ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

MAKİNE ÖĞRENMESİ İLE ACİL TRİYAJ VE HASTANE YATIŞ TAHMİNİ

ABDOUL AZIZ RAZIKOU DILWANI

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ANKARA
2019

Her hakkı saklıdır

TEZ ONAYI

Abdoul Aziz RAZIKOU DILWANI tarafından hazırlanan "**Makine Öğrenmesi ile Acil Triyaj ve Hastane Yatış Tahmini**" adlı tez çalışması 08/08 2019 tarihinde aşağıdaki jüri tarafından oy birliği ile Ankara Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Dr. Öğr. Üyesi Özgür TANRIÖVER
Ankara Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı

Jüri Üyeleri :

Başkan :Doç. Dr. İhsan Tolga MEDENİ
Ankara Yıldırım Beyazıt Üniversitesi Yönetim Bilişim Sistemleri Anabilim Dalı

Üye :Doç. Dr. Mehmet Serdar GÜZEL
Ankara Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı

Üye :Dr. Öğr. Üyesi Özgür TANRIÖVER
Ankara Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı

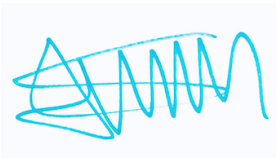
Yukardaki sonucu onaylarım.

Prof. Dr. Özlem YILDIRIM
Enstitü Müdürü V.

ETİK

Ankara Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

08.08.2019



Abdoul Aziz RAZİKOU DİLWANİ

ÖZET

Yüksek Lisans Tezi

MAKİNE ÖĞRENMESİ İLE ACİL TRİYAJ VE HASTANE YATIŞ TAHMİNİ

Abdoul Aziz RAZİKOU DİLWANİ

Ankara Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Danışman : Dr. Öğr. Üyesi Ömer Özgür TANRIÖVER

Acil hasta triyajı hastane acil servis bölümlerinde ve acil durum müdahalesi gereken alanlarda, yüksek riskli hastaları tanımlamak için gittikçe daha fazla kullanılan bir teknik olmuştur. Bu çalışma, hastaların hastaneye yatış gerekliliği ve aciliyet düzeylerini öngörmek için makine öğrenimi (ML) algoritmalarının performansını, iki farklı triyaj sistemi için sınırlı hasta verilerini kullanarak karşılaştırmayı amaçlamaktadır: Bu sistemler Manchester Triyajı Sistemi (MTS) ve Acil Durum Şiddeti Endeksi (ESI). Bu çalışmada, dört çocuk hastanesinde kaydedilen 60.735 hastanın MTS veri setini kullandık. ESI verileri, elektronik sağlık kayıtlarından elde edilen 352.999 erişkin hastayı içermektedir. Bu çalışma iki bölüm halinde alt bölümlere ayrılmıştır : Birinci bölümde, hastaneye yatışı tahmin ettik ve ikinci bölümde triyaj seviyelerini belirledik. Triyaj için, ilk iki seviye (seviye 1 ve 2) ve beş seviye öngördük. Tahmin için hasta yaşı, hayati belirtiler, hastanede kalma durumu, başlıca şikayetleri ve triyaj seviyeleri bilgileri kullanılmıştır. Modeller makine öğrenimi algoritmaları ve teknikleri kullanılarak eğitilmiştir : Lojistik Regresyon (LR), Rastgele Orman (RF), Gradyan Yükseltme (XGBoost), K-Komşular çapraz doğrulama (KNN), otomatik kodlayıcı ve Derin Sinir Ağı (DNN). Her bir modeli eğri altındaki alanı, doğruluk, f-skoru, hassasiyet ve hatırlama açısından karşılaştırdık. Makine öğrenme modellerinin kısıtlı hasta tanı ve şikâyet bilgileri kullanarak hastaneye yatış, triyaj seviye 1 ve 2 tahminin yüksek doğrulukla ve üç kategorili (çok acil, acil ve diğerleri) triyaj sınıfı tahminin kabul edilebilir şekilde öngörebildiği gösterildi.

Ağustos 2019, 69 sayfa

Anahtar Kelimeler : Triyaj, Acil Servis, Makine Öğrenmesi, MTS, ESI, Yaşamsal Belirtiler, Tahmin.

ABSTRACT

Master Thesis

EMERGENCY TRIAGE AND HOSPITALIZATION PREDICTION USING MACHINE LEARNING

Abdoul Aziz RAZIKOU DILWANI

Ankara University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

Supervisor : Asst. Prof. Dr. Ömer Özgür TANRIÖVER

Emergency triage is a technique increasingly being used to identify patients with high risk, in on site emergency situtations and hospital's emergency departments. This study aims to compare the performance of machine learning (ML) algorithms to predict "Hospitalization" and "Triage Levels" by using a set of limited patient information using two different triage system's datasets : Manchester Triage System (MTS) and Emergency Severity Index (ESI). We used a MTS dataset of 60,735 patients recorded in four children hospital. ESI data included 352,999 adult patients recorded from electronic health records. This study is subdivided in two stages : in the first stage we predicted hospitalization and the second stage we predicted triage levels. For Triage levels, we predicted the first two levels which were emergent and very urgent and the rest as another class. Patient Age, Vital Signs, Hospitalization status, Chief Complaints, and triage levels information are used for the prediction. The predictions are done through using machine learning models such as: Logistic Regression(LR), Random Forest (RF), Gradient Boosting (XGBoost), K-Neighbors(KNN), autoencoder and Deep Neural Network (DNN). We evaluated each model by measuring the area under the curve, the accuracy, recall, precision and f-score of classification. Using a set of patient features which can possibly be collected through portable medical sensors, required hospitalization, triage level 1 and level 2 may be predicted with very high accuracy and the other levels can be predicted with acceptable accuracy.

August 2019, 69 pages

Keywords: Triage, Emergency Department, Machine Learning, MTS, ESI, Vital Signs, Prediction.

TEŞEKKÜRLER

Araştırmalarımın her aşamasında bilgi, öneri ve yardımlarını esirgemeyerek çalışmalarıma katkıda bulunan ve yön veren danışman hocam Sayın Dr. Öğr. Üyesi Ömer Özgür TANRIÖVER'e, bu süreç boyunca uzakta olmalarına rağmen yokluklarını bir an hissettirmeyen ve beni destekleyen anneme, babama ve kardeşlerime, manevi desteklerini eksik etmeyen arkadaşlarıma ve son olarak tezimin çeviri aşamasına yardımcı olan Asude BEKCAN'a teşekkür ederim.

Abdoul Aziz RAZIKOU DILWANI

Ankara, Ağustos 2019

İÇİNDEKİLER

TEZ ONAY SAYFASI
ETİK.....i
ÖZET..... ii
ABSTRACT.....iii
TEŞEKKÜR..... iv
KISALTMALAR DİZİNİ.....vii
ŞEKİLLER DİZİNİ..... viii
ÇİZELGELER DİZİNİ.....ix
1. GİRİŞ.....1
2. KURUMSAL TEMELLER VE İLGİLİ ÇALIŞMALAR.....4
2.1 Triyaj Tarihi.....7
2.2 Beş Seviye Triyaj Sistemi Örneği: MTS ve ESI.....8
2.2.1 MTS..... 8
2.2.2 ESI.....9
2.3 Yaşamsal Belirtiler..... 10
2.3.1 Nabız Sayısı..... 11
2.3.2 Solunum Hızı.....11
2.3.3 Kan Basıncı..... 12
2.3.4 Vücut Isısı..... 13
3. MATERYAL VE YÖNTEM..... 14
3.1 Materyal..... 14
3.2 Yöntemler.....15
3.2.1 Veri ön işleme.....16
3.2.2 Öğrenme algoritmaları.....19
3.2.3 Model değerlendirme.....31
4. BULGULAR VE TARTIŞMA.....34
4.1 Araştırma Bulguları..... 34
4.1.1 Hastanede yatış tahmini..... 34
4.1.2 Triyaj tahmini..... 38
4.2 Tartışma..... 45

4.3 Geçerliliğe Yönelik Tehditler.....49
5. SONUÇ.....50
KAYNAKLAR.....52
EKLER.....55
EK 1 MTS Hemen ve Çok Acil Python Betiği.....55
EK 2 Hastaneye Yatış Python Betiği.....56
EK 3 ESI DNN Betiği.....57
ÖZGEÇMİŞ.....58

KISALTMALAR DİZİNİ

ABD	Amerika Birleşik Devletleri
AI	Artificial Intelligence
ATS	Australian Triage Scale
AUC	Area Under the Curve
AUROC	Area Under the Receiver Operating Characteristic Curve
AVG	Average
BP	Blood Pressure
BT	Body Temperature
CATS	Canadian Acuity Triage Scale
DNN	Deep Neural Network
DTAS	Deep Learning based Triage Acuity Score
ED	Emergency Department
ESI	Emergency Severity Index
HR	Heart Rate
INT	Interger
KNN	K Nearest Neighbors
KTAS	Korean Triage and Acuity System
LR	Logistic Regression
MEWS	Modified Early Warning Score
ML	Machine Learning
MSE	Mean Square Error
MTS	Manchester Triage System
NaN	Not a Number
NN	Neural Network
RF	Random Forest
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic
RR	Respiratory Rate
STR	String
XGBOOST	Extreme Gradient Boosting

ŞEKİLLER DİZİNİ

Şekil 2.1 Çocuklarda nefes darlığının aciliyetinin değerlendirilmesine yönelik bir MTS akış şeması örneği....9
Şekil 2.2 Acil durum şiddeti endeksi triyaj algoritması....10
Şekil 3.1 Tahmini modellemede makine öğrenmesi için iş akışı diyagramı....14
Şekil 3.2 Bölünmüş veri mimarisi: eğitim , test....19
Şekil 3.3 Probit ve Logit modelleri için Olasılık Yoğunluk Fonksiyonu....21
Şekil 3.4 RF algoritması mimarisi....23
Şekil 3.5 KNN algoritması mesafe fonksiyonu....24
Şekil 3.6 KNN örnek çizim....24
Şekil 3.7 KNN adımları....25
Şekil 3.8 Karar-ağacı algoritmalarının XGBoost evrimi....28
Şekil 3.9 Sinir ağı mimarisi diyagramı....30
Şekil 3.10 Girdiyi gösteren bir nöron örneği, karşılık gelen ağırlıkları, bir önyargı ve girdilerin ağırlıklı toplamına uygulanan aktivasyon fonksiyonu....30
Şekil 3.11 Sigmoid fonksiyon....31
Şekil 3.12 ESI özellik önem örneği....33
Şekil 4.1 ESI hastaneye yatış ROC eğrisi....35
Şekil 4.2 MTS hastaneye yatış ROC eğrisi....36
Şekil 4.3 ESI hastaneye yatış karışıklık matrisi....37
Şekil 4.4 MTS hastaneye yatış karışıklık matrisi....37
Şekil 4.5 MTS seviye 1 ROC eğrisi....39
Şekil 4.6 ESI seviye 1 ROC eğrisi....40
Şekil 4.7 MTS seviye 2 ROC eğrisi....41
Şekil 4.8 ESI seviye 2 ROC eğrisi....41
Şekil 4.9 ESI 5 seviye karışıklık matrisi....43
Şekil 4.10 MTS 5 seviye karışıklık matrisi....43

ÇİZELGELER DİZİNİ

Çizelge 2.1 MTS ölçeği....8
Çizelge 2.2 Farklı yaşlarda kalp hızı ve solunum hızı....12
Çizelge 2.3 kan basıncı kategorileri....13
Çizelge 3.1 MTS veri çerçevesi istatistik açıklaması....17
Çizelge 3.2 ESI veri çerçevesi istatistik açıklaması....17
Çizelge 3.3 MTS değişkeni....18
Çizelge 3.4 ESI değişkeni....18
Çizelge 3.5 ikili karışıklık matrisi örneği....31
Çizelge 4.1 hastaneye yatış modeli doğruluk, AUC, hassasiyet ve hatırlama....35
Çizelge 4.2 hastaneye yatış duyarlılık ve özgüllüğü....36
Çizelge 4.3 Otomatik kodlayıcı kullanarak hastaneye yatış doğruluğu....38
Çizelge 4.4 Triyaj Seviye 1 tahmini....39
Çizelge 4.5 Otomatik kodlayıcı kullanarak triyaj seviye 1 doğruluğu....40
Çizelge 4.6 Triyaj Seviye 2 tahmini....40
Çizelge 4.7 Otomatik kodlayıcı kullanarak triyaj seviye 2 doğruluğu....41
Çizelge 4.8 Triyaj Seviyeleri 1 ve 2 tahmini....42
Çizelge 4.9 Triyaj 5 Seviye Tahmini....42
Çizelge 4.10 Otomatik kodlayıcı kullanarak triyaj 5 seviye doğruluğu....42
Çizelge 4.11 Hastaneye yatış öngörüsünün karşılaştırmalı çalışması....44

1. GİRİŞ

Savaş zamanında icat edilen triyaj, bugün acil servis (ED) bölümlerinin hastane seviyesindeki ana faaliyetlerinden biri haline gelmiştir. ED, acil bakıma ihtiyacı olup, randevu almadan hastaneye gelen hastalara tıbbi bakım sağlanmasından sorumlu olan hastane bölümüdür. ED, hastanede erişim noktasıdır. Her yıl, ED'ye binlerce kişi kabul edilmektedir. Hastanın durumuna ilişkin, onu kabul edilmesinden önceki durumundan hastaneye yatışına kadar olan tüm kararlar bu bölümden alınır. ED, triyajın başlangıç noktasıdır.

Tıpta, triyaj en kritik durumda olan hastaları birinci basamakta tanımlayıp, bekletmeden zamanında kabul ederek, hastaların yoğunluğunu azaltmak için kullanılan bir tekniktir. Günümüzde dünyada birkaç triyaj sistemi vardır. Bunlar 3. ve 5. seviyede triyaj sistemleridir. 5. seviyede triyaj en çok kullanılandır. Yapılan bir araştırma (Travers. vd, 2002), 3. ve 5. seviyede triyajlarin güvenilirliği ve geçerliliği karşılaştırması sonunda, üçüncü derece acil serviste, 5. seviyede triyaj sisteminin 3. seviyedekininden daha etkili olduğunu göstermiştir. En çok kullanılan 5. seviyede triyaj sistemi şunlardır: ATS, CATS, ESI ve MTS. Bu tez çalışmasında, MTS ve ESI veri tabanları üzerinde çalışılmıştır.

MTS, İngiliz beş seviyeli bir triyajdır. Hastalar beş seviyeye ayrılır (Seviye 5 için Acil Olmayan, Seviye 4 için Standart, Seviye 3 için Acil, Seviye 2 için Çok Acil ve Seviye 1 için Hemen). Her seviye belirli bir renge karşılık gelir (Seviye 1 için kırmızı renk, 10 dakika bekleme süresiyle seviye 2 için turuncu renk, 60 dakika bekleme süresiyle seviye 3 için sarı renk, 120 dakika ile seviye 4 için yeşil renk ve seviye 5 için 240 dakika bekleme süresi ile mavi renk). MTS algoritması karar verme süreci için ayrıştırıcılar ve akış çizelgeleri ile ilişkilidir.

ESI, ABD'de geliştirilen beş seviyeli bir triyaj algoritmasıdır. Tedavi önceliğine, hastalığın şiddeti ve beklenen kaynaklar ihtiyacı temelinde karar verilir. Triyaj algoritması, eğitimli triyaj hemşiresinin özel sorular sorduğu dört karar noktasından oluşur. İlk olarak, hayati tehlike arz eden hastalar (ESI seviye 1 ve 2) belirlenir.

Dayanıksız hastalar tipik olarak ESI triyaj seviyesi 1'e atanır. Potansiyel olarak hayatı tehdit edici semptomları olan, örneğin akut koroner sendromda torasik ağrı veya bilinç kaybı olan ve ayrıca şiddetli ağrı, psikiyatrik bozukluklar veya zehirlenme durumları olan hastalar triyaj seviyesi 2'ye atanır. Kalan seviyeler (3 ile 5 arasındaki seviyeler) beklenen kaynak ihtiyaçları ve hayati bulgular ile tanımlanmaktadır.

Triyaj, hemşireleri, doktorları ve sistemleri içeren dinamik ve karmaşık bir süreçtir. Triyaj hemşireleri triyaj sürecinde çok önemli bir rol oynar çünkü hastalarla ilk temasta olan kişiler onlardır. Karar verme hemşire bilgi ve tecrübesi ile yapılır. Karar verme, triyajın en karmaşık görevlerinden biridir. Karar verme, bir hastayı sağlık durumlarına göre beş seviye arasında triyaj seviyesinden birine atamanın son adımıdır. Hayati bulgular, triyaj karar alma sürecinde dikkate alınan ilk parametrelerdir ve toplamda dört tane başlıca yaşamsal belirti vardır: vücut ısısı (BT), kalp atım hızı (HR), kan basıncı (BP) ve solunum hızı (RR). Bir kişinin yaşamsal belirtileri için normal aralıklar yaş, kilo, cinsiyet ve genel sağlık durumlarına göre değişir. Hasta tanımlandıktan ve sınıflandırıldıktan sonra bakıma atanır. En kritik hastalar hastaneye yatırılır.

Her triyaj sisteminin amacı, en kritik hastayı beklemeden tanımlamaktır. Dünya nüfusunun artması, şüphesiz önümüzdeki yıllarda acil servise girişi artıracaktır. Gelecekteki triyaj sistemine bir çözüm bulmak için çeşitli çalışmalar yapılmıştır. Araştırmalar, biyosensör teknolojisi ve makine öğrenmesinin alanda en yaygın kullanılanlar olduğunu göstermiştir. Kablosuz teknolojilerdeki ve performans vücut biyosensörlerinde kaydedilen ilerlemeler muhtemelen yakın gelecekte sağlık sistemi ve acil servis triyajını değiştirecektir. Vücut sensörleri bireysel yaşamsal belirtileri ölçebilir ve tıbbi verileri aynı anda analiz için bir triyaj sistemine gönderebilir. Bu nedenle, sonuç tahminini nasıl etkileyebileceklerini görmek için bu çalışmada hayati belirtileri göz önüne alıyoruz.

Makine öğrenmesi, makinenin programlanmadan verilerden kendi kendine öğrenme yeteneği olarak tanımlanabilir. Makine öğrenmesi yapay zeka algoritmalarını içerir. Yapay zeka birkaç alanda uygulanır. Bu tez, sağlık alanına, özellikle de triyaja odaklanmaktadır. Günümüzde, sağlık alanında özellikle hastaneye yatış, hastalıklar ve

triyaj tahmini gibi çeşitli araştırmalar yapılmaktadır. Şimdiye kadar elde edilen sonuçların yüksek doğruluğu etkileyicidir. Triyaj, zaman ve kaynak gerektiren bir süreçtir. Daha az kaynak ve zamanla triyajı daha hızlı yapmak için bu çalışmayı yaptık. Tezimiz, sınırlı kaynaklar kullanılarak hastaneye yatış ve triyajın öngörülmesine odaklanmaktadır. Böylelikle, bu çalışma gelecekti triyaj sistemlerinin performansının artışına katkıda bulunacaktır işlemini ve zamanı azaltacaktır.

Bu tez aşağıdaki gibi planlanmıştır. İkinci bölümde ilgili çalışmalar, triyajın tarihi, triyaj sistemi örneğini (MTS ve ESI) ve yaşamsal belirtileri tartışılmaktadır. Üçüncü bölümde çalışmada kullanılan materyal ve yöntemler detaylı olarak ele alınmıştır. Bir sonraki bölüm de (dördüncü) ise sonuçlar, tartışmalar ve geçerliliğe yönelik tehditler sunulmaktadır. Son olarak beşinci bölüm çalışmanın genel sonucu ve yapılabilecek gelecek çalışmalar için tavsiyeler içermektedir.

2. KURUMSAL TEMELLER ve İLGİLİ ÇALIŞMALAR

Bu bölümde, bu alanda yapılan ilgili çalışmalardan bahsedeceğiz. Ayrıca triyaj tarihini, triyajın evrimini ve günümüzdeki triyaj sisteminin (MTS ve ESI) nasıl yürütüldüğünü tartışacağız. Son olarak, triyaj sürecinde en önemli unsur olan dört ana hayati bulgunun tanımını göreceğiz.

Önceki bölümde, triyajın doktorları, hemşireleri ve sistemi içeren karmaşık bir süreç olduğunu gördük. Genel olarak, triyaj aşağıdaki gibi yapılır : hasta ambulans veya kendi imkanları ile acil servise gider. Hemşire yaşını, cinsiyetini, başlıca şikayetlerini sorar ve yaşamsal belirtilerini ölçer. Bilgi toplandıktan sonra hemşire karar verir. Günümüzde dünya nüfusunun artması ile bu teknik yeterli gelmemektedir. ED'ye kabul edilen binlerce hastayla ilgilenmek, triyajı daha hızlı ve etkili hale getirmek için başka çözümler bulmak gerekmektedir. Herhangi bir triyaj sisteminde zaman en önemli unsurdur. Kritik hastaları zamanında tanımlayabilmek hayat kurtarır. Ve bu ML ile mümkün olabilir. ED triyaj tahmini için ML algoritmaları kullanılarak çeşitli çalışmalar yapılmıştır. Araştırmaların çoğu, ML'nin ED triyajını yüksek doğrulukla tahmin edebileceğini göstermektedir.

Derin öğrenme temelli triyaj keskinliği skoru (DTAS), Kore triyaj ve keskinlik sistemi (KTAS), değiştirilmiş erken uyarı skoru (MEWS), rastgele orman ve lojistik regresyon kullanılan bir çalışma da Kwon vd. (2018a) DTAS'ın hastane ölümünü, kritik bakımı ve hastaneye yatışı öngörmede mevcut triyaj keskinliğinden daha doğru olduğu sonucuna varıldı. ED en çok kullanılan ve dinamik hastane bölümü oldu. Bu yüzden ED, her türlü tedavinin başlangıcı için hastane giriş noktasıdır. Tüm triyaj sistemlerinin amacı kritik durumda olan hastaları bekletmeden zamanında belirlemektir. Lakin triyaj süreci birçok kaynağa ihtiyaç duyar ve fazla zaman alır. Ek olarak, bu kaynaklar toplandıktan sonra karar verme süreci en karmaşık görevlerden biridir. Triyaj hemşireleri kararlarını deneyimlerine ve hastaların bilgilerine göre vermektedir. Ancak ML bu prosedürleri kolayca basitleştirebilir ve triyajı daha hızlı bir şekilde yapabilir. ML algoritmaları tahminlerde etkilidir. Mevcut triyaj ile karşılaştırıldığında, Kwon vd. (2018b) çalışması DTAS'ın mevcut triyaj ve keskinliğinden daha doğru olduğunu ve çok merkezli bir veri merkezi kullanılarak doğrulandığını gösterdi. Tahmini değişkenler : yaş, cinsiyet, başlıca

şikayet, semptom başlangıcından ED ziyaretine kadar geçen süre, varış şekli, travma, ilk hayati bulgular ve zihinsel durum. Alıcının ölçülen karakteristik eğrisi altındaki alan (AUROC) sonuçları DTAS için 0.935, KTAS için 0.785, MEWS için 0.810, LR için 0.903 ve RF için 0.910 olarak belirlenmiştir.

ML araştırması Hong. vd (2018c) kullanılarak yapılan hastaneye başvuru öngörüsü, lojistik regresyon, XGBoost ve derin sinir ağı ML modelleri için en iyi AUC sonucunu verdi. Bu araştırma, triyajda toplanan bilgilere ek olarak, ED triyajı sırasında hasta geçmişini de kullanarak, hastaların hastaneye kabulünü tanımlamayı kolaylaştırmasını amaçlamaktadır. Çalışmada kullanılan değişkenler şunlardır : hasta kabulü, demografik özellikler, triyaj değerlendirme, hastanın başlıca şikayeti, hastane kullanım istatistiği, hasta tıbbi öyküsü, ayakta tedavi edilen ilaçlar, geçmiş hayati değerler, geçmiş laboratuvar ve görüntüleme/EKG sayımı sonuçları. Çalışmanın öngörüleri üç bölüme ayrılmıştır : triyaj bilgileri, hasta geçmişi ve veri setinin tam değişkenleri. Araştırmada tüm algoritmalar yüksek bir doğruluğa (% 80'den büyük) ulaşmıştır. Modeller, tam set değişkenleri veri setleri için triyaj bilgileri ve hasta geçmişine göre daha yüksek doğruluk tahmininde bulundu. Çalışma, makine öğrenmesinin triyaj bilgileri ve hasta geçmişi kullanılarak hastaneye kabulünün güçlü bir şekilde tahmin edebileceği sonucuna vardı.

Levin.vd (2018d) makine öğrenmesi temelli Elektronik triyaj (E-triyaj) ve ESI karşılaştırma araştırması göre, E-triyajının hastaları klinik sonuçlarına göre ESI'den daha doğru bir şekilde sınıflandırdığı görülmüştür. ESI, hastaların Seviye 1'den Seviye 5'e sıralandığı ABD 5 Seviye triyaj algoritmasıdır. Seviye 1 ve 2, yüksek riskli öncelik seviyeleri olarak kabul edilir. Araştırmanın amacı, ML modeli kullanılarak sıralanan hastalar arasında daha yüksek risk altındaki hastaları belirlemektir. Genel olarak, ABD triyaj seviyesine atanmış hastalar (örneğin seviye 3) bazen yüksek riskli hastaları içerir. Bu hastalar (yüksek risk) 3. seviyede oldukları gibi genellikle göz ardı edilirler. E-triyaj, kritik bakım, acil durum prosedürü ve hastanın hastaneye yatış ihtiyacını tahmin etmek için hayati belirtiler, başlıca şikayeti ve tahminler gibi aktif tıbbi öykü kullanan bir RF modelidir. E-triyaj, her iki ED'deki ESI ile karşılaştırıldığında ve 0.73 ila 0.92 arasında değiştiğini bir AUC ile karşılaştırıldığında klinik hasta sonuçlarının eşdeğer veya geliştirilmiş tanımlandığını gösterdi. ESI hasta dağılımına göre, E-triyaj, önemli ölçüde

artmış kritik bakım veya acil prosedür riski olan triyaj gerektiren ESI seviye 3 hastalarının% 10'undan fazlasını tanımladı (% 1.7 ESI seviye 3 -% 6.2 artışla triyaj) ve ED'ler arasında hastaneye yatış (% 18.9 ve% 45.4).

Goto.vd (2019a) tarafından yapılan araştırmada, çocukların acil servis trijayı sırasında ML temelli çıkan tahmini klinik sonuçları, ML'nin geleneksel triyaj yaklaşımlarına kıyasla klinik sonuçları ve hastanede yatışı önceden tahmin edebileceğini göstermektedir. Ocak 2007 - Aralık 2015 tarihleri arasında yapılan Ulusal Hastane Gezici Tıbbi Bakım anketinin verilerine göre, bu çalışma kapsamında 6 ila 12 yaş aralığında 52.037 çocuğun ED ziyareti yaptığı bilgisi elde edilmiştir. Öngörülen sonuçlar kritik bakım ve hastaneye yatırmadır. Öngörücüler hastanın mevcut triyaj verileridir: yaş, cinsiyet, varış şekli, hayati belirtiler, ziyaret nedenleri, hastanın ikametgahı, 72 saat öncesinde yapılan ED ziyaretleri ve hasta ek tanısı. Veri kümesi, eğitim verileri için %70'e; testler için% 30'a bölünmüştür. Tahmin sırasında dört makine öğrenme modeli kullanılır: Kement regresyonu, rastgele orman, gradyanı arttırılmış karar ağacı ve derin sinir ağları. Modellerin tahmin performansı, C istatistikleri, ileriye dönük tahmin sonuçları ve karar eğrileri üzerinde değerlendirilir. Eğitimden sonra, her bir model sonuçları geleneksel triyaj bilgisini kullanan referans modelle karşılaştırılır. Tüm makine öğrenim modelleri (RF:0.85 and DNN: 0.85), referans model (C statistic: 0.78) kabiliyetinden daha yüksek bir ayırt edici özellik göstermektedir. Hastaneye yatış öngörüsü için, tüm makine öğrenim yaklaşımları, belirgin bir şekilde daha yüksek ayrımcılık özelliğine sahipti (C istatistiği, derin sinir ağı için 0.80, referans için 0.73) ve konvansiyonel triyaj seviyelerinde 1 ila 3 arasında yatan hasta tedavisine ihtiyaç duymayan az triyajlı çocuklar (Hemen ve acil). Buna karşılık, makine öğrenim modellerinde 1 ila 3 triyaj seviyelerinde daha az sayıda aşırı tetiklenmiş çocuk vardı ve triyaj 4 ve 5'te daha fazla sayıda hastanede yatan çocuk doğru bir şekilde tanımlandı. Çalışmanın sonuçları, makine öğrenmesi temelli triyajın, klinik sonuçları ve hastaneye yatışı öngörmede daha iyi bir ayrımcılığa sahip olduğu ve kritik derecede hasta olan çocukları sıralamada aşağıya çekme ve daha az hasta olan çocukları sıralamada yukarılara çıkarma oranlarının azaldığı sonucuna vardı.

Raita.vd (2019b) araştırması, ML modellerini kullanarak kritik bakım ve hastaneye yatışı tahmin etmeyi sağlamak için yüksek bir performans göstermiştir. Bu çalışmada tahmin

için dört ML modeli kullanılmıştır: kement regresyonu, rastgele orman (RF), gradyanı artırılmış karar ağacı ve derin sinir ağı (DNN). Her yıl ED'ye başvuran hasta sayısının artması bu çalışmaya olumlu yönde katkı sağlamıştır. ED'de yoğunluğun olması kritik bakımdaki bazı hastaların zamanında doktor tarafından görülmemesine sebep olur ve bu ölüm sayısını artırabilir. Bu nedenle çalışma kritik bakım ve hastaneye yatış üzerine odaklanmaktadır. Öngörü değişkenleri olarak hasta yaşı, cinsiyeti, varış şekli, triyaj hayati bulguları, başlıca şikayetleri ve hasta ek tanısı kullanılır. Dört ML modeli için kritik bakım sonucu tahmini performansı kement regresyon için 0.84, RF için 0.85, gradyanı artırılmış karar ağacı için 0.85 ve DNN için 0.86'dır. Söz konusu dört model için hastanede yatış sonucu, 0,81, 0,81, 0,82 ve 0,82 olarak gerçekleşmiştir. Bu araştırmanın sonucuna göre, geleneksel triyaj sistemine kıyasla, ML, kritik bakım ve hastaneye yatış öngörmede yüksek performans göstermektedir.

2.1 Triyaj Tarihi

Triyaj, savaş sırasında icat edilen askeri bir süreçtir. Triyaj, Fransızca "trier" kelimesinden gelir. Triyaj, 1792 yılında Napolyon'un İmparatorluk Muhafızı şefi olan cerrah Dominique Jean Larrey tarafından icat edilen bir sıralama sürecidir. Baron, uçan ambulansların mucididir. Uçan ambulanslar, savaş alanında yaralı askerlerin toplanmasından oluşan atlı vagonlardır. Bu teknik en çok 1972'deki İtalyan kampanyası sırasında kullanılmıştır (Anonymous 2006a). Personel bir doktor, bir levazım subayı, bir komisyon üyesi olmayan memur, bandajları taşıyan seyyar bir görevli ve 24 piyade erini kapsıyordu. Uçan ambulans büyük bir başarıydı ve diğer ordular tarafından da kullanıldı. Bu başarısından sonra, Larrey çok geçmeden tüm Fransız Ordusu için uçan ambulanslar düzenledi (Anonymous 2006b). Larrey'in amacı, herhangi bir acil durumda, yaralılarla ilgilenilmesi için ambulansı savaş hattına mümkün olduğunca yakın yerleştirmekti.
Larry'nin icadı birçok askerin hayatının kurtulmasına yardımcı oldu. Tarama süreci bakım görebilecek ve savaş alanına geri dönebilecek en az kritik durumda olan askerleri belirliyordu. Durumu kritik olan askerler ambulanslar tarafından tahliye ediliyordu. Örneğin, Vietnam savaşı sırasında ABD ordusu, tahliye edilmesi gereken yaralıları taşımak için helikopter kullandı. Sıralama işlemi, bugünkü sıralama işlemine benzerdir ve şu şekildedir: hemen, çok acil, acil, acil olmayan ve standart.

Daha sonra triyaj modernize edildi ve hastanelerde acil servis bölümünde kullanıldı. ABD, İngiltere ve Avrupa gibi bazı ülkeler hastaların yararına modern triyaj sistemleri geliştirmiştir. Günümüzde triyaj, hastanelerde acil servislerin ana faaliyeti haline gelmiştir. Her yıl binlerce hasta dünyadaki hastanelerin acil servislerine başvuruyor ve hastalar ED'ye geldiklerinde sıralanmaktadırlar. Durumlarına göre , sıralanan her hasta bir triyaj seviyesine kabul edilir. En yaralı hastalar hastanede yatırılır. Bu sayede, hastanelerde bekleme zamanın azalması, en zayıf hastaların tespiti ve onlara gerekli bakımın sağlanması, birçok hayatın kurtulmasına vesile olmaktadır.

2.2 Beş Seviye Triyaj Sistemi Örneği: MTS ve ESI

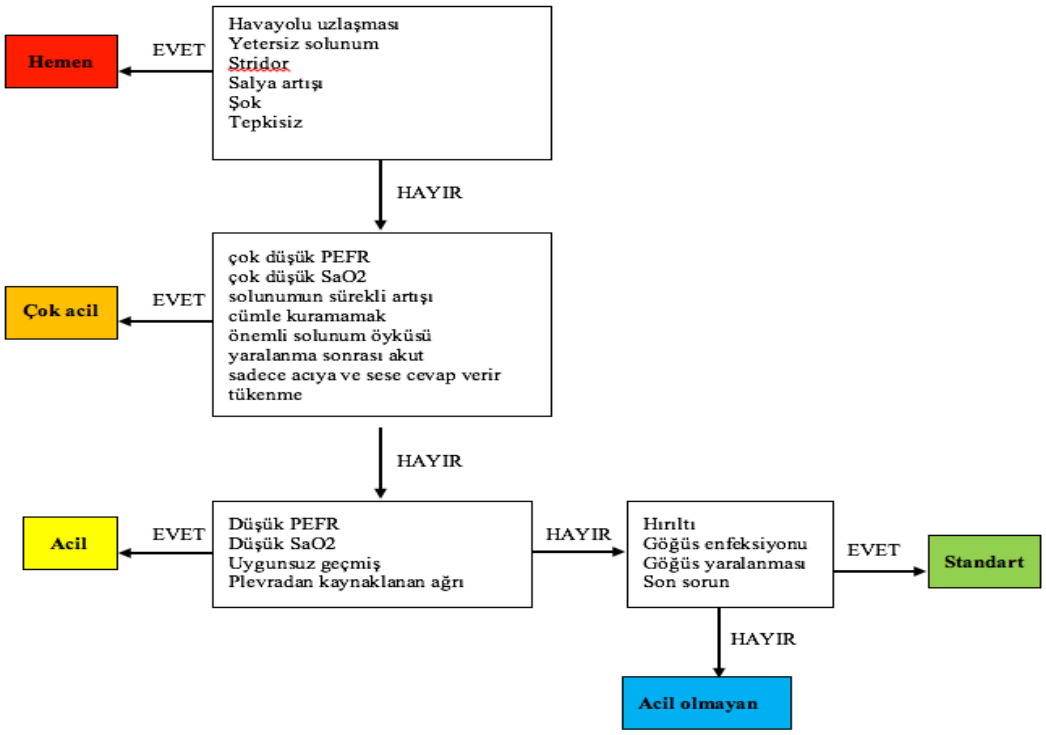
Birkaç tür triyaj sistemi vardır: ED triyajı, yatan hasta triyajı, olay (çoklu zayiat) triyajı, askeri (savaş alanı) triyajı ve afet (kitlesel zayiat) triyajı. Bu tez çalışmasında ED triyajına odaklandık. Önceden de belirttiğimiz gibi, 3 ve 5 seviye triyaj sistemi vardır. Ancak bir çalışma, seviye 5 sıralama sisteminin güvenilirlik açısından seviye 3'ten daha verimli olduğunu göstermiştir.

2.2.1 MTS

MTS, 1997 yılında Manchester triyaj grubu doktorları tarafından icat edilen bir İngiliz 5 seviyeli triyaj sistemidir. Hastalar acil seviyeden acil olmayan seviyeye kadar sıralanır. (Çizelge 2.1). MTS, 52 akış şeması ile ilişkilidir. Her akış şeması, karar verme sürecinde hemşirelere yardımcı olan bir ayrıştırıcı ile ilişkilidir.

Çizelge 2.1 MTS ölçeği (Anonymous 2014a)

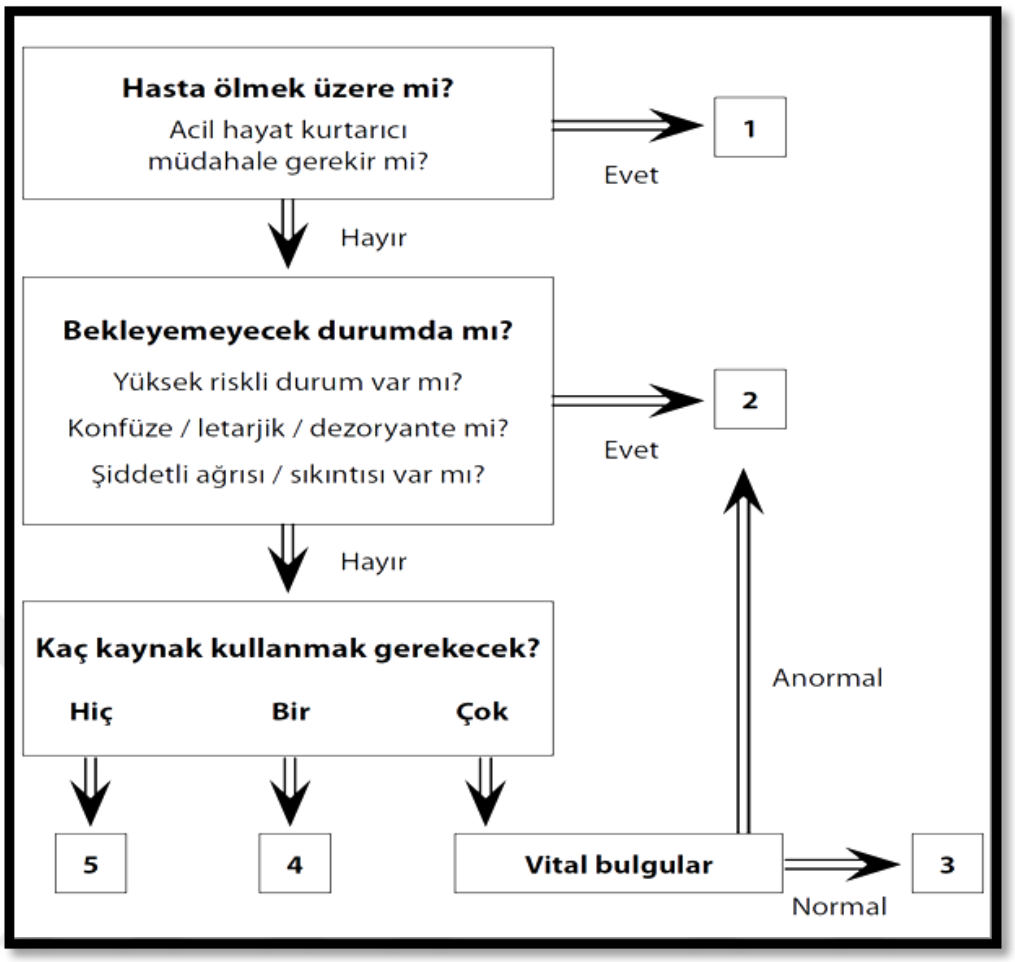
Sayı	Ad	Renk	Max.zaman(dakika)
1	Hemen	Kırmızı	0
2	Çok acil	Turuncu	10
3	Acil	Sarı	60
4	Standart	Yeşil	120
5	Acil olmayan	Mavi	240



Şekil 2.1 Çocuklarda nefes darlığının aciliyetinin değerlendirilmesine yönelik bir MTS akış şeması örneği (Cremeens ve Khorasani 2014b).

2.2.2 ESI

ESI, 1999 yılında ED doktorları Richard Wuerz ve David tarafından ABD'de geliştirilen beş seviyeli bir triyajdır. ESI, hastanın durumunu ve kaynak ihtiyaçlarını değerlendirmek için bir algoritma kullanır. Seviye 1 ve 2 en kritik seviyelerdir. Eğer hasta bu ilk iki seviyeye alınmazsa, triyaj hemşiresinin hastanın hangi seviyeye (3, 4 ve 5) gitmesi gerektiğini belirlemesi için daha fazla kaynak değerlendirilmelidir. Hastanın durumu, yaşamsal belirtilerin istikrarlılığı ve potansiyel yaşam tehdidi, uzuv veya organ tehdidi ile belirlenir. Triyaj hemşiresi, benzer yaralanmaları veya şikayetleri olan hastaların önceki deneyimlerine dayanarak kaynak ihtiyaçlarını tahmin eder (Gilboy.vd. 2012). Kaynak ihtiyaçları, bir yerleştirme kararının (taburcu, kabul veya nakil) ulaşılması için hastanın tüketmesi beklenen kaynak sayısı olarak tanımlanır. Yönlendirildikten sonra triyaj hemşiresi, triyaj hastasını hemen beş seviyeden en uygun olanına yerleştirir.



Şekil 2.2 Acil Durum Şiddeti Endeksi triyaj algoritması(Anonymous. 2016)

2.3 Yaşamsal Belirtiler

Yaşamsal belirtiler bir kişiden ölçülebilen en temel fonksiyonlardır; fiziksel durumlarını ve sağlık durumlarını gösterirler. Ölçümler normalden uzağa gitme eğiliminde olduğunda, fiziksel durumda bir anormallik çıkması olağandır. Tıbbi durumların çoğu hayati belirtilerin ölçümüyle teşhis edilebilir ve özel testlerin yardımıyla da bu ölçümler doğrulanabilir. Her hayati bulgu, özel ekipman kullanımıyla farklı şekillerde ölçülür. Günümüzde teknolojinin gelişimi ile birlikte, hayati belirtiler akıllı saatler biyosensörleri ile de ölçülebilmektedir. Çoğu tıbbi ortamda standart olan dört hayati bulgu vardır:

• Nabız sayısı
• Solunum hızı
• Kan basıncı
• Sıcaklık

2.3.1 Nabız sayısı

Nabız hızı (HR) kalbin atışı sırasında, vuruşların bilek veya boyunda dakikada ölçülen hızdır. Nabız, her atım için arter duvarının genişlemesinden etkilenir. Nabız sayımı için en belirgin noktaları bilek (Radial arter), boyun (Karotid arter), avucun iç kısmı (Brakiyal arter), diz arkası (Popliteal arter) ve ayak bileğinde (Posterior tibial arter) bulunmaktadır (Anonymous, 2018d). Nabız, yaşa göre değişmekle birlikte vücut üzerindeki fiziksel ve psikolojik etkenlere de bağlıdır. Yüksek nabız hızı, vücutta anormal bir durumun olduğunu gösterir ve ayrıca endişe, öfke, heyecan, duygu, kalp rahatsızlıkları, astım, büyük bir öğün ve benzeri gibi diğer nedenlerden de kaynaklanabilir. Bir bireyin nabız hızı vücuttaki çeşitli problemlerin belirlenmesinde yardımcı olabilir, fakat yalnızca bir anormallik teşhisinde kullanılamaz. Nabız, tanı için sadece temel bir araçtır ve bu nedenle sadece birincil tanı için kullanılabilir (Chandrasekaran, 2010a).

2.3.2 Solunum hızı

Solunum hızı (RR), bir kişinin belirli bir süre içinde veya daha resmi olarak aldığı nefes sayısını, birim zamanda nefes alma ve nefes verme ile ilgili göğüs hareketlerinin sayısı olarak tanımlanır. RR, dakika başına nefes birimleri cinsinden ölçülür. Genellikle kişi dinlendiğinde bir dakika boyunca nefes sayısı (göğsün yükselme sayısı) sayılarak ölçülür. Solunum oranları, oksijen talebi arttıkça artacaktır; ayrıca hastalık, yoğun fiziksel aktivite vb. nedeniyle de artış göstermektedir. İstirahat halindeki sağlıklı bir yetişkin için bildirilen ortalama RR genellikle dakikada 12 nefes (12/60 Hz) olarak verilir (Gerard ve Nicholas 1990) ve tahminler dakikada 12-20 nefes arasında değişmekle birlikte, yetişkinlerde, çocuklarda ve bebeklerde solunum hızı daha yüksektir. İnsanlar yaşlandıkça nefes alma oranı düşer. Yavaş oranlarda, tam bir dakika boyunca nefes sayısını sayarak daha doğru sonuçlar elde edilir (Chandrasekaran, 2010b). Çizelge 2.2, değişen yaşlarda kalp atış hızını ve solunum hızını, yaş oranındaki kademeli düşüşü göstermektedir.

Çizelge 2.2 Farklı Yaşlarda Kalp Hızı ve Solunum Hızı(Anonymous, 2010c)

Yaş	Kalp Atış Hızı (Atım / dk.)	Solunum Hızı (nefes / dk)
Yeni doğan	100-160	30-50
0-5 ay	90-150	25-40
6-12 ay	80-140	20-30
1-3 yaş	80-130	20-30
3-5 yaş	80-120	20-30
6-10 yaş	70-110	15-30
11-14yaş	60-105	12-20
14+ yaş	60-100	12-20

2.3.3 Kan basıncı

Kan basıncı (BP), arterlerin, damarların ve kalp odalarının duvarlarına kanın uyguladığı bir kuvvettir. Kan basıncı en önemli yaşamsal belirtilerden biridir ve vücut, kan hacmi ve kalbin kasılma kuvveti ile etkileşime girerek onu korur. Her kalp atışı sırasında, BP sistolik basınç denilen maksimum basınç ile diyastolik basınç denilen minimum basınç arasında değişir. Üst kolun kanı kalpten uzağa taşıyan ana kan damarı olan kol atardamarında (Brachial artery) avuç içinde ölçülür. Bir kişinin kan basıncı genellikle sistolik basınç ve diyastolik basınç değerleri ile ifade edilir. Ortalama sağlıklı bir yetişkinin basınç değerleri, sistol sırasında 120 mmHg ve diyastol sırasında 80 mmHg olarak çıkar. Pompalama Hızı, kan hacmi, direnç, viskozite vb. bir kişinin kan basıncını etkileyen faktörlerden bazılarıdır. Çeşitli nedenlerden dolayı, ortalama kan basıncı herbir kişide farklıdır. Basınç değerleri beş ana bölüme ayrılmıştır (Chandrasekhar, 2010d). Çizelge 2.3, insanların kan basıncı aralıklarına göre kategorilerini göstermektedir.

Çizelge 2.3 kan basıncı kategorileri (American heart association, 2017a)

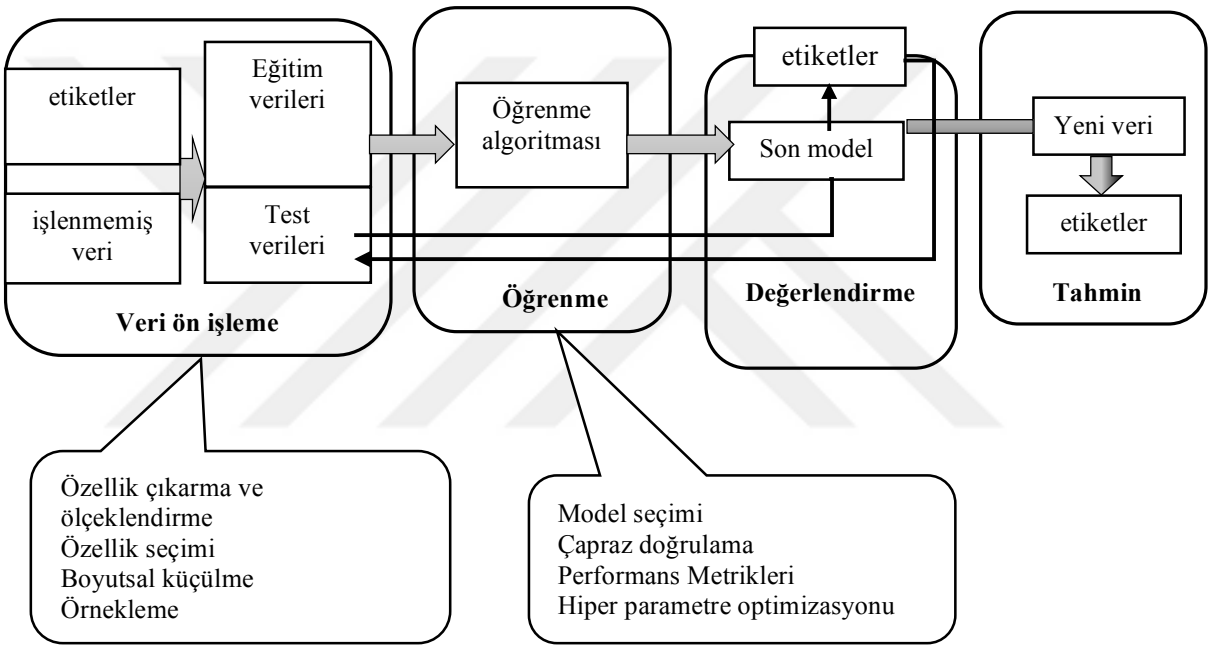
Kategori	Sistolik (mmHg)	Diyastolik (mmHg)
Düşük Tansiyon	<90	<60
Normal	90-120	60-80
Prehipertansiyon	121-139	veya 81-89
Aşama 1 Yüksek tansiyon	140-159	veya 90-99
Aşama 2 Yüksek tansiyon	≥ 160	veya = 100

2.3.4 Vücut ısısı

Vücut ısısı diğer önemli yaşamsal belirtilerden biridir. Vücut sıcaklığı, vücut tarafından üretilen ısı ile vücuttan kaybedilen ısı dengesi ile korunur. Sıcaklık, bireyin normal vücut sıcaklığı ve ölçüm koşulları için bir temel oluşturmak amacıyla kaydedilebilir. Vücut ısısını kontrol etmenin temel nedeni, ateş varlığında veya bireyin normal sıcaklığının önemli derecede üzerine çıkmasının herhangi bir sistemik enfeksiyon veya iltihap ile ilgili olup olmadığını öğrenmektir. Yüksek sıcaklığın diğer nedenleri arasında hipertemi vardır. Sıcaklığın düşüşü (hipotermi) de değerlendirmeye tabi tutulmalıdır. Hastanın sıcaklığının yönünü gözden geçirmek de dikkate değerdir. Hastanın önceki sıcaklığı daha yüksek ise, 38 ° C ateşi her zaman kaygı verici bir işaret göstermez (Anonymous.2018e).

3. MATERYAL VE YÖNTEMLER

Bu bölümde, bu tezde kullanılan farklı materyal ve yöntemleri tartışacağız. İlk aşamada farklı malzemeleri ve ikinci aşamada, tahmin için uygulanan yöntemleri tanımlayacağız. Ayrıca özellik önemine bir örnek verdik. Yöntemler bölümü veri işlemeyi, tahmin için kullanılan algoritmaları ve son olarak modellerin değerlendirilmesinde kullanılan ölçümleri içerir. Makine öğrenme algoritmaları: LR, RF, KNN, XGBoost ve DNN kullanılır. Şekil 3.1 tahmin için kullanılan farklı adımları özetlemektedir.



Şekil 3.1 Tahmini modellemede makine öğrenmesi için iş akışı diyagramı

3.1 Materyal

Kullandığımız sistem Windows 10, işlemci çekirdeği i5 2.5 GHz, ram 16 Gigabayt ve sabit disk 250 gigabayt ‘tır. Kullanılan farklı malzemeler aşağıda tanımlanmaktadır. Anakonda dağılımı altında python versiyon 3.7.2 programlama dilini kullandık. Tensorflow üzerinde derin öğrenme uygulanmaktadır.

Python : 1991 yılında Guido van Rossum tarafından icat edilmiş, yorumlanmış bir programlama dilidir. Makine öğrenimi üzerinde çalışmanıza izin veren birkaç

kütüphanesi vardır. İstenirse başka kütüphaneler de kurulabilir. Bu tez çalışmasında, scikit-learn kütüphanesi ile çalıştık.

Scikit – learn: Numpy ve Scipy gibi python bilimsel kütüphanelerini destekleyen ücretsiz bir python makinesi öğrenme kütüphanesidir. Scikit learn birkaç denetimli ve denetimsiz öğrenme algoritması içerir. Scikit içeren algoritmalar arasında şunları sayabiliriz: regresyon, kümeleme, sinir ağları, vektör makinelerinin desteği, karar ağacı, saf Bayes ve topluluk öğrenmesi.

Anaconda: Piton ve R'de bilimsel hesaplama ve veri bilimi için uygulamaları başlatmayı kolaylaştıran anaconda dağıtımının grafiksel, özgür ve açık kaynak kodlu bir kullanıcı arayüzüdür. Anaconda dili, conda yönetim sistemi tarafından yönetilen birkaç paketi kapsar. Tüm paketler paket yönetim sisteminden eklenebilir veya güncellenebilir. Windows, Mac Os ve Linux'ta mevcuttur. Anaconda, jupyter notebook, spyder, orange3, R studio, glueviz ve qtconsole vs. gibi çeşitli uygulamaları içerir. Jupyter notebook uygulamasını çalışmamız için kullandık.

Jupyter notebook: Gerçek kod, denklem, görselleştirme ve metin içeren belgeler oluşturmanıza ve paylaşmanıza olanak sağlayan açık kaynaklı bir web uygulamasıdır. Ayrıca şunları içerir: veri temizleme ve işleme, sayısal simülasyon, istatistiksel modelleme, veri görselleştirme, makine öğrenmesi ve çok daha fazlası (Anonymous.2019b).

Tensorflow: Çeşitli görevlerde türevlenebilen programlama iş akışı için Google Brain tarafından geliştirilen ücretsiz açık kaynaklı bir yazılım kütüphanesidir. Sinir ağları gibi makine öğrenmesi uygulamaları için de kullanılan sembolik bir matematik kütüphanesidir (Anonymous.2017b).

3.2 Yöntemler

MTS verileri, Mayıs 2007 ile Haziran 2012 arasında kaydedilen dört farklı çocuk hastanesinin acil servis (ED) bölümünden (<= 16 yıl) elde edildi. MTS verileri 60,735 hastayı içeriyor. ESI verileri Yale-New Haven Sağlık Sistemine yapılan hasta ziyaretlerinin elektronik sağlık kayıtlarından elde edilen sadece yetişkinler içindir (<= 100 yıl). ESI veri setine 352.999 hasta dahil edildi. Bu çalışma, Hastaneye yatış için ikili bir sınıflandırma ve triyaj seviyelerinin tahmini için çok sınıflı sınıflandırmadır. MTS

hastaneye yatış tahmini, başlıca şikayet veri türü (dizgiler) türü nedeniyle başlıca şikayetler olmadan yapıldı. MTS için hastaneye yatış veri türü dizgedir (Evet ve Hayır). "Evet", hastanın hastaneye yatırıldığı ve "Hayır", hastanın hastaneye yatırılmadığı anlamına gelir. Bu değişkeni "Evet" için 1, "Hayır" için 0 olarak değiştirdik. Aynı düzenlemeyi ESI veri kümesi düzenleme sütunu için de ("Kabul Et" için 1, "Tahliye et" için 0) yaptık. Triyaj tahmini iki evreye bölünmüştür. İlk etapta triyajı ilk iki seviye (Acil ve Çok Acil) ve ikinci etapta beş seviyeyi tahmin ettik. Verileri içe aktardıktan sonra bir veri çerçevesi (DF) tanımladık. DF, tahmin için kullanılmış olan değişkenleri içerir. Hastanede yatış öngörüsü için hastanın yaşını, yaşamsal belirtilerini ve triyaj seviyelerini öngörücü değişkenler olarak topladık. Diğer veri setleri değişkenleri bu çalışmadan çıkarıldı. Tüm modeller eğitim için %75 ve test verileri için %25 olarak ayrılmıştır. Her tahmin için veriler eğitilir, test edilir ve yerleştirilir. Doğruluk, modele uyduktan sonra her algoritma için hesaplandı. Her bir modeli değerlendirmek için karışıklık matrisini hesapladık ve tahmin edilen her bir sınıf için kesinliği ve hatırlamayı içeren sınıflandırma raporunu içe aktardık (şekil 4.1, 4.2, 4.3 ve 4.4).

AUC, çapraz doğrulama yöntemi kullanılarak hesaplanmaktadır. K-kat çapraz doğrulama, k-eğitim test veri oluşturmasını sistematik olarak oluşturur ve sonuçları bir araya alarak daha iyi bir sonuç elde eder (Link.vd, 2003).

3.2.1 Veri ön işleme

Bu çalışmaya 413,734 hasta (çocuklar ve yetişkinler) dahil edildi. Veri ön işleme, veri sınıflandırma sürecinde önemli bir adımdır. Ön işleme, uygulanmış modelden ML'de daha iyi sonuçlar elde etmek için algoritmaya işlemeden önce verilere uygulanan dönüşümdür. Veri ön işleme için veri setimizde python'da en çok kullanılan iki yöntemi uyguladık: MinMaxScaler ve Standard Scaler.

MinMaxScaler, her bir özelliği belirli bir aralığa ölçekleyerek özellikleri dönüştürür. MinMaxScaler orijinal dağılımın şeklini korur. Orijinal verilere gömülü bilgileri anlamlı şekilde değiştirmez ve aykırı değerlerin önemini azaltmaz (Anonymous.2019c).

Standart Ölçekleyici, ortalamayı kaldırarak ve birim varyansı ölçeklendirerek özellikleri standartlaştırır. Örnek x'in standart puanı şöyle hesaplanır: z = (x - u) / s U, eğitim

örneklerinin ortalaması ise veya with_mean = False ise sıfır; eğer with_std = False ise eğitim örneklerinin standart sapması s'tir (Anonymous.2019d).

Kullanılan veri kümeleri ve komut dosyaları Google veri sürücüsü bağlantısı bulunmaktadır:https://drive.google.com/open?id=1Z4oJlNFuA2equQYuSdgFr9ZwnTvfJkDw.

Çizelge 3.1 MTS veri çerçevesi istatistik açıklaması

Özellik	Mean	Std	Min	Max	%25	%50	%75
Yaş	5.315134	4.577151	0.00	16.00	1.42	3.85	8.60
Orijinal MTS	4.002865	1.254370	1.00	5.00	3.00	5.00	5.00
Solunum hızı	11.621997	16.232045	0.00	124.00	0.00	0.00	0.00
Nabız sayısı	63.952877	64.853184	0.00	260.00	0.00	74.00	123.00
Vücut ısısı	19520557	18.685138	0.00	41.50	0.00	36.00	37.20
Oksijen doygunluğu	45.570347	49.194982	0.00	100.00	0.00	0.00	99.00
Hastane yatış	0.113526	0.317237	0.00	1.00	0.00	0.00	0.00

Çizelge 3.2 ESI veri çerçevesi istatistik açıklaması

Özellik	Mean	Std	Min	Max	%25	%50	%75
Yaş	51.52842	20.34847	18.00	108	34.00	51.00	66.00
ESI	3.014250	0.856387	1.00	5.00	2.00	3.00	4.00
Triyaj_HR	85.52508	16.65524	30.00	244.0	74.00	84.00	96.00
Triyaj_sbp	134.0774	21.62529	45.00	312.0	119.0	132.0	146.0
Triyaj_dbp	80.64998	13.99944	25.00	214.0	72.00	80.00	89.00
Triyaj_rr	17.64542	1.985476	8.00	69.00	16.00	18.00	18.00
Triyaj_o2	97.24293	1.963687	60.00	99.00	96.00	98.00	99.00
Triyaj_o2c	0.043400	0.203756	0.00	1.00	0.00	0.00	0.00
Triyaj_ısısı	98.08729	0.781282	90.00	106.0	97.70	98.10	98.40
Hastane yatış	0.308117	0.461716	0.00	1.00	0.00	0.00	1.00

Çizelge 3.3 MTS değişkeni

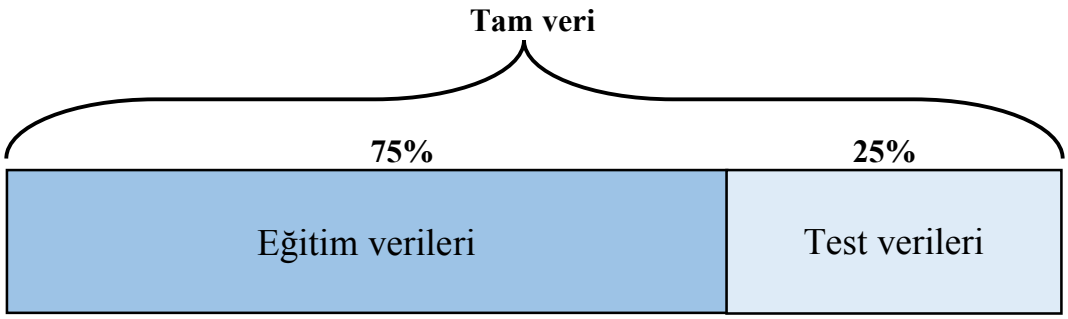
Açıklama	Değişken ismi	Açıklama	Tip
	Yaş	Yaş	Numeric
Yatış	Hastane yatış	Evet (1) ve Hayır (0)	Binary
Hayati belirtiler	Solunum hızı	Triyajda kaydedilen solunum hızı	Numeric
	Nabız sayısı	Triyajda kaydedilen kalp atışı	Numeric
	Vücut ısısı	Triyajda kaydedilen sıcaklık	Numeric
	Oksijen doygunluğu	Triyajda kaydedilen oksijen doygunluğu	Numeric
Triyaj Sistemi	Orijinal MTS	Triyaj hemşiresi tarafından belirlenen MTS seviyesi	Categorical

Çizelge 3.4 ESI değişkeni

Açıklama	Değişken ismi	Açıklama	Tip
	Yaş	Yaş	Numeric
Yatış	Hastane yatış	Giriş (1) veya taburcu (0)	Binary
Hayati belirtiler	Triyaj nabız sayısı	Triyajda kaydedilen sistolik kan basıncı	Numeric
	Triyaj sbp	triyajda kaydedilen diyastolik kan basıncı	Numeric
	Triyaj dbp	triyajda kaydedilen kalp atış hızı	Numeric
	Triyaj solunum hızı	triyajda kaydedilen solunum hızı	Numeric
	Triyaj O2	Triyajda kaydedilen oksijen satürasyonu	Numeric
	Triyaj oksijen	triyajda ek Oksijen cihazı varlığı	Numeric
	Triyaj vücut ısısı	triyajda kaydedilen sıcaklık	Numeric
Triyaj Sistemi	Esi	Triyaj hemşiresi tarafından belirlenen ESI seviyesi	Categorical
Başlıca şikayetler	cc_.....	Hastanın başlıca şikayetleri	Binary

Eğitim ve test seti oluşturma: Verileri eğitim için %75'e, test verileri için %25'e bölüyoruz. Eğitim verileri X adlı bağımsız bir özelliğe sahiptir ve test verileri y adlı bağımlı bir değişkene sahiptir. X Eğitim ve y Eğitim, makine öğrenmesi için model

oluşturma yeteneğine sahip veriler olur. Model oluşturulduktan sonra X Test girişi ve çıktı, y Test değerine eşit olmalıdır. Model çıktısı, y Testi ne kadar yakınsa model de o kadar doğru olur.



Şekil 3.2 Test bölünmüş veri mimarisi

3.2.2 Öğrenme algoritmaları

Tüm algoritmalar python sürüm 3.7.2'de çalıştırıldı. Algoritmalar, scikit learn kütüphanesinden alınır (versiyon 0.20). tüm algoritmalar varsayılan parametrelerle çalıştırılır.

Lojistik regresyon (LR): LR, ayrılabilir doğrusal sınıfa sahip yüksek performanslı bir makine öğrenme sınıflandırma algoritmasıdır. LR ikili sınıflandırma algoritmasıdır ve birkaç sınıfın sınıflandırılmasından genişletilebilir. LR kavramı, şans oranını kullanarak çalışır (Raschka. 2015a). Tahmin için, Scikit learn paketinde bulunan 0.21 lojistik regresyon versiyonunu kullandık. LR Varsayılan parametreler şunlardır:

- penalty (Ceza) isteğe bağlı (varsayılan = ’l2’): Cezada kullanılan normu belirlemek için kullanılır.
- dual (ikili) bool, isteğe bağlı (varsayılan = Yanlış). İkili veya ilkel formülasyon. İkili formülasyon, sadece liblinear çözücü ile l2 cezası için uygulanır.
- tol: şamandıra (float) , isteğe bağlı (varsayılan = 1e-4)
- C: şamandıra (float) , isteğe bağlı (varsayılan = 1.0): Düzenleme gücünün tersi; pozitif bir şamandıra olmalı. Destek vektör makinelerinde olduğu gibi, daha küçük değerler daha güçlü düzenlileşmeyi belirtir.

- fit_intercept (kesişmeye uygun) : bool, isteğe bağlı (varsayılan = Doğru) : Karar işlevine bir sabitin (a.k.a. yanlılığı veya önleme) eklenmesi gerekip gerekmediğini belirtir.
- intercept_scaling (engelleme ölçeklendirme) : float, isteğe bağlı (varsayılan = 1) : Yalnızca 'liblinear' çözücüsü kullanıldığında ve self.fit_intercept, doğru (True) olarak ayarlandığında kullanışlıdır.
- class_weight (sınıf ağırlığı) : dict veya "dengeli", isteğe bağlı (varsayılan = Yok) : Formdaki sınıflarla ilişkili ağırlıklar. Vermezse, tüm sınıfların bir tanesine ağırlık vermesi gerekir.
- random_state (rastgele durum): int, rastgele durum (RandomState) örneği veya Yok, isteğe bağlı (varsayılan = Yok): Verileri karıştırırken kullanılacak sözde rasgele sayı üretecinin tohumu.
- solver (çözücü) : isteğe bağlı (varsayılan = 'liblinear') : Optimizasyon probleminde kullanılacak bir algoritmadır.
- max_iter (en fazla iter): int, isteğe bağlı (varsayılan = 100) : Çözücülerin bir araya gelmesi için alınan azami yineleme sayısı.
- verbose (ayrıntılı): int, isteğe bağlı (varsayılan = 0): Liblinear ve lbfgs çözücüler için, ayrıntılı bir ayrıntı için herhangi bir pozitif sayıya ayarlanmıştır.
- warm_start (sıcak başlangıç) : bool, isteğe bağlı (varsayılan = Yanlış): Doğru olarak ayarlandığında, önceki çağrının çözümünü ilklendirme olarak sığdırmak için yeniden kullanın, aksi takdirde önceki çözümü silin.
- n_jobs (n işleri): int veya Yok, isteğe bağlı (varsayılan = Yok) : çok sınıflı (multi_class) = 'ovr' ise sınıfları paralelleştirirken kullanılan CPU çekirdeği sayısı.

LR, ilgi karakteristiklerinin varlığının olasılığını logit dönüşümünü tahmin etmek için bir formülün katsayılarını (ve standart hatalarını ve önem seviyelerini) üretir:

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_kX_k$$

Burada p, karakteristik özelliğinin var olma olasılığıdır.

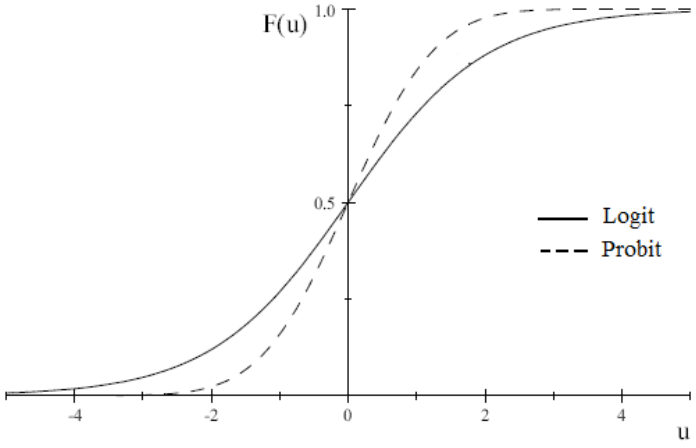
$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristic}{probability\ of\ abscence\ of\ characteristic}$$

ve

$$logit\ (p) = \ln\left(\frac{p}{1-p}\right)$$

Karekök hataların toplamını en aza indirgeyen parametreleri seçmek yerine (sıradan regresyon gibi), lojistik regresyonda tahmin, örnek değerlerin gözlem olasılığını en yükseğe çıkaran parametreleri seçer.

Aşağıdaki grafik (şekil 3.3) , farklı değerler (-4,4) için bir logit ve probit modelinin farkını göstermektedir



Şekil 3.3 Probit ve Logit modelleri için Olasılık Yoğunluk Fonksiyonu (Anonymous.2019e)

Rastgele orman (RF): Rastgele orman algoritması, ML denetimli bir sınıflandırma algoritmasıdır. Rastgele orman, hem sınıflandırma hem de regresyon sorunları için kullanılabilir. Bu çalışmada RF versiyon 0.20 kullandık (boostrap=True, class_weight=None, criterion='gini', max_depth=None, max_features='auto', max_leaf_nodes=None,min_impurity_decrease=0.0, min_impurity_split=None, min_samples_leaf=1, min_samples_splits=2, min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=1, oob_score=False, random_state=None, verbose=0, warm_start=False). Rastgele Orman, ağaçları büyütürken modele ilave rastgelelik katmaktadır. Bir düğümü bölerken en önemli özelliği aramak yerine, rastgele bir özellik alt kümesi içinde en iyi özelliği arar. Bu, genellikle daha iyi bir modelle sonuçlanan geniş bir çeşitlilikle sonuçlanır. Her karar ağacı için, Scikit-learn, yalnızca iki alt düğümü (ikili ağaç) varsayarak Gini önemini kullanarak bir düğüm önemini hesaplar:

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

- ni sub (j) = j düğümünün önemi
- w sub (j) = j düğümüne ulaşan ağırlıklı örnek sayısı

- C sub (j) = j düğümünün kirlilik değeri
- sol (j) = j düğümündeki sol bölmeden alt düğüm
- sağ (j) = alt düğüm j düğümü üzerindeki sağ bölmeden

sub (), abone olarak Orta'da kullanılamıyor.

Bir karar ağacı üzerindeki her bir özelliğin önemi daha sonra şöyle hesaplanır:

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \in allnodes} ni_k}$$

- fi sub (i) = i özelliğinin önemi
- ni sub (j) = j düğümünün önemi

Bunlar daha sonra tüm özellik önem değerlerinin toplamına bölünerek 0 ile 1 arasında bir değere normalleştirilebilir:

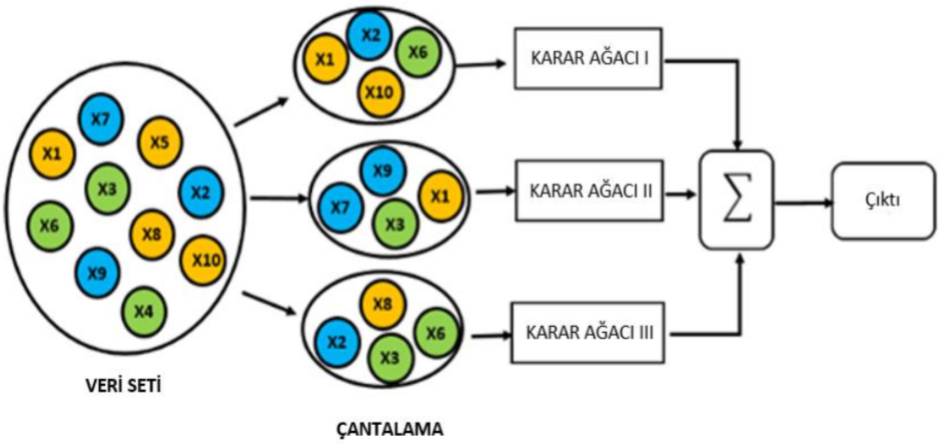
$$normfi_i = \frac{fi_i}{\sum_{j \in all\ features} fi_j}$$

Nihai özellik önemi, Rastgele Orman düzeyinde, tüm ağaçlarda ortalamadır. Özelliğin her bir ağaçtaki önem değerinin toplamı hesaplanır ve toplam ağaç sayısına bölünür:

$$RFfi_i = \frac{\sum_{j \in all\ trees} normfi_{ij}}{T}$$

- RFfi sub (i): Rastgele Orman modelinde tüm ağaçlardan hesaplanan i özelliğin önemi .
- normfi sub(ij) = j ağacında i için normalleştirilmiş özellik önemi
- T = toplam ağaç sayısı

Rastgele orman, bankacılık, ticaret, sağlık gibi birçok alanda kullanılıyor. Sağlık alanında, tıptaki bileşenlerin doğru kombinasyonunu tanımlamak ve bir hastanın tıbbi kayıtlarını kullanarak hastalıkları tanımlamak ve hastanın tıbbi geçmişini analiz etmek için kullanılır (Raschka. 2015b).



Şekil 3.4 RF algoritması mimarisi (Anonymous.2018f)

K en yakın komşular: K en yakın komşular (KNN) algoritması, denetimli ve denetimsiz makine öğrenme algoritmaları sağlayan bir algoritma türüdür. Bu tez çalışmasında KNN 0.21.2 versiyonunu kullandık. KNN versiyon 0.21.2 parametreleri şunlardır:

- n_neighbors (n komşu), int, isteğe bağlı (varsayılan = 5) : K komşu sorgusu için varsayılan olarak kullanılacak komşu sayısı.
- Weights (ağırlıklar) str veya çağrılabilir, isteğe bağlı (varsayılan = "üniforma") : Tahminde kullanılan ağırlık fonksiyonu. Olası değerler: üniforma, mesafe ve istenebilen.
- Algoritma : en yakın komşuları hesaplamak için kullanılan algoritma.
- leaf_size (yaprak büyüklüğü), int, isteğe bağlı (varsayılan = 30) : Yaprak büyüklüğü BallTree veya KDTree'ye geçti.
- p: tamsayı, isteğe bağlı (varsayılan = 2) : Minkowski metriği için güç parametresi. P = 1 olduğunda, bu manhattan_distance (1) kullanmaya eşdeğerdir.
- metrik: string veya çağrılabilir, varsayılan ‘minkowski’: ağaç için kullanılacak mesafe ölçüsü. Varsayılan metrik minkowski'dir ve p = 2 ile standart Öklid metriğine eşittir.
- metric_params (metrik params) : dict, isteğe bağlı (varsayılan = Yok)
- n_jobs (n işleri): int veya Yok, isteğe bağlı (varsayılan = Yok) : Komşular araması için çalıştırılacak paralel işlerin sayısı.

Denetimli komşular temelli öğrenme iki çeşittir: ayrık etiketli verilerin sınıflandırılması ve sürekli etiketli verilerin regresyonu. Komşular temelli sınıflandırma, bir örnek esasına

dayalı öğrenme veya genelleştirmeyen öğrenmenin bir türüdür: Genel bir iç model oluşturmayı denemez, ancak eğitim verilerinin örneklerini saklar. Sınıflandırma, her bir noktanın en yakın komşularının basit çoğunluk oyu ile hesaplanır: Bir sorgu noktası, noktanın en yakın komşularında en fazla temsilciye sahip olan veri sınıfına atanır. Uzaklık hesapları için genelde 3 tip uzaklık fonksiyonu kullanılmaktadır: “Euclidean” Uzaklık, “Manhattan” Uzaklık ve “Minkowski” Uzaklığıdır. Şekil 3.5, her bir fonksiyonun mesafe formülünü gösterir.

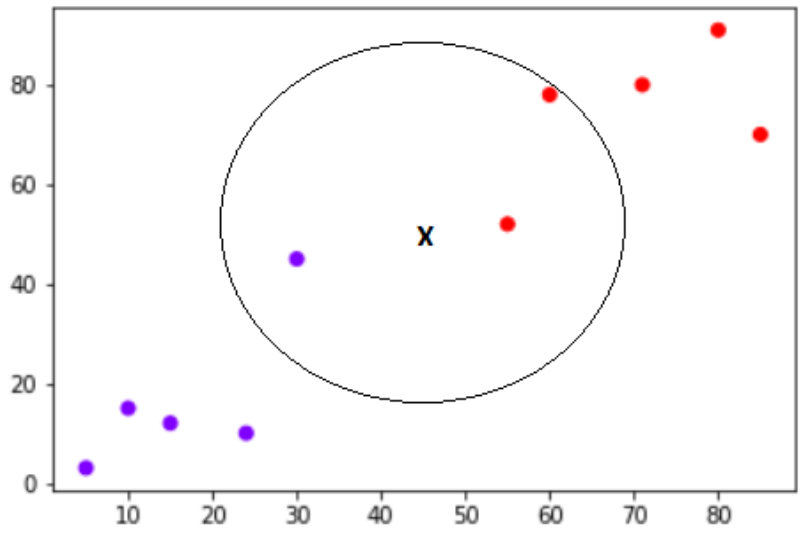
Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Minkowski $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$

Şekil 3.5 KNN algoritması mesafe fonksiyonu (Anonymous.2019f)

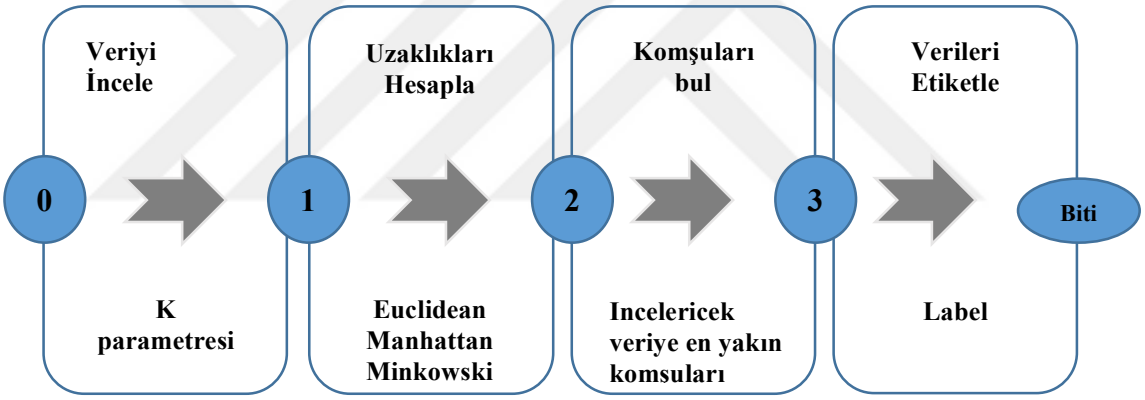
Şekilde 3.6 , en yakın üç noktadan ikisinin "Kırmızı", birinin "Mavi" sınıfına ait olduğu görülmektedir.



Şekil 3.6 KNN örnek çizim (Anonymous.2019)

En yakın sınıflandırıcı, her bir sorgu noktasının en yakın komşusuna dayanan öğrenmeyi uygular; burada k, kullanıcı tarafından belirtilen bir tam sayı değeridir (Anonymous. 2019g). KNN algoritmasının adımları:

- İlk olarak k parametresi belirlenir. Bu parametre verilen bir noktaya en yakın komşuların sayısıdır. Örneğin: k=2 olsun. Bu durumda en yakın 2 komşuya göre sınıflandırma yapılacaktır.
- Örnek veri setine katılacak olan yeni verinin, mevcut verilere göre uzaklığı tek tek hesaplanır. İlgili uzaklık fonksiyonları yardımıyla.
- İlgili uzaklıklardan en yakın k komşu ele alınır. Öznitelik değerlerine göre k komşu veya komşuların sınıfına atanır.
- Seçilen sınıf, tahmin edilmesi beklenen gözlem değerinin sınıfı olarak kabul edilir. Yani yeni veri etiketlenmiş (label) olur.



Şekil 3.7 KNN adımları (Anonymous. 2019h)

XGBoost: "Extreme Gradient Boost" un kısaltması olan XGBoost, 2014 yılında Chen tarafından tanıtıldı. Bu çalışmada XGBoost versiyon 0.90 kullandık. xgboost parametreleri şunlardır:

- Loss (kayıp), isteğe bağlı (varsayılan = "sapma"). Kayıp fonksiyonunun optimize edilmesi. Kayıp için "üstel" gradyan yükseltme AdaBoost algoritmasını kurtarır.
- Learning_rate (öğrenme oranı) : float, isteğe bağlı (varsayılan = 0,1). Öğrenme oranı, her ağacın katkısını öğrenme oranına göre azaltır.

- n_estimators (n tahmin edicileri) : int (varsayılan = 100). Gerçekleştirilecek yükseltme aşamalarının sayısı. Gradyan yükseltme, fazla uydurma için oldukça sağlamdır, bu nedenle büyük bir sayı genellikle daha iyi performans sağlar.
- Subsample (alt örnek): değişken, isteğe bağlı (varsayılan = 1.0). Bireysel temel öğrenicilerin yerleştirilmesi için kullanılacak örneklerin oranı. 1.0'dan küçükse, bu, Stokastik Degrade Artışı ile sonuçlanır.
- Criterion (ölçüt): string, isteğe bağlı (varsayılan = ”friedman_mse”). Bölünmenin kalitesini ölçme işlevi. “friedman_mse” nin varsayılan değeri, bazı durumlarda daha iyi bir yaklaşım sağlayabildiğinden, genellikle en iyisidir.
- min_samples_split (min örnekleri bölünmüş): int, şamandıra (float), isteğe bağlı (varsayılan = 2). Bir iç düğümü bölmek için gereken minimum örnek sayısı.
- min_samples_leaf : int, şamandıra, isteğe bağlı (varsayılan = 1). Bir yaprak düğümünde olması gereken minimum örnek sayısı.
- min_weight_fraction_leaf (min ağırlık fraksiyonu): şamandıra (float), isteğe bağlı (varsayılan = 0.). Bir yaprak düğümünde olması gereken toplam ağırlık (toplam girdi numunelerinin) toplamının minimum ağırlıklı kesridir.
- max_depth (maksimum derinlik): tamsayı, isteğe bağlı (varsayılan = 3). Bireysel regresyon tahmin edicilerin maksimum derinliği. Maksimum derinlik, ağaçtaki düğüm sayısını sınırlar.
- min_impurity_decrease (min kirlilik azalması): şamandıra (float), isteğe bağlı (varsayılan = 0.) Bu bölünme, bu değerden daha büyük veya bu değere eşit olan safsızlığın azalmasına neden olursa bir düğüm bölünecektir.
- min_impurity_split (min safsızlık bölünmesi): şamandıra, (varsayılan = 1e-7). Ağaç büyümesinde erken durma eşiği. Bir düğüm, eğer kirliliği eşiğin üstünde ise, aksi halde bir yapraksa bölünecektir.
- random_state (rastgele durum) : int, RandomState örneği veya yok, isteğe bağlı (varsayılan = Yok). Rastgele durum, rasgele sayı üreteci tarafından kullanılan tohumdur.
- max_features (maksimum özellikler) : int, şamandıra (float), str veya yok, isteğe bağlı (varsayılan = Yok). En iyi bölmeyi ararken dikkat etmeniz gereken özelliklerin sayısı.

- Verbose (ayrıntılı): int, varsayılan: 0. Ayrıntılı çıkışı etkinleştirin. 1 ise, arada sırada ilerleme ve performans yazdırır (ne kadar fazla ağaç frekansı o kadar düşük olur). 1'den büyükse, her ağaç için ilerleme ve performans yazdırır.
- max_leaf_nodes (maksimum yaprak düğümleri) : int veya Yok, isteğe bağlı (varsayılan = Yok). En iyi düğümler kirlilikteki nispi azalma olarak tanımlanır. Yok ise, sınırsız sayıda yaprak düğümü.
- warm_start (sıcak başlangıç): bool, varsayılan: Yanlış. Doğru olarak ayarlandığında, önceki aramanın çözümünü tekrar sığdırmak ve gruba daha fazla tahminci eklemek için yeniden kullanın, aksi takdirde önceki çözümü silin.
- presort: bool veya 'otomatik', isteğe bağlı (varsayılan = 'otomatik'). Uydurmadaki en iyi bölmelerin bulunmasını hızlandırmak için verilerin gösterilip gösterilmeyeceği.
- validation_fraction (doğrulama kesri) : float, isteğe bağlı, varsayılan 0.1. Eğitim verilerinin erken durdurma için belirlenen doğrulama olarak ayrılması oranı. 0 ile 1 arasında olmalı.
- n_iter_no_change (n değişiklik olmaz): int, varsayılan Yok. N değişiklik olmaz, geçerlilik puanı düzelmediğinde eğitimi durdurmak için erken durdurmanın kullanılıp kullanılmayacağına karar vermek için kullanılır. Varsayılan olarak, erken durmayı devre dışı bırakmak için Yok olarak ayarlanmıştır.
- tol: şamandıra, isteğe bağlı, varsayılan 1e-4. Erken durma toleransı. Kayıp (n_iter_no_change) yinelemelerinde (eğer bir sayıya ayarlanmışsa) en azından tol ile iyileşmediğinde, eğitim durur.

XGBoost, gradyanı artırılmış ağaçlar algoritmasının verimli bir uygulamasıdır. Günümüzde en çok kullanılan makine öğrenme algoritmalarından biridir. XGBoost 'un diğer makine öğrenme algoritmalarından daha iyi çözümler sağladığı iyi bilinmektedir. XGBoost, adımı doğrudan çözerek belirlemeye çalışır :

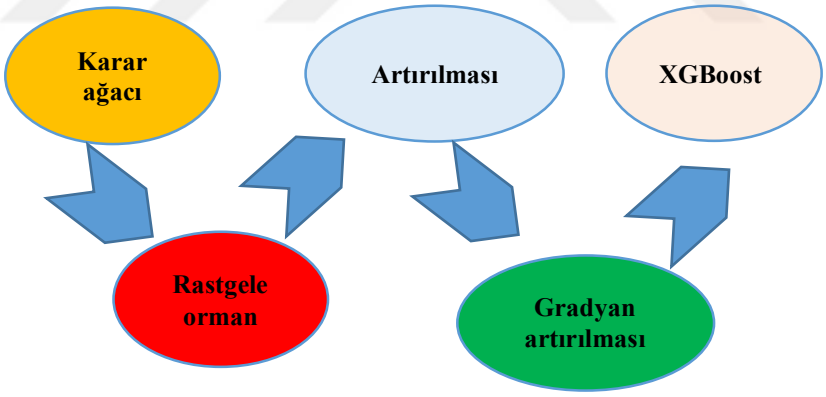
$$\frac{\partial L\left(y, f^{(m-1)}(x)+f_m(x)\right)}{\partial f_m(x)} = 0$$

veri kümesindeki her x için. F (m-1) (x) akımının tahmini değerinin etrafında Taylor fonksiyonunun ikinci dereceden genişlemesini yaparak.

XGBoost, bir yükseltme algoritması ailesine aittir ve gradyan yükseltme çerçevesini özünde kullanır (Anonymous. 2019i). Aslında, kuruluşundan bu yana, yapılandırılmış

verilerle ilgilenmek için “son teknoloji” makine öğrenme algoritması haline gelmiştir. XGBoost algoritması 4 ana geliştirmeye (Anonymous.2019j) sahiptir:

- Regularizasyon: Hem LASSO hem de Ridge regularizasyonu fazla uydurma kullanarak engellenebiliyor.
- Seyreklik Uyumu: Gerçek hayatta veri setleri maalesef bir çok eksik değer bulunduruyor. XGBoost, zayıf öğrenenler ile kayıp değeri eğitim kaybına bakarak en doğru şekilde öğrenebiliyor. Ya da bazen, veri seti belirli bir düzen içerisinde eksik degerlere sahip oluyor (sensor/iletişim hataları vb.) bu durumlarda XGBoost durumu toparlayabiliyor.
- Ağırlıklı Çeyrek Çizim: XGBoost ‘un en büyük avantajlarından biri ağaçlara ayırırken en doğru noktadan ayırabilmek için veri setindeki gözlem noktalarını ağırlıklandırarak kullanıyor olması
- Çapraz doğrulama: XGBoost, kendi içinde çapraz doğrulama uygulaması ile geliyor yani dışarıdan scikit-learn . kullanarak çapraz doğrulama yapmanıza gerek yok. Ayrıca her çalıştırmada kaç iter asyön yapacağını belirtmenize de gerek yok.



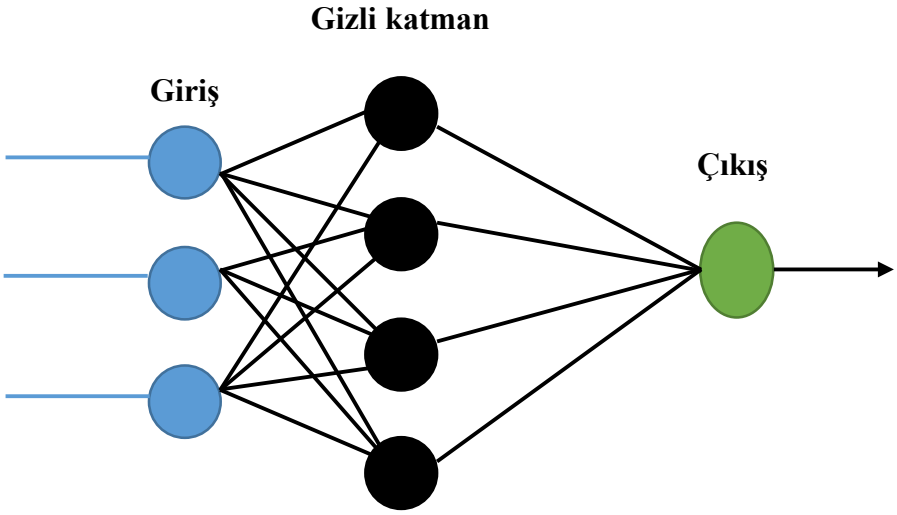
Şekil 3.8 Karar-ağacı algoritmalarının XGBoost evrimi (Anonymous. 2019k)

Derin sinir ağı (DNN): Derin öğrenme, insan beyninin veri işleme ve karar vermede kullanmak için modeller oluşturmadaki çalışmalarını taklit eden yapay bir zeka işlevidir. Derin öğrenme, yapay zekanın (AI) yapılandırılmamış veya etiketlenmemiş verilerden denetimsiz öğrenebilecek ağlara sahip bir makine öğrenmesi alt kümesidir. Derin sinir öğrenme veya derin sinir ağı olarak da bilinir (Anonymous.2019l).

Derin öğrenme ağları, daha yaygın olan tek katlı gizli sinir ağlarından derinlikleriyle ayrılır; yani, verilerin çok aşamalı bir desen tanıma işleminde geçmesi gereken düğüm katmanlarının sayısı. İlk algılayıcılar gibi sinir ağlarının önceki versiyonları sığdı, bir giriş ve bir çıkış katmanından ve aralarında en fazla bir gizli katmandan oluşur.

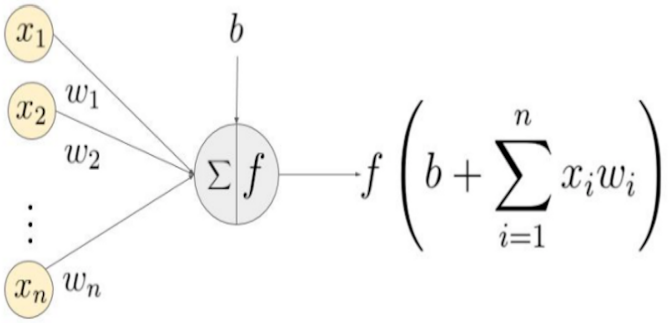
İlk algılayıcılar gibi sinir ağlarının önceki sürümleri, bir girdi ve bir çıktı katmanından ve aralarında en fazla bir gizli katmandan oluşan sığdı. DNN, girdiyi çıktıya dönüştürmek için doğrusal bir ilişki mi yoksa doğrusal olmayan bir ilişki mi olduğu doğru matematiksel manipülasyonu bulur. Ağlar, her çıktının olasılığını hesaplayan katmanlar arasında hareket eder. DNN'ler tipik olarak, verilerin geri döngü olmadan giriş katmanından çıkış katmanına aktığı ileri besleme ağlarıdır. İlk başta, DNN sanal nöronlar haritası oluşturur ve bunlar arasındaki bağlantılara rasgele sayısal değerler veya "ağırlıklar" atar. Ağırlıklar ve girişler çarpılır ve 0 ile 1 arasında bir çıktı döndürür. Şebeke belirli bir modeli tam olarak tanımıyorsa, bir algoritma ağırlıkları ayarlayacaktır. Bu şekilde algoritma, verileri tamamen işlemek için doğru matematiksel manipülasyonu belirleyene kadar bazı parametreleri daha etkili hale getirebilir.

Bu çalışmada, ileri beslemeli (feedforward) sinir ağı oluşturduk. 4 katmanlı, tam bağlantılı (Yoğun sınıf) bir ağ oluşturuyoruz. DNN modeli, tensorflow versiyon 1.14.0'da çalıştırılır. İlk iki katmanda doğrultucu (relu) aktivasyon fonksiyonunu (şekil 3.10) ve çıkış katmanı için sigmoid (şekil 3.11) fonksiyonunu kullandık. Ağ çıkışının 0 ile 1 arasında olmasını sağlamak için bir sigmoid çıkış katmanı kullandık. İlk katman, giriş değişkenlerini kabul etmek için 20 nörona sahiptir. Gizli katmanlar 10 nöron ve çıkış 1 nörondur. Şekil 3.2 sinir ağı mimarisi diyagramının bir örneğini göstermektedir.

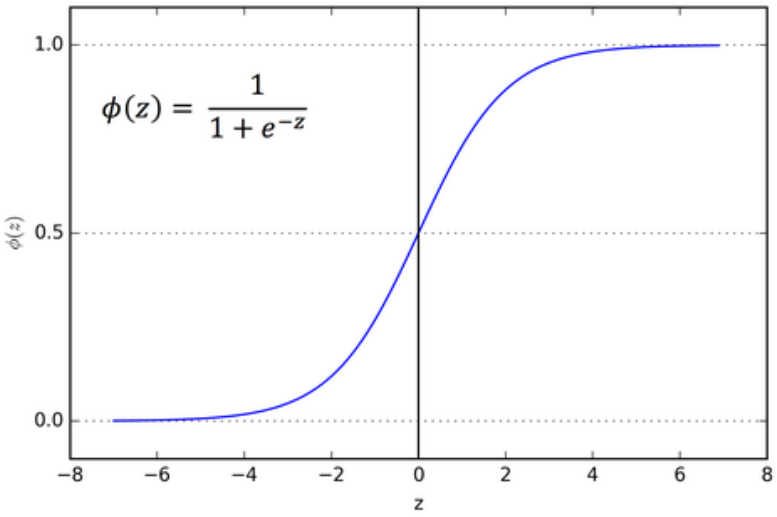


Şekil 3.9 Sinir ağı mimarisi diyagramı

Bir aktivasyon fonksiyonu yapay bir sinir ağının çok önemli bir özelliğidir, temel olarak nöronun aktive edilip edilmemesi gerektiğine karar verirler.



Şekil 3.10 Girdiyi gösteren bir nöron örneği, karşılık gelen ağırlıkları, bir önyargı ve girdilerin ağırlıklı toplamına uygulanan aktivasyon fonksiyonu. (Anonymous. 2018f)



Şekil 3.11 Sigmoid fonksiyon (Anonymous. 2018g)

3.2.3 Model değerlendirme

Model değerlendirme ölçevleri, model ve veri arasındaki uyumu değerlendirmek, farklı modelleri karşılaştırmak, model seçimi bağlamında ve tahminlerin (belirli bir model ve veri setiyle ilişkili) ne kadar doğru olacağını tahmin etmekte kullanılır. Bu tez çalışmasında modellerimizi değerlendirmek için karışıklık matrisi, doğruluk, kesinlik, hatırlama, f skoru, eğri altında, kök ortalama kare hatası, hassasiyet ve özgüllük kullanılmıştır.

Karışıklık matrisi: : Karışıklık matrisi iki veya daha fazla sınıfa sahip bir modelin doğruluğunun bir sunumudur.

Çizelge 3.5 ikili karışıklık matrisi örneği

	Tahmin edilen: evet	**Tahmin edilen: hayır**
Gerçek: evet	Gerçek pozitif: TP	Yanlış pozitif: FP
Gerçek: hayır	Yanlış negatif: FN	Gerçek negatif: TN

Doğruluk: : Sınıflandırma doğruluğu, yapılan tüm tahminlerin bir oranı olarak yapılan doğru tahminlerin sayısıdır. Doğruluk aşağıdaki gibi hesaplanır :

Doğruluk = TP + TN / (TP + FP + TN + FN)

Hassas: Sınıflandırıcı tarafından tahmin edilen pozitif sonuç sayısına bölünen doğru pozitif sonuçların sayısıdır. Hassas formül:

$$\text{Hassas} = TP / TP + FP$$

Hatırlama: İlgili tüm örneklerin sayısına bölünerek doğru pozitif sonuçların sayısıdır.

$$\text{Hatırlama} = TP / TP + FN$$

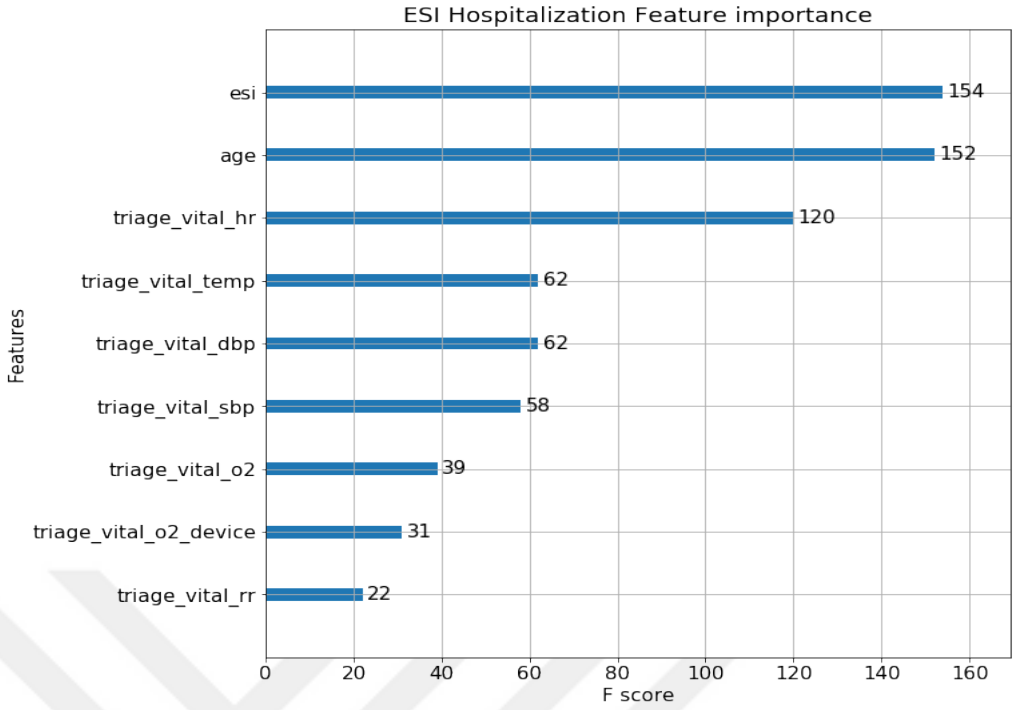
F skoru: F skoru, hassasiyet ve hatırlama arasındaki Harmonik Ortalamadır (Harmonic Mean). F skoru aralığı [0, 1]. Sınıflandırıcımızın ne kadar hassas olduğunu, kaç tane örneği doğru şekilde sınıflandırdığını ve ne kadar sağlam olduğunu anlatır ve önemli sayıda örneği kaçırmaz.

ROC eğrisi: Eğri altındaki alan (AUC) değerlendirme için en yaygın kullanılan ölçümlerden biridir. İkili sınıflandırma problemi için kullanılır. AUC, bir modelin olumlu ve olumsuz sınıflar arasında ayrımcılık yapabilme yeteneğini temsil eder. 1.0 alanı, tüm tahminleri kusursuz bir şekilde yapan bir modeli temsil eder.
RMSE: Ortalama kare hatası (MSE), tahmin (y) ve beklenen (y) arasındaki kare farkının toplamıdır. MSE değerleri belirli bir birimde değildir. Eğer bir model için bir MSE değeri azalmışsa, bu iyidir. RMSE, esasen MSE'nin karekökü'dür.

Duyarlılık: Duyarlılık, hatırlama olarak da adlandırılan gerçek pozitif orandır. Aslında doğru olarak tahmin edilen pozitif (birinci) sınıftan sayı örnekleridir.

Özgüllük: Özgüllük aynı zamanda gerçek negatif oran olarak da adlandırılır. Aslında doğru olarak tahmin edilen negatif (ikinci) sınıftan sayı örnekleridir.

Özelliğin önemi: Tıbbi bir sınıflandırma yapılırken en önemli unsur uygun özelliğin seçimidir. Kötü özelliklerin kullanılması model doğruluğunu azaltabilir. Özellik seçimi genellikle, tüm veri kümelerinden en iyi özellikleri seçerek sistem performansını iyileştirmek için bir ön işleme adımı olarak yapılır.



Şekil 3.12 ESI özellik önem örneği

4. BULGULAR VE TARTIŞMA

Bu bölüm, yaptığımız çalışmanın sonucunda elde ettiğimiz bulgular hakkındaki tartışmaları kapsamaktadır. Bu tezin amacı, makine öğrenme modellerini kullanarak, hastaların, hastaneye yatışını ve sınıflandırılmalarını (triyaj) tahmin etmektir. Tahmin için veri setine dört makine öğrenmesi algoritması (LR, RF, KNN, XGBoost), otomatik kodlayıcı ve derin sinir ağı uygulanmıştır. Bu bölüm araştırmamızın sonucunu, tartışmamızı ve yapılan çalışmanın geçerliliğinin tehdidinin ele almaktadır. İlk başta, elde edilen sonuçları sunacağız. İlk sonuç hastaneye yatış tahmini ve ikinci sonuç triyaj tahmini ile ilgilidir. Triyaj seviyeleri tahmini üç bölüme ayrılır. İlk kısım seviye 1, 2 ve diğerleri için ikili tahmin, ikinci kısım seviye 1 ve 2, artı diğerleri için üç sınıflı tahmin ve son olarak üçüncü kısım ise 5 seviye için çok sınıflı sınıflandırmadır. Daha sonra, tüm sonuçlar tartışma bölümünde detaylı şekilde bahsedilecektir.

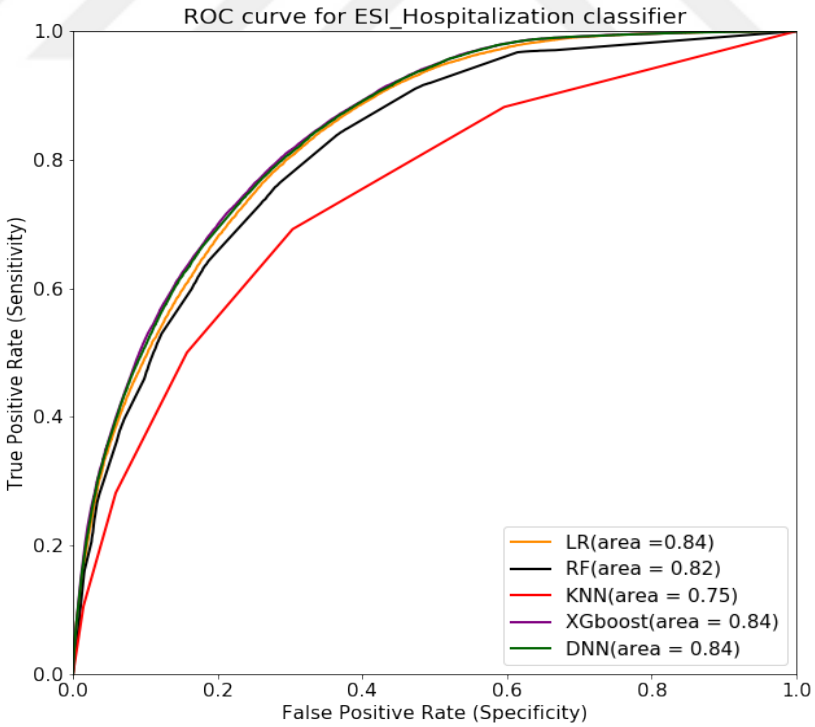
4.1 Araştırma Bulgular

4.1.1. Hastaneye yatış tahmini

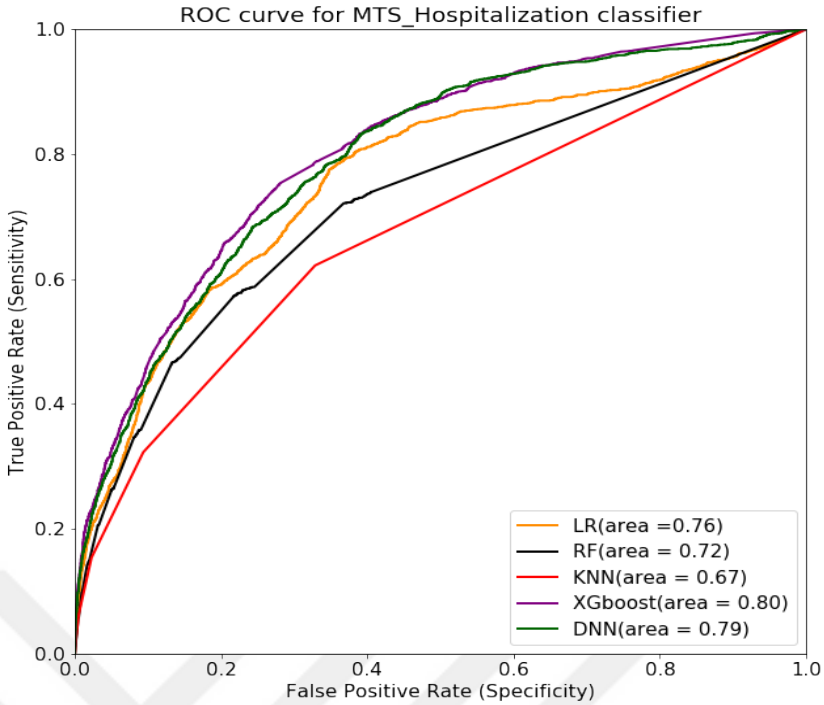
Bu çalışmaya 413.734 ED ziyaretinin bir kombinasyonu dahil edildi. Her bir triyaj sistemi için tüm algoritmalar en yüksek doğruluğa ulaştı (Çizelge 4.1). MTS AUC sonuçları şunlardır : LR için 0.76, RF için 0.72, KNN için 0.67, XGBoost için 0.80 ve DNN için 0.79'dur. Başlıca şikayetleri olmayan ESI için ML modelleri, LR için 0.84, RF için 0.82, KNN için 0.75, XGBoost için 0.84 ve DNN için 0.84 AUC'yi öngörmüştür. Hastanede yatış tahmini doğruluk ortalaması MTS için (0.80 - 0.90) ve ESI için (0.70 - 0.80) arasındadır. LR, XGBoost ve DNN, her iki hastaneye yatış tahmini için en yüksek doğruluğu öngördü. Ancak, başlıca şikayetleri olan ESI hastaneye yatış tahmin doğruluğu, başlıca şikayetleri olmayan tahminlerden daha fazla arttı. Bilindiği gibi, en iyi AUC 1'e yakındır. Alıcı çalışma özellikleri eğrisi veya sadece AUC, doğruluktan (accuracy) daha iyi bir ölçüm sağlar. AUC, sınıflandırma sistemlerini ölçerken ve karşılaştırırken doğruluğu değiştirmelidir. Her iki yatış sınıflamasının özgüllüğü duyarlılıktan daha yüksektir. Bu, her triyaj için öngörülen 0 sınıfının baskınlığından kaynaklanmaktadır.

Çizelge 4.1 hastaneye yatış modeli doğruluk, AUC, hassasiyet ve hatırlama

Model	Doğruluk	AUC	0 için hassasiyet	1 için hassasiyet	0 için hatırlama	1 için hatırlama
MTS hastanede yatış tahmini						
LR	89.08%	0.76	0.87	0.65	0.88	0.54
RF	87.93%	0.72	0.87	0.57	0.97	0.20
KNN	88.22%	0.67	0.90	0.59	0.99	0.10
XGBoost	89.53%	0.80	0.90	0.66	0.94	0.26
DNN	89.24%	0.79	0.90	0.61	0.99	0.10
ESI Başlıca şikayetleri olmadan hastanede yatış tahmini						
LR	77.45%	0.83	0.81	0.66	0.88	0.54
RF	77.09%	0.79	0.81	0.65	0.84	0.58
KNN	73.76%	0.69	0.79	0.58	0.84	0.50
XGBoost	78.34%	0.83	0.82	0.68	0.89	0.54
DNN	77.90%	0.84	0.66	0.82	0.56	0.87
ESI ile başlıca şikayetleri hastanede yatış tahmini						
LR	79.61%	0.86	0.83	0.70	0.89	0.58
RF	78.04%	0.82	0.82	0.67	0.88	0.55
KNN	74.92%	0.75	0.79	0.61	0.86	0.49
XGBoost	79.33%	0.85	0.82	0.70	0.89	0.56
DNN	79.78%	0.86	0.73	0.82	0.53	0.91



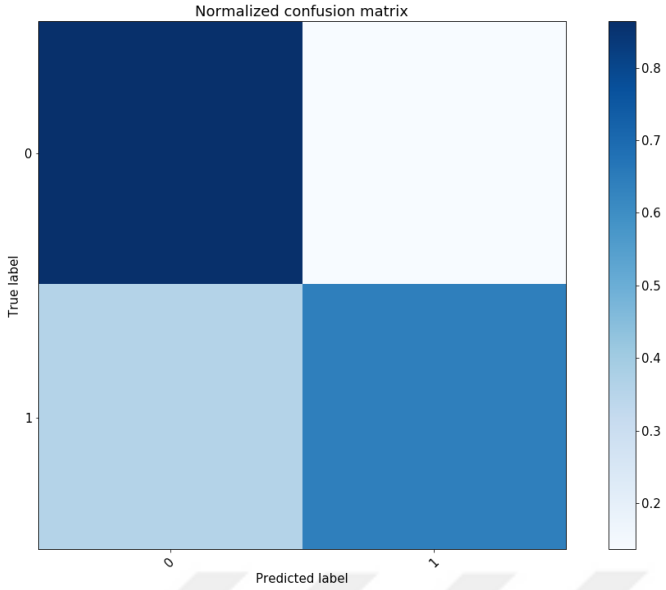
Şekil 4.1 ESI hastaneye yatış ROC eğrisi



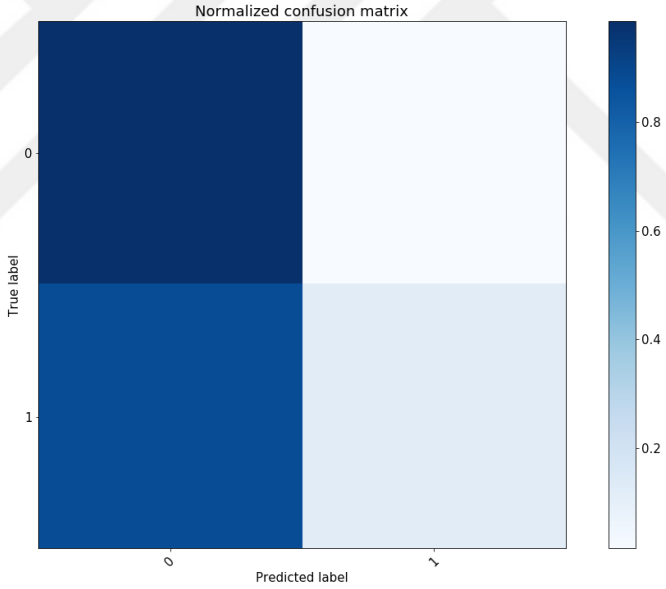
Şekil 4.2 MTS hastaneye yatış ROC eğrisi

Çizelge 4.2 hastaneye yatış duyarlılık ve özgüllüğü

Model	Duyarlılık	Özgüllük
MTS hastanede yatış tahmini		
LR	0.12	0.99
RF	0.20	0.97
KNN	0.097	0.99
XGBoost	0.16	0.98
DNN	0.05	0.99
Başlıca şikayetleri olmadan ESI hastaneye yatış tahmini		
LR	0.71	0.81
RF	0.52	0.84
KNN	0.50	0.84
XGBoost	0.54	0.88
DNN	0.87	0.56
Başlıca şikayetleri ile ESI hastaneye yatış öngörüsü		
LR	0.58	0.89
RF	0.55	0.88
KNN	0.49	0.82
XGBoost	0.56	0.89
DNN	0.91	0.53



Şekil 4.3 ESI hastaneye yatış karışıklık matrisi



Şekil 4.4 MTS hastaneye yatış karışıklık matrisi

Şekil 4.3 ve 4.4, DNN modeli için hastaneye yatış öngörüsünün ikili sınıflandırma standartlaştırılmış karışıklık matrisini gösterir. Öngörülen etiketler 0 ve 1'dir. Sıfır (0) hastanın hastaneye yatmadığı ve bir (1) hastanın hastaneye yatırıldığı anlamına gelir. ESI gerçek-negatif etiketi %80 ve gerçek-pozitif etiketi %70'tir (Şekil 4.3). Yanlış pozitif %20 ila %30 arasındadır, bu modelin gerçek pozitif etiketi doğru tahmin ettiği anlamına gelir. Yanlış-negatif etiket %30 ila %40 arasındadır, bu da %80'den büyük olan gerçek-negatif etikete kıyasla çok düşüktür.

MTS derin sinir ağı hastaneye yatış tahmini karışıklık matrisi, gerçek-negatif etiketi (>%80) ve yanlış-negatif etiketi (%60 -%70) için en yüksek skoru gerçekleştirdi. Gerçek-pozitif etiket (% 20 ila % 40) ve yanlış-pozitif etiket (%10) arasındadır. ESI karışıklık matrisine kıyasla, gerçek-pozitif etiketler model tarafından yanlış tanımlanmıştır. Bu, tahmin sırasında göz önünde bulundurulan değişkenlerden kaynaklanmaktadır. MTS başlıca şikayetler tahmini dikkate alınmamıştır.

Çizelge 4.3 Otomatik kodlayıcı kullanarak hastaneye yatış doğruluk

Veri	**KNN**	**RF**	**LR**	**XGboost**	**DNN**
MTS					
Orijinal	0.8809	0.8835	0.8889	0.8960	0.8894
kodlanmış	0.7801	0.8053	0.8871	0.8669	0.8868
ESI					
Orijinal	0.7483	0.7798	0.7945	0.7920	0.7907
kodlanmış	0.7677	0.7761	0.7854	0.7867	0.7851

4.1.2 Triyaj tahmini

Triyaj tahmini üç aşamaya ayrılır. İlk aşama bir ikili sınıflandırma ve ikinci ve üçüncü aşama çok sınıflı sınıflandırmadır. İlk aşamada seviye 1, seviye 2 ve diğerlerini tahmin ettik. İkinci aşamada ise geri kalanıyla 1. ve 2. seviyeyi tahmin ettik. Son olarak beş seviyeyi tahmin ettik.

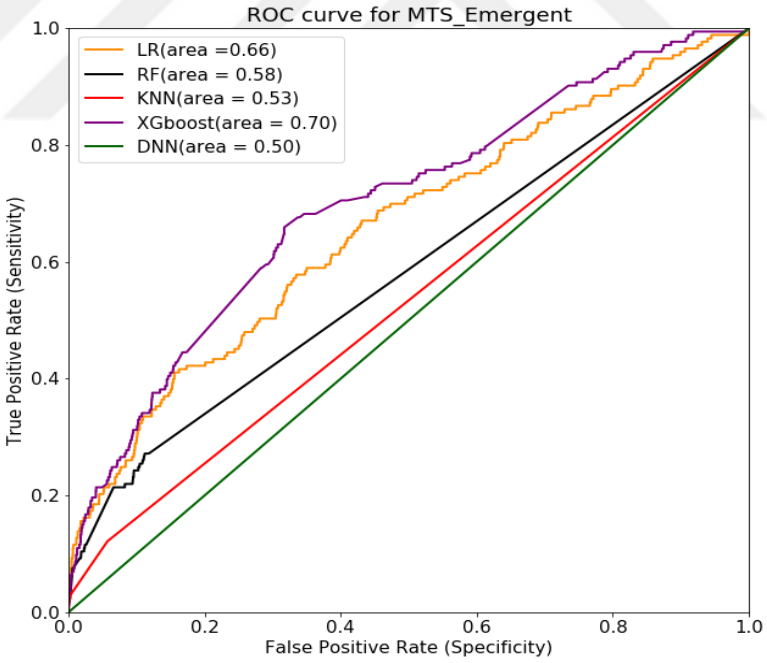
1.seviye AUC için şunlardır (Şekil 4.3 ve Şekil 4.4) : ESI (LR: 0.95, RF: 0.84, KNN: 0.70, Xgboost: 0.94, and DNN: 0.94) ve MTS (LR: 0.66, RF: 0.58, KNN: 0.53, Xgboost: 0.70, DNN: 0.50). Tüm modeller seviye 1 için daha yüksek bir doğruluk öngördü (Çizelge 4.3). Seviye 1 için doğruluk ortalaması şunlardır : MTS için% 98,6 ve ESI için 99,17'dir. Triyaj seviye 2 için öngörülen AUC şunlardır (Şekil 4.5 ve Şekil 4.6): ESI (LR: 0.68, RF: 0.67, KNN: 0.62, Xgboost: 0.71, and DNN: 0.70) ve MTS (LR: 0.74, RF: 0.72, KNN: 0.68, Xgboost: 0.77 and DNN: 0.75). Seviye 2 doğruluk arasında (MTS için 0.84 ila 0.86) ve (ESI için 0.68 ila 0.73).

MTS triyaj seviyesi 1 ve 2 tahmin edilen doğruluk değerleri : LR için %84.44, RF için %83.99, KNN için %83.60, XGboost için %85.43 ve DNN için %84.30'dur. ESI seviye 1 ve 2 doğrulukları : LR için %81.30, RF için % 80.46, KNN için %73.24, XGBoost için %79.37 ve DNN için %67.30'dur. MTS 5 seviye doğruluğu %45 ile %62 arasında

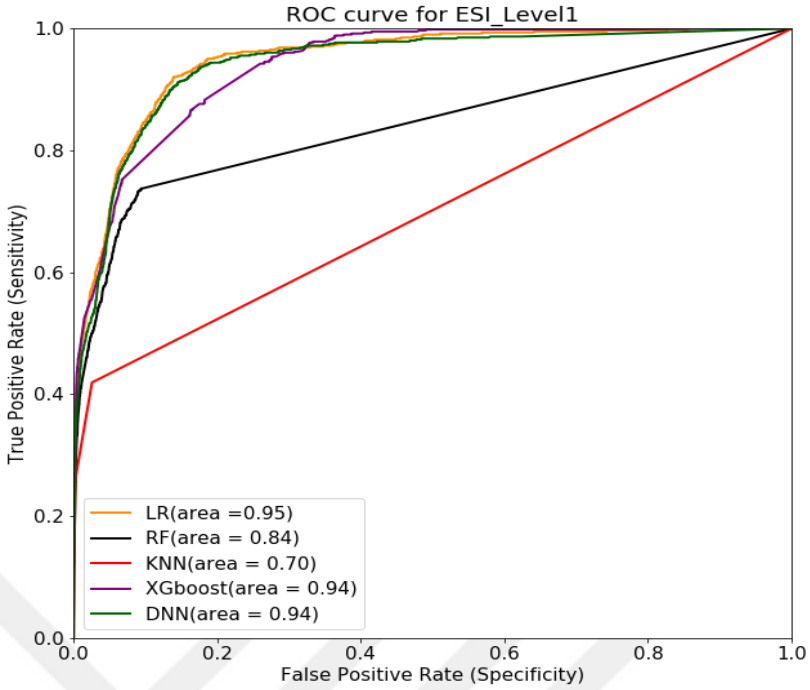
değişmektedir. ESI 5 seviye tahminleri %45 ile %65 arasında değişmektedir. DNN, her iki triyaj tahmininde de maksimum doğruluğu öngördü.

Çizelge 4.4 Triyaj Seviye 1 tahmini

Model	Doğruluk	Kesinlik ortalama	Hatırlama ortalama	F skoru ortalama
MTS				
LR	98.58%	0.97	0.99	0.98
RF	98.55%	0.97	0.99	0.98
KNN	98.58%	0.97	0.99	0.98
XGBOOST	98.57%	0.97	0.99	0.98
DNN	98.72%	0.96	0.98	0.97
ESI				
LR	99.22%	0.99	0.99	0.99
RF	99.13%	0.99	0.99	0.99
KNN	99.15%	0.99	0.99	0.99
XGBOOST	99.21%	0.99	0.99	0.99
DNN	99.20%	0.97	0.98	0.98



Şekil 4.5 MTS seviye 1 ROC eğrisi



Şekil 4.6 ESI seviye 1 ROC eğrisi

Çizelge 4.5 Otomatik kodlayıcı kullanarak triyaj seviye 1 doğruluk

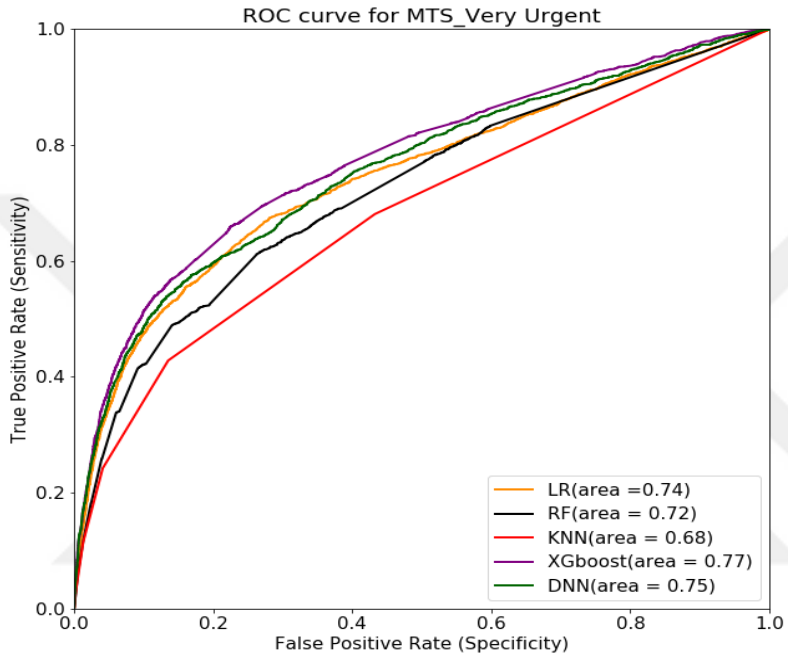
Veri	KNN	RF	LR	XGboost	DNN
MTS					
Orijinal	0.9850	0.9846	0.9851	0.9851	0.9851
kodlanmış	0.9848	0.9837	0.9851	0.9851	0.9851
ESI					
Original	0.9915	0.9913	0.9922	0.9921	0.9920
kodlanmış	0.9889	0.9909	0.9837	0.9920	0.9920

Çizelge 4.6 Triyaj Seviye 2 tahmini

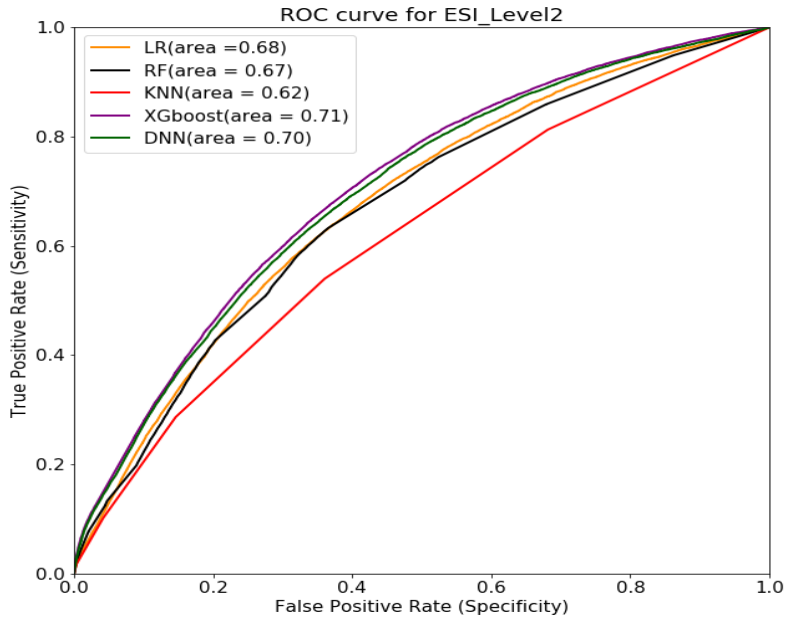
Model	Doğruluk	Kesinlik ortalama	Hatırlama ortalama	F1-score AVG
MTS				
LR	84.17%	0.79	0.84	0.77
RF	85.09%	0.82	0.85	0.83
KNN	84.44%	0.81	0.84	0.82
XGBOOST	86.32%	0.84	0.86	0.84
DNN	85.40%	0.83	0.84	0.83
ESI				
LR	71.83%	0.67	0.72	0.65
RF	71.92%	0.67	0.72	0.65
KNN	68.44%	0.66	0.68	0.67
XGBOOST	73.14%	0.72	0.73	0.65
DNN	72.53%	0.71	0.72	0.71

Çizelge 4.7 Otomatik kodlayıcı kullanarak triyaj seviye 2 doğruluk

Veri	KNN	RF	LR	XGboost	DNN
MTS					
Orijinal	0.8468	0.8503	0.8531	0.8648	0.8249
kodlanmış	0.8435	0.8409	0.8408	0.8605	0.8462
ESI					
Orijinal	0.7183	0.7192	0.7183	0.7314	0.7253
kodlanmış	0.7115	0.7010	0.7019	0.7310	0.7299



Şekil 4.7 MTS seviye 2 ROC eğrisi



Şekil 4.8 ESI seviye 2 ROC eğrisi

Çizelge 4.8 Triyaj Seviyeleri 1 ve 2 tahmini

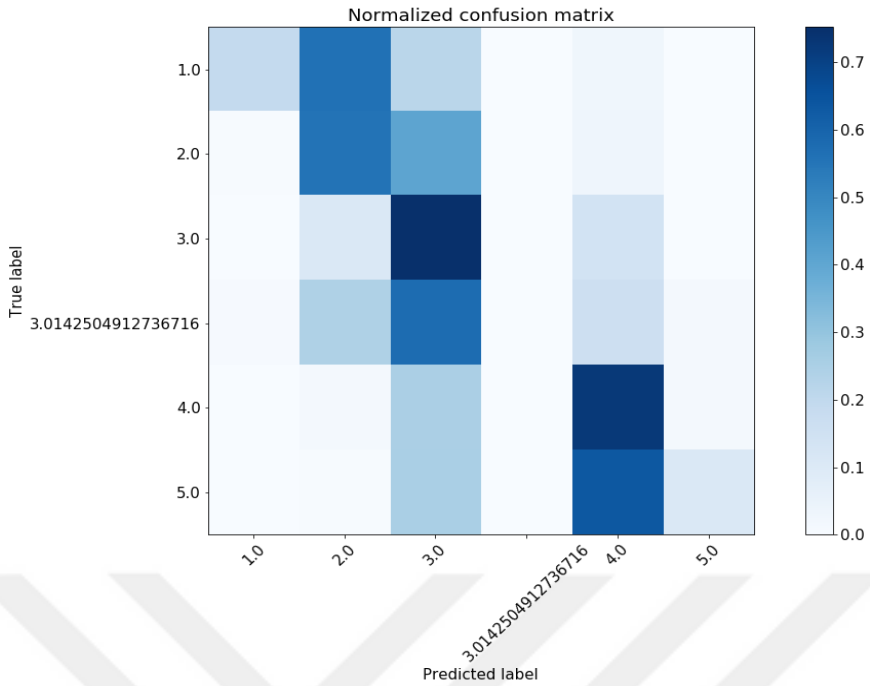
Model	Doğruluk	Kesinlik ortalama	Hatırlama ortalama	RMSE
MTS				
LR	84.44%	0.81	0.84	0.76
RF	83.99%	0.81	0.84	0.77
KNN	83.60%	0.80	0.84	0.78
XGBOOST	85.43%	0.83	0.85	0.73
DNN	84.30%	0.82	0.84	0.76
ESI				
LR	81.30%	0.81	0.81	0.85
RF	80.46%	0.80	0.80	0.86
KNN	73.24%	0.71	0.73	1.02
XGBOOST	79.37%	0.79	0.79	0.89
DNN	81.67%	0.81	0.82	1.24

Çizelge 4.9 Triyaj 5 Seviye Tahmini

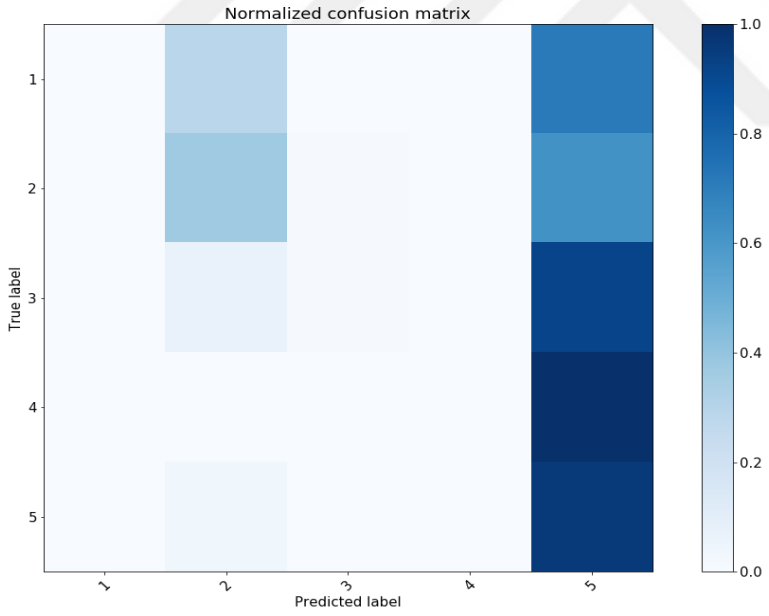
Model	Doğruluk	Kesinlik ortalama	Hatırlama ortalama	F1-skoru ortalama
MTS TRİYAJ SEVİYESİ TAHMİNİ				
LR	65.68%	0.59	0.63	0.62
RF	59.45%	0.55	0.59	0.59
KNN	55.90%	0.53	0.56	0.54
XGBOOST	64.30%	0.61	0.64	0.63
DNN	62.33%	0.57	0.62	0.60
ESI TRİYAJ SEVİYESİ TAHMİNİ				
LR	66.29%	0.66	0.66	0.65
RF	62.15%	0.62	0.62	0.62
KNN	45.42%	0.44	0.45	0.44
XGBOOST	60.82%	0.63	0.61	0.62
DNN	65.89%	0.65	0.66	0.65

Çizelge 4.10 Otomatik kodlayıcı kullanarak triyaj 5 seviye doğruluk

Veri	KNN	RF	LR	XGboost	DNN
MTS					
Orijinal	0.5622	0.6015	0.6284	0.6488	0.6730
Kodlanmış	0.4631	0.8053	0.3016	0.6131	0.6668
ESI					
Orijinal	0.6629	0.6215	0.6629	0.6082	0.6589
Kodlanmış	0.5531	0.6826	0.6815	0.5918	0.6418



Şekil 4.9 ESI 5 seviye karışıklık matrisi



Şekil 4.10 MTS 5 seviye karışıklık matrisi

Çizelge 4.11 Hastaneye yatış öngörüsünün karşılaştırmalı çalışması

Çalışma	Tahmini değişkenler	Modeller	Metrik
Büyük bir ulusal veri seti kullanarak derin öğrenme temelli triyaj ve keskinlik skorunun doğrulanması	yaş, cinsiyet, başlıca şikayet, semptom başlangıcından ED ziyaretine kadar geçen süre, varış şekli, travma, ilk hayati bulgular ve zihinsel durum.	DTAS	AUROC : 0.804
		KTAS	AUROC: 0.681
		NEWS	AUROC: 0.614
		RF	AUROC: 0.738
		LR	AUROC: 0.713
ML modellerini kullanarak kritik bakım ve hastaneye yatışı (Raita.vd 2019)	hasta yaşı, cinsiyeti, varış şekli, triyaj hayati bulguları, başlıca şikayetleri ve hasta ek tanısı kullanılır.	Referans modeli	AUC : 0.69
		Kement regresyonu	AUC : 0.81
		RF	AUC : 0.81
		Degrade artırılmış karar ağacı	AUC : 0.82
		DNN	AUC : 0.82
ML kullanılarak yapılan hastaneye başvuru öngörüsü	Tam veri setleri	LR	AUC : 0.87
		XGBoost	AUC : 0.87
		DNN	AUC : 0.87
Makine öğrenmesi temelli elektronik triyaj, acil durum şiddeti indeksine kıyasla hastaları klinik sonuçlar açısından daha doğru bir şekilde farklılaştırır.	yaşamsal belirtiler, başlıca şikayet ve aktif tıbbi öykü	RF	AUC: 0.73 to 0.92
Acil servis triyajı sırasında çocuklar için klinik sonuçların makine öğrenmesine dayalı tahmini.	Yaş, cinsiyet, variş modu, yaşamsal bulguları, ziyaret sebebi, yatış durumu	RM	C istatistik :0.73
		LR	C istatistik: 0.78
		RF	C istatistik: 0.80
		Xgboost	C istatistik: 0.80
		DNN	C istatistik: 0.80
Makine öğrenmesi ile triyaj ve hastane yatış tahmini	MTS: Hasta yaşı, yaşamsal bulguları ve triyaj seviyeleri	LR	AUC : 0.76
		RF	AUC : 0.72
		XGBoost	AUC : 0.80
		KNN	AUC : 0.67
		DNN	AUC : 0.79
	ESI: Hasta yaşı, yaşamsal bulguları ve triyaj seviyeleri	LR	AUC : 0.83
		RF	AUC : 0.79
		KNN	AUC : 0.69
		XGBoost	AUC : 0.83
		DNN	AUC : 0.84

Çizelge 4.11, bu çalışmanın ve hastaneye yatış öngörüsü ile ilgili diğer çalışmaların karşılaştırmasını özetlemektedir. DNN'yi diğer ML algoritmaları (RF, KNN, XGBoost ve LR) ile hastaneye yatış ve hasta triyajını sınırlı bilgi ile tahmin etmek için uyguladık. Tüm makine öğrenme modelleri, her iki triyaj sistemi (MTS ve ESI) için yüksek bir doğruluk öngörmüştür. Önceki çalışmalar, triyajda daha fazla hasta kaydedilmiş hasta bilgisi kullanılarak hastaneye yatış öngörmüştür. ROC eğrisi analizinden elde ettiğimiz sonuçlar, ilgili çalışmalarla karşılaştırıldığında, ML'nin aynı sonuçları ve hatta hastaneye yatış için daha da fazlasını öngörebileceğini göstermiştir. Örneğin, Kwon vd. (2018a) çalışmalarında (çizelge 4.11) LR ve RF sırasıyla 0.713 ve 0.738 AUROC tahmin ederken, araştırmamız aynı modeller için 0.83 ve 0.79 tahmin etti. Hong. vd (2018c) araştırmaları, tam veri setini kullanılarak yapılan hastaneye başvuru öngörüsü üzerine tüm modeller için bir AUC 0.87 öngördü: LR, RF ve XGBoost. Hasta sınırlı bilgisini kullanarak LR, RF ve XGBoost için 0,83, 0,79 ve 0,83 elde ettik. Aradaki fark, AUC sonuçlarımıza yakın olan sadece 0.4'tür. MTS AUC, acil servis triyajı sırasında çocuklar için klinik sonuçların öngörülmesine dayalı Makine öğrenmesi üzerine yapılan Goto.vd (2019a) çalışmasına kıyasla, gradyan artırma algoritması için aynı sonucu öngördü. Bulgularımız ML'nin hastanede yatışı sınırlı hasta bilgileri ile öngörebileceğini göstermektedir. Bu sonuçlar gelecekteki araştırmacılara akıllı elektronik triyaj sistemi uygulamak için katkıda bulunabilir. Bu sistem, triyaj sürecini ve tesis karar verme sürecini azaltacaktır.

4.2 Tartışma

Bu tez çalışmasında, ML algoritmalarının iki farklı triyaj veri setindeki performans karşılaştırmasını yaptık : MTS ve ESI. Bu araştırma triyaj sistemlerinin karşılaştırmalı bir incelemesi değildir. Araştırmanın asıl amacı, hastaların sınırlı sayıdaki bilgilerini kullanarak hastaneye yatış ve triyaj seviyelerini tahmin etmektir. Geleneksel olarak triyaj, zaman, başlıca şikayetler gibi hasta bilgilerini, tarihi laboratuvarları, görüntülemeyi, yaşamsal bulguları ve karar almak için diğer kaynakları gerektiren karmaşık ve dinamik bir süreçtir. Karar verme, triyaj sürecinde en önemli görevlerden biridir. Karar verme, triyaj alanında bilgili hemşireler tarafından alınır. Triyaj hataları ölüm oranını artırabilir. Bu prosedürlerin tümü zaman gerektirir. Her triyaj sisteminin ilk amacı hastanelerdeki tıkanıklığı azaltmak ve hayati tehlikesi olan hastaları tanımlayarak ölümlerin önüne

geçmektir. Hastaların hastaneye yatışı, hemen triyajdan sonra yapılır. İşlem süresini azaltmak için, en önemli özellikleri göz önünde bulundurarak, hastaneye yatış ve triyaj tahmininde bulunmayı öngörmek için makine öğrenmesi (ML) modellerini kullandık. Bu Modeller, hastaneye yatış tahmini için daha yüksek bir özgüllük ve triyaj için daha yüksek hassasiyet göstermektedir.

Hastaneye yatış öngörüsü (Çizelge 4.1), tüm algoritmalar yüksek doğruluk seviyesine ulaştı (>% 70). AUC ESI hastaneye yatış öngörüsü için (Çizelge 4.1), MTS'den daha yüksek bir skor kaydetmektedir. Bu, tahmin sırasında kullanılan özelliklerden kaynaklanmaktadır. MTS hayati belirtiler kaydedilen şudur: sıcaklık, oksijen, doygunluğu, solunum hızı ve kalp hızı. Kaydedilen ESI yaşamsal belirtileri sıcaklık, oksijen doygunluğu, solunum hızı, kalp atış hızı, sistolik kan basıncı, diyastolik kan basıncı ve tamamlayıcı oksijenin varlığıdır. ESI triyaj hayati belirtilerde kaydedilen üç değişkene daha sahiptir (Sistolik kan basıncı, Diyastolik Kan Basıncı ve Ek Oksijen Varlığı). Bu üç değişkeni veriden kaldırıp algoritmayı çalıştırdığımızda öncekinden daha az doğruluk elde ettik. Dolayısıyla, bu üç değişkenin varlığı, algoritmanın performansını artırarak tahmin sonucunu etkiler. ESI ROC eğrisinin (şekil 4.1) MTS ROC eğrisinden (şekil 4.2) daha hassas olmasının nedeni budur. MTS hastaneye yatış tahmini sırasında, başlıca şikayetleri dikkate alınmaz. Sınıflandırmada sadece hasta yaşı, yaşamsal belirtiler ve triyaj düzeyleri göz önünde bulundurulmuştur. İlgili çalışma bölümünde, ML sınıflandırma araştırmalarının çoğunun tahmin için daha fazla hasta bilgileri göz önüne aldığını gördük. Başlıca şikayetlerin MTS öngörüsünden çıkarılmasına rağmen, MTS doğruluğu ESI doğruluğundan daha büyük (başlıca şikayetleri olan ve olmayan). Doğruluk açısından, ML modelleri çocukların hastaneye yatışlarını öngörmede daha doğrudur. Geleneksel olarak bir hasta bir hastane acil sevisine (ED) başvurduğunda, triyaj hemşireleri önce hastanın yaşamsal belirtilerini, yaşını ve diğer kişisel bilgilerini alır. Yaşamsal belirtiler, her türlü triyaj sürecinde göz önünde bulundurulan ilk unsurlardır ve ne kadar önemli olduklarını daha önce belirttik. Günümüzde teknolojinin ilerlemesiyle, hayati belirtileri gerçek zamanlı olarak hassas bir şekilde ölçebilen ve bu tıbbi verileri analiz için uygun bir sisteme gönderen vücut sensörleri vardır. Bu teknoloji gelecekteki triyaj sistemi sürecinde herhangi bir tıbbi triyaj tahmini için kullanılabilir ve triyaj işlemi sırasında hastanelerdeki tıkanıklık süresini kısaltabilir. DNN, XGBoost ve LR, hastanede yatış tahmini için kullanılan en iyi modeller oldu. Önceki çalışmalara kıyasla, makine

öğrenmesinin (ML) sınırlı hasta bilgisi ile hastaneye yatışı tahmin edebildiği sonunca vardık.

ESI seviye 1, tüm ML modelleri için %99'luk bir yüksek doğruluk ortalaması öngördü. Çizelge 4.4 , kesinlik, hatırlama ve f skorunun iyi tanımlandığını göstermektedir. Xgboost, LR ve DNN için eğri analizi altındaki alan, triyaj seviyesi 1'i tahmin etmek için daha yüksek hassasiyet (Şekil 4.6) göstermiştir. Özel olarak, ML, en kötü hastayı ortaya çıkan düzeyde öngörmüştür. ESI seviye 2'nin eğri analizi (şekil 4.8) , ML modellerinin ESI seviye 1 eğrisi ile karşılaştırıldığında yetişkinler için seviye 2'yi öngörmede daha az duyarlı olduğunu göstermiştir. Seviye 2 tahmini sırasında, hatırlama ve f skoru, öngörü sınıfı için tanımlanmamıştır (seviye 2) ve bazı modellerde (LR, RF, DNN) sıfıra ayarlanmıştır. ML modelleri, en kötü yetişkinin hastanın seviyesini öngörmede daha doğru ve hassastır (1 ila 3).

MTS seviye 1 (hemen) eğrisi analizi (Şekil 4.5), çok acil seviyeden (0,68 ila 0.74) düşük bir hassasiyet (0,50 ila 0,70 arasında) olduğunu göstermektedir. ESI birinci seviye ile karşılaştırıldığında MTS birinci seviye , AUC analizinde ESI'nin birinci seviyesinin MTS'den daha duyarlı olduğunu göstermiştir. Bunu, ML modellerinin çocukların acil durum seviyesi 1'in öngörülmesinde daha az doğru olduğu anlamına gelir. Ancak, tüm modeller birinci seviye için% 98'in doğruluğunu öngördü. Her ML modelinin sınıflandırma raporu, hassasiyet, hatırlama ve f skoru için düşük bir skor gösterdi. MTS sırasında birinci seviye sınıf tahmini, hassasiyet, hatırlama ve f skoru DNN, RF ve KNN tarafından tanımlanmamıştır. ML modelleri, hasta çocukların seviyesini triyaj altında tahmin etmede daha kesin ve hassastır (3 ila 5).

MTS ve ESI tahmini için triyaj seviyeleri (1 ve 2) yüksek doğruluk verdi (Çizelge 4.8). LR ve DNN modelleri, her iki triyaj için en yüksek doğruluğu tahmin eder. İlk iki triyaj seviyesi (Acil ve Çok Acil) en kritik seviye olarak bilinir. Her triyaj sisteminin mimarisine göre hareket edilmelidir. Örneğin seviye 1 için başvuran bir hasta beklemeden doktora görünmeli ve ikinci seviyedeki hasta 10 dakika içinde görülmelidir. ML, bu iki seviyeyi kabul edilebilir bir doğrulukla öngördü (%73 -% 85). Kesinlik ve hatırlama, modelin öngörmeyi ne kadar iyi yaptığını gösterdi (Çizelge 4.8). Yukarıda tartışıldığı gibi, triyaj karar verme sürecinde daha fazla hasta bilgisine ve hemşire bilgisine ihtiyaç duyar. Burada, bu iki seviyeyi sadece hasta sınırlı bilgisini kullanarak tahmin ettik.

Gerçek hayatta, hasta sınırlı bilgisi triyaj kararını zorlaştırır. 5 seviyeli karar verme, 2 veya 3 seviyeden daha karmaşıktır. Doğruluk MTS için % 55ila %65 ve ESI için % 45ila %65 arasındadır (Çizelge 4.9). Bu doğruluk tıbbi tahmin için çok düşüktür. MTS, LR ve XGBoost modelleri en yüksek doğruluğa sahiptir. MTS 5 seviyeleri tahmini sırasında bazı hatalar meydana geldi. Kesinlik ve f skoru kötü tanımlanmıştır ve öngörülen numunelere sahip olmayan etiketlerde sıfıra ayarlanmıştır. Verimdeki bazı etiketler 'y_pred'de görünmüyor. Özellikle, bu durumda, '2', '3' ve '4' etiketlerinin LR, KNN ve XGBoost için hiçbir zaman öngörülememesi : bu etiket için hesaplanacak hiçbir f skoru olmadığı ve bu nedenle bu vaka için f-skorunun sıfır olduğu anlamına gelir. NAN değerinin baskınlığı, her iki tahminin doğruluğunu da etkiler.

Şekil 4.9 ve 4.10, DNN için triyaj 5 tahmin seviyelerini gösterdi. ESI'nin karışıklık matrisi (şekil 4.9), 3. ve 4. seviyelerin doğru tahmin edildiğini (>%70) ve seviye 1 ve 5'in yanlış tanımlandığını (<% 10) göstermiştir. Bununla birlikte, MTS karışıklık matrisinin sadece seviye 5'i doğru tahmin ettiği görülmüştür (% 80 ila% 90). Makine öğrenim modelleri 3 ila 5 seviyesindeki çocuklar ve 1 ila 3 seviyesindeki yetişkinler için daha duyarlıdır. MTS derin sinir ağı hastaneye yatış tahmini karışıklık matrisi, gerçek-negatif etiketi (>%80) ve yanlış-negatif etiketi (%60 -%70) için en yüksek skoru gerçekleştirdi. Gerçek-pozitif etiket (% 20 ila % 40) ve yanlış-pozitif etiket (%10) arasındadır. ESI karışıklık matrisine kıyasla, gerçek-pozitif etiketler model tarafından yanlış tanımlanmıştır. Bu, tahmin sırasında göz önünde bulundurulan değişkenlerden kaynaklanmaktadır. MTS başlıca şikayetler tahmini dikkate alınmamıştır.

Tahmin sonucunu arttırmak için otomatik kodlayıcı yöntemleri uyguladık. Otomatik kodlayıcı kullanmanın doğruluğu azalttığı sonucuna vardık. Otomatik kodlayıcı, bir DNN türüdür. Otomatik kodlayıcı işlemi, kodlama ve kod çözme verilerinden oluşur. Kodlama işlemi verileri belirtir. Çıktı bilgileri kod çözücü yöntemleriyle elde edilmektedir. Tıbbi verilerde otomatik kodlayıcıyın kullanarak, orijinal verilerle düşük doğruluk elde ettik. Bu, verilerin sıkıştırıldığı ve veri kaybına neden olduğu kodlama işlemidir. Derin otomatik kodlayıcı test sürecinde sıklıkla başarısız oldu. Ancak RF ile MTS 5 seviyesinin tahmini için %80 doğruluk elde ettik (Çizelge 4.10). Bunun birkaç nedeni olabilir: Otomatik kodlayıcı her zaman girişin ortalamasını verebilir, giriş setini her zaman tam olarak yeniden yapılandırabilir, bu iki kusuru sinsi yollarla birleştirebilir ya da eğitim

setinde mükemmel bir şekilde çalışabilir, fakat örneklem dışı kötü bir şekilde başarısız olabilir.

4.3 Geçerliliğe Yönelik Tehditler

İçsel geçerlilik : Bu çalışmada, triyaj sistemlerinden kaydedilen iki farklı veri setinde iki tür tahmin (yatış ve triyaj) yaptık : MTS ve ESI. MTS 60,735 çocuk ED ve ESI 352,999 yetişkin ED içermektedir. ESI veri kümesi, başlangıçta çıkardığımız R verileridir. Çıkarma işlemi sırasında verilerin bir kısmı kaybolur. Çalışmanın ana amacı, hastanın hastaneye yatışını tahmin etmek ve triyaj birincil kaydedilen bilgisiyle bilinen triyajı belirlemektir : yaşamsal belirtiler ve yaş (diğer hasta bilgileri dikkate alınmadan). Hastaneye yatış tahmini için %80'den fazla bir doğruluk elde ettik. Ancak, yukarıda da belirtildiği gibi, bu deneyimin gerçek hayatta başka bir veri setiyle aynı sonucu vereceğini garanti edemeyiz. Triyaj, triyaj hemşirelerinin, hastaların durumlarına göre birinci basamak tedaviye yönlendirebilmeleri için hasta bilgilerini topladıkları bir işlemdir. Bu deneyimi onaylamak için gerçek hastalarla test edilmelidir.

Dış geçerlilik : Bu çalışma sadece iki triyaj sistemi veri setiyle ilgilidir. Diğer triyaj sistemleri veri seti ile sonuç farklı olabilir. MTS veri kümesi, çocukların veri ve ESI veri yetişkin verileridir. Yaşın ML kullanarak triyaj kararını etkileyip etkilemediğini değerlendirmek ve bu hipotezi doğrulamak için ayrı bir çalışma yapılmalıdır. Bir sonraki çalışmada her triyaj sistemi için aynı veri türü göz önünde bulundurulmalıdır: her veri seti için yetişkinler veya çocuklar. Çalışma henüz gerçek bir hastanede test edilmemiştir.

5. SONUÇ

Bu çalışmada, acil servis departmanında (ED) kaydedilen iki farklı triyaj sistemi (MTS ve ESI) veri setinde ML modellerini kullanarak hastaneye yatış ve triyaj tahminini yaptık. Amacımız, hastaların sınırlı olan mevcut bilgilerini kullanarak hastaneye yatışı ve triyajı öngörmektir. Çalışmanın ikinci bölümünde, daha önce yapılan ilgili çalışmalarda triyaj tahmin sırasında çok fazla hasta bilgisi kullanıldığını gördük. Bu araştırmayı yürütmek için iki sebep bizi motive etti. Triyaj, hasta kaynaklarını, zamanlarını ve triyaj hemşirelerinin deneyimlerini gerektiren bir süreç olduğu için ilk sebep triyaj sürecini azaltmaktır. İkinci motive edici sebep, biyosensör teknolojisinin yaşamsal belirtileri ölçmek ve triyaj işlem süresini azaltmak için kullanılması ve bu teknolojinin ilerlemesinin tıp alanında bir devrim yaratmasıdır. Bu akıllı sensörler yaşamsal belirtileri ölçebilir ve bilgileri gerçek zamanlı olarak görselleştirebilir. ML ile, makine kendini verilerden öğrenebilir ve sonuçları gerçeğe yakın büyük bir doğrulukla daha hızlı tahmin edebilir. Aynı zamanda, portatif tıbbi sensörler yoluyla toplanabilen hasta özellikleri kullanılarak, hastaların hastaneye yatışı çok yüksek doğrulukla ve acil durum seviyeleri de kabul edilebilir şekilde tahmin edilebilir. Bu nedenle, gelecekte akıllı bir triyaj sistemi oluşturmak için sensörler ML teknikleriyle birleştirilebilir. Bu çalışma triyajı daha hızlı ve verimli hale getirmek için triyaj alanındaki diğer araştırmaların başlangıcı olmalıdır.

Bu çalışma bazı sınırlamalar sunmaktadır. Hastanede yatış ve triyajı öngörmek için sadece hasta yaşı, yaşamsal belirtiler, başlıca şikayetler ve triyaj seviyeleri kullanılmıştır. Hastanın triyajda kaydedilen diğer bilgileri hariç tutulmuştur. Ancak, her triyaj sisteminin karar verme süreci için kendi algoritması ve prosedürü vardır. Bu çalışmanın amacı, triyaj sisteminin performansını karşılaştırmak değil, triyajda kaydedilen sınırlı bilgiyi kullanarak hastaneye yatış ve triyajı öngörmektir. ESI için hastanede yatış tahmininde MTS'den yüksek AUC sonuçları elde etsek bile, ESI'nin MTS'den daha iyi olduğu sonucuna varamayız. ESI için başlıca şikayetler tahmin sırasında dikkate alınmıştır. ML kullanarak triyajda hastaneye yatış kabulünü öngörme üzerine yapılan benzer bir çalışmada, sadece triyaj bilgilerinin kullanılmasının yerine, hastanın geçmiş bilgilerinin de eklenmesinin tahmini performansı arttırdığını göstermektedir.

Gelecek çalışmalar : ML modellerinin yetişkinler için triyaj sistemi seviyelerini (1'den 3'e kadar) ve çocukların veri setine yönelik seviyeleri (3'ten 5'e) tahmin etmekte daha

duyarlı olduğunu bulduk. Bu bulguları doğrulamak için, her durumda çocuk ve yetişkin veri setlerini kullanan iki veya daha fazla triyaj sisteminde karşılaştırmalı bir çalışma yapılmalıdır. Gelecekteki çalışma, doğruluğu artırmak ve eksik olan değerleri işlemek için başka teknikler kullanmayı içerecektir.

Gelecekteki araştırmacılar için sağladığımız başlıca öneriler şunlardır:

- Bu çalışmada kullanılan algoritmalar ve yöntemler aynı zamanda başka veri işleme yöntemleri ve Hiper tüner parametreleri uygulanarakta kullanılabilir.
- Özellikle triyaj seviyelerinin tahmini için eksik değeri olmayan veri kümesi ile çalışma yapılabilir. Triyaj, tüm değişkenlerin aralarında güvenilir olduğu karmaşık bir görevdir. Gerçek hayatta bile, hasta eksik değeri ile triyaj karar sürecini yapmak zordur. Bir hastayı beş seviyeden birine sokmak için, kaydedilen tüm hasta bilgileri arasında bir ilişki vardır (Yaş, hayati belirtiler, başlıca şikayetler ve diğerleri). Bazı bilgiler eksik olduğunda karar verme süreci zorlaşmakta ve tahmin doğruluğu azalmaktadır.
- Otomatik kodlayıcılar, geleneksel sinir ağı, üst düzey derin öğrenme çerçevesi (Keras) ve ML algoritmaları gibi diğer teknikler, doğruluğu arttırmak için gelecekteki tahminlerde kullanılabilir. Mesela, Keras, DL modelinin geliştirilmesine ikinci düzeyde bir soyutlama sağlayan en popüler üst seviye derin öğrenme çerçevesidir. Herhangi bir tahminden önce, uygun özelliğin seçimi de yapılması gereken en önemli adımlardan biridir. Deneyimlerimizde doğruluğu azaltan özellikler elde ettik. Tıbbi veriler üzerinde çalışırken öngörmedeki en önemli unsur, özelliğin önemidir.
- Farklı triyaj sistemlerinin çocuk veri setlerinde ML modellerini kullanarak hastaneye yatış ve triyaj tahmini ile ilgili karşılaştırmalı bir çalışma yapılabilir.
- Farklı yetişkin triyaj sistemleri veri setlerinde ML modellerini kullanarak hastaneye yatış ve triyaj tahmini ile ilgili karşılaştırmalı bir çalışma yapılabilir.

KAYNAKLAR

Anonymous .2006. Web Sitesi : http://www.trauma.org/archive/history/larrey.html,
Erişim Tarihi : 01.05.2019

Anonymous.2014. Web Sitesi : http://healthindisasters.com/images/Books/Emergency-Triage--Manchester-Triage-Group-Third-Edition.pdf
Erişim Tarihi : 01.05.2019

Anonymous. 2016. Web Sitesi : http://azkurs.org/acil-triyaj-egitim-rehberi-ocak-2015-icindekiler.html
Erişim Tarihi : 01.05.2019

Anonymous . 2017a. Web Sitesi : https://www.dentalcare.com/en-us/professional-education/ce-courses/ce490/blood-pressure-categories
Erişim Tarihi : 01.05.2019

Anonymous. 2017b. Web Sitesi : https://en.wikipedia.org/wiki/TensorFlow
Erişim Tarihi : 01.05.2019

Anonymous. 2018a. Web Sitesi : http://www.buzzle.com/articles/normal-pulse- rate.html
Erişim Tarihim : 01.05.2019

Anonymous. 2018b.Web Sitesi : https://en.wikipedia.org/wiki/Vital_signs#Temperature
Erişim Tarihi : 01.05.2019

Anonymous.2018c.WebSitesi:https://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
Erişim Tarihi : 01.05.2019

Anonymous. 2018d. Web Sitesi : https://towardsdatascience.com/scale-standardize-or-normalize-with-scikit-learn-6ccc7d176a02
Erişim Tarihi : 01.05.2019

Anonymous. 2018e. Web Sitesi : https://jupyter.org/
Erişim Tarihi: 01.05.2019

Anonymous. 2018f. Web Sitesi : https://medium.com/@srnghn/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3,
Erişim Tarihi: 01.05.2019

Anonymous. 2019a. Web Sitesi : https://scikit-learn.org/stable/modules/neighbors.html
Erişim Tarihi : 01.05.2019

Anonymous. 2019. Web Sitesi: https://www.saedsayad.com/k_nearest_neighbors.htm
Erişim Tarihi : 01.05.2019

Anonymous. 2019. Web Sitesi: https://stackabuse.com/k-nearest-neighbors-algorithm-in-python-and-scikit-learn/
Erişim Tarihi : 01.05.2019

Anonymous.2019b.WebSitesi:https://www.datacamp.com/community/tutorials/xgboost-in-python
Erişim Tarihi : 01.05.2019

Anonymous. 2019c. Web Sitesi: https://www.researchgate.net/figure/a-Cumulative-Distribution-Function-of-Probit-and-Logit-Models-b-Probability-Density_fig1_265404717
Erişim Tarihi : 01.05.2019

Anonymous. 2019d. Web Sitesi : https://veribilimcisi.com/2017/07/18/lojistik-regresyon/
Erişim Tarihi : 01.05.2019

Anonymous. 2019e. Web Sitesi : https://medium.com/@ahmetkuzubasli/neden-ilk-adresiniz-xgboost-75010416a583
Erişim Tarihi : 01.05.2019

Anonymous. 2019f. Web Sitesi: https://www.investopedia.com/terms/d/deep-learning.asp
Erişim Tarihi : 01.05.2019

Anonymous. 2019g.Web Sitesi : https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html
Erişim Tarihi : 01.05.2019

Anonymous. 2019h. Web Sitesi : https://scikit-learn.org/stable/modules/neighbors.html,
Erişim Tarihi : 01.05.2019

Anonymous. 2019i. Web Sitesi : https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
Erişim Tarihi : 01.05.2019

Cremeens, M ve Khorasani, S. 2014. FMTS : A fuzzy implementation of the Manchester triage System.

Chandrasekaran, V. 2010. Measuring vital signs using smart phones. Master thesis, University of North Texas, Department of computer science, 130, North Texas.

C. X. Ling, J. Huang, and H. Zhang, "AUC: A better measure than accuracy in comparing learning algorithms," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2671, pp. 329–341, 2003.

D. A. Travers, A. E. Waller, J. M. Bowling, D. Flowers, and J. Tintinalli, "Five-level triage system more effective than three-level in tertiary emergency department," *J. Emerg. Nurs.*, vol. 28, no. 5, pp. 395–400, 2002.

Gilboy, N., Tanabe, P., Travers, D., and Rosenau, A. 2012. Emergency Severity Index(ESI), A triage tool for department care version 4. AHRQ,7, USA.

J. myoung Kwon, Y. Lee, Y. Lee, S. Lee, H. Park, and J. Park, "Validation of deep-learning-based triage and acuity score using a large national dataset," *PLoS One*, vol. 13, no. 10, pp. 1–10, 2018.

Raschka,S. 2015. Python Machine Learning. Pack Publishing Ltd, 56, UK.

S. Levin, M. Toerper, E. Hamrock, JS. Hinson, S. Barnes, H. Gardner, A. Dugas, B. Linton, T. Kirsch and G. Kelen, "Machine-learning based electronic triage more accurately differentiates patients with respects to clinical outcomes compared with the emergency severity index," Elsevier, vol. 71, no. 5, pp. 565–574, 2018.

T. Gerard and A. Nicholas, Principles of Anatomy and Physiology, Harper-Collins, New York, sixth edition, 1990.

T. Goto, C. Camargo, M. Faridi, R. Freishtat and K. Hasegawa, "Machine learning based prediction of clinical outcomes for children during emergency department triage", JAMA Network open, vol: 2(1), pp. e186937, 2019

W. S. Hong, A. D. Haimovich, and R. A. Taylor, "Predicting hospital admission at emergency department triage using machine learning," *PLoS One*, vol. 13, no. 7, pp. 1–13, 2018.

EK 1

MTS : Hemen ve çok acil python betiği

```
import xlrd
import pandas as pd
import numpy as np
import sklearn.metrics import classification_report, confusion_matrix
from sklearn.metrics import roc_curve, auc
import math
from sklearn.metrics import accuracy_score

//IMPORT THE DATASET
data = pd.read_excel('triage.xlsx')

//DEFINING A DATA FRAME df
df=data[['Age','Respiratory_rate','Heart_rate','Temperature','Oxygen_saturation','Original_MTS']]
df = df.fillna(0)
mymap = {'Standard':0, 'Urgent':0, 'Very urgent':2, 'Non urgent':0, 'Emergent':1}
df = df.applymap(lambda s: mymap.get(s) if in mymap else s)

//DEFINING PREDICTING DATA
y = df.Original_MTS
X=df[['Age','Respiratory_rate','Heart_rate','Temperature','Oxygen_saturation']]

// TRAIN_TEST_SPLIT
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)

// LOGISTICREGRESSION
from sklearn.linear_model import LogisticRegression
model = LogisticRegression ()
model.fit (X_train, y_train)
y_pred_class = model.predict (X_test)

// XGBOOST
from xgboost import XGBClassifier
model = XGBClassifier ()
model.fit.(X_train, y_train)
y_pred_class = model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred_class )
print ("Accuracy: %.2f%%" %(accuracy * 100) )

print (classification_report (y_test, y_pred_class))
 from sklearn import metrics
confusion = metrics.confusion_matrix (y_test, y_pred_class)
print (metrics.confusion_matrix(y_test, y_pred_class))
```

EK 2

Hastaneye Yatış Python Betiği

```
import xlrd
import pandas as pd
import numpy as np
import sklearn.metrics import classification_report, confusion_matrix
from sklearn.metrics import roc_curve, auc
import math
from sklearn.metrics import accuracy_score

//IMPORT THE DATASET
data = pd.read_excel('triage.xlsx')

//DEFINING A DATA FRAME df
df=data[['Age','Respiratory_rate','Heart_rate','Temperature','Oxygen_saturation','Origina
l_MTS',Hospitalisation]]
df = df.fillna(0)
mymap = {'Standard':5, 'Urgent':3, 'Very urgent':2, 'Non urgent':4, 'Emergent':1}
df = df.applymap(lambda s: mymap.get(s) if in mymap else s)

//DEFINING PREDICTING DATA
y = df.Hospitalisation
X=df[['Age','Respiratory_rate','Heart_rate','Temperature','Oxygen_saturation','Original_
MTS']]

// TRAIN_TEST_SPLIT
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)

// LOGISTICREGRESSION
from sklearn.linear_model import LogisticRegression
model = LogisticRegression ()
model.fit (X_train, y_train)
y_pred_class = model.predict (X_test)

// XGBOOST
from xgboost import XGBClassifier
model = XGBClassifier ()
model.fit.(X_train, y_train)
y_pred_class = model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred_class )
print ("Accuracy: %.2f%%" %(accuracy * 100) )

print (classification_report (y_test, y_pred_class))
 from sklearn import metrics
confusion = metrics.confusion_matrix (y_test, y_pred_class)
print (metrics.confusion_matrix(y_test, y_pred_class))
```

EK 3

ESI: DNN Betiği

```
import xlrd
import os
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
import tensorflow as tf
from sklearn import metrics
from keras.models import Sequential
from keras.layers.core import Dense, Activation
from keras.callbacks import EarlyStopping
from keras.callbacks import ModelCheckpoint
tf.logging.set_verbosity(tf.logging.ERROR)
path = "./Data/"
filename = os.path.join(path,"clean.csv")
df = pd.read_csv(filename,na_values=['NA','?'])
disposition = encode_text_index(df,'disposition')
num_classes = len(disposition)
df=data[['age','triage_vital_hr','triage_vital_sbp','triage_vital_dbp','triage_vital_rr','triage_vital_o2',
'triage_vital_o2_device',
'triage_vital_temp', 'disposition']]
mymap = {'Admit':1,'Discharge':0}
df=df.applymap(lambda s: mymap.get(s) if s in mymap else s)
df=df.fillna(df.mean())
x, y = to_xy(df,'disposition')
# Split into train/test
x_train, x_test, y_train, y_test = train_test_split( x, y, test_size=0.25, random_state=42)

# Build network
model = Sequential()
model.add(Dense(20, input_dim=x.shape[1], activation='relu'))
model.add(Dense(10, activation='relu'))
model.add(Dense(y.shape[1],activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam')
monitor = EarlyStopping(monitor='val_loss', min_delta=1e-3, patience=5, verbose=1,
mode='auto')
checkpointer = ModelCheckpoint(filepath="best_weights.hdf5", verbose=0,
save_best_only=True) # save best model
model.fit(x_train,y_train,validation_data=(x_test,y_test),callbacks=[monitor,checkpointer],verbos
e=0,epochs=1000)
model.load_weights('best_weights.hdf5') # load weights from best model

# Measure accuracy
pred = model.predict(x_test)
pred = np.argmax(pred,axis=1)
y_compare = np.argmax(y_test,axis=1)
score = metrics.accuracy_score(y_compare, pred)
print("Final accuracy: {}".format(score))
```

ÖZGEÇMİŞ

Adı Soyadı : Abdoul Aziz RAZİKOU DİLWANI

Doğum Yeri : Nijer

Doğum Tarihi : 26.07.1990

Medeni Hali : Bekar

Yabancı Dili : Fransızca, İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise : Niamey (Nijer), Bombey Lisesi (2007 - 2010)

Lisans : Nijer İslam Üniversitesi, Bilgisayar Bilimi ve Bilgi Teknolojisi bölümü (2011 - 2014)

Nijer Afrika Teknoloji Enstitüsü, ağ sistemi, Yönetim Bilişim Bolümü (2010 - 2013)

Yüksek Lisans : Ankara Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği (Haziran, 2019)

Konferans

Razikou, D.A. 2019. Emergency triage and hospitalization prediction with state of the art machine learning models. IDRC 2019, 26 -28 Haziran, Eskişehir.