti için etkili bir aykırı veri temizleme ve özellik seçme yöntemi sunmaktadır. Çalışmada en başarılı yöntem kNN olmuştur. Uzaklık tabanlı yapılarına bağlı olarak Mahalanobis uzaklık tabanlı aykırı değer tespiti ve k-NN yönteminin iyi eşleştiği sonucuna ulaşılmıştır (Sağbaşı, 2021).

Akıllı telefon tabanlı doğrudan geçişli kimlik doğrulamanın veri güvenliği gücünü gerçekçi bir şekilde ölçmek amaçlanmıştır. Bir cep telefonu uygulaması kullanılarak 50 farklı denekten ivmeölçer verileri toplanmıştır, farklı sınıflandırma algoritmaları kullanılarak oluşturulmuş ve Subspace KNN kullanılarak %97,2 oranında başarılı sınıflandırma elde edilmiştir. Cinsiyete göre sınıflandırmada %99,7 başarı oranı elde edilmiştir (Aydemir ve Karşoğlu, 2021).

Bireylere bağlı sensörlerden gelen insan hareket sinyallerini tahmin etme sağlık görevini ele almaktadır. Hareket sinyallerini tekrarlayan grafik yöntemini kullanarak görüntü formatlarına dönüştürmüş ve derin öğrenme modeline girdi olarak kullanmıştır, yöntemin sonuçları, CNN modelinin sonuçlarından daha iyi performans gösteriyor ve ayrıca, tahmin edilen hareket sinyalleri aracılığıyla potansiyel kazara düşmeleri daha hızlı tespit etmek için kullanılabileceğini doğrulanırmış (Kim, 2021).

İnsan eylemi tespiti için elverişli ve kapalı tekrarlayan birimleri birleştiren yeni bir hibrit derin sinir ağı modeli (CNN-GRU) önermektedir. Bu model, WISDM veri kümesinde başarıyla doğrulanmıştır ve AutoML ile oluşturulmuş Inception Time ve Deep ConvLSTM gibi diğer son teknoloji derin sinir ağı modellerinden önemli ölçüde daha iyi bir sonuç elde etmiştir (Gupta, 2021).

Makine Öğrenimi ve derin öğrenme yaklaşımlarının ilerlemesi ile akıllı telefon ve akıllı saatteki donanımlı sensörleri kullanan insan etkinliği tanıma ile ilgili sorunları çözmek için uygulanan metodolojiler hakkında bir inceleme sunmaktadır (Lateef ve Abbas, 2021).

kNN, SVM, HMM SVM, SVM kNN, NB, LR, DT, RF, MLP, DNN, LSTM, CNN-LSTM, CNN BiLSTM gibi birçok klasik makine ve derin öğrenme algoritmasını araştırmıştır ve karşılaştırmıştır. İnsan aktivitesi tespiti için ResNet. Jiroskoplar ve ivmeölçerler gibi sensörlerden toplanan bağlamsal kullanıcı eylemlerini içeren bu yazıda, insan etkinliği tanıma için farklı klasik makine öğrenimi ve derin öğrenme teknikleri araştırılmış ve karşılaştırılmıştır. En iyi tanıma performansına sahip sınıflandırıcıyı seçmek için teknikler arasında karşılaştırmalı bir analiz yapılır. Deneysel sonuçlar, kurulan Derin Sinir Ağı (DNN) modelinin, halka açık UCI-HAR veri setinde %96,81'e varan doğruluk ve 0,03'e varan ortalama mutlak hataya ulaştığını göstermektedir. Bu yöntem, bu çalışmada gerçekleştirilen insan etkinliği tahmin sınıflandırma yöntemleri arasında en iyi performansı göstermiştir (Bozkurt, 2021).

Cep telefonu ve cihazın ivmeölçeri için Android yazılımı kullanılarak IoT tabanlı insan hareketi algılaması yapılmıştır. İlk olarak, veriler toplanmıştır ve algılanabilir hareketler için ön işleme tabi tutulmuştur. Bu özellik, elde edilen veri kümesinden çıkarılmıştır. Destek Vektör Makineleri, Rastgele Orman ve K En Yakın Komşu algoritmaları kullanılarak yapılan hareketler, elde edilen bilgilere göre sınıflandırılmıştır. Sınıflandırma başarıları belirlenmiş ve en başarılı sınıflandırma algoritması, web tabanlı bir uygulama aracılığıyla nesneleri gerçek zamanlı olarak sınıflandırmak için kullanılmıştır (Erin, 2020).

Çalışmalarında, yalnızca akıllı verileri kullanan stres algılama çalışmalarını gözden geçirmekte ve sınıflandırmada kullanılan veri kaynakları, veri türleri ve makine öğrenmesi yöntemlerini ele almıştır. Bu çalışmalar veri kaynaklarına göre beş başlık altında incelenmiştir. Araştırma sonucunda akıllı telefon uygulamaları, hareket sensörleri, arama sıklığı ve kısa mesaj gibi bilgilerin stresi algılamada önemli rol oynadığı tespit edilmiştir (Sağbaş, 2020).

Çalışmalarında, insan hareket algılaması için ivmeölçer verilerinden zaman ve frekans alanlarında çeşitli öznitelikler çıkarılmıştır. Amaç, elde edilen öznitelikleri bir topluluk sınıflandırıcısı ile değerlendirerek hataları belirlemektir. Sınıflandırma aşamasında tek bir sınıflandırıcı kullanmak yerine parçacık sürüsü optimizasyonu ile beş sınıflandırıcıdan en uygun olanı seçilmiş ve sonuçlar birleştirilmiştir (Aydın ve Aşıcı, 2020).

Araştırmanın bir parçası olarak insan performans takip sistemi tanıtılmıştır. WISDM (Kablosuz Sensör Veri Madenciliği) ham verileri, sürekli zaman dalgacık

fonksiyonlarına ve gizli Markov modellerine dayalı özellikleri çıkarmak ve bu özelliklerin insan etkinliği tespiti üzerindeki etkisini incelemek için kullanılmıştır. Bu çalışma çerçevesinde, köklü makine öğrenmesi algoritmaları kullanılarak gerçek zamanlı bir uygulama örneği gerçekleştirilmiştir. Test sonuçları ile gerçek zamanlı uygulama arasındaki farklar araştırılmış ve bu nedenle insan aktivitesi tespitinde karşılaşılan sorunlar tartışılmış ve çözüm önerileri getirilmiştir (Eren, 2019).

Verilen çalışmada kullanılan materyal, 19-48 yaşları arasında günlük yaşamlarını sürdüren 30 gönüllünün verilerini içermektedir. Veriler, yerleşik atalet sensörlerine sahip, bele takılan bir akıllı telefonda elde edilmiştir. Her denek beline Samsung Galaxy S II akıllı telefon takarak “yürüme”, “merdiven tırmanma”, “merdiven tırmanma”, “oturma”, “ayakta durma” ve “ulaşma” olmak üzere altı aktivite gerçekleştirmiştir. Bu veri setinde K En Yakın Komşu, Rastgele Orman, Ekstra Ağaçlar ve XGBoost sınıflandırma algoritmaları kullanılarak oluşturulan modellerin performansı 5 kat çapraz doğrulama tekniği ile değerlendirilmiştir. Bu teknik ile eğitim verileri bu modellere 5 kez giriş verisi olarak gönderilmiş ve modellerin performansı test verilerine karşı analiz edilmiştir. Modellerin performansı ortalama doğruluk ölçülerek karşılaştırılmıştır. Elde edilen sonuçlar, tüm modellerin başarılı bir performans göstermesine rağmen XGBoost tabanlı modelin 0.80 ortalama doğrulukla diğer algoritmalara göre daha başarılı olduğunu göstermiştir (Akyol ve Karacı, 2019).

UKSB hücreleri, oluşturulan veri seti kullanılarak ivmeölçer ve jiroskop verileri toplanarak akıllı telefonlarda oluşturulan altı insan eylemini sınıflandırmak için eğitilmiş ve test edilmiştir. Bu testlerde UKSB hücreleri farklı giriş parametreleri almış ve en yüksek başarı oranı 0.91 olmuştur (Metin ve Karasulu, 2019).

Farklı makine öğrenimi sınıflandırma yöntemlerini kullanan akıllı telefon sensörleri ile insan etkinliği tespitine odaklanmıştır. Akıllı telefon ivmeölçerlerinden ve jiroskoplardan gelen veriler, insan etkinliği tanıma olarak sınıflandırılmıştır. Kullanılan yaklaşımların sonuçları verimlilik ve doğruluk açısından karşılaştırılmıştır (Bülbül, 2018).

2. MATERYAL VE METOT

2.1. Makine Öğrenimi

2.1.1. Makine Öğrenimi Tanımı

Makine öğrenimi (MÖ), bilgisayar sistemlerinin verileri insanlara benzer şekillerde anlamlandırmasına izin veren bilgisayar bilimi alanıdır. Basit bir ifadeyle makine öğrenimi, ham verilerden kalıpları çıkarmak için bir algoritma veya yöntem kullanan bir tür yapay zekadır. Makine öğreniminin temel merkezi, bilgisayar sistemlerinin açık bir biçimde programlanmadan ya da insan müdahalesine gerek kalmaksızın tecrübe edilerek öğrenilebilmesidir.

2.1.2. Makine Öğrenimi İhtiyacı

İnsanlar düşünebildikleri, değerlendirebildikleri ve karmaşık problemleri çözebildikleri için dünyadaki en zeki ve gelişmiş canlılardır. Öte yandan, yapay zeka henüz başlangıç aşamasında ve bu yüzden birçok açıdan insan zekasını henüz aşamamıştır. Bu da bir makine neden öğrenmeli sorusunu ortaya çıkarmaktadır. Bu sorunun en iyi cevabı ise verimli ve ölçekli verilere dayalı kararlar almasıdır diyebiliriz.

Son zamanlarda şirketler, gerçek dünyadaki çeşitli görevleri ve sorunları çözmek maksadı ile verilerden maksimum verimli bilginin çıkarılması için yapay zeka, derin öğrenme ve makine öğrenimi gibi yeni teknolojiler için ciddi yatırımlarda bulunmaktadır. Bu yüzden özellikle sürecin hızlandırılması için makinelerin verdiği kararlar veri odaklı olarak değerlendirilebilir. Veri odaklı bu kararlar, doğası gereği programlanamamakta olan sorunlar için programlama mantığı gibi kullanılabilmektedir. Esasında insan zekası olmazsa olmazdır fakat bir başka açıdan insanlığın sorunları büyük ölçekte verimlilik ile çözmesi gerekmektedir. Bu yüzden makine öğrenimi hayatımızın her aşamasında gereklidir.

2.1.3. Makine Öğrenmenin Nedenleri ve Zamanı

Makine öğrenimi, başka bir soruyu ortaya çıkarmaktadır. Makinenin hangi senaryolarda öğrenmesini sağlamamız gerekiyor? Veriye dayalı kararları verimli ve uygun ölçekte vermek için makinelere gereksinim duyulan durumlar ortaya çıkabilir. Makinelerin öğrenmesine izin verilmesi daha etkili sonuçlar almayı sağlayabilir. Bu durumların bazılarını aşağıda yer verilmektedir:

- **İnsan uzmanlığı eksikliği**

Bir makineye ait öğrenme ve veriye dayanan kararlar almasına ilişkin ilk senaryo, insan uzmanlığının yeterli olmadığı alanlardır. Örneğin, makine öğrenmesinin, erişilemeyen alanlar veya uzamsal gezegenlere ait alanlarda karar alabilmesidir.

- **Dinamik senaryolar**

Dinamik bir doğaya sahip olan, yani zaman içinde değişiklik gösteren birtakım senaryolar söz konusudur. Söz konusu senaryolar ile davranışlar için bir makinenin öğrenmesi, veri temelli karar alması gerekir.

- **Uzmanlığı Hesaplamalı Görevlere Dönüştürmede Zorluk**

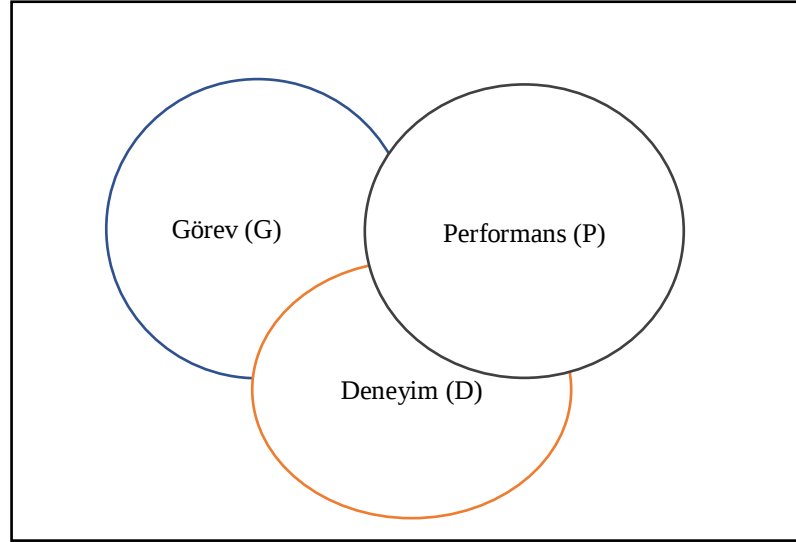
İnsanların uzmanlığa sahip olduğu ancak bu uzmanlığı hesaplama görevlerine çeviremediği birkaç alan olabilir. Bu gibi durumlarda makine öğrenimine ihtiyacımız vardır. Örnekler arasında konuşma tanıma, bilişsel görevler yer almaktadır.

2.1.4. Makine Öğreniminin Modelleri

Makine öğrenim modelinin tartışmasından önce Profesör Mitchell'in verdiği ve aşağıda yer verilen resmi ve makine öğrenimi tanımının anlaşılması oldukça önemlidir. “Bir bilgisayar programı P ile ölçülen T'deki görevlerdeki performansı, E deneyimiyle geliyorsa, bir görev sınıfı (T) ve bir performans ölçüsü (P) ile ilgili olarak E deneyiminden öğrenir.” Yapılan bu tanımlamada esasında bir öğrenme, algoritmasının temel bileşenleri olan deneyim (D), performans (P) ve görev (G) parametrelerine odaklanmaktadır. Tanımlama bu açıdan daha sade ifade edilebilir: Makine öğrenimi, şunları sağlayan öğrenme algoritmalarından meydana gelen bir yapay zeka alanıdır:

- Performanslarını geliştir (P)
- Bazı görevleri yerine getir (G)
- Deneyim (D)

Buna dayanarak aşağıda verilen diyagram, bir Makine Öğrenimi Modelini temsil etmektedir:



Şekil 2.1. Makine Öğrenimi Modeli

Şimdi bunları daha ayrıntılı olarak tartışalım.

- **Görev (G):** problem bakımından G, görevini çözümümesi gereken gerçek dünya problemi olarak ifade edilebilir. Problemi belli bir yerde bulunan evler arasında en iyi fiyata sahip olanın bulunması ya da en iyi pazarlama stratejisinin tespiti gibi bir şey olabilmektedir. Ancak, makine öğreniminden bahsettiğimizde, görevin tanımı farklıdır çünkü makine öğrenimini çözmek zordur. Geleneksel bir programlama yaklaşımı kullanan tabanlı görevler. Bir görev G, sistemin veri noktalarıyla çalışmak için izlemesi gereken süreci temel alıyorsa makine öğrenimi tabanlı görev olarak adlandırılır. Makine öğrenimi tabanlı görev örnekleri arasında kümeleme, yapılandırılmış ek açıklama, sınıflandırma, transkripsiyon ve regresyon bulunmaktadır.

- **Deneyim (D):** Deneyim, modele ya da algoritmaya sağlanan veri noktalarından sağlanan bilgiyi ifade etmektedir. Veri setinin sağlanması sonrasında model yinelemeli olarak bazı içsel kalıpları yürütmekte ve öğrenmektedir. Bu biçimde sağlanan bilgi deneyim olarak ifade edilmektedir. İnsan öğrenmesine benzer biçimde bu durum insanın çeşitli niteliklerinden öğrenmesi ve deneyim elde etmesi olarak düşünülebilir. Takviyeli, denetimsiz ve denetimli öğrenme, deneyim kazanmanın yöntemlerindendir. Makine öğrenimi modeli ya da algoritmadan sağlanan denetim, G görevinin çözümlenmesinde kullanılmaktadır.

- **Performans (P):** Bir makine öğrenimi algoritması bir görevi gerçekleştirmeli ve zaman içinde deneyim kazanmalıdır. Bir makine öğrenimi algoritmasının

beklendiği gibi çalışıp çalışmadığının bir ölçüsü, performansdır. P temel olarak modelin bir görevi ne kadar iyi gerçekleştirdiğini söyleyen nicel bir ölçüdür, G deneyimine bağlıdır, D. kesinlik puanı F-skoru, karışıklık matrisi, kesinlik, hatırlama gibi makine öğrenimi performansını anlamaya yardımcı olan birçok ölçü vardır. Bu, ölçüye hassasiyet örnek olarak verilebilir.

2.1.5. Makine Öğrenimi Uygulamaları

Makine öğrenimi hızlı büyüyen teknolojidir ve araştırmacılara göre yapay zeka ve MÖ'nün en parlak döneminde yaşadığımız söylenebilmektedir. Geleneksel yaklaşımlarla çözilemeyen birçok karmaşık gerçek dünya problemini çözmek için kullanılır. ML'nin bazı uygulamaları aşağıda yer almaktadır.

- Duygu analizi,
- Hata algılama ve önleme,
- Hava tahmini,
- Borsa analizi ve tahmini,
- Konuşma sentezi,
- Konuşma tanıma,
- Müşteri segmentasyonu,
- Nesne tanıma,
- Dolandırıcılık tespiti,
- Dolandırıcılık önleme,
- Online alışverişte müşteriye ürün tavsiyesi.

2.2. Kullanılan Makine Öğrenme Yöntemleri

Makine öğrenimi sınıflandırma yöntemleri, insan vücuduna yerleştirilen akıllı telefonlar veya sensörler kullanılarak insan hareketlerini sınıflandırmak için kullanılır. Bu çalışmada kullanılan makine öğrenmesi yöntemleri alt başlıklar halinde açıklanmıştır.

2.2.1. Denetimli Makine Öğrenme Yöntemleri

Örneklerden öğrenme olarak da adlandırılan denetimli makine öğrenimi, doğru yanıtlarla bir eğitim veri kümesinden öğrenme sağlar. Söz konusu algoritma, olası tüm girdilere doğru yanıt verecek şekilde genelleştirir. Bu nedenle, tüm denetimli sınıflandırma yöntemleri başlangıçta eğitim verileri için bazen ulaşılması zor olabilen

doğru amaçlara sahip olmalıdır. Girdi verilerinin tüm olası örneklerine sahip olunsaydı, bunları büyük bir arama tablosuna koyabilmek mümkün olurdu ve böylece makine öğrenimine hiç gerek kalmayacaktı. Fakat hayatımızda böyle bir durum söz konusu değildir.

Yapılan bu çalışmada, yöntemlerin performanslarını karşılaştırmak amaçlanmıştır. Seçilen Python kütüphanesi scikit-learn, çok çeşitli yöntemleri desteklemektedir ve bu çalışma kapsamında hepsini incelemek mümkün değildir. Bunun yerine, Lineer modeller, Destek vektör makineleri gibi en azından temel yöntemleri farklı bir perspektiften seçmek ve incelemek gerekmektedir. Bu çalışmada kullanılan denetimli makine öğrenme yöntemleri; Lojistik regresyon, Destek vektörü sınıflandırıcı (DVC) ve Karar ağaçları (KA) gibi yöntemlerdir. Bu yöntemler alt başlıklarda kısaca açıklanmıştır.

2.2.1.1. Lojistik Regresyon

Lojistik regresyon, kategorik bir bağımlı değişkenin olasılığını tahmin etmek için kullanılan makine öğrenimi sınıflandırma algoritmasında kullanılan bir yöntemdir.

İstatistikte lojistik regresyon, başarı-başarısızlık, kazanma-kaybetme, canlı-ölü veya sağlıklı-hasta gibi belirli bir kategori veya olayın olasılığını modellemek için kullanır. Lojistik regresyon tahmine dayalı bir modeldir. Elde edilen verileri tanımlamaya ve bağımlı ikili değişken ile çok kategorili değişkenler arasındaki ilişkiyi açıklamaya yardımcı olur. Lojistik regresyon bir sigmoid işlevi kullanır. İşlev, olasılıksal modellere dayalıdır ve her olayın olasılığı,

$$Pr(Y_i = 1|X_i) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)} \quad (2.1)$$

eşitliği ile hesaplanır.

Lojistik regresyon (logit model), uygulama verileri oturma, ayakta durma, uzanma, yürüme, merdiven inip çıkma gibi insan aktivitelerinin olasılığını modellemek için kullanılmaktadır. Amaç vektör Y tanımlayıcıları, özellik matrisi X sensör değerlerini ve kompanzasyon değeri 1 'i temsil eder. En iyi model rasgele arama yöntemi ile elde edilir. Kabul edilen on sekiz β değeri için parametreler 50'den 400'e kadar 50'lik adımlarla değişir. Rastgele arama yöntemi, modeli eğitmek için farklı parametreler kullanmalıdır ve ardından en iyi modeli oluşturulan sonuçla

karşılaştırılır. Lojistik regresyon modelini oluşturmak için kullanılan kayıp fonksiyonu, tahmin edilecek 18 kategori olduğundan çapraz entropi kategorisi kullanılmıştır.

Anomali tespiti için sınıflandırılması için en uygun modellerden biri Lojistik regresyondur. Lojistik regresyon, bağımlı değişkenlere dayalı bir algoritmadır. Günümüzün en popüler ve en çok kullanılan algoritmalarından biridir. Bu isimde regresyon terimi geçmektedir ancak lojistik regresyon bir sınıflandırma algoritmasıdır. Muhtemel uygulamalardan biri, logistic regresyon sınıfında Python kütüphanesi scikit-learn tarafından gerçekleştirilmiştir.

2.2.1.2. Destek Vektörü Sınıflandırıcısı (DVS)

Destek vektörü sınıflandırıcısı (DVS), sınıflandırma için kullanılan denetimli makine öğrenimi yöntemidir. Scikit-learn kütüphanesinde de uygulanan Destek vektör makineleri (DVM'ler) kümesinden biridir.

DVM'ler 1992'de Vapnik tarafından önerilmiştir ve esas olarak büyük veri kümelerinde çok daha iyi sınıflandırma performansları nedeniyle popüler olmuştur. Pahalı bir hesaplama işlemi olan veri matrisinin tersinin bulunmasını içerdiklerinden, büyük veri kümelerinde iyi çalışmazlar. Algoritma, veri kümesini sınıflara bölen bir hiperdüzlem veya bir dizi hiperdüzlem oluşturur. İki boyutta ve iki sınıfta, hiperdüzlem basitçe bir karar sınırını oluşturan çizgidir. En yakın nokta ile karar sınırı arasındaki mesafeye marj denir. Karar sınırına en yakın her sınıftaki veri noktalarına destek vektörleri denir. DVM için en iyi hiperdüzlem, destek vektöründeki veri noktaları arasında en büyük mesafeye sahip olandır.

DVM'ler, yalnızca sınıflandırma için değil, aynı zamanda regresyon ve aykırı değer tespiti için de bir dizi yöntem içerir. Genellikle 20 destek vektörü olarak adlandırılan karar fonksiyonundaki eğitim noktalarının bir alt kümesi, bu yöntemlerin hafıza verimliliğini sağlar. DVM'ler, karar işlevi için birkaç çekirdek işlevini destekler. Kullanıcı tanımlı bir çekirdek işlevi de kullanılabilir.

DVC sınıfı, scikit-learn kütüphanesi tarafından DVM'lerin somut bir uygulamasıdır. DVC sınıfının çekirdeği adlı parametrelerle ayarlanabilen çekirdek işlevi aracılığı ile yöntem tarafından başka bir karar sınırı bulunur. Örnek olarak, özelliklerden birinin doğrusal bir çizgisi veya bir polinomu verilebilir. Scikit-learn'de DVM'lerin sağlanan

çekirdek işlevleri, doğrusal ve polinom yapıdadır. Burada polinom fonksiyonunun derecesi, DVC sınıfının derecesine sahip parametre tarafından belirtilir.

DVC yöntemi, diğer denetimli yöntemlere benzer şekilde, uyum yönteminin girdisi olarak eğitim örneklerini tutan X dizisini ve etiketleri tutan Y dizisini alır. Bizim durumumuzda, Y dizisi yalnızca iki sınıfın değerlerini içerir.

2.2.1.3. Karar Ağaçları (KA)

Karar ağaçları (KA), sınıflandırma ve regresyon için kullanılan parametrik olmayan denetimli bir öğrenme yöntemidir. Veri özelliklerinden çıkarılan basit karar kurallarını öğrenerek bir amaç değişkenin değerini tahmin eden bir model oluşturulur.

Karar ağacı analizi genel itibarıyla farklı pek çok alanda kullanıma uygun, tahmine dayanan bir modelleme aracıdır. Karar ağaçları, farklı şartlara uygun biçimde veri setinin bölünmesini sağlayan algoritmik bir yaklaşımla meydana getirilebilir. Karar ağaçları, denetimli algoritmalar kategorisinde yer alan en güçlü algoritmalar. Hem regresyon hem sınıflandırma görevlerinde kullanıma uygundur. Verilerin bölündüğü karar düğümleri ile sonuçların yer aldığı yapraklar, bir ağacın iki bileşenidir.

Karar ağaçlarının anlaşılması, yorumlanması ve görselleştirilmesi kolaydır. Verileri karar ağacına göre tahmin etmenin maliyeti, ağacı eğitmek için kullanılan veri kaydı sayısı açısından logaritmiktir. Karar ağaçlarının en önemli dezavantajlarından biri, verilerdeki küçük değişikliklerin sonucu olabilen, ağaçlar arasındaki önemli farklılıkları içerir. Çözüm, daha sonra açıklanacak olan topluluk yöntemlerinden birini kullanmaktır. Ağacın en yüksek derinliğini ayarlayarak çözülebilecek başka bir sorun da aşırı uyum olabilir.

Karar düğümleri, herhangi bir karar vermek için kullanılır ve birden fazla şubeye sahiptir, oysa yaprak düğümleri bu kararların çıktısıdır ve başka şube içermez.

$$Entropi(x) = - \sum P(x) \log_2 [P(x = k)] \quad (2.2)$$

$P(x = k)$, bir hedef özelliğin belirli bir k değeri alma olasılığını belirtir.

Gini İndeksi, her bir sınıfın olasılıklarının karelerinin toplamının birden çıkarılmasıyla hesaplanır. Elde edilen bilgiler, daha büyük bölümlerin tercih edildiği ve uygulanmasının daha kolay olduğu Gini Index'in aksine, farklı değerlere sahip daha küçük bölümleri tercih eder.

$$Gini = 1 - \sum [P(x = k)]^2 \quad (2.3)$$

2.2.1.4. Rastgele Orman Algoritması

Rastgele orman algoritması regresyon ve sınıflandırmada kullanılan denetimli bir öğrenme algoritması olup söz konusu algoritma temelde sınıflandırma sorunlarında tercih edilmektedir. Bir orman ağaçlardan meydana gelmektedir ve ağacın daha fazla olması daha sağlam bir ormanı ifade eder. Rastgele orman algoritması da benzer şekilde veri örnekleri üzerinde karar ağaçları oluşturur. Oluşturulan her bir karar ağacından tahmin alır ve sonunda oylama yoluyla en iyi çözümü seçer. Sonucun ortalamasını alarak uyumu azalttığı için tek bir karar ağacından daha iyi olan bir topluluk yöntemidir.

Rastgele orman algoritması, scikit-learn learning kütüphanesinde yer alan rastgele karar ağaçlarına dayalı torbalama algoritmalarından biridir. Rastgele orman algoritması, özellikle ağaçlar için geliştirilmiş bir perturb-and-combine tekniğidir. Sınıflandırıcıya rastgelelik eklenerek, bir dizi sınıflandırıcı oluşturulur. Bireysel tahminler topluluğu, ortalama bir tahmin üretmek için birleştirilir. Bununla birlikte, rastgele orman sınıflandırıcısı, bu çalışmada eğitim için kullanılan torbalama algoritmalarının somut ve denetimli bir uygulamasıdır.

Algoritma, scikit-learn Python kitaplığındaki Rastgele Orman Sınıflandırıcısı sınıfı aracılığıyla kullanılabilir. Kurucunun n tahmin edicileri parametresi, eğitim setinden bir toplulukta oluşturulan karar ağaçlarının sayısını belirtmek için kullanılabilir. Sınıflandırıcı ayrıca, eğitim örneklerini içeren seyrek veya yoğun bir X dizisi ve 22 etiketi içeren bir Y dizisi ile sağlanmalıdır. Her ağacın inşası sırasında, her bir düğümün en iyi bölünmesi ya tüm girdi özelliklerinden ya da yapıcının parametrelerinden max boyutundaki özelliklerin rastgele bir alt kümesinde bulunur. Bunun nedeni, orman tahmincilerinin varyansını azaltmaktır. Dahili bir düğümü bölmek için gereken minimum örnek sayısı, minörnek bölünmesi adı verilen bir parametre ile tanımlanır. Bir karar ağacı, dahili düğümü yinelemeli olarak bölümlere ayırır.

2.2.2. Değerlendirme Teknikleri

Makine öğrenimi modeli performansını ölçmek için çeşitli yöntemler kullanılır:

- Karışıklık matrisi (Confusion matrix)
- Sınıflandırma Raporu
 1. Doğruluk (Accuracy)
 2. Kesinlik (Precision)
 3. Duyarlılık (Recall)
 4. F1-Değeri (F1-Measure)

• **Karışıklık matrisi:** Bu yöntem sayesinde eğitilen tüm modelleri karşılaştırmak mümkündür. Problem bir sınıflandırma problemi olarak kabul edildiğinde, önce dört temel kriter sunulmalıdır: öğretilen modelin doğruya karşı yanlış ve pozitif karşı negatif. Örnek olarak, pozitif (1) ve negatif (0) sınıfları olan bir ikili atama alabiliriz. Bir modelin gerçek pozitif (TP), pozitif sınıfı doğru bir şekilde tahmin eden sonuçtur ve gerçek negatif (TN), tahmin edilen şiddetli negatif sınıftır. Yanlış pozitif (FP), pozitif sınıfı yanlış tahmin eden bir modelin sonucudur ve yanlış negatif (FN), yanlış tahmin edilen negatif sınıftır.

Karışıklık matrisi, bir sınıflandırma modelinin bilinen gerçek değerlere sahip test verileri üzerindeki performansını tanımlamak için yaygın olarak kullanılan bir tablodur.

		Gerçek	
		Pozitif	Negatif
Tahmin	Pozitif	TP	FP
	Negatif	FN	TN

Şekil 2.2. Karışıklık matrisi ve bileşenleri

Yukarıdaki şekil, karışıklık matrisini ve bir karışıklık matrisinin bileşenlerini göstermektedir.

• **Sınıflandırma Raporu:** Sınıflandırma raporu, hatırlama, F1 puanı ve kesinlikten oluşmaktadır. Sınıflandırma raporu, sınıflandırma modellerini karşılaştırmak için kullanılır.

1. Doğruluk (Accuracy): Doğruluk değeri, modeldeki doğru tahmin edilen alanların tüm veri setine oranına göre hesaplanır. Eğer biz 100 mağazadan oluşan bir kitlede 40 mağazanın churn edip etmediğini doğru tahmin edersek Doğruluk değerimiz %40 olacaktır

$$\text{Doğruluk} = \frac{TP+TN}{TP+FB+TN+FN} \quad (2.4)$$

2. Kesinlik (Precision): Pozitif olarak tahmin ettiğimiz kaç değerın gerçekte pozitif olduğunu gösterir.

$$\text{Kesinlik} = \frac{TP}{TP+FB} \quad (2.5)$$

3. Duyarlılık (Recall): Aktivitenin ne kadarını pozitif olarak tahmin etmemiz gerektiğini ölçer.

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (2.6)$$

4. F1 puanı(f1-skoru): F1 değeri bize Kesinlik (Precision) ve Duyarlılık (Recall) değerlerinin harmonik ortalamasını göstermektedir. Basit bir ortalama yerine harmonik ortalama olmasının sebebi ise uç durumları göz ardı etmememiz gerektiğidir. Eğer basit bir ortalama hesaplaması olsaydı Kesinlik değeri 1 ve Duyarlılık değeri 0 olan bir modelin F1 Değeri 0.5 olarak gelecektir ve bu yanıltıcı olacaktır.

$$F1 = 2 * \frac{\text{Kesinlik} * \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (2.7)$$

3. UYGULAMA VE BULGULAR

3.1. Python Tarihçesi ile Makine Öğrenmesinin Kullanımı

Python, üst düzey bir dilin yeteneklerine sahip popüler bir nesne yönelimli programlama dilidir. Öğrenmesi kolay sözdizimi ve taşınabilirlik özellikleri onu günümüzde erişilebilir kılmaktadır. Aşağıdaki durumlar bize Python'a bir giriş sağlamaktadır.

- Python, Hollanda'daki Stichting Mathematisch Centrum'da Guido van Rossum tarafından geliştirilmiştir.
- 'ABC' adlı programlama dilinin ardılı olarak yazılmıştır.
- İlk versiyon 1991'de yayınlanmıştır ve Python adı Guido van Rossum tarafından Monty Python's Flying Circus (TV) adlı bir gösteriden seçilmiştir.
- Açık kaynaklı bir programlama dilidir, yani onu özgürce indirebilir ve program geliştirmek için kullanabilmektedir.
- Python programlama dili, Java ve C özelliklerine sahiptir. #C koduna sahiptir ve diğer yandan nesne yönelimli programlama için Java gibi sınıflar ve nesneler bulunmaktadır.
- Yorumlanmış bir dildir, yani bir Python programının kaynak kodu önce bayt kodu haline getirilir ve daha sonra Python sanal makinesi tarafından yazılmaktadır.

Bu çalışmada uygulanan tüm araçlar için Python programlama dili seçilmiştir. Bu karar, dilin kullanımının kolay olması, serbestçe erişilebilir olması, kullanım alanı için birçok faydalı kütüphanenin bulunması ve bilimsel hesaplama için varsayılan bir dil haline gelmesi nedeniyle alınmıştır. Python üst düzey bir kitap dili olduğu sürece, araçlar işletim sistemi türüne bağlı değildir. Tanınmış makine öğrenimi kitaplıklarının çoğu Python hakkında yazılmıştır.

Scikit-learn, açık kaynaklı bir makine öğrenimi kitaplığıdır. Model uydurma, veri ön işleme gibi araçlarla birlikte denetimli ve denetimsiz makine öğrenimi için birçok model sağlar. Bu çalışma için en uygun kütüphanedir. Bu kitaplık, numpy (bilimsel bilgi işlem) ve scipy (matematik, bilim ve mühendislik) kitaplıklarının üzerine inşa edilmiştir. Bu kitaplık, model kalıcılığı sunmaktadır. Böylece eğitilmiş bir model kolayca bir dosyaya kaydedilebilmektedir ve herhangi bir zamanda veri tahmini

için yüklenebilir. Bazı güvenlik ve bağımlılık sınırlamaları vardır, ancak eğitilen modeli kaydetmek ve gerektiğinde kullanmakta fayda vardır.

3.2. Veri Kümesi

Bulunan KU-HAR veri seti, toplanan HAR. İnsan Aktivitesi Tanıma (IAT), makinelerin insan hareketlerini tanıma yeteneğini ifade eder. Bu veri seti, akıllı telefon sensörleri (ivmeölçer ve jiroskop) kullanılarak 75'i erkek 15'i kadın toplam 90 katılımcı tarafından kaydedilen 18 farklı aktivite hakkında bilgi içermektedir. Doğrudan katılımcılardan toplanan 1945 ham aktivite örneğini ve onlardan çıkarılan 20750 alt örneği içermektedir (data.mendeley, 2020).

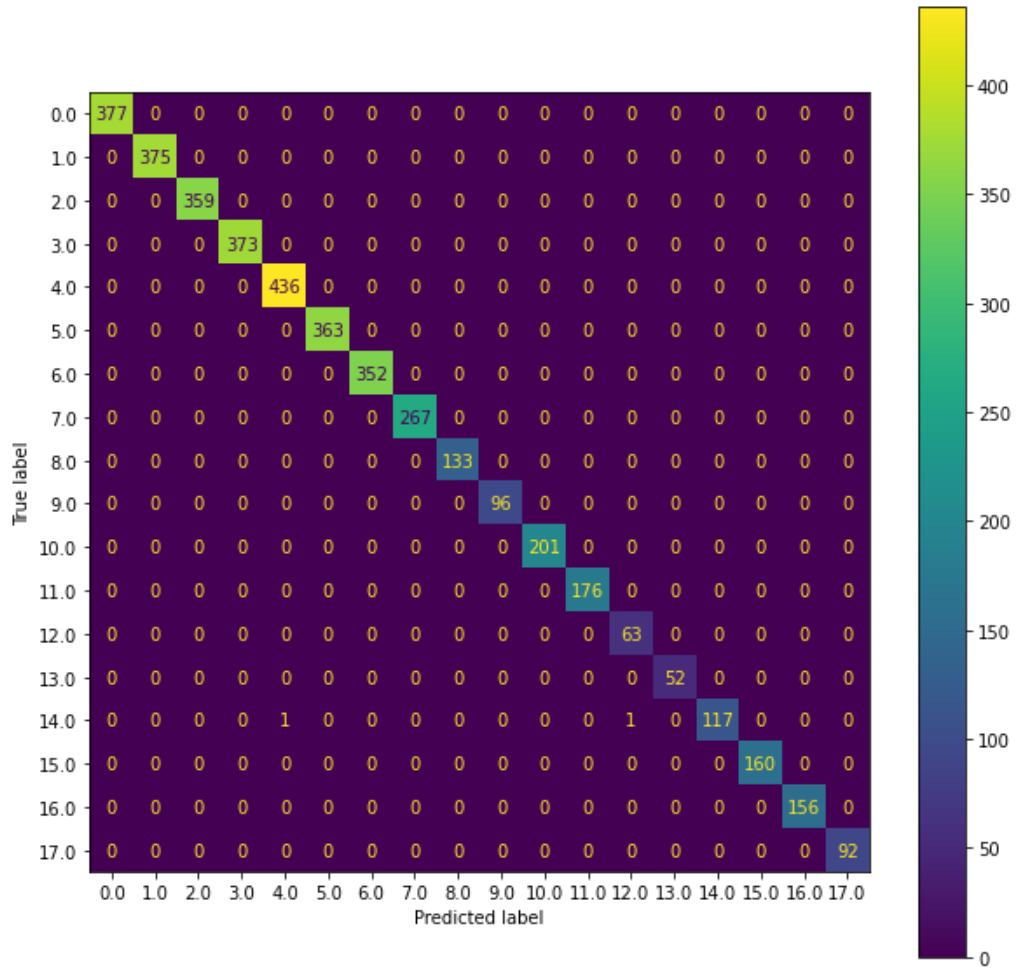
Aktiviteler:

0. Durmak (Stand) → Ayakta dururken (1 dakika)
1. Oturmak (Sit) → Hareketsiz oturma (1 dakika)
2. Konuş-otur → Otururken el hareketleriyle konuşma (1 dakika)
3. Konuşmak- durmak → Ayakta dururken veya yürürken el hareketleriyle konuşmak (1 dakika)
4. Ayağa kalk-otur → Tekrar tekrar ayağa kalkıp oturarak (5 kez)
5. Sermek → Hareketsiz bırakma (1 dakika)
6. Uzanma → Tekrar tekrar ayağa kalkıp uzanma (5 kez)
7. Seçmek (Pick) → Yerden bir nesne almak (10 kez)
8. Zıplamak → Art arda atlama (10 kez)
9. Şınav → Tam şınav çekme (5 kez)
10. Mekik → Mekik yapma (5 kez)
11. Yürü → Yürü 20 metre (≈ 12 s)
12. Geriye yürüme → 20 metre geriye doğru yürüme (≈ 20 s)
13. Yürüyüş dairesi → Dairesel bir yolda yürümek (≈ 20 s)
14. Koş → Koşu 20 metre (≈ 7 sn)
15. Merdiven yukarı → Bir dizi merdivene çıkma (≈ 1 dakika)
16. Merdiven-aşağı → Bir dizi merdivenden inme (≈ 50 s)

17. Masa tenisi → Masa tenisi oynamak (1 dakika)

3.3. Uygulama Yöntemi

Bu çalışmada, bir akıllı telefonun ivme ölçer ve jiroskopundan gelen sinyallerden oluşan bir veri setinden elde edilen veriler dört farklı makine öğrenmesi yöntemi (Lojistik Regresyon, Doğrusal SVC'nin (Destek Vektör Sınıflandırıcısı), Karar Ağacı, Rastgele orman) ile sınıflandırmıştır ve sonuçları karşılaştırılmak için Python 3.8 kullanılmıştır. Her bir yönetim için *karmaşıklık matrisi* hesaplanmıştır.



Şekil 3.1. Lojistik Regresyon yönteminden elde edilen karmaşıklık matrisi

Lojistik Regresyon performansını ölçmek için, doğruluk (Accuracy), kesinlik (Precision), duyarlılık (Recall), F1 Değeri hesaplanmıştır.

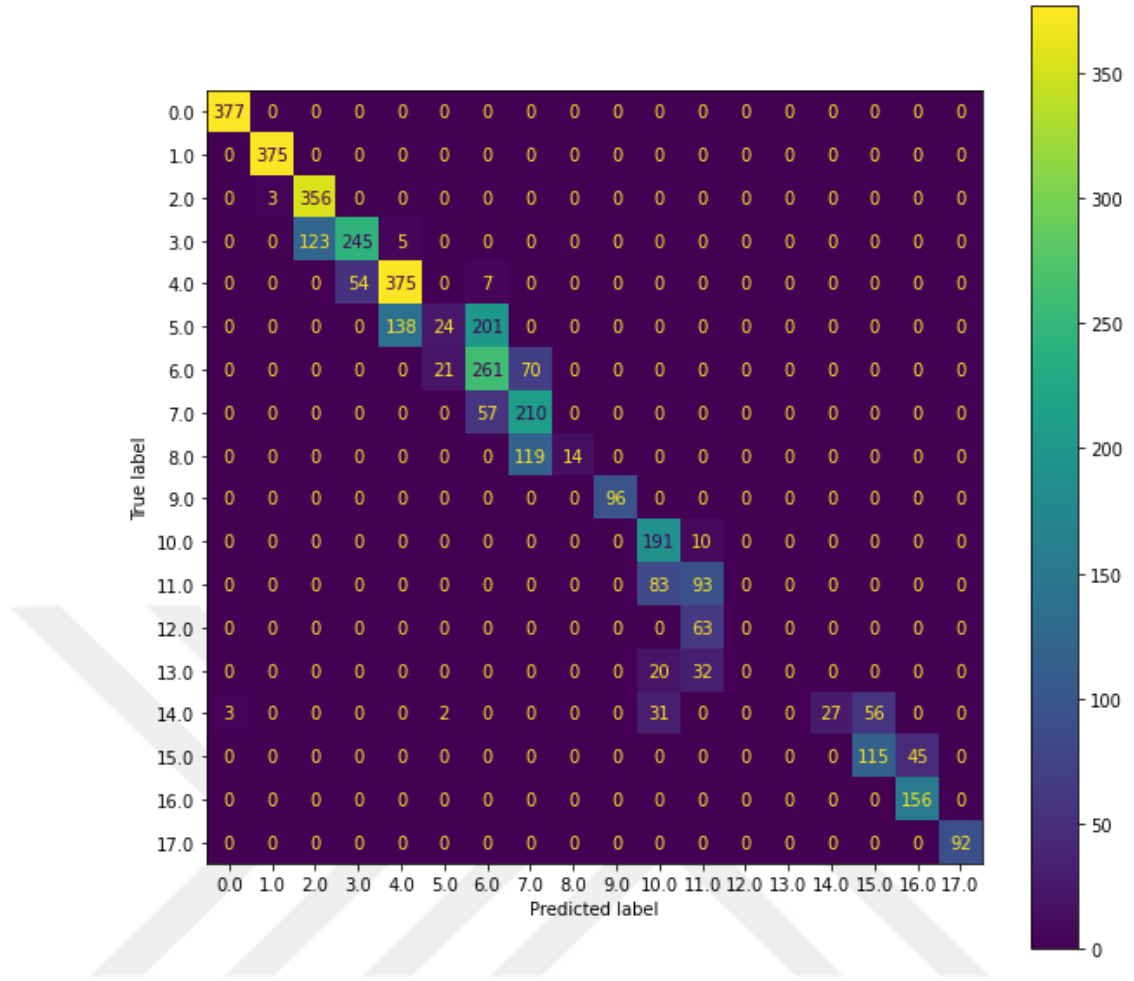
Örneğin 2. aktivitede (2.0) kesinlik değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.5 numaralı formül kullanılmıştır ve TP, FB değerini tablodan anlayabiliriz. TB= 359, FB= 0, sonuç =1 ifade etmektedir. Diğer aktivitelerde bu

şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.1’de gösterilmiştir.

Örneğin 14. aktivitede (14.0) duyarlılık değeri hesaplamak için yukarıdaki 2.2.2 bölümünde 2.6 numaralı formül kullanılmıştır ve TP, FN değerini tablodan anlayabiliriz. TB= 117, FN= 2, sonuç= 0.98 ifade etmektedir. Diğer aktivitelere bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.2’de gösterilmiştir.

Örneğin 17. aktivitede (17.0) F1 değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.7 formülü kullanılmıştır ama öncesi kesinlik değeri ve duyarlılık değeri bulmamız lazım. Kesinlik değeri için 2.5 formülü kullanılmıştır, sonuç = 1 elde edilmiştir. Duyarlılık değeri için 2.6 formülü kullanılmıştır, sonuç= 1 elde edilmiştir. 2.7 formülü kullanılmıştır sonuç=1 elde edilmiştir. Diğer aktivitelere bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.3’de gösterilmiştir.

Bu modelin doğruluk değeri bulmak için 2.4. numaralı formülü kullanarak sonuç 0.999 elde edilmiştir.



Şekil 3.2. DVS'nin (Destek Vektör Sınıflandırıcısı) yönteminden elde edilen karmaşıklık matrisi

Doğrusal DVC'nin (Destek Vektör Sınıflandırıcısı) performansını ölçmek için doğruluk (Accuracy), kesinlik (Precision), duyarlılık (Recall), F1 değeri hesaplamak gerekmektedir.

Örneğin 4. aktivitede (4.0) kesinlik değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.5 numaralı formül kullanılmıştır ve TP, FB değerini tablodan anlayabiliriz. TB= 375, FB= 5+138= 143, 0.72 sonucuna ulaşılmaktadır. Diğer aktivitelerde bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.1'de gösterilmiştir.

Örneğin 15. aktivitede (15.0) duyarlılık değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.6 numaralı formül kullanılmıştır ve TP, FN değerini tablodan anlayabiliriz. TB= 115, FN= 45, sonuç= 0.72 ifade etmektedir. Diğer aktivitelerde bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.2'de gösterilmiştir.

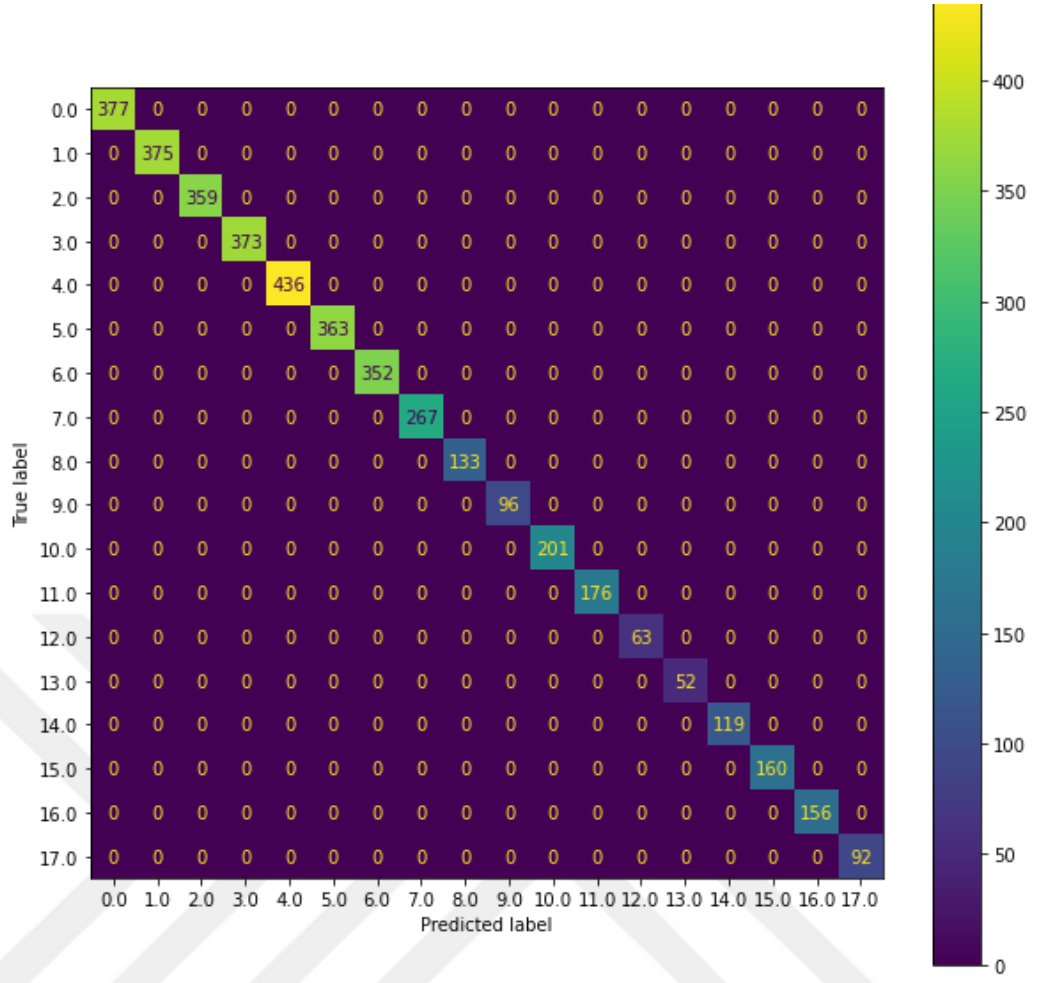
Örneğin 16. aktivitede (16.0) F1 değeri hesaplamak için yukardaki 2.2.2 bölümünden 2.7 formülü kullanılmıştır ama öncesi kesinlik değeri ve duyarlılık değeri bulmamız lazım. Kesinlik değeri için 2.5 formülü kullanılmıştır, sonuç= 0.78 elde edilmiştir. Duyarlılık değeri için 2.6 formülü kullanılmıştır, sonuç= 1 elde edilmiştir. 2.7 formülü kullanılmıştır, sonuç= 0.87 elde edilmiştir. Diğer aktivitelere bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.3’de gösterilmiştir.

Bu modelin doğruluk değerini bulmak için 2.4 numaralı formülü kullanarak sonuç 0.72 elde edilmiştir.

Karar Ağacı performansını ölçmek için doğruluk (Accuracy), kesinlik (Precision), duyarlılık (Recall), F1 değeri hesaplanmalıdır.

Örneğin 5. aktivitede (5.0) kesinlik değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.5 numaralı formül kullanılmıştır ve TP, FB değerini tablodan anlayabiliriz. TB= 363, FB= 0,1 sonucuna ulaşılmaktadır. Diğer aktivitelere bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.1’de gösterilmiştir.

Örneğin 11. aktivitede (11.0) duyarlılık değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.6 numaralı formül kullanılmıştır ve TP, FN değerini tablodan anlayabiliriz. TB=176, FN=0, sonuç=1 ifade etmektedir. Diğer aktivitelere bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.2’de gösterilmiştir.



Şekil 3.3. Karar Ağacı yönteminden elde edilen karmaşıklık matrisi

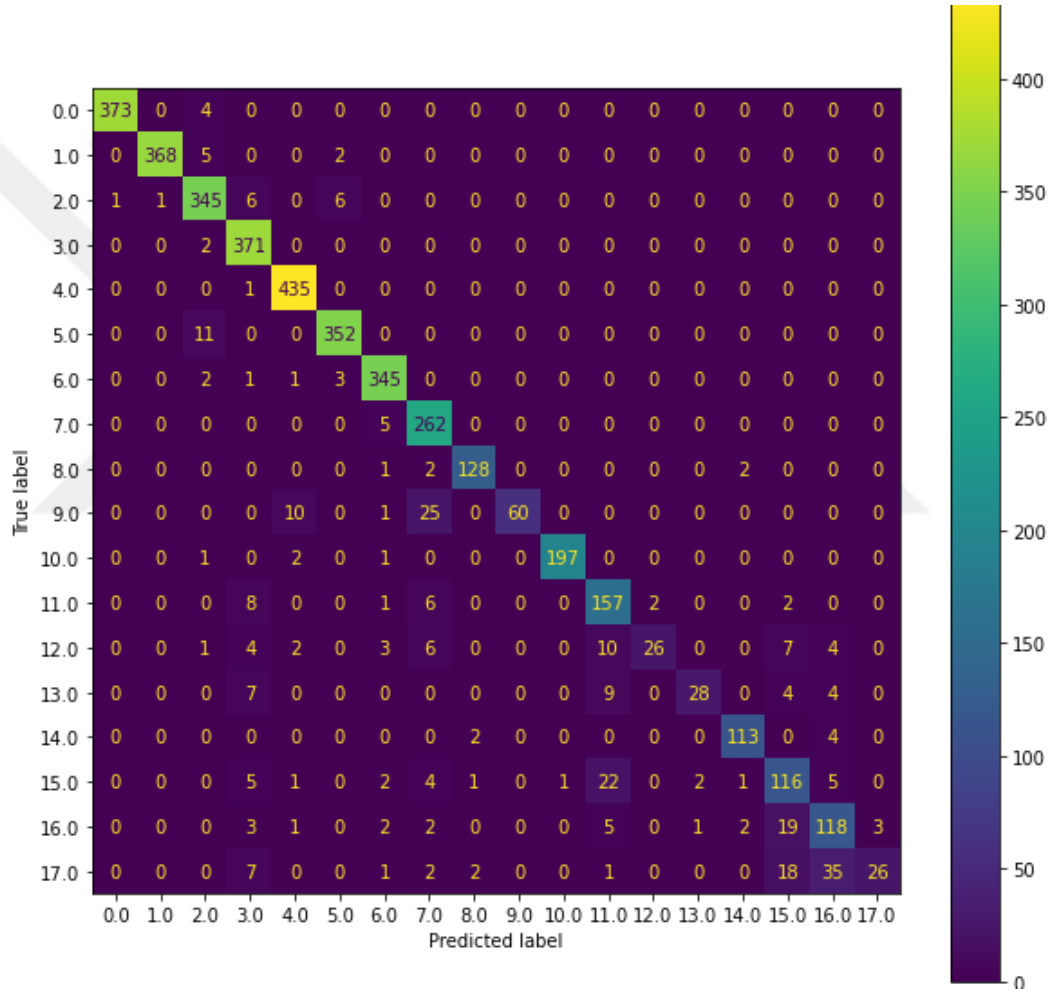
Karar Ağacı performansını ölçmek için doğruluk (Accuracy), kesinlik (Precision), duyarlılık (Recall), F1 değeri hesaplanmalıdır.

Örneğin 5. aktivitede (5.0) kesinlik değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.5 numaralı formül kullanılmıştır ve TP, FB değerini tablodan anlayabiliriz. TB= 363, FB= 0,1 sonucuna ulaşılmaktadır. Diğer aktivitelerde bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.1’de gösterilmiştir.

Örneğin 11. aktivitede (11.0) duyarlılık değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.6 numaralı formül kullanılmıştır ve TP, FN değerini tablodan anlayabiliriz. TB= 176, FN= 0, sonuç= 1 ifade etmektedir. Diğer aktivitelerde bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.2’de gösterilmiştir.

Örneğin 13. aktivitede (13.0) F1 değeri hesaplamak için yukardaki 2.2.2 bölümünden 2.7 formülü kullanılmıştır ama öncesi kesinlik değeri ve duyarlılık değeri bulmamız lazım. Kesinlik değeri için 2.5 formülü kullanılmıştır, sonuç =1 elde edilmiştir. Duyarlılık değeri 2.6 için formülü kullanılmıştır, sonuç=1 elde edilmiştir. 2.7 formülü kullanılmıştır, sonuç=1 elde edilmiştir. Diğer aktivitelerde bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.3’de gösterilmiştir.

Bu modelin doğruluk değerini bulmak için 2.4 numaralı formülü kullanarak sonuç 1 elde edilmiştir.



Şekil 3.4. Rastgele orman yönteminden elde edilen karmaşıklık matrisi

Rastgele orman performansını ölçmek için doğruluk (Accuracy), kesinlik (Precision), duyarlılık (Recall) ve F1 değeri hesaplanmalıdır.

Örneğin 1. aktivitede (1.0) kesinlik değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.5 numaralı formülü kullanılmıştır ve TP, FB değerini tablodan anlayabiliriz. TB= 368, FB= 1,1 sonucuna ulaşılmaktadır. Diğer aktivitelerde bu

şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.1’de gösterilmiştir.

Örneğin 9. aktivitede (9.0) duyarlılık değeri hesaplamak için yukarıdaki 2.2.2 bölümünden 2.6 numaralı formül kullanılmıştır ve TP, FN değerini tablodan anlayabiliriz. $TB = 60$, $FN = 10 + 1 + 25 = 36$, sonuç= 0.62 ifade etmektedir. Diğer aktivitelerde bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.2’de gösterilmiştir.

Örneğin 15. aktivitede (15.0) F1 değeri hesaplamak için yukardaki 2.2.2 bölümünden 2.7 formülü kullanılmıştır ama öncesi kesinlik değeri ve duyarlılık değeri bulmamız lazım. Kesinlik değeri için 2.5 formülü kullanılmıştır, sonuç= 0.70 elde edilmiştir. Duyarlılık değeri için 2.6 formülü kullanılmıştır, sonuç=0.72 elde edilmiştir. 2.7 formülü kullanılmıştır, sonuç= 0.71 elde edilmiştir. Diğer aktivitelerde bu şekilde hesaplanmaktadır. Bu hesaplamalar Sonuç 3.4 bölümünde Tablo 3.3’de gösterilmiştir

Bu modelin doğruluk değerini bulmak için 2.4 numaralı formülü kullanarak sonuç 0,92 elde edilmiştir.

3.4. Sınıflandırma Sonuçları

Faaliyetleri sınıflandırmak için dört sınıflandırıcı kullanılmıştır. Bu sınıflandırıcılar Lojistik regresyon (LR), Doğrusal SVC (Destek Vektör Sınıflandırıcısı), Karar Ağacı, Rastgele orman yöntemleridir. Kesinlik ve F-skoru, Duyarlılık sınıflandırıcıların performansını değerlendirmek için kullanılmıştır. Dört sınıflandırıcıya ait veriler aşağıda yer almaktadır.

Tablo 3.1. Kesinlik değerleri

Aktiviteler	LR	Doğrusal SVC	Karar Ağacı	Rastgele Orman
Durmak (1 dk)	1.00	1.00	1.00	1.00
Oturmak (1 dk)	1.00	0.99	1.00	1.00
Otururken el hareketleri ile konuşmak (1dk)	1.00	0.74	1.00	0.93
Ayakta dururken veya yürürken el hareketleriyle konuşmak (1dk)	1.00	0.82	1.00	0.90
Tekrar tekrar ayağa kalkmak ve oturmak (5 kere)	1.00	0.72	1.00	0.96
Yatmak (1 dk)	1.00	0.51	1.00	0.97
Tekrar tekrar ayağa kalkmak ve uzanmak (5 kere)	1.00	0.50	1.00	0.95
Yerden bir nesneyi almak (10)	1.00	0.53	1.00	0.85

Tablo 3.1. (devam)

Zıplamak (10)	1.00	1.00	1.00	0.96
Şınav (Tam şınav yapmak) (5 kere)	1.00	1.00	1.00	1.00
Oturmak – ayakta (5 kere)	1.00	0.59	1.00	0.99
Yürümek (20 metre, 12 saniye)	1.00	0.47	1.00	0.77
Geriye doğru yürümek (20 metre, 20 saniye)	0.98	0.00	1.00	0.93
Yürüyüş çemberi (20 saniye)	1.00	0.00	1.00	0.90
Koşmak (20 metre. 7 saniye)	1.00	1.00	1.00	0.96
Üst yürüme (1 dk)	1.00	0.67	1.00	0.70
Aşağı yürüme (50 saniye)	1.00	0.78	1.00	0.69
Masa Tenisi (1 dk)	1.00	1.00	1.00	0.90
Ortalama	1	0.68	1	0.91

Kesinlik değerleri incelendiğinde LR ve karar ağacı modeli en yüksek değere sahiptir.

Tablo 3.2. Duyarlılık değerleri

Aktiviteler	LR	Doğrusal SVC	Karar Ağacı	Rastgele Orman
Durmak (1 dk)	1.00	1.00	1.00	0.99
Oturmak (1 dk)	1.00	1.00	1.00	0.98
Otururken el hareketleri ile konuşmak (1dk)	1.00	0.99	1.00	0.96
Ayakta dururken veya yürürken el hareketleriyle konuşmak (1dk)	1.00	0.66	1.00	0.99
Tekrar tekrar ayağa kalkmak ve oturmak (5 kere)	1.00	0.86	1.00	1.00
Yatmak (1 dk)	1.00	0.07	1.00	0.97
Tekrar tekrar ayağa kalkmak ve uzanmak (5 kere)	1.00	0.74	1.00	0.98
Yerden bir nesneyi almak (10)	1.00	0.79	1.00	0.98
Zıplamak (10)	1.00	0.11	1.00	0.96
Şınav (Tam şınav yapmak) (5 kere)	1.00	1.00	1.00	0.62
Oturmak – ayakta (5 kere)	1.00	0.95	1.00	0.98
Yürümek (20 metre, 12 saniye)	1.00	0.53	1.00	0.89
Geriye doğru yürümek (20 metre, 20 saniye)	1.00	0.00	1.00	0.41
Yürüyüş çemberi (20 saniye)	1.00	0.00	1.00	0.54
Koşmak (20 metre. 7 saniye)	0.98	0.23	1.00	0.95
Üst yürüme (1 dk)	1.00	0.72	1.00	0.72
Aşağı yürüme (50 saniye)	1.00	1.00	1.00	0.76
Masa Tenisi (1 dk)	1.00	1.00	1.00	0.28
Ortalama	1	0.65	1	0.83

Duyarlılık değerleri olarak karşılaştırıldığında LR ve karar ağacı modeli en yüksek sonucu vermiştir.

Tablo 3.3. F-skoru değerleri

Aktiviteler	LR	Doğrusal SVC	Karar Ağacı	Rastgele Orman
Durmak (1 dk)	1.00	1.00	1.00	0.99
Oturmak (1 dk)	1.00	1.00	1.00	0.99
Otururken el hareketleri ile konuşmak (1dk)	1.00	0.85	1.00	0.95
Ayakta dururken veya yürürken el hareketleriyle konuşmak (1dk)	1.00	0.73	1.00	0.94
Tekrar tekrar ayağa kalkmak ve oturmak (5 kere)	1.00	0.79	1.00	0.98
Yatmak (1 dk)	1.00	0.12	1.00	0.97
Tekrar tekrar ayağa kalkmak ve uzanmak (5 kere)	1.00	0.59	1.00	0.97
Yerden bir nesneyi almak (10)	1.00	0.63	1.00	0.91
Zıplamak (10)	1.00	0.19	1.00	0.96
Şınav (Tam şınav yapmak) (5 kere)	1.00	1.00	1.00	0.77
Oturmak – ayakta (5 kere)	1.00	0.73	1.00	0.99
Yürümek (20 metre, 12 saniye)	1.00	0.50	1.00	0.83
Geriye doğru yürümek (20 metre, 20 saniye)	1.00	0.00	1.00	0.57
Yürüyüş çemberi (20 saniye)	1.00	0.00	1.00	0.67
Koşmak (20 metre. 7 saniye)	0.99	0.37	1.00	0.95
Üst yürüme (1 dk)	1.00	0.69	1.00	0.71
Aşağı yürüme (50 saniye)	1.00	0.87	1.00	0.72
Masa Tenisi (1 dk)	1.00	1.00	1.00	0.43
Ortalama	1	0.61	1	0.85

F-skoru değerleri incelendiğinde LR ve karar ağacı modeli en yüksek değere sahiptir.

Akıllı telefonların ivmeölçer ve jiroskop sensörlerinden alınan veriler için kesinlik değerleri incelendiğinde LR ve karar ağacı modeli en yüksek değere sahiptir. Duyarlılık değerleri olarak karşılaştırıldığında LR ve karar ağacı modeli en yüksek sonuçları vermiştir. F-skoru için en yüksek sonuç veren algoritma LR ve karar ağacı algoritmasıdır.

4. SONUÇ VE ÖNERİ

Bu çalışmada, akıllı telefon sensörleri kullanılarak toplanan 18 farklı günlük yaşam aktivitesi hakkında bilgi içeren KU-HAR veri setini sunmakta ve açıklamaktadır.

Yapılan çalışmada, 18 ila 34 yaşları arasındaki 90 katılımcıdan toplanan bir dizi HAR verisi yer almaktadır. Oluşturulan veri seti KU-HAR olarak isimlendirilmiştir ve 18 farklı sınıfa ait 1945 adet ham aktivite örneğini içermektedir. Bu HAR verilerini toplamak için yerleşik akıllı telefon sensörlerini (ivmeölçer ve jiroskop) kullanılmıştır. Gerçekleşen zaman serisi örneklerine ek olarak, veri kümesi, her bir kullanıcıya karşılık gelen aktivitenin 3 saniyelik verilerini içeren ve bağımsız olarak gerçekleşen 20750 alt örneği içermektedir. Önerilen veri setinin kalitesini kanıtlamak için bazı sınıflandırma sonuçları da sağlanmıştır. Elde edilen sonuçlar, Karar Ağacı ve Lojistik Regresyon alt örneklemeleri yaklaşık %99 doğrulukla sınıflandırabildiğini göstermektedir. Bu veri seti, akıllı telefonların ve diğer akıllı cihazların yeni etkinlikleri belirlemesini sağlayacak ve araştırmacıların pratik HAR verilerine dayalı daha ince modeller geliştirmelerine yardımcı olacaktır.

Diğer araştırmacıların bu veri setinden yararlanabileceğini, çalışmalarına dahil edebileceğini ve yeni alanlarda farklı fikirlerle çalışmaya yön verebileceklerdir. Belirli sınıflardaki örnek sayısını artırmak ve günlük yaşamda gerçekleştirilen yeni aktiviteleri dahil etmek için etkinlik veri setleri genişletilebilir. Çalışma tamamlandıktan sonra, bu veriler KU-HAR veri setinin gelecek sürümlerine eklenebilmektedir.

KAYNAKLAR

- Abdulhussien, A. A., Hassen, O. A., Gupta, C., Virmani, D., Nair, A., & Rani, P. (2022). *Health monitoring catalogue based on human activity classification using machine learning*. International Journal of Electrical and Computer Engineering, 12(4), 3970.
- Brownlee, J. (2022, August 22). *How to calculate precision, recall, F1, and more for Deep learning models*. MachineLearningMastery.com. Retrieved February 23, 2023, from <https://machinelearningmastery.com/how-to-calculate-precision-recall-f1-and-more-for-deep-learning-models>.
- Akalın, B., Demirbaş, M. B. (2022). *Rehabilitasyon hizmetlerinde yapay zekâ uygulamaları*. Acta Infologica .
- Akyol, K., & Karacı, A. (2019) *İnsan Etkinliğini Tanıma İçin Makine Öğrenmesi Algoritmalarının Karşılaştırmalı Analizi* Comparative Analysis Of Machine Learning Algorithms For Human Activity Recognition.
- Alobaidi, H., Clarke, N., Li, F., & Alruban, A. (2022). Real-world smartphone-based gait recognition. *Computers & Security*, 113, 102557.
- Anderson, W., Choffin, Z., Jeong, N., Callihan, M., Jeong, S., & Sazonov, E. (2022). Empirical study on human movement classification using insole footwear sensor system and machine learning. *Sensors*, 22(7), 2743.
- Arjovsky, M., Chintala, S., Bottou, L. (2017). Wasserstein generative adversarial networks. *34th International Conference on Machine Learning*, Australia, 214-223.
- Aydemir, E., Karşlıoğlu, İ. (2021). *Akıllı telefonların ivmeölçer sensörü yardımıyla yürüyüş deseni analizi*. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 21(2), 283-299.
- Aydın, İ., Aşıcı, B. (2020). İnsan hareketlerinin tanınması için parçacık sürü optimizasyonu tabanlı topluluk sınıflandırıcı yöntemi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 32(2), 381-390.
- Ayna Altuntaş, S. (2022). *Makine öğrenmesi teknikleri ile yürüyüş verileri kullanılarak Parkinson hastalığının ön tespiti*, Basılmamış Yüksek Lisans Tezi. Trakya Üniversitesi Fen Bilimleri Enstitüsü.
- Bhattacharya, D., Sharma, D., Kim, W., Ijaz, M. F., & Singh, P. K. (2022). Ensem-har: an ensemble deep learning model for smartphone sensor-based human activity recognition for measurement of elderly health monitoring. *Biosensors*, 12(6), 393.
- Bozkurt, F. (2022). A comparative study on classifying human activities using classical machine and deep learning methods. *Arabian Journal for Science and Engineering*, 47(2), 1507-1521.
- Bülbül, E., Çetin, A., & Doğru, I. A. (2018). Human activity recognition using smartphones. *2nd international symposium on multidisciplinary studies and innovative technologies*, ss. 1-6.
- Comeau, J., Costantino, S., Wiseman, P. (2016). *A guide to accurate fluorescence microscopy colocalization measurements*. Biophysical Journal, 91(12), 4611-4622.
- Costes, S., Daelemans, D., Cho, E., Dobbin, Z., Pavlakakis, G., Lockett, S. (2004). Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophysical Journal*, 86(6), 3993-4003.

- Çalışkan, D., Demir, Ö. (2022). Derin Öğrenme Yöntemleri ile Şüpheli Davranış Tespiti. *International Periodical of Recent Technologies in Applied Engineering*, 3(1), 26-41.
- Eren, K. K. (2019). *Makine öğrenmesi yöntemleri ile insan aktivite tespit sistemi tasarımı*, Basılmamış Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü.
- Gade, S., Lokare, S., Gadekar, T., & Khoje, K.(2022) Human Activity Recognition System Using Smartphone Data Sensors with Python and Machine Learning.
- GeeksforGeeks. (2023, January 10). *Decision tree*. GeeksforGeeks. Retrieved January 27, 2023, from <https://www.geeksforgeeks.org/decision-tree/>
- Gençaslan, S., Utku, A., Akçayol, M. A. (2022). Derin öğrenme tabanlı video üzerinde olay sınıflandırma. *Politeknik Dergisi*, 1(1).
- Gupta, S. (2021). *Deep learning based human activity recognition (HAR) using wearable sensor data*. *International Journal of Information Management Data Insights*, 1(2), 100046.
- Hayat, A., Morgado-Dias, F., Bhuyan, B. P., & Tomar, R. (2022). *Human activity recognition for elderly people using machine and deep learning approaches*. *Information*, 13(6), 275.
- Khan, Y. A., Imaduddin, S., Prabhat, R., & Wajid, M. (2022). Classification of human motion activities using mobile phone sensors and deep learning model. *8th International Conference on Advanced Computing and Communication Systems*, 1(1), 1381-1386.
- Lateef, R. A., & Abbas, A. R. (2022). *Human activity recognition using smartwatch and smartphone: a review on methods, applications, and challenges*. *Iraqi Journal of Science*, 363-379.
- Lehmussola, A. (2008). Synthetic images of high-through put microscopy for validation of image analysis methods. *Proceedings of the IEEE*, 96(8), 1348-1360.
- Logistic Regression Analysis*. Logistic Regression Analysis- an overview | ScienceDirect Topics. (n.d.). Retrieved September 29, 2021, from <https://www.sciencedirect.com/topics/medicine-and-dentistry/logistic-regression-analysis>.
- Machine learning tutorial in PDF*. Tutorials Point. (n.d.). Retrieved December 5, 2022, from https://www.tutorialspoint.com/machine_learning/machine_learning_pdf_version.htm
- Metin, İ. A., Karasulu, B. (2019). İnsan aktivitelerinin sınıflandırılmasında tekrarlayan sinir ağı kullanan derin öğrenme tabanlı yaklaşım. *Veri Bilimi*, 2(2), 1-10.
- Moradi, B., Aghapour, M., & Shirbandi, A. (2022, May). *Compare of machine learning and deep learning approaches for human activity recognition*. In 2022 30th International Conference on Electrical Engineering (ICEE) (pp. 592-596). IEEE.
- Murphy, R. (2016). Building cell models and simulations from microscope images. *Methods*, 96, 33-39.
- PlumX. (n.d.). Retrieved December 6, 2022, from https://plu.mx/plum/a?mendeley_data_id=45f952y38r&theme=plum-bigen-theme
- Predicting human motion signals using modern deep learning techniques ...* (n.d.). Retrieved December 13, 2022, from https://www.researchgate.net/publication/356994688_Predicting_Human_Motion_Signals_Using_Modern_Deep_Learning_Techniques_and_Smartphone_Sensors
- Real-time human activity detection based on accelerometers and internet ...* (n.d.). Retrieved December 13, 2022, from <https://www.researchgate.net/publication/348887029>

Real Time Human Activity Detection Based on Accelerometers and Internet of Things

- Sağbaşı, E. A., & Ballı, S. (2022) *Davranışsal Biyometri ve Makine Öğrenmesi Kullanılarak Akıllı Telefon Yetkilendirmesi Üzerine Kısa Bir İnceleme A Short Review on Smartphone Authorization Using Behavioral Biometrics and Machine Learning*.
- Sağbaşı, E. A., Körükoğlu, S., Ballı, S. (2020). Stress detection via keyboard typing behaviors by using smartphone sensors and machine learning techniques. *Journal of Medical Systems*, 44(4), 1-12.
- Svoboda, D., Kašík, M., Maška, M., Hubený, J., Stejskal, S., & Zimmermann, M. (2007). On simulating 3d fluorescent microscope images. In *Computer Analysis of Images and Patterns: 12th International Conference, CAIP 2007, Vienna, Austria, August 27-29, 2007. Proceedings 12* (pp. 309-316). Springer Berlin Heidelberg.
- Wikimedia Foundation. (2021, July 4). *Confusion matrix*. Wikipedia. Retrieved September 29, 2021, from https://en.wikipedia.org/wiki/Confusion_matrix.
- Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1912-1920).
- Xiong, W., Wang, Y., Ong, S. H., Lim, J. H., & Jiang, L. (2010, September). Learning cell geometry models for cell image simulation: An unbiased approach. In *2010 IEEE International Conference on Image Processing* (pp. 1897-1900). IEEE.
- Zhao, T. (2007). Automated learning of generative models for subcellular location: building blocks for systems biology. *Cytometry Part A*, 71(12), 978–990

ÖZ GEÇMİŞ

Dina ALMOMAYEZ, ilkokul, ortaokul ve lise eğitimini Irak'ta tamamladı. 2012 yılında başladığı Al-Mansour Üniversitesi Fen Fakültesi, Bilgisayar Bilimi ve Bilişim Sistemi Bölümü'nden 28.07.2016 tarihinde mezun oldu. 2020 yılında OMÜ LEE İstatistik Ana Bilim Dalı Yüksek Lisans programına girdi. Yaşadığı ülke nedeniyle iyi derecede İngilizce, Arapça ve Türkçe bilmektedir. Temel ilgi alanları ise makine öğrenmesi, Python ve yapay zekadır.

İletişim Bilgileri

ORCID ID : 0000-0003-0743-3602

Yayınlar:

1. Almomayez, Dina, "Makine Öğrenmesi Kullanarak Akıllı Telefon Algılayıcıları ile İnsan Hareketlerinin Sınıflandırılması", 7. Uluslararası "Başkent" Fen, Sosyal ve Sağlık Bilimleri Kongresi, Türkiye, 27-28 Ağustos 2022, ss. 336-336.