

T.C.

İSTANBUL AYDIN ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ



MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE PARKİNSON HASTALIĞININ
TEŞHİS EDİLMESİ

YÜKSEK LİSANS TEZİ

Raife Göksu KÜÇÜK

Bilgisayar Mühendisliği Ana Bilim Dalı

Bilgisayar Mühendisliği Programı

Ağustos 2019

T.C.

İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE PARKİNSON HASTALIĞININ
TEŞHİS EDİLMESİ

YÜKSEK LİSANS TEZİ

Raife Göksu KÜÇÜK

Bilgisayar Mühendisliği Ana Bilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Dr. Öğr. Üyesi İlknur DÖNMEZ

Ağustos 2019



T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ

Yüksek Lisans Tez Onay Belgesi

Enstitümüz Bilgisayar Mühendisliği Ana Bilim Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı Y1613.010010 numaralı öğrencisi **Raife Göksu KÜÇÜK**'ün "**MAKİNE ÖĞRENMESİ YÖNTEMİYLE PARKINSON HASTALIĞININ TAŞIŞ EDİLMESİ**" adlı tez çalışması Enstitümüz Yönetim Kurulunun 25.07.2019 tarih ve 2019/15 sayılı kararıyla oluşturulan jüri tarafından *eg.bilgi* ile Tezli Yüksek Lisans tezi olarak *...kabul* edilmiştir.

Öğretim Üyesi Adı Soyadı

İmzası

Tez Savunma Tarihi : 07/08/2019

1)Tez Danışmanı: Dr. Öğr. Üyesi İlknur DÖNMEZ

.....
[Signature]

2) Jüri Üyesi : Dr. Öğr. Üyesi Meryem UZUN PER

.....
[Signature]

3) Jüri Üyesi : Dr. Öğr. Üyesi Ahmet GÜRHANLI

.....
[Signature]

Not: Öğrencinin Tez savunmasında **Başarılı** olması halinde bu form **imzalanacaktır**. Aksi halde geçersizdir.

YEMİN METNİ

Yüksek Lisans tezi olarak sunduğum “Makine Öğrenmesi Yöntemleri ile Parkinson Hastalığının Teşhis Edilmesi” adlı çalışmanın, tezin proje safhasından sonuçlanmasına kadarki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurulmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya ’da gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim. (14/06/2019)

Aday / İmza
Raife Göksu Küçük

ÖNSÖZ

Çalışma konusunun belirlenmesinde ve çalışmanın gerçekleştirilmesinde zaman ayırarak değerli bilgileri ile bana yol gösteren tez danışmanım Dr. Öğr. Üyesi İlknur DÖNMEZ' e ve çalışmam boyunca beni hep destekleyen aileme ve arkadaşlarıma sonsuz teşekkürlerimi sunarım.

Haziran 2019

Raife Göksu Küçük





İÇİNDEKİLER

	<u>Sayfa</u>
ÖNSÖZ.....	vii
İÇİNDEKİLER.....	ix
KISALTMALAR.....	xi
ÇİZELGE LİSTESİ.....	xiii
ŞEKİL LİSTESİ.....	xv
ÖZET.....	xvii
ABSTRACT.....	xix
1 GİRİŞ.....	1
1.1 Tezin Amacı.....	1
1.2 Literatür Araştırması.....	2
1.3 Konu ile ilgili kavramlar.....	2
1.3.1 Parkinson Hastalığı.....	2
1.3.2 Nedenleri.....	3
1.3.2.1 Çevresel Nedenler.....	3
1.3.2.2 Genetik Nedenler.....	3
1.3.3 Parkinsonizm Türleri.....	3
1.3.4 Bulgular ve Belirtiler.....	3
1.3.4.1 Motor Belirtiler.....	4
1.3.4.1.1 Konuşma Bozuklukları.....	4
1.3.4.2 Nöropsikiyatrik Belirtiler.....	5
1.3.4.3 Diğer Belirtiler.....	5
1.4 Hipotez.....	5
2 TEZDE KULLANILAN YÖNTEMLER.....	8
2.1 Makine Öğrenmesi.....	8
2.1.1 Denetimsiz Öğrenme.....	9
2.1.2 Denetimli Öğrenme.....	10
2.1.3 Yarı Denetimli Öğrenme.....	12
2.1.4 Takviyeli Öğrenme.....	12
2.1.5 Diğer Kavramlar.....	12
2.1.6 Test Aşaması.....	12
2.1.7 Aşırı Öğrenme.....	12
2.1.8 Çapraz Doğrulama.....	13
2.1.9 Performans Değerlendirme.....	13
2.1.9.1 Karışıklık Matrisi.....	13
2.1.9.2 F-Ölçümü.....	14
2.1.9.3 Öğrenme Eğrisi.....	15
2.2 Makine Öğrenmesinin Tıp Alanındaki Önemi.....	15
2.3 Kullanılan Makine Öğrenmesi Yöntemleri.....	16
2.3.1 Destek Vektör Makineleri.....	16
2.3.1.1 Doğrusal Destek Vektör Makineleri.....	16
2.3.1.2 Doğrusal Olmayan Destek Vektör Makineleri.....	17
2.3.2 K-En Yakın Komşu.....	18

2.3.3	Karar Ağacı.....	18
2.3.4	Rastgele Orman.....	19
2.3.5	Gauss Süreci.....	20
2.3.6	Logistic Regresyon.....	22
2.3.7	Naive Bayes.....	22
3	PROJE MODELİ.....	24
3.1	Veri Seti.....	24
3.2	Öznitelik Seçimi.....	26
3.3	Gauss Süreci Uygulama Adımları	28
4	DEĞERLENDİRME.....	30
5	SONUÇLAR.....	36
	KAYNAKLAR.....	38





KISALTMALAR

PH:	Parkinson Hatalığı
MDVP:	Multi Dimensional Voice Program
NHR:	Noise-to-Harmonic Ratio
PPE:	Pitch Period Entropy
TP:	True Positive
TN:	True Negative
FP:	False Positive
FN:	False Negative
DVM:	Destek Vektör Makineleri
RBF:	Radial Basis Function





ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 2.1: Karışıklık Matrisi Gösterimi.....	11
Çizelge 3.1: Öznitelik Eşleştirme ve Açıklamaları.....	22
Çizelge 3.2: Rastgele Orman Karışıklık Matrisi.....	22
Çizelge 3.3: Rastgele Orman Performans Değerlendirmesi.....	23
Çizelge 4.1: Sonuçlar.....	30





ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 1.1: Parkinson Hastalığı Duruş Bozukluğu.....	4
Şekil 2.1: Kümeleme.....	8
Şekil 2.2: Sınıflandırma.....	9
Şekil 2.3: Regresyon Grafiği.....	9
Şekil 2.4: Öğrenme Eğrisi Grafiği Örneği.....	13
Şekil 2.5: Doğrusal Ayrılabilme Durumlarında Optimum Hiper-Düzlem ve Destek Vektörleri.....	15
Şekil 2.5: Doğrusal Olmayan Destek Vektör Makineleri Doğrusal Ayırma Gösterimi.....	16
Şekil 2.6: Karar Ağaçlarından Rastgele Orman Oluşturulması.....	18
Şekil 2.7: Çan Eğrisi Grafiği.....	19
Şekil 3.1: Gini indeksi ile oluşturulan karar ağacı.....	25
Şekil 3.2: Önem Sırasına Göre Öznitelikler.....	25
Şekil 3.3: Sınıflandırıcının eklenmesi ve eğitim-test verilerinin seçilmesi.....	26
Şekil 3.4: Modelin eğitilmesi ve sonuçların alınması.....	26
Şekil 4.1: Gauss Süreci Modeli Doğruluk Oranı Grafiği.....	28
Şekil 4.2: Gauss Süreci Öğrenme Eğrisi Grafiği.....	28
Şekil 4.3: Lojistik Regresyon Öğrenme Eğrisi Grafiği.....	29
Şekil 4.4: Naive Bayes Öğrenme Eğrisi Grafiği.....	30
Şekil 4.5: Destek Vektör Makinesi Öğrenme Eğrisi Grafiği.....	30
Şekil 4.6: Rastgele Orman Öğrenme Eğrisi Grafiği.....	31
Şekil 4.7: Karar Ağacı Öğrenme Eğrisi Grafiği.....	31
Şekil 4.8: K En Yakın Komşu Öğrenme Eğrisi Grafiği.....	28



MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE PARKİNSON HASTALIĞININ TEŞHİS EDİLMESİ

ÖZET

Parkinson Hastalığı beyinde dopamin üreten hücrelerin kaybından kaynaklanan yavaş ilerleyen nörolojik bir hastalıktır. İlk kez 1817 yılında James Parkinson isimli bir İngiliz doktor tarafından tanımlanmıştır. Sinir hücre kaybı sonucu gelişen hastalıklar arasında Alzheimer'dan sonra en sık görülen hastalık Parkinson'dur ve bu hastalık ortalama 60 yaşlarında ortaya çıkmaktadır. Dünyada ortalama 5 milyon, Türkiye'de ise 150 bin Parkinson Hastası olduğu tahmin edilmektedir. Parkinson Hastalığı Titreme, hareketlerde yavaşlama, denge bozuklukları, konuşma, yazma ve yürüme bozuklukları gibi hayat kalitesini önemli ölçüde olumsuz etkileyen bir hastalıktır. Hastalığın erken evrelerinde teşhis edilebilmesi bu semptomların kontrol altına alınabilmesi için çok önemlidir. Bu çalışmada 23'ü Parkinson hastası olan 31 denekten alınan biyomedikal ses ölçümlerinden oluşan bir veri seti üzerinde çeşitli makine öğrenmesi yöntemleri uygulanmıştır. Buna bağlı sonuçlar ortaya koyulmuş ve tartışılmıştır.

Anahtar Kelimeler: *Makine Öğrenmesi, Parkinson, Hastalık Teşhisi*



DIAGNOSIS OF PARKINSON'S DISEASE WITH MACHINE LEARNING METHODS

ABSTRACT

Parkinson's disease is a slowly progressive neurological disorder that is associated with the loss of dopamine-producing cells in the brain. It was first described in 1817 by a British doctor named James Parkinson. Among the diseases that develop as a result of nerve cell loss, the most common disease after Alzheimer's disease is Parkinson's, and this disease occurs at an average age of 60 years. Approximately 5 million in the world, in Turkey is estimated to be 150 thousand Parkinson's disease. Parkinson is a disease that significantly affects quality of life such as tremors, deceleration in movements, balance disorders, speech, writing and gait disturbances. Diagnosis at the early stages of the disease is very important to control these symptoms. In this study, various machine learning methods were applied on a set of biomedical sound measurements taken from 31 subjects ,23 with Parkinson's disease. The related results were presented and discussed.

Keywords: *Machine Learning, Parkinson, Disease Diagnosis*



1. GİRİŞ

Dopamin, substansiya nigra ile vücut hareketlerini kontrol eden diğer beyin bölgeleri arasında mesajlar ileten bir kimyasaldır[1]. Dopamin üretimi yapan hücrelerin %60 ila 80'i kaybedildiğinde Parkinson hastalığının titreme, hareketlerde yavaşlama, denge bozuklukları, konuşma, yazma, yürüme bozuklukları, psikolojik bozukluklar ve uyku bozuklukları gibi motor ve motor-dışı semptomları ortaya çıkar. Parkinson hastalığı olan birçok kişide çeşitli konuşma bozuklukları olduğu görülmektedir. En sık bildirilen konuşma bozuklukları ise sesin zayıf, boğuk, nazone veya monoton çıkması, bozulmuş artikülasyon, yavaş veya hızlı konuşma, konuşmayı başlatmada zorlanma, seslerin vurgu ve ritminin bozulması, kekemelik ve seste titreme olarak sayılabilir [2]. Bu parkinsonizm belirtileri farklı nedenlere bağlı olarak da ortaya çıkabilir ve bazen hastalara yanlışlıkla Parkinson teşhisi konulabilmektedir. Bu nedenle farklı sebeplere bağlı parkinsonizm belirtilerinin tedavilerinde hasta tedaviye yanıt vermeyebilir. Ayrıca parkinson hastalığı için uygulanan semptomik tedaviler genellikle erken dönemlerde etkili olabilmektedir. Dolayısıyla Parkinson'da erken evrelerde kesin tanı koyulması ve doğru tedavinin belirlenmesi hastanın yaşam kalitesini mümkün olan en iyi seviyede tutmak için çok önemlidir.

Hastalardan alınan medikal ses ölçüm verileri üzerinde uygulanan öğrenme teknikleri ile erken tanı mümkün olmaktadır.

Tezin birinci bölümde Parkinson Hastalığı, nedenleri çeşitli belirtileri ve insan hayatı üzerindeki etkilerinden bahsedilmiştir. Tezin ikinci bölümünde makine öğrenmesinin tarihi genel uygulanması ve tıp alanında makine öğrenmesi üzerinde durulmuştur. Tezin üçüncü bölümünde modelin detayları anlatılmaktadır. Dördüncü ve beşinci bölümler değerlendirme ve sonuç bölümleridir.

1.1 Tezin Amacı

Bu çalışmada Parkinson Hastalığı olan kişilerin ses kayıtlarından faydalanılmıştır. Oluşturulan makine öğrenmesi yöntemleri ile herhangi bir bireyin Parkinson hastası olup olmadığı tespit edilmeye çalışılmıştır.

Yapılan uygulamalarda Max A. Little tarafından hazırlanan veri seti kullanılmıştır. Veri seti 23'ü Parkinson hastası olan 31 denekten alınan biyomedikal ses ölçümlerinden oluşmaktadır. Çokboyutlu ses analiz programı MDVP ile yapılan ölçümler temel frekans (Fo), temel frekanstaki düzensizlik (Jitter), ses sinyallerindeki şiddet değişimleri (Shimmer), doğrusal olmaya temel ölçüm sıklığı (PPE), gürültü harmonik oranı (NHR) gibi birçok farklı değer içermektedir. Veri setinde Parkinson hastalığı bulunan kişiler 1 sağlıklı kişiler ise 0 olarak gösterilmiştir. Parkinson hastalığının teşhis edilebilmesi için veri seti içerisindeki 22 öznitelik kullanılarak makine öğrenmesi algoritmaları uygulanmış ve sonuçlar karşılaştırılmıştır.

1.2 Literatür Araştırması

Max A. Little tarafından 2009 da yapılan çalışmada destek vektör makinesi yöntemi ile 91.4% başarı oranı elde edilmiştir. Marius Ene ise 2008 yılında olasılıksal sinir ağları yöntemleri ile %81 oranında başarı elde etmiştir. Mehmet Fatih Çağlar, Burcu Toprak 2010 senesinde ÇokKatmanlı Algılayıcı (ÇKA) ve Radyal Tabanlı Fonksiyon Ağları (RTFA). Diğer metot olarak ise Dilsel Kuvvetli Adaptif Sinir-Bulanık Sınıflayıcı (DKASBS) kullanarak %95,38 başarı oranını yakalamışlardır. M. Serdar Başçıl ve arkadaşları 2012 yılında yaptıkları çalışmada olasılıksal sinir ağları yöntemleri ile %95 başarıya ulaşmışlardır. S. BOUCHIKH ve arkadaşları Relief öznitelik seçim algoritması ve destek vektör makinesi yöntemlerini kullanarak %96,88 doğruluk oranına ulaşmıştır. Özalbizal 2014 yılında birkaç farklı makine öğrenmesi yöntemi ile çalışmış KNN yöntemi ile %96,97 başarı oranı elde etmiştir.

1.3 Konu ile ilgili Kavramlar

Makine öğrenmesinin uygulandığı Parkinson hastalığı ile ilgili kavramlar anlatılmaktadır.

1.3.1 Parkinson Hastalığı

Parkinson Hastalığı motor sistemi büyük ölçüde etkileyen uzun süreli dejeneratif bir hastalıktır. Bu hastalıkta en çok öne çıkan hareket bozuklukları bradikinezi olarak tanımlanan istemli hareketlerin ileri derecede yavaşlaması buna bağlı olarak kaslarda zayıflama, istirahat halindeyken titreme ve postural dengesizliktir. Birçok farklı hastalıkta bu tip parkinsonizm türünde hareket bozuklukları görülebilir.

1.3.2 Nedenleri

Parkinson hastalığının nedenleri iki grupta toplanabilir.

1.3.2.1 Çevresel nedenler

Pestisit maruziyeti ve kafa travması öyküleri Parkinson Hastalığı ile ilişkilendirilmektedir. Ayrıca hiç sigara içmemek ve hiç kafein tüketmemek Parkinson Hastalığının gelişme riskindeki küçük artışlarla ilişkilendirilir[3].

1.3.2.2 Genetik nedenler

Araştırmalara göre Parkinson Hastalığının çevresel ve genetik faktörlerin etkileşiminin bir sonucu olduğunu göstermektedir. PH'li bireylerin yaklaşık% 15'inin hastalığı olan birinci derece akrabaları vardır, [4] ve PH'li kişilerin% 5 - 10'unun belirli genlerden birindeki mutasyon nedeniyle oluşan hastalık formlarına sahip oldukları bilinmektedir.

1.3.3 Parkinsonizm türleri

Parkinson Hastalığı, en sık görülen parkinsonizm şeklidir ve bazen “idiyopatik parkinsonizm” olarak adlandırılır, bu da tanımlanabilen bir nedeni olmayan parkinsonizm anlamına gelir [5]. Tanımlanabilir parkinsonizm nedenleri ise toksinler, metabolizmadaki bozulma, ilaçlardan kaynaklı yan etkiler, enfeksiyonlar ve felçtir. Bazı nörodejeneratif rahatsızlıklar da parkinsonizm belirtilerine neden olabilir bunlar Parkinson art sendromları olarak adlandırılır.

En sık görülen nörodejeneratif hastalıklardan biri de Alzheimer'dır. Fakat Alzheimer çoğunlukla hafıza kaybı ile ortaya çıkar ve Parkinson Hastalığı bulguları Alzheimer'da normal değildir. Bu hastalıkların erken evrelerinde Parkinson hareketleri Alzheimer ise zihinsel işleyişi etkiler.

Demans hastalığının bir alt tipi olan Lewy cisimcikli demans Parkinson Hastalığı demansı olarak bilinir ve Parkinson Hastalığı ile çok yakın patolojik benzerliklere sahiptir[6].

1.3.4 Bulgular ve belirtiler

Parkinson Hastalığının en belirgin bulguları motor fonksiyonlardır. Bunun yanında otonomik yetmezlik veya ruh hali, davranış, düşünce değişiklikleri gibi psikiyatrik sorunlar, koku alma duyusunda farklılaşma, uyku problemleri gibi motor olmayan semptomlara da neden olur. Bu motor dışı semptomlar da tanı konulmasına yardımcı olur.

1.3.4.1 Motor belirtiler

Sıklıkla karşılaşılan motor belirti ellerin istirahat halinde iken titremesi ve bu durumun derin uyku ve istemli hareketler esnasında ortadan kaybolmasıdır. Bu titreme tipik olarak işaret parmağın ve başparmağın dairesel olarak birbirine dokunması şeklindedir.

Diğer motor belirtiler ise ayakları sürüyerek, küçük adımlarla ve kollar sabit olarak yürüme öne eğik bir duruş, monoton ve alçak sesle konuşma, donuk yüz ifadesi ve el yazısında küçülme olarak sıralanabilir. 1886 yılında nörolog araştırmacı ve sanatçı Sir William Richard Gowers, Parkinson Hastalığı duruş bozukluğunu resmetmiştir. Bu eser Şekil 1.1’de verilmiştir. [7]



Şekil 1.1: Parkinson Hastalığı Duruş Bozukluğu

1.3.4.1.1 Konuşma bozuklukları

Parkinson Hastalığında görülen motor belirtilerden biri ise dizartri olarak tanımlanan konuşma bozukluklarıdır. Dizartri konuşurken kullanılan kasların zayıf olması sebebi ile bu kasların koordine ve kontrolünde güçlük yaşanmasına bağlı olarak ortaya çıkar.

Parkinson dizartrisinde dikkati çeken klinik özellikler; ses perdesinin monotonluğu, vurgulamalarda azalma, sesin iniş çıkışında azalma, sözcüklerin ağızda yuvarlanması, konuşurken uygun olmayan duraklamalar, kullanılan cümlelerin kısa tutulması, sesin tırmalayıcı, pürüzlü ve soluklu nitelikte olması, perde düzeyinde genellikle tizliğe doğru değişme ve konuşmada hızlanma-yavaşlama değişiklikleri şeklinde sıralanmaktadır[8].

1.3.4.2 Nöropsikiyatrik belirtiler

Parkinson Hastalığı hastalar üzerinde en az motor semptomlar kadar olumsuz etkilere sahip olan nöropsikiyatrik problemlere sebep olabilmektedir. En sık karşılaşılan nöropsikiyatrik belirtiler depresyon, ilgisizlik, yorgunluk, kaygı bozuklukları, düşünce bozuklukları ve davranış bozukluklarıdır. Bunlara ek olarak hafıza bozuklukları, düşünmede zorluklar, algı ve zaman tahmininde bozulma gibi durumlar meydana gelebilir. Bu gibi psikiyatrik bozukluklar motor bozuklukların şiddetinin artmasına sebep olmaktadır[9].

Bunlara ek olarak dopamin seviyesini arttıran bazı ilaçların yan etkileri olarak cinsel davranış bozuklukları, kumar alışkanlığı, aşırı yemek yeme, aşırı alışveriş yapmak söylenebilir[10]. Ek olarak aşırı para harcamaya eğilimi gibi dürtü kontrol bozuklukları ortaya çıkabilmektedir.

1.3.4.3 Diğer belirtiler

Parkinson Hastalığının motor ve nöropsikiyatrik belirtilere ek olarak bazı farklı olumsuz etkileri de görülmektedir.

Uyku bozukluğu bu problemlerin başında gelir ve ilaçların etkisi ile bu sorun artabilir. Gündüzleri ani uyku hali, uykusuzluk veya uyku sırasında kendine veya çevreye zarar verme gibi sorunlar yaşanabilir. Bu belirtiler motor veya motor dışı belirtilerin başlamasından yıllar önce ortaya çıkabilir. Uyku bozuklukları dışında yağlı cilt, aşırı terleme, idrar kaçırma, mide bağırsak problemleri ve görme, koku alma duyularında rahatsızlıklar ortaya çıkabilmektedir[11].

1.4 Hipotez

Çalışmamızda kullandığımız Parkinson veri seti için yapılmış çalışmalar arasında “Gaussian Process” yöntemi ile yapılmış herhangi bir çalışma bulunmadığı görülmüştür. Ayrıca Gauss Süreci modelinin temeli normal dağılıma dayanmaktadır ve doğrusal olmayan regression problemlerinde başarı göstermektedir. Ayrıca veri seti yapısı itibariyle çok büyük bir hacme sahip olmadığı için de Gauss Süreci yöntemi ile daha iyi sonuçlara ulaşılabileceğini öngörüyoruz.



2. TEZDE KULLANILAN YÖNTEMLER

2.1 Makine Öğrenmesi

Makine Öğrenmesi bilgisayar sistemlerinin bir görevi gerçekleştirmek üzere öğrenmeye ve çıkarıma dayalı algoritmaları ve istatistiksel modelleri kullanması ve buna bağlı bir sonuç üretmesi işlemidir. Yapay zekânın alt dallarından biridir. Makine öğrenmesinde makine öngöründe bulunmak veya karar vermek için açıkça programlanmadan, eğitim verisi denilen örnek verileri kullanarak matematiksel bir model oluşturulur[12]. Bu model ise geleceğe yönelik tahminde bulunmak için bir tahmin edici görevi görür.

Makine Öğrenmesi kavramının tarihi Alan Turing'in 1950 yılında yazdığı, bugün Turing testi olarak bilinen, makinelerin düşünüp düşünemediğine dair bir deneyin yer aldığı Bilgisayar ve Zekâ adlı makalesine kadar uzanır. Makine öğrenmesi ismi 1959'da Arthur Samuel tarafından bulunmuştur [13]. Samuel'in dama oyununu kullanan çalışması Makine öğrenmesi alanında yapılan önemli dönüm noktalarından biridir[14]. Bu çalışmada ezberci öğrenme ve genelleme yolu ile öğrenme gibi metotlar tasarlanmış ve bilgisayarın başlangıç statüsünden turnuva oyuncusu statüsüne kadar kendini geliştirmesi sağlanmıştır. Önemli dönüm noktalarından biri de 1957 yılında Frank Rosenblatt tarafından yapay sinir ağı 'Perceptron' tanımının yapılmasıdır. İkinci kırılma noktası ise 1986 yılında Geoffrey Hinton, David Rumelhart ve Ronald Williams tarafından ileri sürülen çok katmanlı ağ yapısı ve 'Geriye Yayılım' (Backpropagation) teorisidir. 1995 yılında ise Vladimir Vapnik ve Corinna Cortes tarafından Destek Vektör Makineleri yöntemi ortaya atılmıştır. Bu çalışma ile Makine Öğrenmesi konusunda yapılacak yeni çalışmalara ışık tutmuşlardır.

Günümüzde ise teknolojinin geldiği noktada hemen hemen her alanda büyük veri karşımıza çıkmaktadır. E-ticaret uygulamalarından birçok veri, tıp alanında kullanılan cihazlardan, sosyal medyadan, borsa işlemlerinden alınan veriler, uydu verileri ve daha birçok farklı alandan elde edilebilecek görüntü, video, ses, yazı, sayısal, grafiksel içerikli veriler analiz edilebilmektedir. Tüm bu veriler geçmişteki veriler kullanılarak o anki verilerin değerlendirilmesinde ve yorumlanmasında kullanılır. Bu verilerin elle

analiz edilmesi mümkün olmadığı için makine öğrenmesi metotları geliştirilmeye ve kullanılmaya başlanmıştır.

Makine öğrenmesi ile geçmişteki veriler istatistiksel yöntemler kullanılarak analiz edilir verinin modeli çıkartılır ve yeni gelen veriler de bu modele göre analiz edilir ve sonuca ulaşılır. Büyük miktarda karmaşık verinin içerisinde anlamlı verinin çıkartılması işlemine veri madenciliği de denir.

Makine Öğrenmesi uygulamalarının gelişimi ile çok farklı öğrenme yöntemleri ortaya çıkmıştır. Bu yöntemler kullanılan algoritmalara göre aşağıdaki şekilde sınıflandırılabilir.

- Denetimsiz Öğrenme
- Denetimli Öğrenme
- Yarı-Denetimli Öğrenme
- Takviyeli Öğrenme

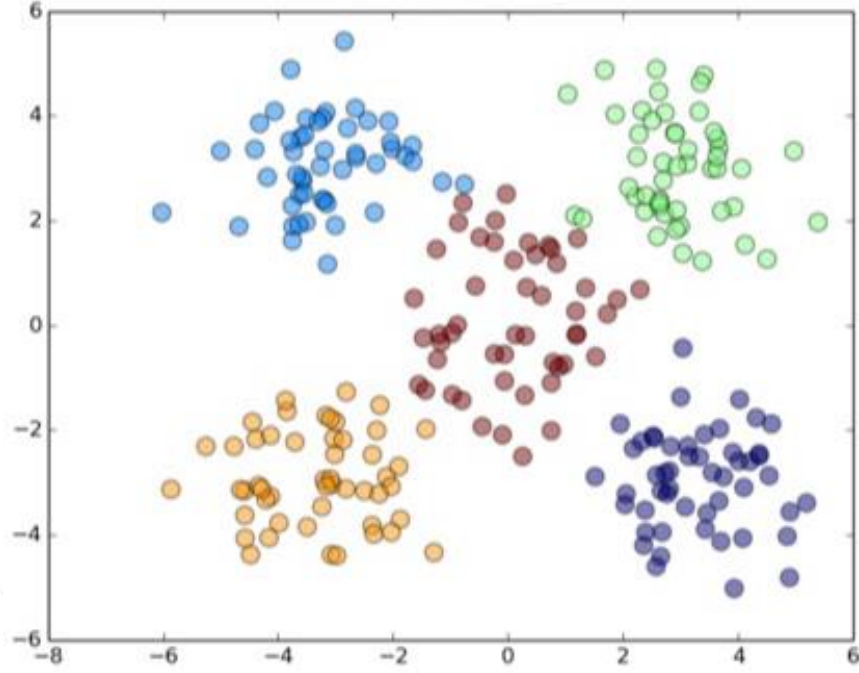
Ayrıca Makine Öğrenmesinde farklı uygulama alanlarına ve veri setlerine göre farklı sonuçlar beklenmektedir. Bu nedenle ulaşılmak istenen bu farklı sonuçlara göre farklı yöntemler ortaya çıkmıştır ve Makine öğrenmesi bu yöntemlere göre de sınıflandırılabilir.

- Sınıflandırma
- Kümeleme
- Regresyon
- Öznitelik Seçimi/Çıkarımı
- İlişki Belirleme

2.1.1 Denetimsiz öğrenme

Denetimsiz öğrenmede makine çıktı verileri olmadan yalnızca girdi verilerle uzaklık, benzerlik gibi ölçütlerle sonuçlar çıkarmaya çalışır. Denetimsiz öğrenme çoğunlukla olasılık teorisine dayanan kümeleme işlemlerinde kullanılır.

Kümeleme farklı kategorilere ait benzer verileri benzer olmayan kümelere dâhil ederek birbirinden ayırma işlemidir. Tahmin edici modellerde düzenli veri grupları oluşturulması için kümeleme işlemi ön veri işleme metodu olarak da kullanılabilir. Şekil 2.1'de makine öğrenmesi yöntemi ile elde edilmiş bir kümeleme örneği verilmiştir.

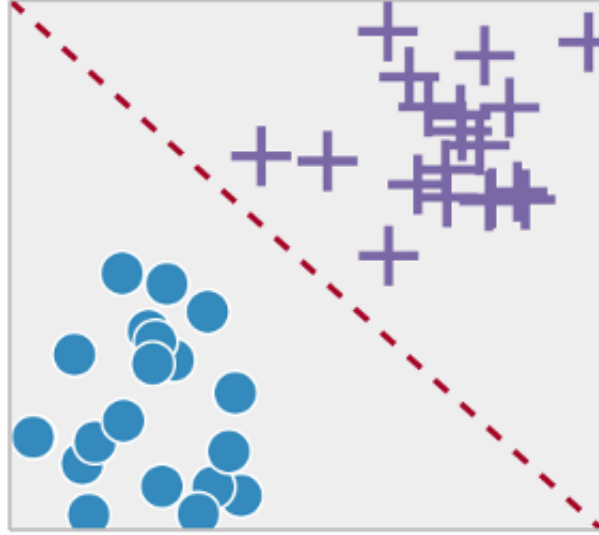


Şekil 2.1: Kümeleme

2.1.2 Denetimli öğrenme

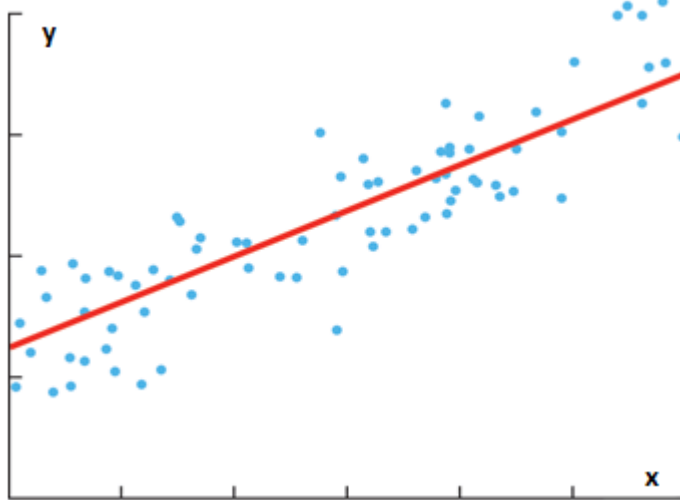
Denetimli öğrenmede veri setinin girdileri ve çıktılarının ne olması gerektiği bilinir. Denetimli öğrenme verileri ve o verilerden çıkartılmış sonuçları eğitim verisi olarak makineye vererek, girdi ve çıktı veriler arasındaki bağlantıyı çözümlemesini ve bir fonksiyon üretmesini sağlamaktır. Denetimli öğrenme çıktıları regresyon veya sınıflandırma olabilir.

Sınıflandırma verilerin hangi sınıfa dâhil olduğunun bulunması işlemidir. Sınıflandırma veriyi tanımlı olan sınıflar arasında dağıtmaktır. Sınıflandırma algoritmaları eğitim verilerini öğrenir bu dağılıma göre test verilerini doğru bir şekilde sınıflandırmaya çalışırlar. Örneğin tahlil sonuçlarına göre bir insanın hasta veya sağlıklı olup olmadığını tespit etmek için daha önceki hasta ve sağlıklı insanların tahlil sonuçları kullanılarak kişinin hangi sınıfa dâhil olduğu belirlenebilir. Şekil 2.2'de sınıflandırma işlemine bir örnek verilmiştir.



Şekil 2.2: Sınıflandırma

Regresyon iki veya daha fazla değişken arasındaki ilişkiyi ölçmek için kullanılan analiz yöntemidir. Geçmiş verilerin sınıflarının sürekli sayılardan oluştuğu durumlarda kullanılmaktadır. Eğer tek değişken kullanılıyorsa buna tek değişkenli regresyon analizi çok değişken var ise çok değişkenli regresyon analizi denilmektedir. Regresyonda değişkenlerden biri bağımlı biri ise bağımsız değişken olarak belirlenmelidir. Bağımlı değişken diğer değişkenlerden etkilenmektedir. Bağımsız değişkenler ise diğer değişkenlerden etkilenmemektedir. Şekil 2.3’de bir regresyon örneği verilmiştir.



Şekil 2.3: Regresyon Grafiği

2.1.3 Yarı denetimli öğrenme

Eğitim verilerinin büyük çoğunluğu etiketlenmemiş verilerden ve bir kısmı da etiketlenmiş verilerden oluşuyorsa bu tip verileri kullanan öğrenme metotlarına yarı denetimi öğrenme denir. Bu tip veriler için pratik bir yöntem olabilmektedir.

2.1.4 Takviyeli öğrenme

Takviyeli öğrenmede çevreden alınan geri beslemeler algoritmaya yol gösterir. Bu yöntemde sistem bulunduğu ortamdan veriler alarak öğrenir ve karar verir. Genellikle robotik, hastalık teşhisi, oyun programlama gibi alanlarda kullanıma uygundur.

2.1.5 Diğer kavramlar

1. Öznitelik Seçimi / Çıkarımı: Büyük veri üzerinde analiz yapılabilmesi için gerekli kaynak miktarını daha verimli hale getirir şekilde azaltmayı amaçlamaktadır. Tüm veri içerisinde sınıflandırmada belirleyici olan özellikler altkütmesi seçilir veya özelliklerin birleşiminden yeni özellikler çıkartılabilir.
2. İlişki Belirleme: Büyük veri yığınları arasındaki ilişkileri belirleme işlemidir. Örneğin hangi ürünlerin hangi ürünler ile birlikte satılabileceğini veya A ürünün satın alan bir müşterinin B ürününü de satın alıp almayacağını belirleme işlemidir.

2.1.6 Test aşaması

Makine Öğrenmesinde bir modelin öğrenme aşaması tamamlandığında öğrenilen model test edilir. Test aşamasında öğrenilen modelin kullanılan veri üzerindeki başarı oranı ölçülür. Test aşamasında veri setinin eğitim aşamasında kullanılmayan %70'lik bölümü kullanılır. Bu aşamada modelin hiç görmediği veriler üzerindeki doğruluk oranı analiz edilir.

2.1.7 Aşırı öğrenme

Bazı durumlarda modelin eğitim verisi üzerindeki doğruluk oranı çok yüksek iken test verisi üzerindeki doğruluk oranı çok düşük olabilir. Bu durum modelin eğitim verisi üzerindeki her durumu ezberlemesi ve test aşamasında bu durumların kopyasını aramasından kaynaklanmaktadır. Buna aşırı öğrenme denir. Kullanılan veri setinin tek

düze olması aşırı öğrenme ihtimalini arttırmaktadır. Bu nedenle veri seti çeşitlendirilerek aşırı öğrenmenin önüne geçilebilir.

2.1.8 Çapraz doğrulama

Bu yöntemde veri seti parçalara bölünür farklı farklı eğitim-test veri setleri oluşturularak modelin ayrı ayrı bu veri setleri üzerinden doğrulukları hesaplanır. Bu sonuçların aritmetik ortalaması alınarak genel doğruluk oranı elde edilir. Bu yöntem ile tüm veri seti üzerinde hem eğitim hem de test işlemi yapıldığı için başarı oranı üzerinde etkili bir yöntemdir. Veri setinin çeşitlenmesini sağladığı için aşırı öğrenme durumunun önüne geçilmiş olur.

2.1.9 Performans değerlendirme

Makine öğrenmesi çalışmalarında başarı değerlendirmesi olarak modelin kaç tane durumu doğru tahmin ettiğinin bilinmesi yeterli değildir. Modelin performansının doğru bir şekilde değerlendirilmesi için sıklıkla karışıklık matrisi (confusion matrix) kullanılmaktadır.

2.1.9.1 Karışıklık matrisi

Karışıklık matrisi tahminlerin doğruluğunu belirlemek için kullanılan bir ölçüm aracıdır. Veri setinde var olan durumlar ile modelin doğru ve yanlış tahminlerinin sayısını gösterir. Tahmindeki hedef sayısına göre NxN boyutunda bir matris olabilir. Modelin pozitif ve negatif olarak tahminde bulunacağı bir veri seti için karışıklık matrisi aşağıdaki gibi olmalıdır. Çizelge 2.1’de bir karışıklık matrisine ait satır ve sütunlar gösterilmiştir. Bu gösterimlerin anlamları aşağıda ayrıntılı olarak açıklanmıştır.

Çizelge 2.1: Karışıklık Matrisi Gösterimi

		Var olan Durum	
		Pozitif Durumlar	Negatif Durumlar
Tahmin	Pozitif	TP	FP
	Negatif	FN	TN

TP: Modelin doğru tahmin ettiği pozitif değer sayısını gösterir.

TN: Modelin doğru tahmin ettiği negatif değer sayısını gösterir.

FP: Modelin yanlış tahmin ettiği pozitif değer sayısını gösterir.

FN: Modelin yanlış tahmin ettiği negatif değer sayısını gösterir.

Bu değerlere göre karışıklık matrisinden bazı oranlar hesaplanabilir.

- Doğruluk oranı: Modelin hangi sıklıkla doğru tahmin yaptığının ölçümüdür. Denklem 2.1 de verilmiştir.

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

- Yanlış sınıflandırma oranı: Modelin ne sıklıkla yanlış tahminde bulunduğu ölçümüdür. Denklem 2.2’de verilmiştir.

$$\frac{FP + FN}{TP + TN + FP + FN} \quad (2.2)$$

- İsbet oranı (recall): Modelin ne kadar pozitif değeri doğru tahmin ettiğinin ölçümüdür. Bu değer mümkün olduğu kadar yüksek çıkması istenir. Denklem 2.3’de verilmiştir.

$$\frac{TP}{TP + FN} \quad (2.3)$$

- Hassasiyet (precision): Tüm sınıflardan ne kadar doğru tahmin yapıldığının ölçümüdür. Pozitif tahmin edici değer olarak da adlandırılır. Denklem 2.4’de verilmiştir.

$$\frac{TP}{TP + FP} \quad (2.4)$$

2.1.9.2 F-ölçümü

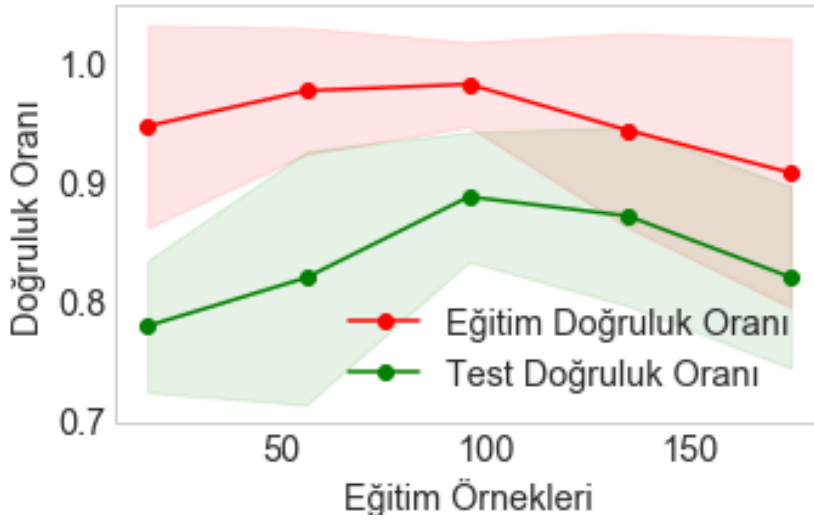
F-ölçümü istatistik biliminde bir skorlama kavramıdır. Bilgisayar bilimlerinde ise test edilen verinin doğruluğunu ölçmek için kullanılır. Bilgi çıkarımı için isbet oranı (recall) ve hassasiyet oranının (precision) harmonik ortalaması kullanılır. Denklem 2.5’de gösterilmiştir.

$$\frac{2 \cdot \text{Hassasiyet} \cdot \text{İsabet Oranı}}{\text{Hassasiyet} + \text{İsabet Oranı}}$$

(2.5)

2.1.9.3 Öğrenme eğrisi

Öğrenme eğrisi makine öğrenmesi ile büyük veriler üzerinden hastalıkların teşhis edilmesinde çok sık kullanılan bir araçtır [15]. Eğitim süresince belirli aralıklarla ölçüm yapılarak modelin eğitim ve test süreçleri grafiğe aktarılır. Şekil 2.4'te örnek bir modelin eğitim ve test sürelerinde ne şekilde davrandığı gözlemlenebilir. Elde edilen grafiğe bakılarak model ile ilgili daha doğru bilgilere ulaşılabilir.



Şekil 2.4: Öğrenme Eğrisi Grafiği Örneği

Şekilde örnek bir öğrenme eğrisi grafiği gösterilmiştir. Bu grafiğe göre eğitim aşamasında model yüksek performans göstermiştir ve test aşamasında ise doğruluk oranında veri sayısı arttıkça yükselme gözlemlenmektedir.

2.2 Makine Öğrenmesinin Tıp Alanındaki Önemi

Makine Öğrenmesi büyük verinin karşımıza çıktığı birçok alanda verilerin analiz edilmesi ve anlamlandırılması adına büyük fayda sağlamaktadır. Diğer tüm alanlardaki uygulamaların yanı sıra insan sağlığı hayati önem taşımaktadır. Bununla birlikte günümüzde neredeyse tüm tıbbi cihazların dijital hale gelmesi ile uygun veri kolaylıkla elde edilmekte, tıp alanında yapılan makine öğrenmesi çalışmaları giderek artmaktadır. Özellikle kanser veya kalp hastalıkları gibi ölümcül hastalıkların ve insan

hayatını olumsuz yönde etkileyen ağır semptomlara sahip hastalıkların erken teşhisinde tahmin gücü yüksek algoritmaların kullanımı önem kazanmaktadır. Tıp alanında yapılmış pek çok farklı öğrenmeye dayalı çalışma ile hastalıklara teşhis koyulabilmekte, hayati risk seviyeleri belirlenebilmektedir.

2.3 Kullanılan Makine Öğrenmesi Yöntemleri

2.3.1 Destek vektör makineleri

Destek vektör makineleri iki sınıfa ait verilerin en uygun biçimde birbirlerinden ayrılması için geliştirilmiş makine öğrenmesi algoritmalarından biridir. Sınıflandırma bir düzlem üzerinde bulunan örnekler arasına bir sınır çizilerek yapılır. Sınır her iki gurubun üyelerine en uzak mesafede olmalıdır. Diğer makine öğrenmesi yöntemlerine göre en önemli avantajı sınıflandırma problemini kareli optimizasyon problemine dönüştürüp çözmesidir. Bu şekilde öğrenme sırasında işlem sayısı azaltılmış olur ve diğer tekniklere göre daha hızlı sonuç verir. Bu nedenle büyük hacimli verilerde kullanımı uygundur. Destek vektör makineleri eğitim esnasında karşılaşılmamış değerleri de sınıflandırmada başarılıdır. Destek vektör makineleri verinin doğrusal olarak ayrılabilme ve ayrılama durumuna göre iki şekilde uygulanır.

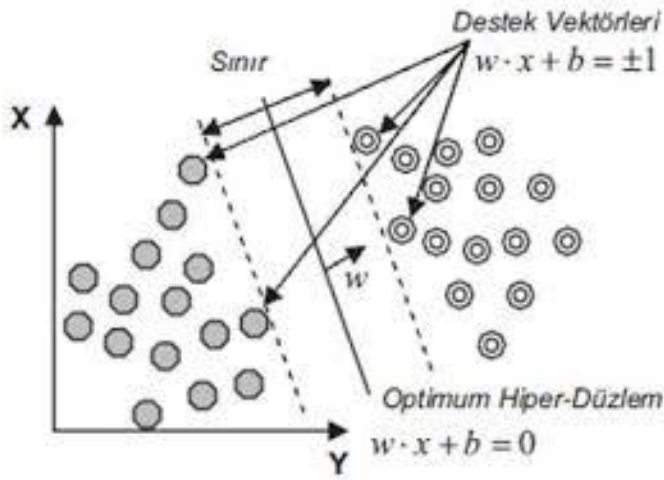
2.3.1.1 Doğrusal destek vektör makineleri

Doğrusal olarak ayrılabilen veriler için iki sınıfı birbirinden ayıran hiper-düzlem Denklem 2.6 da verilmiştir. Bu denklemde n elemandan oluşan veri kümesinin $X = \{x_i, y_i\}, i = 1, 2, \dots, n$ olduğu var sayılsın. $y_i \in \{-1, 1\}$ Etiket değerleri, $x_i \in \mathcal{R}^d$ özellikler vektörüdür, w ağırlığı, x verileri, b ise eğitim terimini ifade eder. Bu parametrelerin alacağı değerler hiper-düzlemin pozisyonunu belirler.

$$F(x) = w^T \cdot x + b = \sum_{i=1}^n w_i \cdot x_i + b$$

(2.6)

Doğrusal ayrılabilme durumlarında optimum hiper-düzlem ve destek vektörleri şekil 2.4'te verilmiştir.



Şekil 2.5: Doğrusal Ayrılabilme Durumlarında Optimum Hiper-Düzlem ve Destek Vektörleri

2.3.1.2 Doğrusal olmayan destek vektör makineleri

Birçok veri doğrusal olarak ayrılabilmeye müsait değildir. Bu durumda doğrusal olmayan sınıflandırma kullanılmaktadır. Veriler çekirdek fonksiyonu adı verilen fonksiyonlardan geçirilerek özellik uzayına taşınırlar ve burada sınıflandırılırlar. En sık kullanılan çekirdek fonksiyonları denklemi 2.7 de verilmiştir.

Radyal tabanlı çekirdek fonksiyonu (RBF)

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2.7)$$

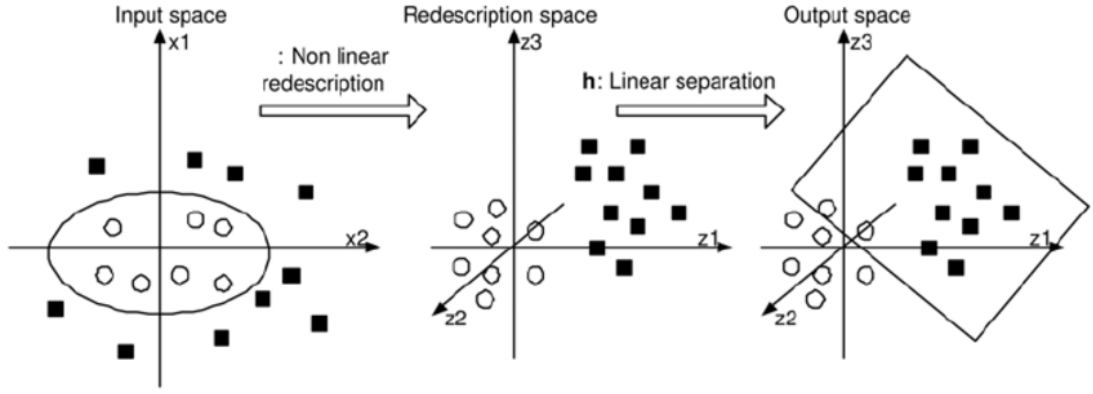
Polinom çekirdek fonksiyonu denklemi 2.8’de gösterilmiştir.

$$K(x_i, x_j) = (x_i^T \cdot x_j)^d \quad (2.8)$$

Doğrusal çekirdek fonksiyonu denklem 2.9’da verilmiştir.

$$K(x_i, x_j) = (x_i^T \cdot x_j) \quad (2.9)$$

Şekil 2.5’te doğrusal olmayan sınıflandırma grafiği ve bu grafiğin üç boyutlu uzayda doğrusal düzlem ile ayrılması gösterilmiştir.



Şekil 2.5: Doğrusal Olmayan Destek Vektör Makineleri Doğrusal Ayırma Gösterimi

2.3.2 K-en yakın komşu

K en yakın komşu sınıflandırma ve regresyon problemlerinde kullanılan en temel makine öğrenmesi metodudur. Bu yöntemde eğitim aşaması yer almaz, tembel bir öğrenme yöntemidir. Bu nedenle büyük veri setlerinde kullanıma uygun değildir. Sınıflandırılacak olan her bir yeni değer için veri setindeki en yakın mesafedeki k adet örneğe bakılır. Bu k tane örnek hangi sınıfta daha fazla yer alıyorsa yeni değer de o sınıfa dahil olur. Bu çalışmada k örnek sayısı 3 olarak alınmıştır. Mesafe ölçümü için aşağıda verilen Öklid uzaklık hesaplaması kullanılmaktadır denklem 2.10'da verilmiştir.

$$d(i, j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2} \quad (2.10)$$

2.3.3 Karar ağacı

Karar ağacı Veri Madenciliği alanında en sık kullanılan sınıflandırma yöntemlerinden biridir. Karar Ağacında amaç veri seti içerisindeki tüm elemanları sınıflandırmaktır. Bunun için veri setindeki öznitelikler düğümleri oluşturur ve her düğümde belirlenen kritere göre veri ikiye bölünür. Daha sonra özellik vektörleri incelenir ve en yüksek bilgi kazancına sahip düğüm dallanma yapmak için kullanılır. Bu şekilde bütün elemanlar sınıflandırılincaya kadar devam edilir. Ağacın yaprakları ise sınıf etiketlerini oluşturur.

Karar ağacı uygulamasında ayırt edici nitelikleri belirlemek için her nitelik için bilgi kazancı (information gain) ölçülür. Ölçüm için karmaşıklık (entropi) değeri kullanılır.

Entropi beklenmeyen durumların ve belirsizliklerin olma olasılığı ile ilgilidir. Entropi denklemi 2.11’de verilmiştir.

$$H = - \sum P(x) \log P(x) \quad (2.11)$$

Bu denklemde $P(x)$ bir sınıfa ait grubun yüzdesini H ise entropiyi belirlemektedir.

Karar ağacında entropi değerinin en az olduğu bölünmeler yapılması gerekmektedir. Bunun için 2.12’deki denklemle bilgi kazancı hesaplanır.

$$Gain(S, D) = H(S) - \sum_{V \in D} \frac{|V|}{|S|} H(V) \quad (2.12)$$

Bu denklem D özelliğinin S orijinal veri kümesi için bilgi kazancını verir. Bilgi kazancının en yüksek olduğu özellik ise kök olarak belirlenir ve bu düğümden dallanma yapılır.

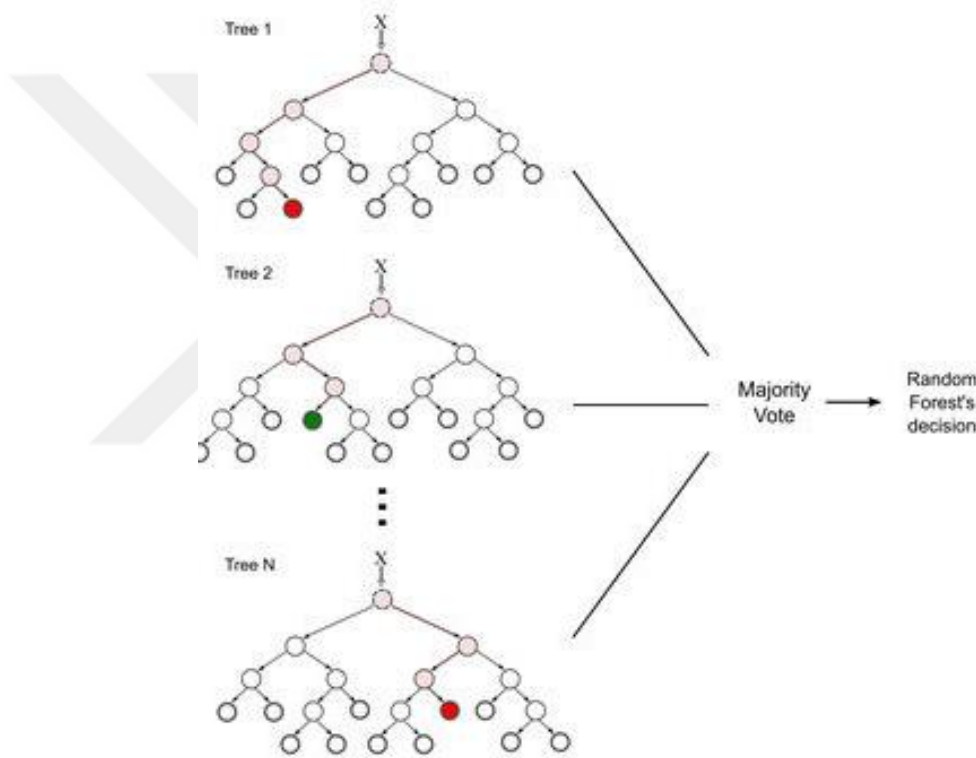
2.3.4 Rastgele orman

Rastgele Orman eğitim ve test olarak ayrılan veri seti üzerinde karar ağacı algoritmasının n defa uygulanması ile daha iyi tahminler yapılmasını sağlayan bir modeldir. Hem sınıflandırma hem de regresyon için kullanılabilir. İstenilen sayıda ağaç oluşturulabilmesi ve en önemli özneliği belirlemek için kullanılabilmesi nedeni ile diğer regresyon ve sınıflandırma araçlarına göre avantajlıdır.

Rastgele Orman algoritmasında kullanıcı tarafından ağaç sayısı (N) ve her bir düğümden kullanılacak değişken sayısı (m) olmak üzere iki parametre alınır. Daha sonra eğitim veri setinin $2/3$ lük kısmı öğrenme (in Bag) $1/3$ ü de test verisi (Out of bag) olarak seçilir. Her düğüm için m değeri tüm değerler arasından rastgele seçilir ve bu değerler arasındaki en iyi dal belirlenir. Bunun için GINI indeksi kullanılır. Denklemi 2.13’de gösterilmiştir.

$$GINI(T) = 1 - \sum_{j=1}^n (P_j)^2 \quad (2.13)$$

Denklemden T veri setinin tamamını, P_j her verinin kendisinden küçük ve büyük eleman sayılarına bölümünün karesini, n ise seçilen veriyi göstermektedir. Algoritma her bir düğümde GINI indeksini hesaplar, GINI indeksinin küçük olduğu değişkeni seçerek diğer düğüme geçer. GINI indeksi sıfır olduğunda düğüm homojendir ve dallanma sona erer. Böylelikle ağaçlar oluşturulur. Her bir ağacın ağırlığı OOB hata oranına göre belirlenir. En düşük hata oranına sahip ağaç en yüksek ağırlığı, en yüksek hata oranına sahip ağaç ise en düşük ağırlığı alır [16]. Daha sonra her bir ağaç ağırlığına göre sınıflandırma oyu kullanır. Son olarak Rastgele Orman algoritmasında bu oylar toplanır. Şekil 2.6'da Rastgele Orman yapısının oluşturulması gösterilmiştir.



Şekil 2.6: Karar Ağaçlarından Rastgele Orman Oluşturulması

2.3.5 Gauss süreci

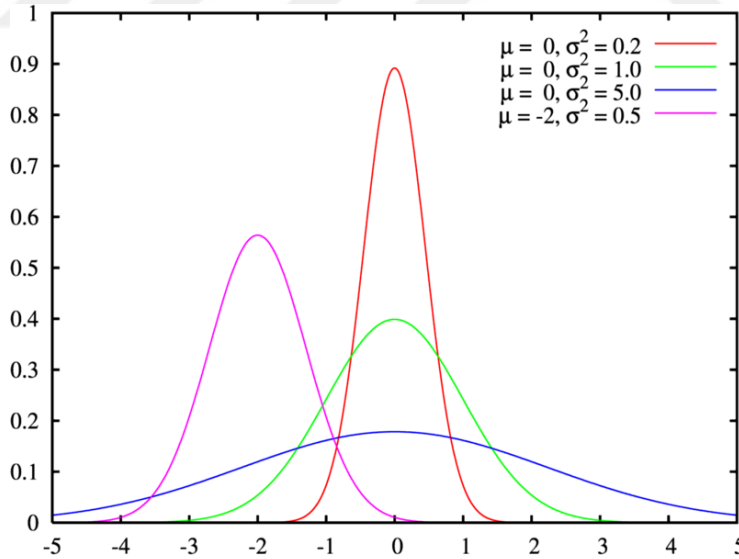
Gauss Süreci hem regresyon hem de sınıflandırma problemleri için kullanılabilen güçlü bir yöntemdir. Çok çeşitli makine öğrenmesi problemlerinde kullanılmıştır. Esnek ve parametrik olmayan yapısı nedeniyle çok tercih edilir. Parametrik olmayan modeller, modelin boyutunun doğrudan veri boyutuna bağlı olduğu, modelin seçimi ve adaptasyonu modellemek için kullanılan basit istatistiksel model teknikleridir[17]. Gauss Süreci ile makine öğrenmesi uygulamasında tembel öğrenmeye dayanır ve eğitim verilerinde karşılaşılmayan noktaların tahmini için çekirdek fonksiyonunu

kullanılır. Bu nedenle diğer çekirdek fonksiyonlu yöntemler ile birçok ortak özelliği vardır. Dezavantajı ise tahminde bulunmak için tüm özellik bilgisini kullanır. Bu nedenle özelliklerin sayısı arttıkça bu model verimliliğini kaybeder. Çok büyük veriler için uygun değildir.

Gauss Süreci, gauss olasılık dağılımının bir genellemesidir[18]. Gauss olasılık dağılımı bir diğer adıyla normal dağılım iki parametre ile tanımlanabilir: μ , aritmetik ortalama ve σ^2 varyansdır. Standart normal dağılım ile verileri analiz etmek için Carl Friedrich Gauss olasılık yoğunluk fonksiyonunu tanımlamıştır[19]. Bu fonksiyonun grafiği çan şeklinde olduğu için çan eğrisi olarak da adlandırılır. Denklem 2.14'de verilmiştir.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.14)$$

Formüldeki μ değeri ortalama değer anlamına gelir. Çan eğrisinin tepe noktasının x değerini belirler. Şekil 2.7'de örnek olarak μ değerlerine karşılık tepe noktaları görülmektedir.



Şekil 2.7: Çan Eğrisi Grafiği

Kesikli olmayan değerlere sahip öznitelikler için o öznitelik uzayındaki örneklerin Gauss (normal) Dağılımına sahip olduğu kabul edilir.

Her öznitelik için ortalama ve varyans tek tek hesaplanır. Bu değerler denklem 2.14 kullanılarak hesaplanır ve Gauss dağılım elde edilir.

2.3.6 Lojistik regresyon

Lojistik Regresyon sıklıkla kategorik verilerin sınıflandırılmasında kullanılan istatistiksel bir yöntemdir. Bu yöntemde bağımsız değişkenlerin sonuç değişkenler üzerindeki etkileri olasılık olarak hesaplanır[20]. Sınıflandırıcı değişkenin iki sonuçlu olduğu durumlarda ikili lojistik regresyon analizi kullanılır. Bu modelde sınıflandırıcı değişken bağımlı değişken olarak belirlenir ve bağımsız değişkenlerle arasındaki ilişki incelenerek sınıflandırmada kullanılır.[21] Lojistik regresyon, s, bağımsız x değişkeninin $-\infty$ ile $+\infty$ arasında değerler alabilen doğrusal işlevi olmak üzere,

$$f(s) = \frac{e^s}{1 + e^s} = \frac{1}{1 + e^{-s}}$$

(2.15)

2.15 de verilen eşitsizlik ile ifade edilir[22].

2.3.7 Naive bayes

Naive Bayes algoritması, verilen bir veri setindeki değerlerin sıklığını ve kombinasyonlarını sayarak bir olasılık kümesini hesaplayan basit bir olasılık sınıflandırıcısıdır[23]. Bayesian sınıflandırıcısı veri içerisindeki öznelikler arası bazı ilişkiler verilmemiş olsa bile sınıflandırma problemlerinde çok iyi performans göstermektedir [24]. Bu model özellikle büyük veriler için kullanışlı ve etkili bir yöntemdir. Naive Bayes tembel öğrenme algoritmasıdır bir eğitim aşaması yoktur. Elemanları sınıflandırmak için hedef değişken ile bağımlı değişkenler arasındaki ilişkiyi analiz eder. Naive Bayes sınıflandırıcısının alt yapısını Bayes Teoremi oluşturur. Bayes Teoremi denklemi 2.16'da verilmiştir.

$$P\left(\frac{C_j}{x}\right) = \frac{P\left(\frac{x}{C_j}\right) \cdot P(C_j)}{P(x)}$$

(2.16)

C_j : Veri Kümesindeki sınıf sayısı

$P\left(\frac{x}{C_j}\right)$: j sınıftan bir olayın x olma olasılığıdır.

$P\left(\frac{C_j}{x}\right)$: x olan bir örneğin sınıf j den olma olasılığıdır.

$P(C_j)$: Sınıf j 'nin ilk olasılığıdır.

$P(x)$: Herhangi bir örneğin x olma olasılığıdır.

Naive Bayes sınıflandırıcı için C sınıfı tahmin edilirken tüm özellikler için yukarıdaki denklem hesaplandığında 2.17'deki eşitlik elde edilir.

$$Y' \leftarrow \operatorname{argmax}_{y_j} \frac{P(y_j) \prod_{i=1}^m P(x = x_i | y_i)}{\sum_{j=1}^J P(y_j) \prod_{i=1}^m P(x = x_i | y_i)} \quad (2.17)$$

Her bir sınıf için bu eşitlik uygulanıp olasılık hesaplanır. En büyük olasılığa sahip sınıf hedef sınıf olarak belirlenir.

3. PROJE MODELİ

Bu çalışma jupyter notebook kullanılarak python dilinde yapılmıştır. Python programlama dili için veri yapıları ve veri analizi araçları sağlayan, csv ve text dosyalarından veri okuma olanağı sağlayan pandas kütüphanesi ve bilimsel hesaplamaların kolaylıkla yapılması için kullanılan bir matematik kütüphanesi olan numpy kütüphanesi kullanılmıştır. Ayrıca makine öğrenmesi yöntemlerini uygulamak için ise karar ağaçları, rastgele ormanlar, doğrusal regresyon, gauss süreci, destek vektör makineleri gibi birçok yöntemi içeren scikit-learn kütüphanesi kullanılmıştır. Matplotlib ve seaborn kütüphaneleriyle ise modellere ait grafiklerin çizimleri gerçekleştirilmiştir. (Ana modelin gauss olduğunu ve diğer modellerle kıyaslandığını genel olarak anlat.)

Veri setinin %30'u eğitim verisi %70'i ise test verisi olarak kullanılmıştır. Verinin eğitimi, testi ve doğruluğunun hesaplanması, öğrenme eğrilerinin çıkarılması her bir model için tek tek uygulanmıştır. Veri seti üzerinde öncelikle Gauss süreci yöntemi uygulanmış bu yöntem ile elde edilen doğruluk oranı, Karar Ağacı, Rastgele Orman, Destek Vektör Makineleri, K-En Yakın Komşu ve Naive Bayes yöntemlerinden elde edilen doğruluk oranları ile karşılaştırılmıştır. Rastgele Orman yöntemi ile aynı zamanda öznitelik seçimi uygulaması da yapılmıştır.

3.1 Veri Seti

Max Little tarafından hazırlanan veri seti MDVP ses analiz programı ile alınan, 22 öznitelikten oluşan ses ölçümlerini içermektedir. Bu ölçümler 23'ü Parkinson hastası olan toplam 31 denekten, altışar veya yedişer defa alınarak toplamda 22x195 boyutlarında bir veri seti oluşturulmuştur. Çalışmada öncelikle veri seti içerisindeki her bir özneliğin anlamı ve işlevi belirlenmiştir. Başlıca kullanılan değerler aşağıda kısaca açıklanmıştır.

Temel Frekans: Sesin bir frekansı, boyu, periyodu ve hızı bulunmaktadır. Sesin birim zamandaki (genellikle saniye) titreşim sayısına "frekans" denir[25]. Bu öznitelik kişinin temel frekansına ait ortalama değeri içermektedir.

Jitter: Bu öznelik temel frekansta meydana gelen bozulmaları içermektedir. Analiz edilen ses örneğindeki her periyodun, kendinden sonraki periyot ile farkı hesaplanarak ölçülür.

Göreceli Düzensizlik Ortalaması (RAP): Kontrollü veya kontrolsüz perde değişimlerinde, kişinin sesini aynı perdede tutamaması gibi durumlarda jitter değerlerini etkilememesi için kullanılan bir jitter hesaplama yöntemidir. Burada 3 periyotluk bir düzeltme faktörü uygulanır. Analiz edilen ses örneğinde 3 periyot arasında periyottan periyoda olan değişikliği analiz eder[26].

Perde Düzensizlik Oranı (PPQ): Bu değer RAP değerinden farklı olarak 3 yerine 5 periyotluk düzeltme uygulanarak hesaplanmaktadır.

Shimmer: Burada ses sinyallerindeki çok kısa süreli şiddet değişimleri ölçülmektedir. “Shimmer” olarak adlandırılan şiddet pertürbasyonu dB veya % olarak ifade edilir.

Gürültü-Harmonik Oranı (NHR/HNR): Harmonik enerjinin gürültü enerjisine ve gürültü enerjisinin harmonik enerjiye olan oranlarını ifade etmektedir.

Veri setine ait öznelikler Çizelge 3.1’de birinci sütunda verilmiştir. Öznelik isimleri uygulamada hatalara sebep olmaması için ikinci sütundaki gibi düzenlenmiş ve bu şekilde kullanılmıştır. Üçüncü sütunda ise özneliklerin açıklamaları verilmiştir. Bu çalışmada 22 öznelik bağımsız değişken olarak ve hastalık durumunu gösteren “status” ise bağımlı değişken olarak belirlenmiştir.

Çizelge 3.1: Öznelik Eşleştirme ve Açıklamaları

MDVP:Fo(Hz)	idx2	ortalama ses temel frekansı
MDVP:Fhi(Hz)	idx3	maksimum ses temel frekansı
MDVP:Flo(Hz)	idx4	minimum ses temel frekansı
MDVP:Jitter(%)	idx5	temel frekans değişimi
MDVP:Jitter(Abs)	idx6	
MDVP:RAP	idx7	
MDVP:PPQ	idx8	
Jitter:DDP	idx9	
MDVP:Shimmer	idx10	
MDVP:Shimmer(dB)	idx11	
Shimmer:APQ3	idx12	
Shimmer:APQ5	idx13	
MDVP:APQ	idx14	
Shimmer:DDA	idx15	Gürültünün tonal bileşenlere oranı
NHR	idx16	
HNR	idx17	dinamik karmaşık ölçüm
RPDE	idx18	
DFA	idx19	Sinyal fraktal ölçeklendirme katsayısı
spread1	idx20	doğrusal olmayan temel frekans değişimi
spread2	idx21	
D2	idx22	dinamik karmaşık ölçüm
PPE	idx23	doğrusal olmayan temel frekans değişimi
status	Label1	hasta ve sağlıklı durumu (1,0)

Veri setini analize uygun hale getirmek için 195'e 25 boyutlarındaki veri seti “.csv” uzantılı dosya olarak düzenlenmiştir.

3.3 Öznitelik Seçimi

Öznitelik seçimi var olan veri setinin yerine kullanılabilir bir alt veri setini elde etmek amacı ile uygulanır. Böylelikle daha az öznitelik kullanılarak aynı veya yakın sonuçlar elde edilebilmektedir. Bu çalışmada veri seti içerisinde en önemli özniteliklerin belirlenebilmesi ve bir alt veri seti oluşturulabilmesi için Rastgele Orman modeli uygulanmıştır.

Rastgele Orman ile elde edilen karışıklık matrisi Çizelge 3.2'de verilmiştir.

Çizelge 3.2: Rastgele Orman Karışıklık Matrisi

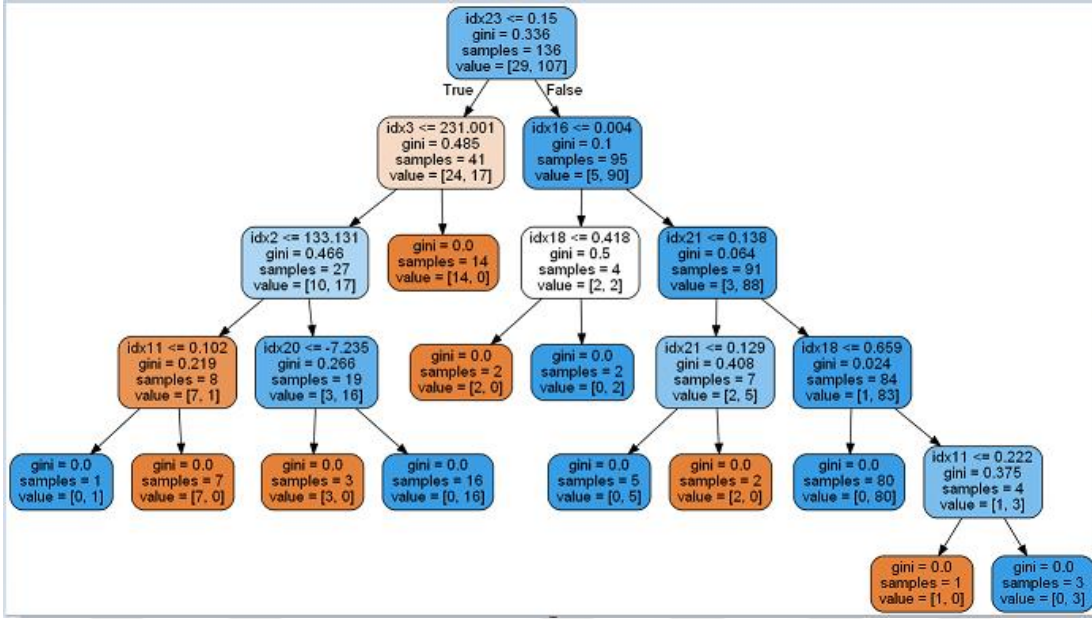
TAHMİN \ GERÇEK	GERÇEK	
	TP=13	FP=6
TAHMİN	FN=0	TN=40

Rastgele Orman modeli ile hasta olan kişiler %93 doğruluk oranı ile belirlenebilmiştir. Karar ağacı yöntemine göre doğruluk oranı artış göstermiştir. Modele ait performans değerlendirme sonuçları Çizelge 3.3'de verilmiştir.

Çizelge 3.3: Rastgele Orman Performans Değerlendirmesi

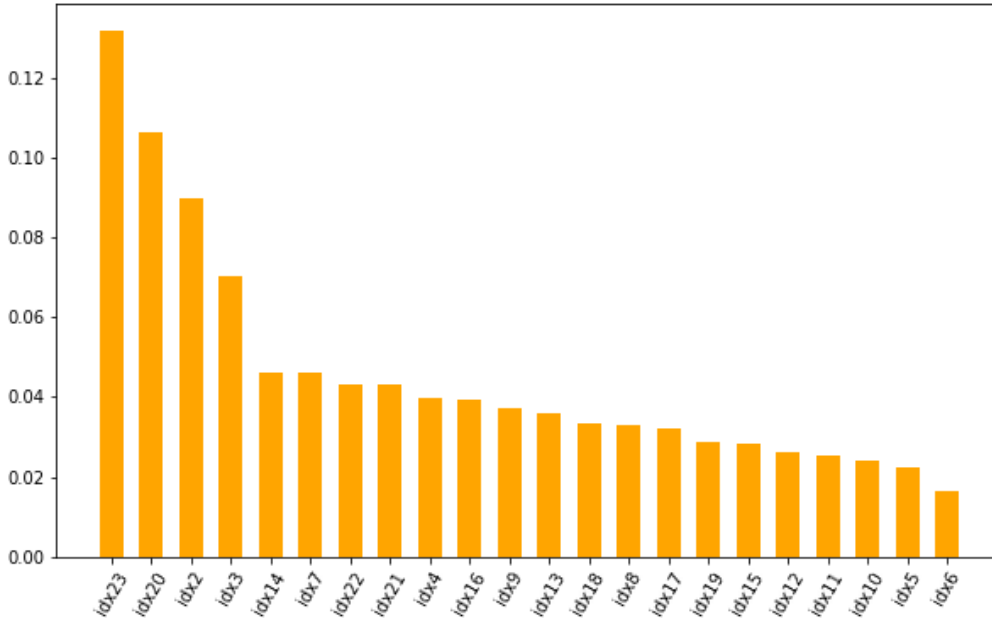
	Precision	Recall	f1-score	Support
0	1.00	0.68	0.81	19
1	0.87	1.00	0.93	40

Gini indeksi kullanılarak oluşturulan ağaç yapısı ise aşağıdaki şekilde gösterilmiştir.



Şekil 3.1: Gini indeksi ile oluşturulan karar ağacı

Veri setindeki özelliklerin önem sırası en yüksekte en düşüğe doğru Şekil 3.2’de verilmiştir.



Şekil 3.2: Önem Sırasına Göre Öznitelikler

Kullanılan orjinal veri setinin çok büyük hacimli bir veri seti olmaması nedeni ile yalnızca en önemli olan ilk 4 (PPE, spread1, Fo, Fhi) öznitelik seçilerek işlem yapıldığında başarı oranlarının ciddi bir düşüş gösterdiği görülmüştür. Öznitelikler içerisinde ilk 10’u seçildiğinde de aynı şekilde tüm özniteliklerin kullanıldığı duruma

göre sonuçlarda düşüş gözlemlenmiştir. Bu nedenle çalışmanın devamında 22 özneliğin tümü kullanılarak makine öğrenmesi yöntemleri uygulanmıştır.

3.4 Gauss Süreci Uygulama Adımları

Veri seti uygun formata getirildikten sonra bağımlı ve bağımsız değişkenler indekslenmiş ve uygulamaya yüklenmiştir. Veri setinin %30 eğitim verisi olarak seçilmiştir. Aşırı öğrenmeyi engellemek ve daha verimli sonuçlar almak için k-fold çapraz doğrulama yöntemi kullanılmıştır. Bu kısımda veri seti 10 parçaya ayrılarak çapraz doğrulama metodu uygulanmıştır. Kullanılacak modelin belirlenmesi için Sklearn kütüphanesinden GaussianProcessClassifier modeli uygulamaya eklenmiştir. Gauss Süreci ile sınıflandırma işlemi için RBF çekirdek fonksiyonu parametresi 1 olarak belirlenmiştir. Bu işlemlerin yapıldığı kod bölümü şekil 3.3'te verilmiştir.

```
In [28]: from sklearn.gaussian_process import GaussianProcessClassifier

In [29]: X_train, X_test, Y_train, Y_test = train_test_split(X_scaled, Y,
test_size=0.3, random_state=seed_pi)
```

Şekil 3.3: Sınıflandırıcının eklenmesi ve eğitim-test verilerinin seçilmesi

Öğrenme verisi üzerinde “Fit” fonksiyonu kullanılarak oluşturulan örnek üzerinden modelin öğrenmesi sağlanmıştır. Model eğitildikten sonra ise her bir veri için tahmin edilen bilgileri elde etmek için test verisi üzerinde “predict” fonksiyonu kullanılmıştır. Ayrıca doğruluk oranı hesaplatılmıştır. Bu işlemlerin yapıldığı kod bölümü şekil 3.4'te verilmiştir.

```
In [41]: for clf, name in zip(classifiers, names):
clf.fit(X_train, Y_train)
r2_train=[]
r2_test=[]
r2_overall=[]
figure = plt.figure(figsize=(25, 20))
i=1
for clf, name in zip(classifiers, names):
results = clf.predict(X_test)
r2_train.append(clf.score(X_train,Y_train))
r2_overall.append(clf.score(X_scaled,Y))
acc_i=clf.score(X_test,Y_test)
r2_test.append(round(acc_i,4))
plt.subplot(6, 5, i)
```

Şekil 3.4: Modelin eğitilmesi ve sonuçların alınması

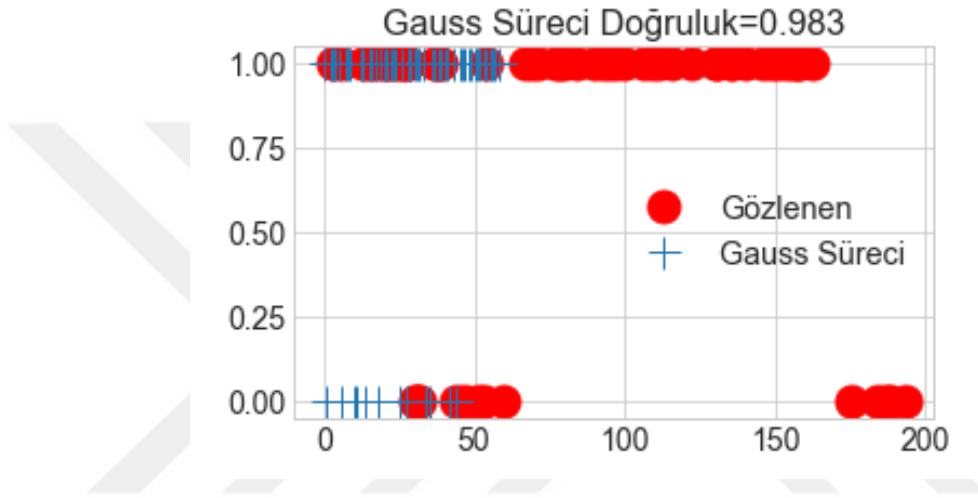
Elde edilen sonuçların değerlendirilmesi için doğruluk grafiđi ve öğrenme eğrisi grafiđi kullanılmıştır. “Predict” fonksiyonu ile elde edilen sonuç değerler grafik 4.1’de “+” işareti ile gösterilmişlerdir.

Diđer modeller için de aynı veri seti üzerinden aynı adımlar izlenerek sonuçlar alınmıştır. Elde edilen sonuçlar karşılaştırılmış ve veri seti için en uygun model belirlenmiştir.



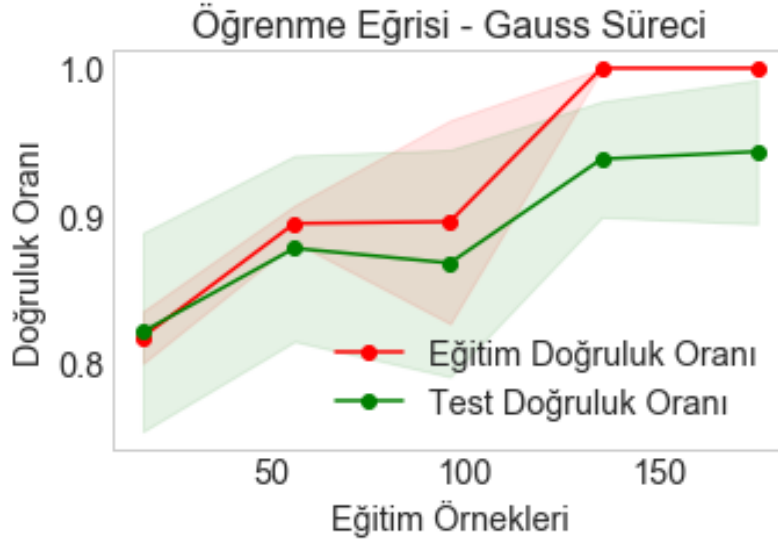
4. DEĞERLENDİRME

Gauss süreci modeli ile veri seti üzerinde yapılan analiz sonucunda %98 doğruluk oranına ulaşılmıştır. Şekil 4.1 de Gauss Süreci için elde edilen doğruluk oranı grafiği verilmiştir. Elde edilen grafiklerin X eksenlerinde örnek sayısı Y eksenlerinde ise doğruluk değeri yer almaktadır.



Şekil 4.1: Gauss Süreci Modeli Doğruluk Oranı Grafiği

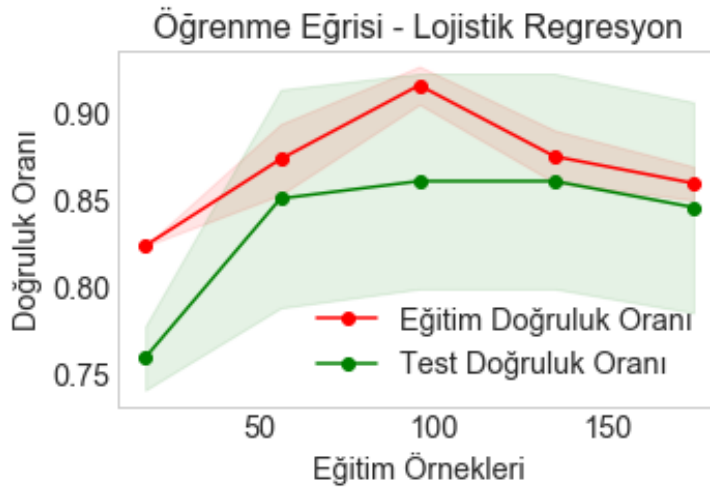
Şekil 4.2 de ise Gauss Süreci modeline ait öğrenme eğrisi grafiği verilmiştir. Grafikte eğitim ve test doğruluk değerleri kayda değer şekilde artış göstermiştir. Gauss Süreci öğrenme eğrisi grafiğinde eğitim ve test aşamalarında modelin veri seti üzerinde verimli bir şekilde çalıştığı gözlemlenmiştir ve doğruluk değeri diğer modellere göre yüksektir.



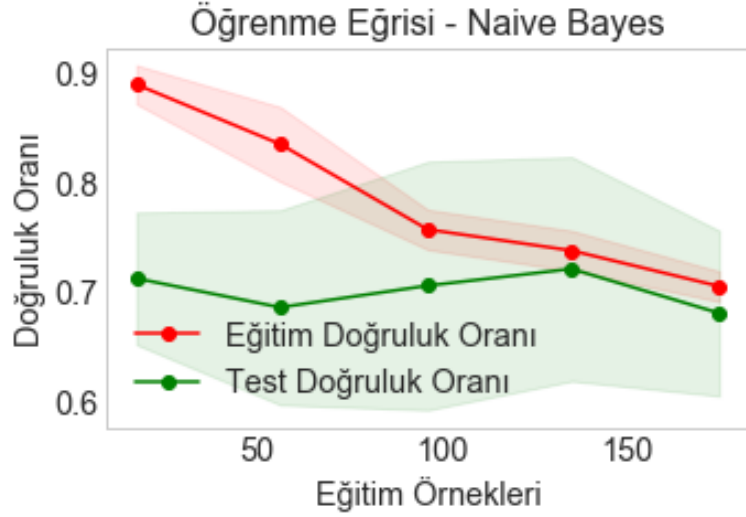
Şekil 4.2: Gauss Süreci Öğrenme Eğrisi Grafiği

Çalışmada kullanılan diğer modeller için doğruluk oranları ve öğrenme eğrisi grafikleri elde edilmiştir. Her bir model ile elde edilen sonuçlar Gauss Süreci modeli ile kıyaslanmıştır.

Şekil 4.3 ve Şekil 4.4’de Lojistik Regresyon ve Naive Bayes modellerinde eğitim ve test aşamalarında doğruluk değerleri düşüş göstermektedir. Lojistik Regresyon modeli için eğitim doğruluk oranında bir miktar yükselme ardından ani bir düşüş gözlenirken, Naive Bayes modelinde ise doğruluk değeri giderek düşen bir eğri ortaya çıkmıştır. Bu grafiklere bakılarak bu modellerde az öğrenme (under fitting) durumu oluşmuş olduğu görülmektedir. Lojistik Regresyon modelinde elde edilen doğruluk oranı %84 Naive Bayes ile elde edilen doğruluk oranı %74 olmuştur.

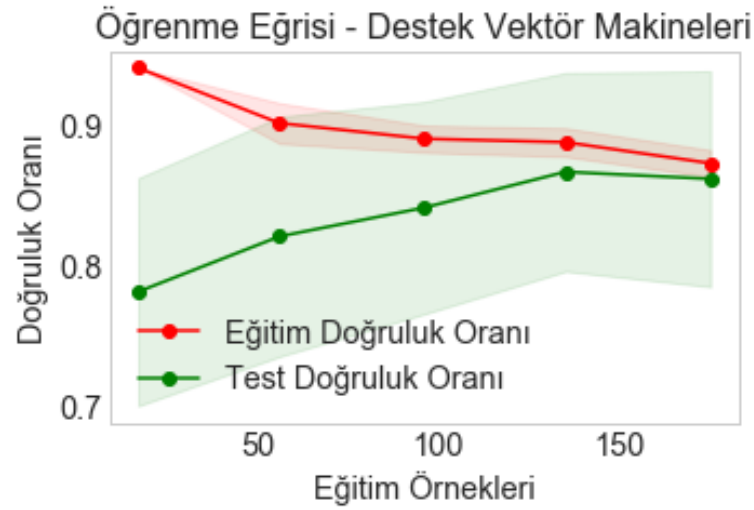


Şekil 4.3: Lojistik Regresyon Öğrenme Eğrisi Grafiği



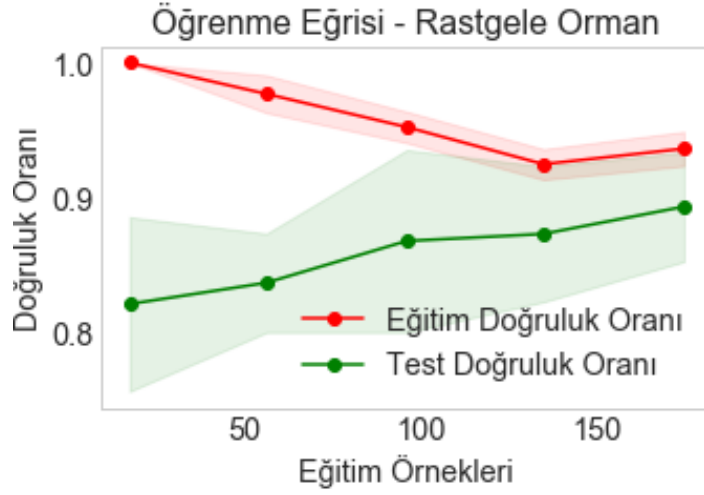
Şekil 4.4: Naive Bayes Öğrenme Eğrisi Grafiği

Şekil 4.5'te DVM yönteminde test aşamasında doğruluk değerinde Gauss sürecine göre düşük bir oran elde edilmiştir. Eğitim aşamasında doğruluk değerinin giderek düştüğü görülmektedir. Bu grafik bu modelde az öğrenme sonunu ile karşılaştığımız anlamına gelmektedir. DVM yöntemi ile elde edilen doğruluk oranı %89'dur.

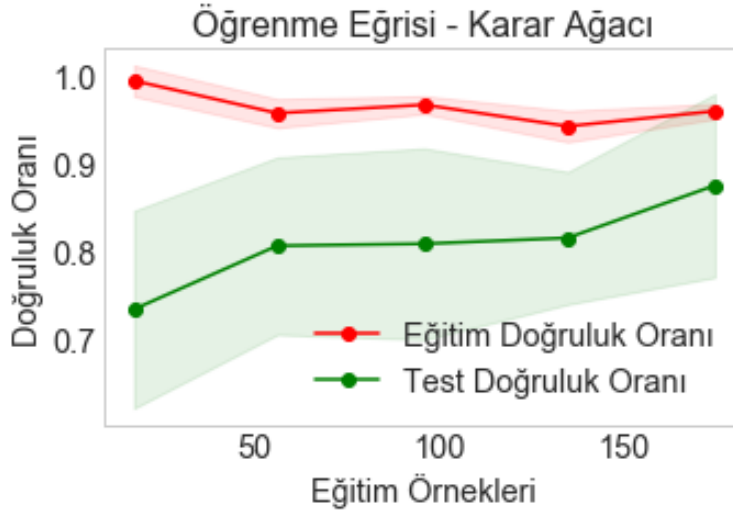


Şekil 4.5: Destek Vektör Makinesi Öğrenme Eğrisi Grafiği

Şekil 4.6 ve Şekil 4.7'da Rastgele Orman ve Karar Ağacı modellerinde diğer modellere göre öğrenme eğrilerinde belirgin bir iyileşme söz konusudur. Rastgele orman yöntemi ile %93 Karar Ağacı yöntemi ile ise %89 oranında başarı elde edilmiştir.

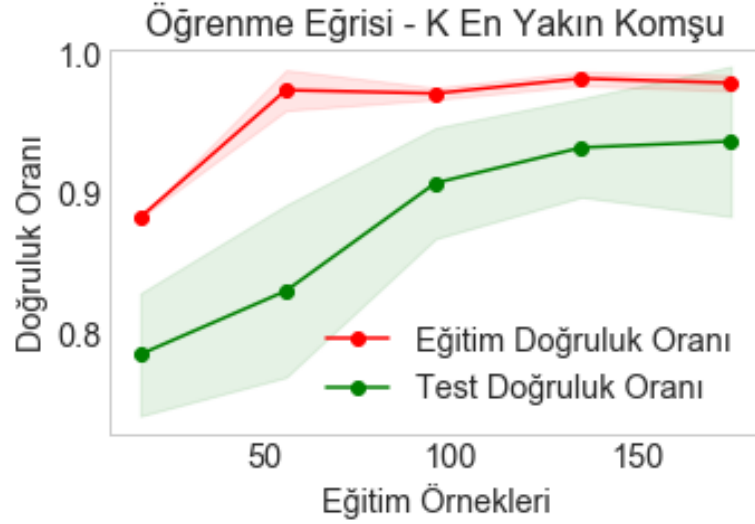


Şekil 4.6: Rastgele Orman Öğrenme Eğrisi Grafiği



Şekil 4.7: Karar Ağacı Öğrenme Eğrisi Grafiği

Şekil 4.8’de ise K En Yakın Komşu yöntemi ile öğrenme eğrisi grafiği ideale en yakın sonuçları vermiştir. Bu modelde doğruluk oranı %93 ile diğer modellere göre daha yüksek çıkmıştır. Fakat Gauss Süreci modelinde doğruluk oranı çok daha yüksektir.



Şekil 4.8: K En Yakın Komşu Öğrenme Eğrisi Grafiği

Çalışmada uygulanan bütün makine öğrenmesi yöntemleri ile elde edilen başarı oranları Çizelge 5.1’de toparlanmıştır. En Yüksek sonuçlar sırası ile Gauss Süreci, Rastgele Orman ve K-En Yakın Komşu yöntemleri ile elde edilmiştir. Veri seti üzerinde uygulanan tüm modeller içerisinde en düşük sonuç ise Naive Bayes yöntemi ile elde edilmiştir.

Çizelge 4.1: Sonuçlar

Yöntem	Sonuç
Gauss Süreci	0,983
Naive Bayes	0,746
Destek Vektör Makineleri	0,898
Rastgele Orman	0,932
K-En Yakın Komşu	0,932
Lojistik Regresyon	0,847
Karar Ağacı	0,898

Tüm modellere bakıldığında Gauss Süreci modelinin bu veri seti üzerinde %98 ile en yüksek doğruluk oranını verdiği görülmüştür.



5. SONUÇLAR

Parkinson Hastalığının erken evrelerde ve yüksek doğruluk oranı ile tespit edilmesi hastaların yaşam kalitesini mümkün olan en iyi düzeyde tutmak için çok önemlidir. Bu çalışma ile Parkinson Hastalığının hastalardan alınan ses verileri kullanılarak yüksek doğruluk oranlarında tespit edilmesinin mümkün olduğu görülmüştür. Elimizdeki veri seti 23 Parkinson Hastası ve 8 sağlıklı insandan 6 veya 7 defa alınan ses örneklerinden oluşuyordu. Veri setinin yapısı itibari ile hasta insanlardaki ses anormalliklerinin tespiti için Gauss Süreci modeli uygun bir yöntemdi. Veri seti üzerinde farklı modeller ile çalışmalar yapılmış, daha önce yapılan çalışmalar arasında KNN yöntemi kullanılarak %96 oranında başarı elde edilmişti, fakat Gauss Süreci modeli ile yapılan herhangi bir çalışma söz konusu değildi.

Daha önce yapılan çalışmalardan farklı olarak Gauss Süreci modeli ile %98 doğruluk oranıyla daha yüksek bir başarı yakalanmıştır. Şekil 5.1'de Gauss Süreci modeli sonucunda elde edilen doğruluk oranı gösterilmiştir.

Bununla beraber K-En Yakın Komşu ve Rastgele Orman yöntemleri ile 0,93 oranında başarı elde edilmiştir. Bu yöntemler ile de elde edilen başarı oranları daha önce yapılan birçok çalışmaya göre daha yüksektir.

Çalışmamızda ayrıca Rastgele Orman algoritması ile öznitelik seçme işlemi yapılmıştır. Bunun sonucunda doğrusal olmayan temel frekans değişimi olan üç değerden ikisi PPE ve Spread1 değerlerinin en önemli öznitelikler olduğu belirlenmiştir. Bu iki değer in ardından ortalama ses temel frekansı (MDVP:F₀(Hz)) ve maksimum ses temel frekansı (MDVP:F_{hi}(Hz)) değerlerinin diğer ölçümlerden hatırı sayılır şekilde daha yüksek öneme sahip olduğu görülmüştür. Bütün öznitelikler içerisinden belirlenen bu en önemli öznitelikler üzerinde aynı makine öğrenmesi modelleri tekrarlandığında başarı oranlarında düşüş olduğu gözlemlenmiştir. Diğer tüm öznitelikler ile beraber Parkinson Hastalığının daha yüksek doğruluk oranı ile teşhis edildiği görülmüştür.



KAYNAKLAR

- [1] Karabudak R.(2014) Parkinson Hatalığı. Türk Nöroloji Derneği
- [2] UĞUREL B, ERANIL Ö, AKTEKİN B, DOĞAN M. (2012) Erken Başlangıçlı Parkinson Hastalığında Psikojenik Disfoni: Olgu Sunumu. Yeditepe Üniversitesi Hastanesi Nöroloji Anabilim Dalı, İstanbul, Türkiye
- [3] Samii A, Nutt JG, Ransom BR (May 2004). "Parkinson's disease". Lancet. 363 (9423): 1783–93.
- [4] Nuytemans K, Theuns J, Cruts M, Van Broeckhoven C (July 2010) [18 May 2010]. "Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update". Human Mutation. 31 (7): 763–80.
- [5] Samii A, Nutt JG, Ransom BR (May 2004). "Parkinson's disease". Lancet. 363 (9423): 1783–93
- [6] ERANIL Ö, AKTEKİN B, DOĞAN M (2012). Psychogenic Dysphonia on Early-Onset Parkinson's Disease: A Case Report Burcu UĞUREL
- [7] Riggare S. (2014) "The Patient Perspective" from The Journal of Parkinson's Disease Alındığı Tarih: 12.04.2019 Adres: <http://www.riggare.se/about-parkinsons-disease/>
- [8] Darley FL, Arenson AE, Brown JR. In (1975), Motor Speech Disorders. Philadelphia: W.B. Saunders Co. 294-295.
- [9] BOZDAĞ M, DOĞU U. (2016) Neuropsychiatric Symptoms in Parkinson's Disease Nöroloji AD. Mersin Üniversitesi Tıp Fakültesi, Mersin -Türkiye Klinikleri J Neurol-Special Topics. 9(4):28-33
- [10] Noyce, A. J.; Bestwick, J. P.; Silveira-Moriyama, L.; Hawkes, C. H.; Giovannoni, G.; Lees, A. J.; Schrag, A. (2012). "Meta-analysis of early nonmotor features and risk factors for Parkinson disease". Annals of Neurology. 72 (6): 893–901.

- [11] Jankovic J (April 2008). "Parkinson's disease: clinical features and diagnosis". *Journal of Neurology, Neurosurgery, and Psychiatry*. 79 (4): 368–76.
- [12] Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*
- [13] Samuel, Arthur (1959). "Some Studies in Machine Learning Using the Game of Checkers". *IBM Journal of Research and Development*. 3 (3): 210–229.
- [14] ŞENER S. (t.y) Yapay Zeka, Makine Öğrenimi ve Derin Öğrenme Arasındaki Farklar, Alındığı Tarih: 25.04.2019 <https://www.endustri40.com/yapay-zeka-makine-ogrenimi-ve-derin-ogrenme-arasindaki-farklar/>
- [15] BROWNLEE J. (2019). A Gentle Introduction to Learning Curves for Diagnosing Machine Learning Model Performance. Alındığı Tarih: 14.03.2019 Adres: <https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/>
- [16] AKMAN M, GENÇ Y, ANKARALI H, (2011). Random Forests Yöntemi ve Sağlık Alanında Bir Uygulama. *Türkiye Klinikleri J Biostat*.
- [17] SULEIMAN N.(2016). Değişim Noktasının Belirlenmesine Parametrik Olmayan Bayesgil Yaklaşım. Gazi Üniversitesi Fen Bilimleri Enstitüsü.
- [18] Rasmussen, C.E.; Williams, C.K.I (2006). *Gaussian Processes for Machine Learning*. MIT Press.
- [19] Gauss, Carolo Friderico (1809). *Theoria motvs corporvm coelestivm in sectionibvs conicis Solem ambientivm* [Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections]
- [20] Baş, S. & Uzun, A. (2018). "Tedarik Zincirinde Müşteri Siparişlerinin Lojistik Regresyon Analizi İle Değerlendirilmesi", Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 11(3), 67-81.
- [21] Burmaoğlu S, Oktay E, Özen Ü (2009). "Birleşmiş Milletler Kalkınma Programı Beşeri Kalkınma Endeksi Verilerini Kullanarak Diskriminant Analizi Ve Lojistik Regresyon Analizinin Sınıflandırma Performanslarının Karşılaştırılması" 28-30.
- [22] Gök, M, Atuntaş, V, "Regresyon Analizi", (Ed. Akçetin E, Çelik, U, Gök, M.) "Rapidminer ile Veri Madenciliği", 85-86, Pusula yayıncılık, Ankara, 2017.
- [23] Domingos P, Pazzani M (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, *Machine Learning*, 29, 103-130.

[24] Reprinted from Nature, Vol. 218, No. 5136, pp. 19-22, April 6, 1968
<https://stacks.stanford.edu/file/druid:jt687kv7146/jt687kv7146.pdf>

[25] Solmaz S.(2017) Ses-Frekans Nedir? Alındığı Tarih: 12.03.2019 Adres:
<https://medium.com/sercansolmaz/ses-frekans-nedir-cb24477d83fa>

[26] Serkan B.(2018) Ses Analiz Programlarının KBB Pratiğinde Kullanım Alanları.
Alındığı Tarih: 12.03.2019 Adres: https://rinoskop.com/wp-content/uploads/2018/04/07_ses_analiz_programlarinin_KBB.pdf





ÖZGEÇMİŞ



Ad-Soyad : Raife Göksu KÜÇÜK
Doğum Tarihi ve Yeri: 02.02.1993 / MALATYA
E-posta : goksu93.kucuk@gmail.com

ÖĞRENİM DURUMU:

Lisans : Fırat Üniversitesi Bilgisayar Mühendisliği

MESLEKİ DENEYİM :

2016-2019 :Bilgi İşlem Yetkili Yardımcısı /Austriacard Türkiye Kart Operasyonları Aş.
2019 :Teknik Danışman /Mechsoft Türkiye