

**KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

YÜKSEK LİSANS TEZİ

**EMG VERİLERİNE DAYALI OLARAK MAKİNE ÖĞRENMESİ
YÖNTEMLERİ İLE POLİNÖROPATİ TEŞHİSİ**

NİMET TANIK

KOCAELİ 2021

KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

YÜKSEK LİSANS TEZİ

EMG VERİLERİNE DAYALI OLARAK MAKİNE ÖĞRENMESİ
YÖNTEMLERİ İLE POLİNÖROPATİ TEŞHİSİ

NİMET TANIK

Prof.Dr. Adnan KAVAK

Danışman, Kocaeli Üniversitesi

.....

Prof.Dr. Kerem KÜÇÜK

Jüri Üyesi, Kocaeli Üniversitesi

.....

Prof.Dr. Celal ÇEKEN

Jüri Üyesi, Sakarya Üniversitesi

.....

Tezin Savunulduğu Tarih: 23.06.2021

ÖNSÖZ VE TEŞEKKÜR

Bu tez çalışması, Sağlık Bilgi Sistemlerinde EMG verilerine dayanarak yaşam kalitesini oldukça düşüren polinöropati hastalığını makine öğrenmesi teknikleri kullanılarak teşhisinin yapılabilmesi amacıyla gerçekleştirilmiştir.

Tez çalışmamda desteğini esirgemeyen, çalışmalarına yön veren, bana güvenen ve yüreklendiren danışmanım Prof.Dr. Adnan KAVAK' a sonsuz teşekkürlerimi sunarım.

Tez çalışmam boyunca EMG verileri ile polinöropati konularında destek veren Balıkesir Devlet Hastanesi Nöroloji uzmanlarına ve aynı zamanda bu çalışmada verilerin incelenmesi ve analiz aşamasında yardımını esirgemeyen Balıkesir Devlet Hastanesinde görevli sağlık çalışanlarına teşekkür ediyorum.

Yüksek lisans öğrenimimde beni cesaretlendiren ve her aşamasında bana güç veren en büyük destekçim, her aşamada sıkıntılarımı ve mutluluklarımı paylaşan sevgili eşim A. Hadi TANIK' a teşekkürü bir borç bilirim.

Okul hayatım boyunca destek, anlayış ve sabırlarını esirgemeyen sevgili aileme teşekkürlerimi sunarım.

Ağustos – 2021

Nimet TANIK

İÇİNDEKİLER

ÖNSÖZ VE TEŞEKKÜR	i
ŞEKİLLER DİZİNİ.....	iii
TABLolar DİZİNİ	iv
ÖZET.....	vi
ABSTRACT.....	vii
GİRİŞ	1
1. PROBLEM TANIMI:POLİNÖROPATİ HASTALIĞI VE TEŞHİS YÖNTEMLERİ.....	2
1.1 Elektromiyografi (EMG).....	2
1.2 Sinir Sistemi	2
1.3 Polinöropati Hastalığı.....	3
2. MAKİNE ÖĞRENMESİ İLE POLİNÖROPATİ TEŞHİSİ	9
2.1 Makine Öğrenmesi	9
2.2 Makine Öğrenme Türleri.....	10
2.2.1 Denetimli makine öğrenmesi.....	11
2.2.2 Denetimsiz makine öğrenmesi	12
2.2.3 Yarı-denetimli makine öğrenmesi	13
2.2.4 Pekiştirerek öğrenme	14
2.3 Problem Çözümünde Uygulanılan Makine Öğrenme Algoritmaları	15
2.3.1 Destek vektör makinesi algoritması(SVM).....	15
2.3.2 Karar ağaçları	18
2.3.3 K- en yakın komşu algoritması(KNN)	20
2.3.4 Naive bayes algoritması(NB)	21
2.3.5 Rastsal orman algoritması	22
2.4 EMG Verilerinin Sınıflandırılması.....	23
3. DENEYSEL ÇALIŞMALAR	25
3.1 Problem Çözümünde Kullanılan Yazılım Programı.....	25
3.2 Uygulama Veri seti Analiz Bilgisi	27
3.3 Yöntem	29
3.3.1 Normalizasyon işlemi	29
3.3.2 Çapraz doğrulama modeli	30
3.3.3 Karışıklık matrisi	31
3.4 Araştırma Bulguları	33
3.4.1 J48 algoritması.....	33
3.4.2 Random forest algoritması	36
3.4.3. Naive bayes algoritması	38
3.4.4. IBK algoritması	40
3.4.5. SMO algoritması	42
3.5 Araştırma Bulgularını Değerlendirme	44
4.SONUÇLAR VE ÖNERİLER	46
KAYNAKLAR	48
KİŞİSEL YAYINLAR VE ESERLER	52
ÖZGEÇMİŞ	53

ŞEKİLLER DİZİNİ

Şekil 1.1. Üst ekstremitte eldeki sinir görünümü	4
Şekil 1.2. Üst ekstremitte koldaki sinir görünümü	4
Şekil 1.3. Alt ekstremitte sinir görünümü	5
Şekil 1.4. EMG çekimi yapıldıktan sonra polinöropati hastasına ait veri	6
Şekil 1.5. EMG çekimi yapıldıktan sonra normal hastaya ait veri	7
Şekil 2.1. Makine Öğrenme türleri	10
Şekil 2.2. Denetimli öğrenme modeli süreci	11
Şekil 2.3. Sınıflandırma ve Regresyon Algoritması	12
Şekil 2.4. Denetimsiz öğrenme modeli	13
Şekil 2.5. Yarı denetimli öğrenme modeli	14
Şekil 2.6. Pekiştirerek öğrenme modeli	14
Şekil 2.7. Destek Vektör Makinesi Algoritması	17
Şekil 2.8. Karar Ağacı Yapısı	19
Şekil 2.9. KNN Algoritması	21
Şekil 2.10. Rastsal Orman Çalışma Yöntemi	23
Şekil 2.11. EMG Verilerinin Sınıflandırılmasında Kullanılan Akış Şeması	23
Şekil 3.1. WEKA aracının ara yüzü	25
Şekil 3.2. Kullanılan EMG veri setinden örnek kayıtlar	28
Şekil 3.3. K-Katlamalı Çapraz Doğrulama modeli	31
Şekil 3.4. Karışıklık Matrisi	31
Şekil 3.5. WEKA pre-process ekranı	33
Şekil 3.6. J48 Algoritması sonucunda çıkan karışıklık matrisi	34
Şekil 3.7. J48 Algoritmasının sınıflandırma sonuçları	35
Şekil 3.8. J48 Algoritmasının ROC Eğrisi sonuçları	35
Şekil 3.9. Random Forest Algoritması sonucunda çıkan karışıklık matrisi	36
Şekil 3.10. Random Forest algoritması sınıflandırma sonucu	37
Şekil 3.11. Random Forest algoritması ROC eğrisi	37
Şekil 3.12. Naive Bayes Algoritması sonucunda çıkan karışıklık matrisi	38
Şekil 3.13. Naive Bayes algoritması sınıflandırma sonucu	39
Şekil 3.14. Naive Bayes algoritması ROC eğrisi	39
Şekil 3.15. IBK Algoritması sonucunda çıkan karışıklık matrisi	40
Şekil 3.16. IBK algoritması sınıflandırma sonucu	41
Şekil 3.17. IBK algoritması ROC eğrisi	41
Şekil 3.18. SMO Algoritması sonucunda çıkan karışıklık matrisi	42
Şekil 3.19. SMO algoritması sınıflandırma sonucu	43
Şekil 3.20. SMO algoritması ROC eğrisi	43
Şekil 3.21. EMG veri seti ile makine öğrenmesi başarı oranları grafiği	45

TABLolar DİZİNİ

Tablo 1.1.EMG Alt ve Üst Ekstremitte Motor ve Duyu Sinir İleti Normal Hızları	9
Tablo 3.1.EMG veri setinin özellik bilgileri	27
Tablo 3.2.MG veri seti ile makine öğrenmesi başarı oranları.....	44



SİMGELER VE KISALTMALAR DİZİNİ

Kısaltmalar

EMG	:	Elektromiyogram
DT	:	Karar ağaçları
FN	:	Yanlış Negatif
FP	:	Yanlış Pozitif
KNN	:	K-En Yakın Komşu
MSS	:	Merkezi Sinir Sistemine
NB	:	Naif Bayes
PSS	:	Periferik Sinir Sistemi
RF	:	Rassal Orman
SML	:	Denetimli Makine Öğrenimi
SVM	:	Destek Vektör Makinesi
TN	:	Doğru Negatif
TP	:	Doğru Pozitif
UML	:	Denetimsiz Makine Öğrenimi

EMG VERİLERİ KULLANILARAK MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE POLİNÖROPATİ TEŞHİSİ

ÖZET

Erken ve doğru teşhis ile en verimli tedavi uygulamalarına karar destek noktasında kullanılmaya başlanan bilgi teknolojileri, sağlık hizmeti sunucuları ve hastaları için hız, zaman, doğruluk ve kaliteli sağlık hizmeti sunmaya destek olmaktadır. Yine bu alanda, ortaya çıkan hastaya özgü verileri farklı algoritma ve bilgi teknolojileri kullanan modeller geliştirerek, tıpta daha verimli bir şekilde uygulama alanı oluşturabiliriz.

Bu çalışmada, tıbbi bilgi sistemleri veri tabanlarında toplanan yapılandırılmış elektronik tıbbi kayıtlara dayanarak polinöropati hastalığı riskini belirlemek için makine öğrenmesi yöntemleri uygulamaktır. EMG (Elektromiyogram) analizine dayalı polinöropati tespiti için, 100 kişiye (72 sağlıklı, 28 polinöropati) ait toplam 2200 gerçek veri seti kullanılmış olup, özellik arz eden veriler yaygın olarak kullanılan Makine Öğrenmesi algoritmalarına uygun olacak şekilde tekrar yapılandırılmıştır. EMG verileri, sınıflandırılmadan önce veri ön işleme, veri gruplandırma ve öznitelik çıkarma aşamalarından geçirilmiştir. Bu aşamada profesyonel sağlık hizmeti sunucularından destek alınarak verimliliğin en üst seviyeye çıkarılması amaçlanmıştır. Bir sonraki aşamada ise Karar Ağaçları, Naive Bayes, K-NN Algoritması, Destek Vektör Makineleri (SVM) Algoritması ve Rastsal Orman sınıflandırma algoritmaları ile veriler sınıflandırılmıştır. Çoklu öznitelik vektörlerinin kullanılması durumunda Naive Bayes algoritması sınıflandırıcıları, diğer sınıflandırıcılara göre EMG verilerini daha yüksek doğrulukla sınıflandırmış olup toplam doğruluk %80 olarak tespit edilmiştir. Ortaya çıkan sonuçlar sağlık hizmeti sunucuları ile paylaşarak verimliliğin memnun edici seviyede olduğu teyit edilmiştir. Tahmine dayalı modellerin sağlık alanında daha fazla uygulama alanı bulması sağlık hizmeti veren ve alan açısından önemli faydalar sağlayacağı tespit edilmiştir.

Anahtar Kelimeler: EMG, K-NN Algoritması, Naive Bayes, Polinöropati, SVM.

MACHINE LEARNING BASED POLYNEUROPATHY DIAGNOSIS USING ELECTROMYOGRAM DATA

ABSTRACT

Information technologies, which are used at the decision support point for the most efficient treatment applications with early and accurate diagnosis, support to provide speed, time, accuracy and quality health care for healthcare servers and patients. Again, in this field, we can create a more efficient application area in medicine by developing models that use different algorithms and information technologies.

The aim of this study is to implement machine learning methods to determine the risk of polyneuropathy based on structured electronic medical records collected in medical information systems databases. A total of 22,000 actual data sets belonging to 1,000 people (720 healthy, 280 polyneuropathies) were used for polyneuropathy detection based on EMG analysis and the featured data was reconfigured to match commonly used Machine Learning algorithms. In this study, EMG(Electromyogram) data was passed through the data preprocessing, data grouping and attribute extraction stages before being classified. At this stage, it is aimed to maximize efficiency by getting support from professional healthcare providers. In the next stage, data is classified with Decision Tree, Naïve Bayes, K-NN Algorithm, Support Vector Machine (SVM) Algorithm and Random Forest classification algorithms. In the case of multiple attribute vectors, Naive Bayes algorithm classifiers classified EMG data with higher accuracy than other classifiers, with total accuracy at 80%. The resulting results were shared with health care providers and the efficiency was confirmed to be satisfactory. It has been determined that predictive models will find more applications in the field of health, providing health care and providing undeniable benefits in terms of the field.

Keywords: EMG, K-NN Algorithm, Naïve Bayes, Polyneuropathy, SVM.

GİRİŞ

Gelişen bilgi teknolojileri hayatın her alanında uygulama imkânı bulmakla beraber, insanlığı yakından ilgilendiren sağlık alanında da sağlık hizmeti sunucuları ve hastaları için hız, zaman, doğruluk ve kaliteli sağlık hizmeti sunma kapsamında oldukça kritik roller üstlenmektedir. Erken ve doğru teşhis ile en verimli tedavi uygulamalarına karar destek noktasında kullanılan bilgi teknolojileri, her geçen gün daha fazla ürünle ortaya çıkmaktadır.[1]

Makine öğrenmesini, sağlık alanında hastalık teşhisi, kişisel bakım desteği, ilaç keşfi, medikal görüntü analizi ve robotik sağlık asistanları geliştirilmesi gibi araştırma çalışmalarına konu olduğunu görmekteyiz [2-6]. Özellikle insanlığın farklı hastalık türleriyle her zaman karşılaşması ve erken teşhis ve tedavi kapsamında, Yapay Zekâ ve Makine öğrenmesi tekniklerinin sağladığı faydalar sağlık hizmeti sunucuları için daha popüler hale gelmiştir. Makine öğrenme algoritmalarının sağlık alanlarında kullanılmaya devam edilmesi ile hem doktorların iş yükleri azaltılacak hem de hastalara daha kaliteli sağlık hizmeti sağlanacaktır.

Bu tez çalışmasının amacı; makine öğrenmesi teknikleri kullanılarak yaşam kalitesini oldukça düşüren polinöropati hastalığı tespiti hedeflenmiştir. Bu amaçla devlet hastanesinden alınan 72'si sağlıklı, 28'ni polinöropati hastası olan yani toplamda 100 kişiden oluşan bir veri seti elde edilmiştir. Bu veri setinde Elektromiyografi (EMG) kullanılarak uzman hekimler tarafından belirlenen her bir veri grubunda 22'şer tane alt veri bulunmaktadır. Önerilen bu sistem ile sağlık kurumlarında daha hızlı ve kesin sonuçlar elde edilebilmesi için uygun bir arayüz programı ile desteklenmesi gerekmektedir.

1. PROBLEM TANIMI: POLİNÖROPATİ HASTALIĞI VE TEŞHİS YÖNTEMLERİ

Bu bölümde araştırması yapılan çalışmanın temel bilgileri, tanımlamaları ve tez çalışması doğrultusunda bilgi edinmek için yapılan literatür araştırmalarından bahsedilmiştir. İlk olarak EMG'nin tanımının ne olduğu, hangi yollar ile yapılabileceği anlatılmıştır. Sonraki aşamada polinöropati hastalığının detay tanımlaması yapılmıştır

1.1. Elektromiyografi (EMG)

EMG, kas çalışması esnasında ortaya çıkan elektrik sinyalleridir. Kas dokusu sinirlerin yapısına benzer olarak elektriksel potansiyeller üretir. Bunlara kas aksiyon potansiyeli denir. Yüzey EMG, bu kas aksiyon potansiyellerinde bulunan mevcut bilgilerin kaydedilmesi için bir yöntemdir. EMG sinyalini kaydederken, sinyalin aslına uygunluğunu etkileyen iki ana konu vardır. Birincisi sinyal-gürültü oranıdır. Genel olarak gürültü, istenilen EMG sinyaline ait olmayan elektrik sinyali olarak tanımlanır. İkincisi ise EMG sinyaline ait herhangi bir frekans bileşeninin değiştirilmemesidir. EMG sinyalleri, doğrudan cilde monte edilen elektrotlardan elde edildiği için, cilt altında bulunan kasların aksiyon potansiyellerinin birleşimidir. Bu aksiyon potansiyelleri rastgele aralıklarla meydana geldikleri için herhangi bir anda EMG sinyali pozitif veya negatif voltaja sahip olabilir. Kas aksiyon potansiyelleri doğrudan kas içine yerleştirilmiş tel veya iğne elektrotlar kullanılarak elde edilebilir [7].

1.2. Sinir Sistemi

Sinir sistemi, vücudun iç ve dış ortamlardaki olayları algılama ve yanıtlama sistemidir. Dokunma hissi veren reseptörler çevremizde meydana gelen değişiklikler ile ilgili olarak Merkezi Sinir Sistemine (MSS) bilgi gönderir. MSS bu uyarılara hem hareket hem de endokrin sistemden bir miktar hormon değişikliği ile yanıt verir. Sinir sistemi anatomik olarak iki ana bölüme ayrılır. MSS ve Periferik Sinir Sistemi

(PSS). MSS beyin ve omurilikte bulunan sinirlerdir. PSS ise MSS'nin dışında kalan sinir sistemini oluşturmaktadır. [8]

1.3.Polinöropati Hastalığı

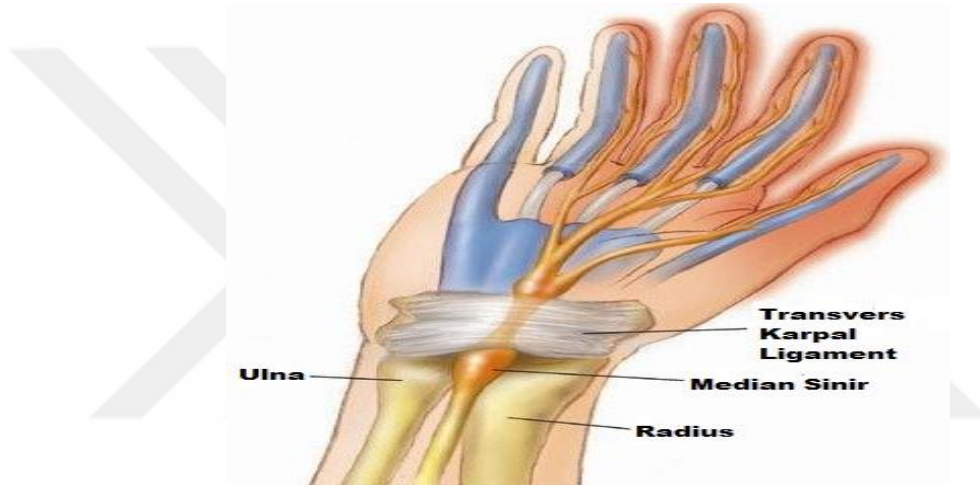
Bu çalışmada yaşam kalitesini oldukça düşüren polinöropati hastalığını makine öğrenmesi teknikleri ile ele aldık. Polinöropati, çoklu periferik (çevre) sinirlerin hasar görmesi durumudur, periferik nöropati olarak da bilinir. Beyin ve omurilik dışındaki sinirler, periferik sinirler olarak sınıflandırılır. [9]

Polinöropatide birden fazla periferik sinir aynı anda etkilenir. Duyularımızdan ve hareket kabiliyetimizden sorumlu olan sinirlerin yanı sıra otonom sinirler de hasar görebilir. Polinöropati hastalığında çoğunlukla çok sayıda sinirin eş zamanlı tutulduğu görülür yani simetriktr. Sıklıkla bacaklarda daha belirgin olarak görülen polinöropati hastalığının nedenleri çok ayrıntılı incelemelere rağmen ancak %65-70'inde hastalığın nedeni ortaya konulabilir.[10] Bu hasara yol açan en önemli etkenlerin başında ise diyabet hastalığı periferik nöropatinin en sık nedenidir. Yüksek kan şekeri ve aşırı salınan insülin etkileri ile metabolik yan maddelerin sinirleri üzerinde birikimi, oksijen yetersizliği, sinirlerin kanlanma bozukluğu nedenleri ile nöropati meydana gelir. Şeker hastalığında yalnızca sinir sistemi değil böbrekler, göz de hastalıklardan etkilenir.[11] Tüm diyabetik hastaların yaklaşık olarak yarısında nöropati gelişimi görülmektedir. Son olarak yapılan TURDEP çalışması sonuçlarına göre ülkemizde şeker hastalığı görülme sıklığı %14'tür. Yani bu demek oluyor ki her 100 kişiden 14'ünde şeker hastalığı tanısı mevcuttur. Bu demek oluyor ki bu hastaların yarısında nöropati hastalığı gelişecek olması çok ciddi bir tablo ile karşı karşıya olduğumuzu göstermektedir.[10]

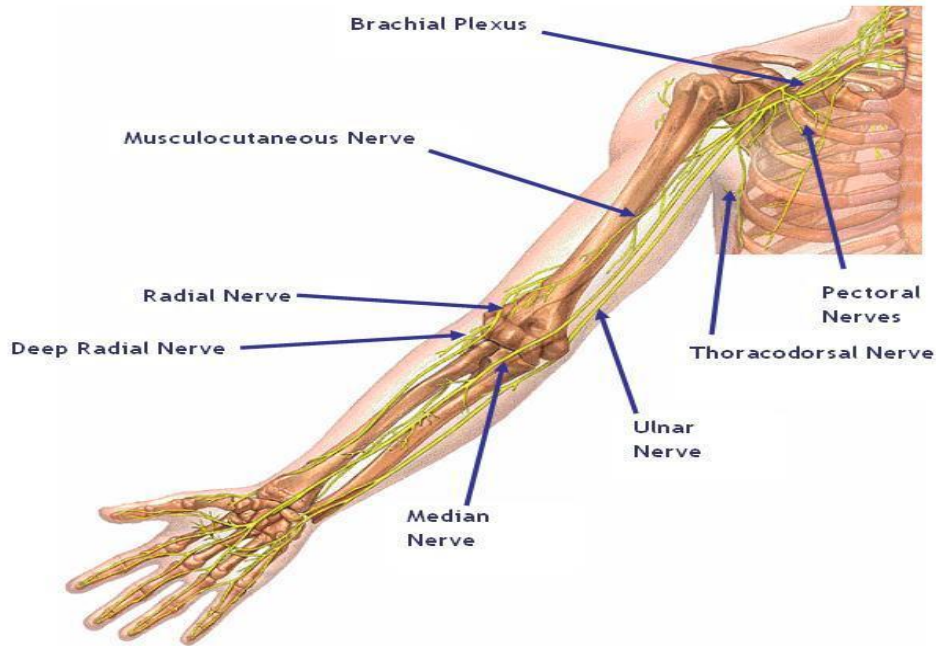
Polinöropati hastalığının bir diğer nedeni ise yetersiz beslenme veya kalitesiz beslenme sonucunda oluşan B1, B6, B12, E vitamini, ve Folik asit eksikliği sonucunda ortaya kişilerde sinir sağlığı özellikle duysal sinirlerde ciddi anlamda hasara yol açabilmektedir. Genel olarak söylemek gerekirse diyabetes mellitus, otoimmün hastalıklar, böbrek yetmezliği, tiroid hastalığı, malign hematolojik hastalıklar, HIV enfeksiyonu, ilaç bağımlılığı, serebrovasküler hastalıklar periferik nöropatiye neden olabilecek sebeplerdendir.[9]

Nöromasküler hastalık olan Polinöropatinin doğru tanısı için son yirmi yıl içinde Elektromiyografi (EMG), arařtırmacılar ve klinisyenler tarafından yaygın olarak kullanılmaktadır.

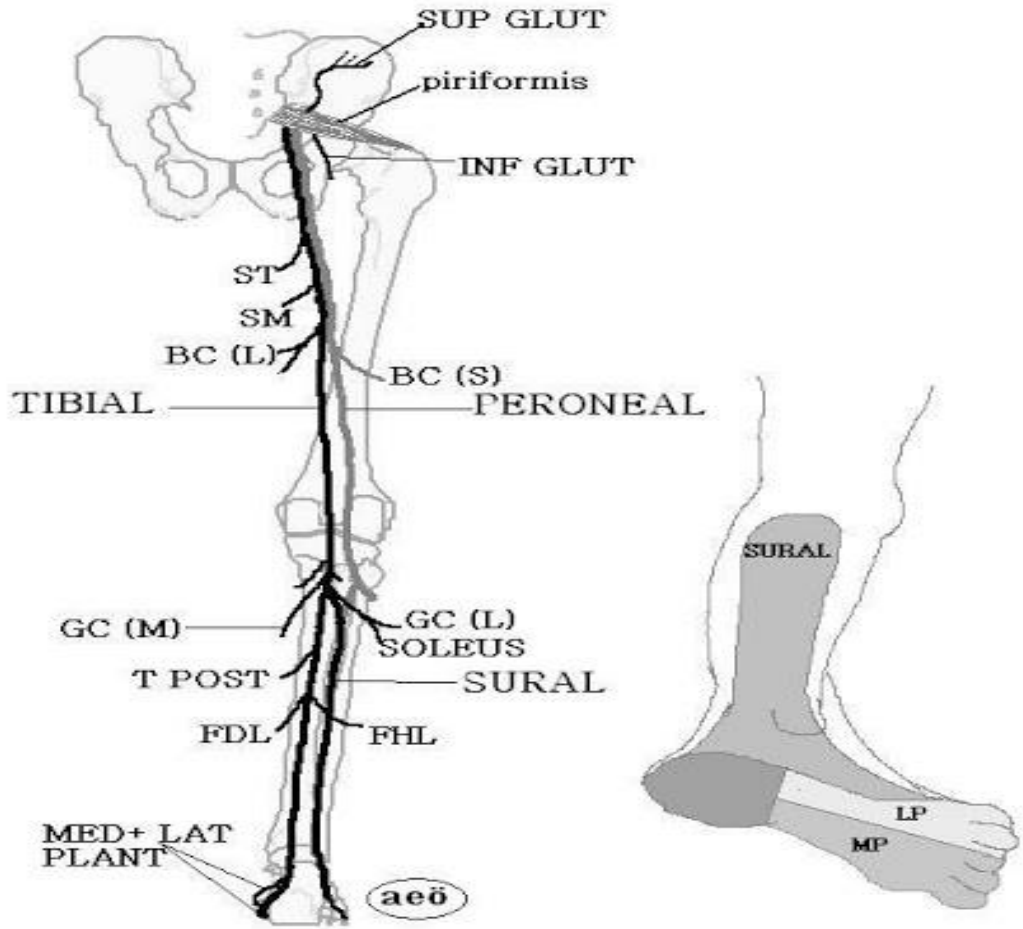
EMG çekiminde üst ekstremitede median, ulnar ve radial sinirler, alt ekstremitede peroneal, tibial, femoral ve sural sinir noktaları üzerinde EMG elektrotları yerleřtirilmektedir. Vücudumuzun belli bölgeleri üzerine yerleřtirilen elektrot ile elektrik akımı verilerek sinirler uyarılır ve bundan sonra sinirde yayılan aktivite bir diđer elektrot aracılıđıyla kaydedilir. Ortaya çıkan sayısal deđerlerin anlamlandırılması ilgili hekim tarafından yapılır. [11]



Şekil 1.1. Üst ekstremitede eldeki sinir görünümü



Şekil 1.2. Üst ekstremitede koldaki sinir görünümü



Şekil 1.3. Alt ekstremitte (peroneal, tibial, femoral ve sural sinir)

Yukarıdaki Şekil 1.1.de üst ekstremitte eldeki sinir görünümü (median, ulnar ve radial sinirler) resimdeki noktalar gösterilmiştir. Bunun yanında yine üst ekstremitte koldaki sinir görünümü (median, ulnar ve radial sinirler) noktaları Şekil 1.2. de gösterilmiştir. Bunların dışında Alt ekstremitte olarak (peroneal, tibial, femoral ve sural sinir) noktaları olarak gösterilmiştir.

Aşağıdaki Şekil 1.4 de hastaneye gelen hastanın hekim tarafından tıbbi muayenesi sonucu polinöropati şüphesiyle Emg çekimi yapılmış ve bunun sonucunda yukarıda da görüldüğü gibi belli değerler doğrultusunda hekim tarafından kesin tanı ile polinöropati teşhisi konulmuş hastanın EMG sonuç görüntüsü mevcuttur.

Aşağıdaki Şekil 1.5 de ise yine hastaneye polinöropati belirtisiyle gelen hastanın Emg çekimi yapılmış ve bunun sonucunda belli değerler doğrultusunda hekim tarafından normal olarak değerlendirilmiş olan hastanın EMG sonuç raporunun görüntüsü mevcuttur.

Summary EMG data

Motor CV

Test	Stimulation site	Lat., ms	Ampl., mV	Dur., ms	Area, mV×ms	Stim., mA	Stim., ms	Dist., mm	Time, ms	Vel., m/s
right, Abductor digiti minimi, Ulnaris, C8 T1										
4	wrist	2,84	6,69	6,96	22,5	19	0,2	80		
	elbow	8,04	6,13	7,44	20,4	34	0,2	180	5,2	34,6
right, Abductor pollicis brevis, Medianus, C8 T1										
2	wrist	6,0	8,75	5,76	25,9	34	0,2	80		
	elbow	12,1	6,48	6,4	21,5	34	0,2	210	6,12	34,3
right, Tibialis anterior, Peroneus, L4 L5 S1										
8	head of fibula	3,92	0,842	5,56	2,9	82	0,5	80		
	popliteal fossa		0			82	0,5			
right, Abductor hallucis, Tibialis, I4 L5 S1										
6	medial malleolus	4,64	1,99	6,72	5,3	60	0,5	70		
	popliteal fossa		0			60	0,5			
right, Extensor digitorum brevis, Peroneus, I4 L5 S1										
	head of fibula		0			84	0,5			

Sensory CV

Test	Site	Lat., ms	Ampl., µV	Dur., ms	Area, nV×s	Stim., mA	Stim., ms	Dist., mm	Time, ms	Vel., m/s
right, n. Ulnaris V dig.										
3	wrist	2,65	12,3	1,16	8,0	11	0,1	120	2,65	45,4
right, n. Medianus										
1	wrist	3,23	49,7	2,81	30,5	9	0,1	140	3,23	43,4
right, n. Suralis, S1-S2										
7	1		0			2	0,1			

SONUÇ: SENSORİ MOTOR POLİNÖROPATİ

21.12.2020

Şekil 1.4. EMG çekimi yapıldıktan sonra çıkan polinöropati hastasına ait veri

Summary EMG data

Motor CV

Test	Stimulation site	Lat., ms	Ampl., mV	Dur., ms	Area, mV×ms	Stim., mA	Stim., ms	Dist., mm	Time, ms	Vel., m/s
right, Abductor digiti minimi, Ulnaris, C8 T1										
4	wrist	2,44	10,4	5,72	24,9	18	0,2	80		
	elbow	5,88	9,74	5,96	24,1	27	0,2	180	3,44	52,3
right, Abductor pollicis brevis, Medianus, C8 T1										
2	wrist	2,68	7,21	7,4	26,9	15	0,2	80		
	elbow	6,68	5,59	7,32	21,5	62	0,2	210	4,0	52,5
right, Abductor hallucis, Tibialis, I4 L5 S1										
6	medial malleolus	3,4	17,8	6,96	46,3	36	0,5	70		
right, Extensor digitorum brevis, Peroneus, I4 L5 S1										
5	sole of the foot	3,28	6,19	6,56	21,6	88	0,5	70		

Sensory CV

Test	Site	Lat., ms	Ampl., μ V	Dur., ms	Area, nV×s	Stim., mA	Stim., ms	Dist., mm	Time, ms	Vel., m/s
right, n. Ulnaris V dig.										
3	wrist	2,24	22,1	3,8	17,9	17	0,1	120	2,24	53,6
right, n. Medianus										
1	wrist	1,92	25,2	2,92	19,9	14	0,1	140	1,92	72,9

SONUÇ. NORMAL İNCELEME

18.2.2021

Şekil 1.5. EMG çekimi yapıldıktan sonra çıkan normal hastaya ait veri

Bu çalışmada kullanılan EMG data veri seti, polinöropati hastalığı şüphesi ile gelen hastaların verileri ile oluşturulmuştur. EMG verisi oluşturulurken motor sinir ve duyu sinir olmak üzere iki ileti sinir grubuna ayrılmaktadır.

Tablo 1.1. EMG Alt ve Üst Ekstremitte Motor ve Duyu Sinir İleti Normal Hızları

		MOTOR İLETİ NORMAL DEĞERLERİ			DUYU İLETİ NORMAL DEĞERLERİ		
		LATANS (msn)	AMPLITUD (mV)	İLETİ HIZI (m/sn)	LATANS (msn)	AMPLITUD (mV)	İLETİ HIZI (m/sn)
ÜST EKSTREMİTE	MEDIANUS	$\leq 4,2$	≥ 4	≥ 50	$\leq 3,5$	≥ 15	≥ 40
	ULNARIS	$\leq 3,4$	≥ 4	≥ 50	$\leq 3,1$	≥ 15	≥ 40
	RADIALIS	$\leq 2,9$	≥ 2	≥ 50	$\leq 2,9$	≥ 10	≥ 40
ALT EKSTREMİTE	PERONEAL	$\leq 6,5$	≥ 2	≥ 40	$\leq 4,4$	≥ 6	≥ 40
	TIBIAL	$\leq 5,8$	≥ 4	≥ 40	$\leq 4,4$	≥ 5	≥ 40
	FEMORAL	≤ 6	$\geq 4,1$	≥ 40	$\leq 2,8$	≥ 10	≥ 40

Erken ve doğru teşhis ile en verimli tedavi uygulamalarına karar destek noktasında kullanılmaya başlanan bilgi teknolojileri, sağlık hizmeti sunucuları ve hastaları için hız, zaman, doğruluk ve kaliteli sağlık hizmeti sunmaya destek olmaktadır. Elektronik tıbbi kayıtlara (EMG verileri) dayanarak polinöropati riskini belirlemek için makine öğrenmesi yöntemleri uygulanan bu çalışmada polinöropati hastalığının teşhisinde kullanılan sınıflandırma algoritmaları sırasıyla Decision Tree (Karar Ağaçları), Naive Bayes, K-NN algoritması, Support Vector Machine (SVM) Algoritması ve Random Forest algoritmalarına girilen verilerin sınıflandırılması sağlanmıştır.

2. MAKİNE ÖĞRENMESİ İLE POLİNÖROPATİ TEŞHİSİ

2.1 Makine Öğrenmesi (Machine Learning)

Makine öğrenmesi, çağımızın nesne tanıma, görüntü işleme, yüz tanıma, ses tanıma, iris tanıma, sanal gerçeklik, artırılmış gerçeklik, pazarlama, müşteri hizmetleri, uydu görüntüleri, yer bilim ve sağlık gibi pek çok farklı alanda kullanılan yapay zekânın en geniş alt koludur.

Temelde makine öğrenmesi genel anlamıyla; öğrendiği bilgiyi saklayarak yani tıpkı insan ve hayvan beyninde olduğu gibi, sakladığı bilgileri deneyim kazanıp yeni durumlara uyarlayarak kullanması beklenir.

Makine öğrenmesi bir diğer deyişle yapay öğrenme; bilgisayarlara daha önceki deneyimlerinde tanıtılan bilgiyi özümseyerek yeni gelen bilgiler için tekrar programlamaya ihtiyaç duymayan, hali hazırda var olan verileri inceleyip sürekli kendi performansını iyileştiren ve bu deneyimlerle bilinmeyene dair tahminlerde bulunan önemli bir bilgisayar algoritmasıdır.

Makine öğrenmesi algoritması doğru bir şekilde tahmin yapabilmesi ve sonrasında doğru kararlar alabilmesi için iki adet veri parçasına ihtiyaç duymaktadır. Bunlardan ilki örnek veri yani test verileri, bir diğeri ise eğitim verileridir.

Turing makinelerin de insanlar gibi düşünebileceği ve yeni durumlar içi kararlar alabileceği fikrini ilk ortaya atan kişidir (1950). Çalışmalarında, makinelerin de tıpkı canlılar gibi düşünerek öğrenebileceğinden bahsetmiş ve yapay zeka alanında temel oluşturabilecek çıkarımlarda bulunmuştur. [12]

Mekanik öğrenme diğeri bir adıyla makine öğrenmesi, bilgisayarın elde ettiği verileri ve bunlardan edindiği deneyimleri kullanarak daha sonra karşılaştığında onu yeni bir duruma uyarlayıp karar verebilmesi ve problemlere yeni çözümler getirebilmesidir. [13]

Makine öğrenmesi (Machine Learning) süreci için bazı kurallar bulunmaktadır. Öğrenenin insan ya da makine olmasına bakılmaksızın öğrenme sürecinde benzerlikler mevcuttur. Ayrıca birbiri ile bağlantılı dört tane unsur bulunmaktadır. Bunlar:

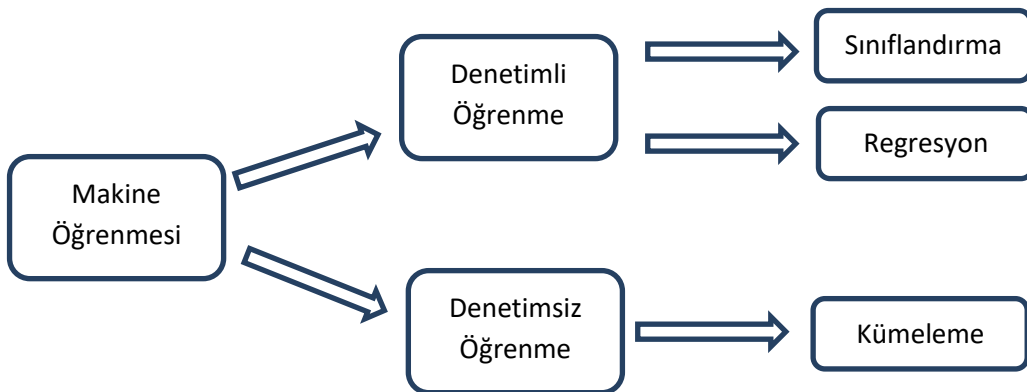
- Depolama: Görsel deneylerden, bellekten ve çağrışımlardan faydalanır.
- Soyutlama: Depoladığı data'ları genişletir ve kavramsal araçlara dönüştürür.
- Genelleştirme: Bir önceki aşamadaki kavramsal araçları kullanır ve genel ifadelere dönüştürür.
- Değerlendirme: Edinilen kavramların uygun olup olmadığını test etmek ve tahmin için ön bildirim sistemi kullanılır [13]

2.2 Makine öğrenme türleri

Makine öğrenmesi algoritmalarının da farklı öğrenme türleri bulunmaktadır. Öğrenmenin nasıl gerçekleştireceğini belirleyen sistem ve bu sistemde kullanılan öğrenme algoritması öğrenme türlerine göre değişkenlik göstermektedir [14]. Makine öğrenme türleri aşağıdaki gibi kategorize edilerek gösterilmiştir.

- Denetimli Makine Öğrenmesi (SML: Supervised Machine Learning)
- Denetimsiz Makine Öğrenme (UML: Unsupervised Machine Learning)
- Yarı-Denetimli öğrenme (Semi-Supervised Learning)
- Pekiştirmeli öğrenme (Reinforcement Learning)

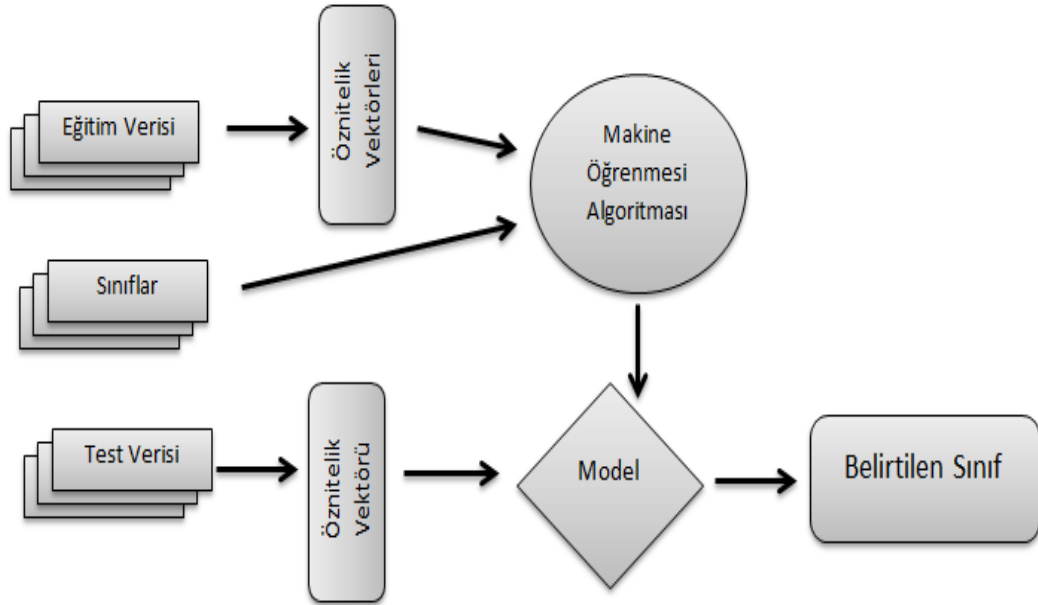
Bu makine öğrenme türlerinden en çok kullanılan denetimli ve denetimsiz öğrenme türleridir.



Şekil 2.1. Makine Öğrenme türleri

2.2.1. Denetimli Makine Öğrenmesi (SML: Supervised Machine Learning)

Denetimli öğrenme (gözetimli öğrenme); sisteme eğitim veri seti ile test veri setinin yüklenmesiyle başlar. Veri setinde örnek olarak etiketlenmiş girdi veri seti ile çıktı veri seti arasındaki kombinasyonları inceler ve bu süreçte, girdiler ve çıktılar arasındaki ilişkiyi tanımlayan bir fonksiyon ya da algoritma öğrenilir. Önceden bilinen çıktılar üstünde sınıflama gerçekleştirildikten sonra sonuçları henüz bilinmeyen veri kümelerinde olası sonuçlar tahmin edilmeye çalışılır. Denetimli öğrenme modeline ait süreç Şekil2.2.de gösterilmiştir. [15]

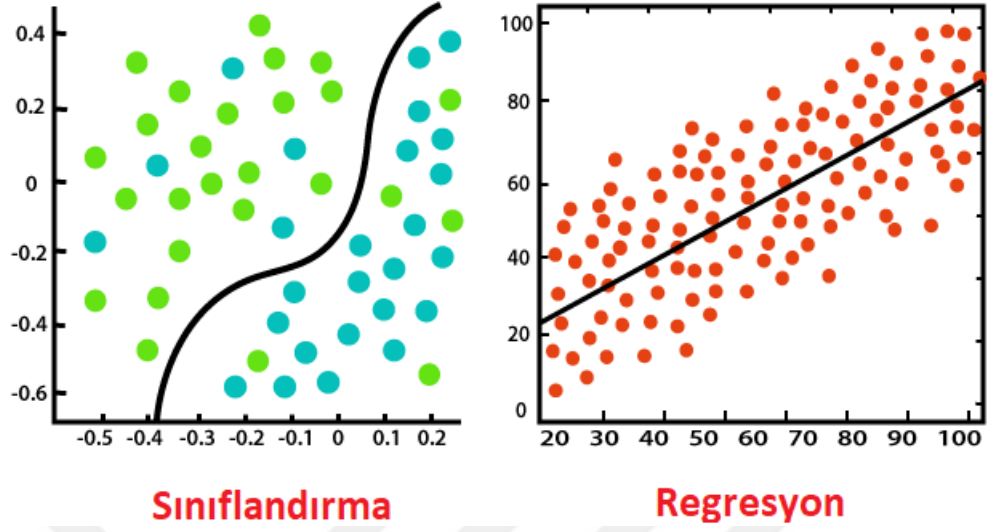


Şekil 2.2. Denetimli öğrenme modeli süreci

Başka bir deyişle gözetimli öğrenme; bilinen örnekleri yani girdi ve çıktı kombinasyonlarını kullanarak bir model oluşturup yeni girilen bir girdiye uygun olarak bir çıktı üreten yani çıktı değişkenini tahmin eden bir makine öğrenimi yaklaşımıdır.

Denetimli öğrenme yaklaşımı iki ana gruba ayrılmaktadır. Bunlar;

- Sınıflandırma
- Regresyon



Şekil 2.3. Sınıflandırma ve Regresyon Algoritması

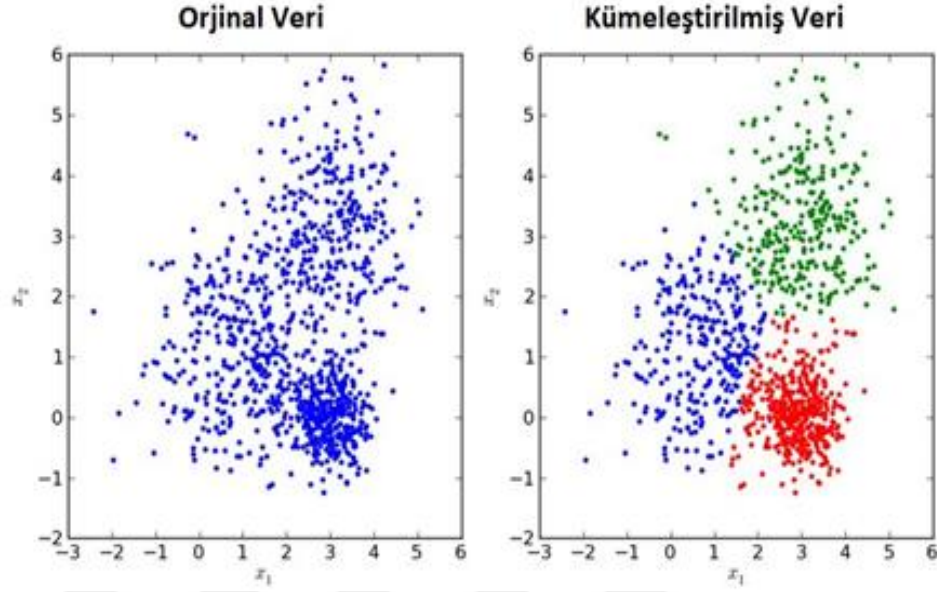
Günümüzde en çok problemlerin çözümünde kullanılan denetimli öğrenme algoritmaları Destek Vektör Makineleri (SVM), Karar ağaçları (DT), K-En Yakın Komşu Algoritması (KNN), Naive Bayes (NB) ve Random Forest olarak karşımıza çıkmaktadır. Bu algoritmalarından tez çalışması içerisinde kullanılanlar Yararlanılan Makine Öğrenmesi Algoritmaları kısmında detaylandırılacaktır.

2.2.2. Denetimsiz Makine Öğrenme (UML: Unsupervised Machine Learning)

Denetimsiz (gözetimsiz) makine öğrenme modeli gözlemlere bağlı bir makine öğrenmesi tekniğidir. Bu yöntem denetimli makine öğrenmeye kıyasla çıktı verilerini kullanmaz, daha fazla yanılma payı olan sonuçlar ortaya çıkmaktadır. Öğrenme işlemini yalnızca girdi verileri üzerinden sağlamaya çalışır.

Denetimsiz makine öğrenmede, sınıfları tanımlayan bir öğretici söz konusu değildir. Bu yüzden sistem kendi sınıfını bulmak durumundadır. Bu yöntemlere kendi kendine öğrenebilen modeller de denilmektedir. Denetimsiz öğrenmede çıktıları olan bir eğitim seti kullanılmamaktadır ve eğitim için sadece girdilerin olması yeterlidir. Ayrıca çokça kullanılan amaçlarından bazıları kümeleme, ilişkilendirme boyut indirgeme ve olasılık yoğunluk tahmini yöntemleri ile yapılmaktadır.

Denetimsiz Makine Öğrenmesi (Unsupervised Machine Learning)



Şekil 2.4. Denetimsiz öğrenme modeli

2.2.3. Yarı-Denetimli öğrenme (Semi-Supervised Learning)

Oluşturulacak öğrenme modelimizin içerisinde az sayıda etiketlenilmiş yani çıktıları belli olan veri ve buna karşılık çok daha fazlaca etiketlenilmemiş yani çıktıları belli olmayan veri varsa hem denetimli hem de denetimsiz öğrenme oldukça yetersiz kalabilir. Böyle zamanlarda en uygun yöntem elde bulunan az sayıdaki etiketlenilmiş verilerden yola çıkarak etiketlenilmemiş veriler hakkında, bilgi sahibi olmaya çalışmaktır. Bunu yapabilmek için “class” yöntemi kullanılmalı yani veriler sınıflandırılmalıdır. Bu şekilde yapılan öğrenmeye yarı denetimli öğrenme denilir. [14]

Yarı denetimli öğrenme modelinin denetimli öğrenme modelinden farkı etiketlenilmemiş verilerin fazla olmasıdır. Denetimli öğrenmede etiketlenilmiş veri sayısı fazla, tahmin edilmeye ihtiyaç duyulan veri sayısı azken yarı denetimlide tam tersi bir durum söz konusudur.

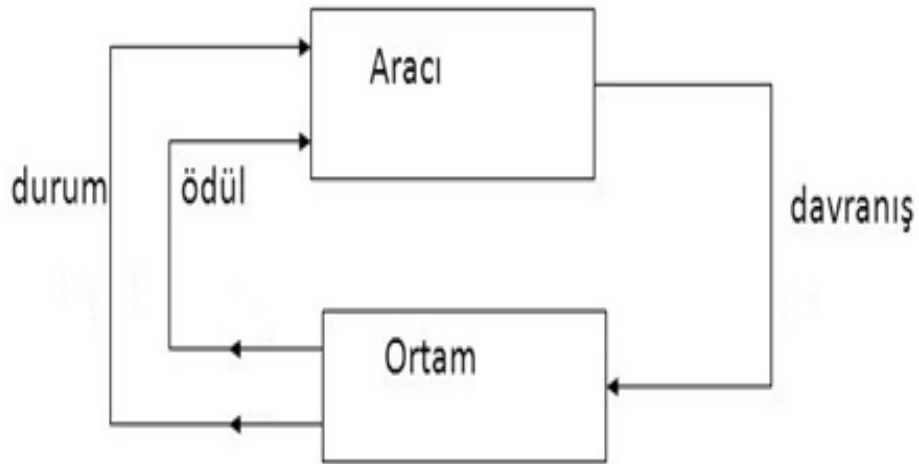


Şekil2.5. Yarı denetimli öğrenme modeli

2.2.4. Pekiştirerek öğrenme (Reinforcement Learning)

Pekiştirmeli öğrenme modelinde, herhangi bir ön bilgiye ihtiyaç duyulmadan, sadece öğrenme sürecinde yapılan durumların amaca ne kadar hizmet ettiğini gösteren bir önleme sistemi mevcuttur. Öğrenme sürecinde algoritmaların kullanıldığı sistemler tek bir olaya odaklandırılıyor bu da en yüksek değerli icraat olarak adlandırılıyor. İçerisinde bulunduğu sistem eğer en yüksek değerden daha az bir değer ya da hiçbir değer almadıysa bu durumu değerlendirmeye almamaktadır. Bu durum sonucunda buradaki öğrenme durumu sonsuza yakın olarak değerlendirilmektedir.

- Öğrenen, karar veren Aracı
- Etkileşim içinde olduğu Ortam



Şekil2.6. Pekiştirerek öğrenme modeli

2.3.Problem Çözümünde Uygulanan Makine Öğrenme Algoritmaları

Bu makalede kullanılan sadece denetimli öğrenmede sınıflandırma algoritmalarından yararlanıldığından dolayı bu başlık altında sadece denetimli öğrenme algoritmalarından bahsedilecektir.

2.3.1. Destek Vektör Makinesi Algoritması (Support Vector Machine)

Destek vektör makineleri (SVM), 1960'lı yılların sonunda Vladimir Vapnik ve Alexey Chervonenkis tarafından geliştirilmiş bir denetimli makine öğrenme algoritmasıdır. Bu öğrenme yöntemi kontrollü sınıflandırma algoritması olup istatistiksel öğrenme teorisine dayalı bir algoritmadır [12].

SVM, yüksek boyutlu verilerde sıkça kullanılan ve veri üzerinde eğitilerek uygulanabildiği için denetimli makine öğrenmesi kapsamındadır.

SVM, uygun değerde bir çizgi bulunmasını sağlayarak sınıfların ayrılması işlemini yapar. Çizilen çizgilerin sınıflara en uzak noktadan geçecek şekilde çizilmesi gerekir. İlk kez iki sınıflı doğrusal verilerin modellenmesi için kullanılmıştır [15]. Zamanla gelişen SVM, ikiden fazla sınıflı ve doğrusal olmayan verilerin modellenmesinde de kullanılmıştır. SVM'nin çalışma prensibi en uygun karar fonksiyonun tahmin edilmesine (hiper-düzlemin tanımlanmasına) dayanır [16].

Destek Vektör Makineleri, sınıflama yapma işlemini kareli optimizasyon işlemine çevirerek oluşturmaktadır. Böylece öğrenme sırasında yapılan işlemlerde nicelik azalmaktadır. Bu sayede diğer algoritmalara nazaran daha hızlı çözümler üretilebilmektedir. Bu becerisi sayesinde hacmi büyük veri setlerinde oldukça kullanışlıdır [17,18].

Bu tez çalışması içerisinde SVM algoritması sınıflandırma alanında ihtiyaç olduğu için kullanılmıştır. Bu algoritma, sınıflandırılacak iki grup arasında bir sınır çizerek bu grupları birbirinden ayırmaya yardımcı olmaktadır. Bu sınır, iki sınıfta elemanlarına en uzak noktadan çizilmektedir.

Doğrusal SVM sınıflandırıcısı iki veri grubunu ayırmak için optimum düzlemi bulmaya çalışır.

$$g(x) = \omega^T x + \omega_0 = 0 \quad (2.1)$$

Bir noktanın hiper düzleme uzaklığı olan w ağırlık vektörü hesaplanır.

$$z = \frac{|g(x)|}{\|w\|} \quad (2.2)$$

Düzlemin iki grubu ayırmak için iki grubun en yakın noktaları olan w_1 ve w_2 arasındaki uzaklığı hesaplanır.

$$\frac{1}{\|w_1\|} + \frac{1}{\|w_2\|} = \frac{2}{\|w\|} \quad (2.3)$$

$g(x)$ doğrusu için w_1 1 değerine ve w_2 ise -1 değerine eşittir.

$$w^T x + \omega_0 \geq 1, \forall x \in w_1$$

$$w^T x + \omega_0 \leq -1, \forall x \in w_2$$

Her bir x_i noktası için y_i (w_1 için 1 ve w_2 için -1) noktası hesaplanır. Denklem 2.4 minimalime edilerek Denklem 2.5 tatmin edilir.

$$J(w, w_0) \equiv \frac{1}{2} \|w\|^2 \quad (2.4)$$

$$y_i (w^T x_i + w_0) \geq 1, i = 1, 2, \dots, N \quad (2.5)$$

Denklem 2.4' ü minimalime etmek için Lagrange çarpanı kullanılır.

$$\frac{\partial}{\partial w} \mathcal{L}(w, w_0, \lambda) = 0 \quad (2.6)$$

$$\frac{\partial}{\partial w_0} \mathcal{L}(w, w_0, \lambda) = 0 \quad (2.7)$$

$$\lambda_i \geq 0, i = 1, 2, \dots, N \quad (2.8)$$

$$\lambda_i [y_i (w^T x_i + w_0) - 1] = 0, i = 1, 2, \dots, N \quad (2.9)$$

Lagrange çarpanı Denklem 2.10' da verilmiştir. Denklem 2.10'daki Lagrange çarpanını Denklem 2.6 ve Denklem 2.7 ile birleştirilince hiper düzlem için en iyi uzaklık Denklem 2.11 elde edilir.

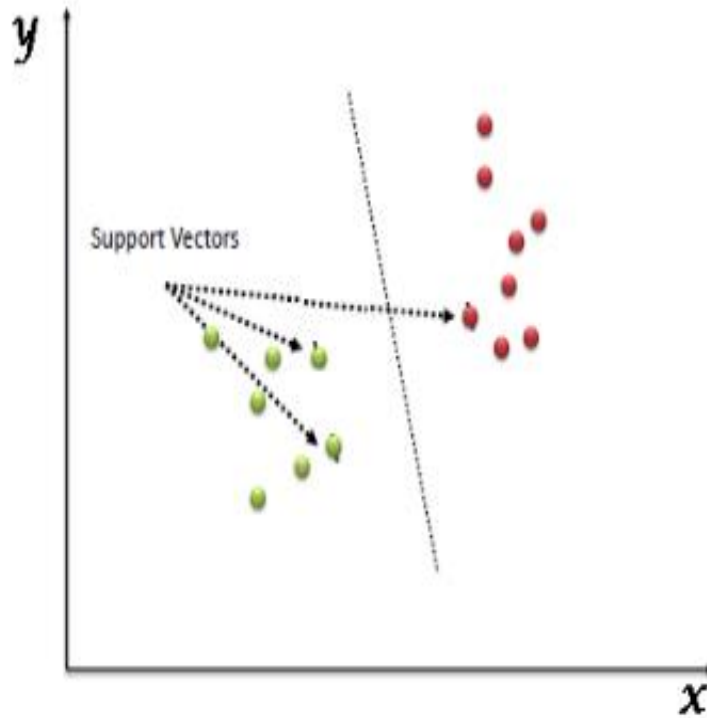
$$\mathcal{A}(w, w_0, \lambda) = \frac{1}{2}w^T w - \sum_{i=1}^N \lambda_i [y_i (w^T x_i + w_0) - 1] \quad (2.10)$$

$$w = \sum_{i=1}^N \lambda_i y_i x_i \quad (2.11)$$

İdeal hiper düzlem sınıflandırıcı olan destek vektör makinesi hesaplanmış olur.

SVM algoritmasının kullanılmasında avantajlar;

- Yüksek doğruluk oranı
- Karmaşık karar sınırları modelleyebilme
- Çok sayıda bağımsız değişken ile çalışabilme
- Diğer birçok yöntemle göre aşırı oturma sorununun az olması.



Şekil 2.7. Destek Vektör Makinesi Algoritması

2.3.2. Karar ağaçları

Karar ağaçları, kök düğümünden başlayarak karar düğümlerine doğru ilerleyen ve etiketlenmiş yapraklarda son bulan bir sınıflandırıcıdır. Karar ağaçları denetimli (gözetimli) makine öğrenmesi kapsamında yer almaktadır [18]. Belirli bir kayıt için bir tahmin kararı alınana kadar ağaç yapılarında alınan kararlar geçerlidir. Karar ağaçları, sınıflandırma ve regresyon sorunları için veriler üzerinde eğitilir. Karar ağaçları hızlı ve kesindir ve makine öğrenmesinde büyük bir favoridir [19,20].

İki sınıflı bir karar ağacı oluşturmak için her bir düğüm t , eğitim kümesi X 'in belli bir X_t alt kümesi ile ilişkilidir. Bir düğümü ayırmak, X_t alt kümesini, X_tN ve X_tY olarak ikiye ayırmaktır.

$$X_tY \cap X_tN = \emptyset$$

$$X_tY \cup X_tN = X_t$$

Bir ayırma sonucunda w_1 ve w_2 sınıfları X_tY alt kümesine ait sınıfları, w_3 ve w_4 ise X_tN alt kümesine ait sınıfları oluşturur. Düğümleri en iyi şekilde ayırmak için, bir ayırma kriteri belirlenir. Düğüm olması ve dallanma sonucu oluşan düğümlerin saf olmama oranının en düşük olmasına çalışılır. $P(w_i|t)$, X_t alt kümesindeki bir vektörün, t düğümündeki, ait olduğu w_i , $i = 1, 2, \dots, M$ sınıflarına ait olasılığıdır. Düğümün saf olmaması $I(t)$ olarak Denklem 2.12'de verilmiştir.

$$I(t) = -\sum_{i=1}^M P(w_i|t) \log_2 P(w_i |t) \quad (2.12)$$

Saf olmama oranındaki düşüşü bulmak için, X_tY düğümündeki olasılık N_{tY}/N_t ve X_tN düğümündeki olasılık N_{tN}/N_t olarak alınır. Saf olmamadaki düşüş Denklem 2.13' de verilmiştir.

$$\Delta I(t) = I(t) - \frac{N_{tY}I(t_Y)}{N_t} - \frac{N_{tN}I(t_N)}{N_t} \quad (2.13)$$

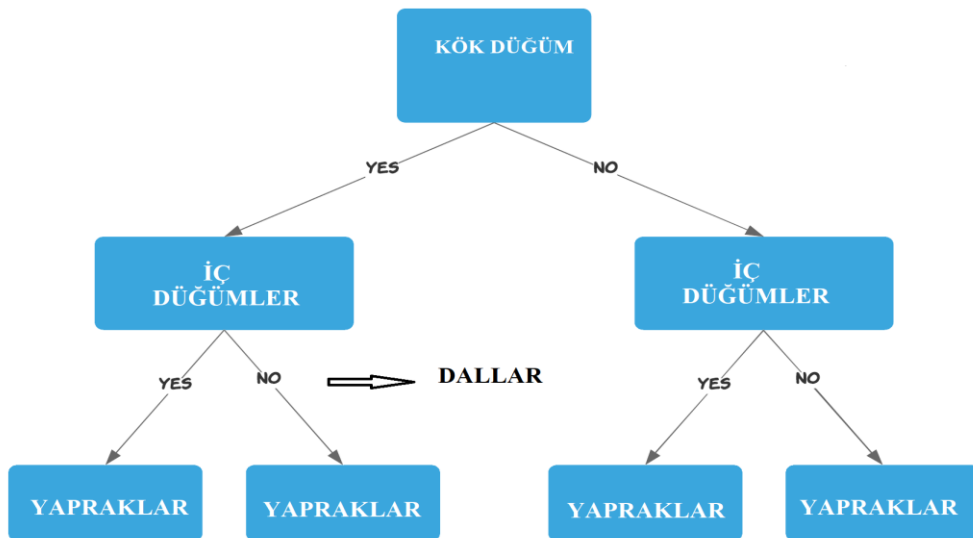
Bir düğüm oluşturulduktan sonra, düğümün hangi sınıfa ait olduğu Denklem 2.14’deki çoğunluk kuralı ile belirlenir.

$$j = \underset{i}{\operatorname{argmax}} P(w_i | t) \quad (2.14)$$

Karar ağaçlarının avantajları;

- Kolay uygulanabilmesi
- Veri tabanı sistemleri ile kolaylıkla birleştirilebilmesi
- Kolay yorumlanabilmesi
- Maliyetinin düşük olması
- Tüm değişkenlere (nitel, nicel, sürekli, kesikli) uygulanabilen algoritmalara sahip olması
- Güvenilir sonuçlar vermesi gibi sebeplerden dolayı en çok tercih edilen algoritmalardandır [21].

Algoritma kök, karar düğümleri, dallar ve yapraklardan oluşmaktadır. İşlemin başladığı yer kök kısmıdır. Karar düğümlerinde nitelik sınıma işlemleri gerçekleşir. Dallarda sınıma sonuçları görülebilir ve son olarak yapraklar ise sınıfları temsil eder [13]. Karar ağaçları tümevarım yöntemini kullanarak büyük veriyi küçük gruplara böler. Basit bir karar ağacı yapısı;



Şekil2.8.Karar Ağacı Yapısı

En popüler karar ağacı algoritmaları:

- Sınıflandırma ve Regresyon Ağacı (CART)
- Tekrarlı İkिलikçi Ağacı (ID3)
- C4.5 ve C5.0 (güçlü bir yaklaşımın farklı sürümleri)
- Ki-kare Otomatik Etkileşim Tespiti (CHAID)
- Karar Kökü
- M5
- Koşullu Karar Ağaçları

2.3.3. K-En Yakın Komşu Algoritması (KNN)

KNN algoritması, benzer verilerin birbirine yakın olduğunu varsayarak bu varsayım üzerinden ilerleyen bir denetimli makine öğrenimi algoritmasıdır. Bu algoritma, veri setine eklenecek bir verinin hali hazırda veri setinde bulunan diğer verilere olan uzaklığını hesaplamakta ve bu hesaplama sonucunda veriyi sınıflandırmaktadır [22].

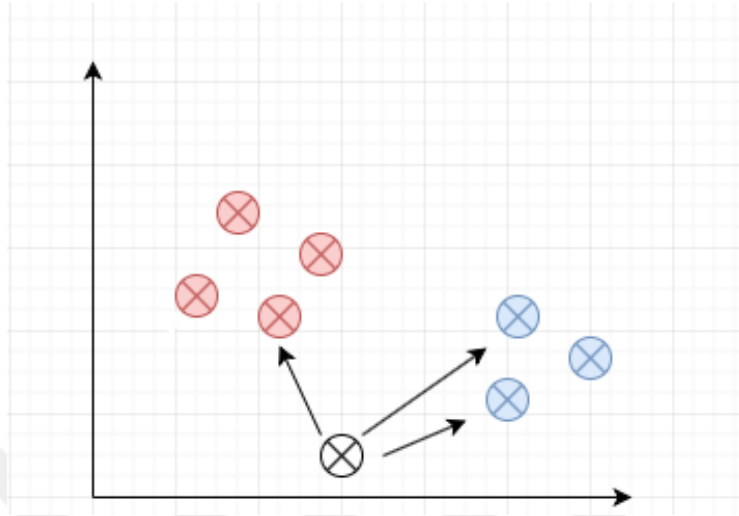
Sınıflandırılacak veri ile yakın komşuları olan veriler için, tek tek Öklid uzaklığı hesaplanır. İki nokta olan a ve b arasındaki Öklid uzaklığı Denklem 2.15’de verilmiştir. İki nokta olan a ve b noktaları x ve y koordinatları ile belirtilmiştir. Uzayda a noktası x_1, y_1 olarak b noktası ise x_2, y_2 olarak ifade edilmiştir. Veri Öklid uzunlukları en küçük olan k adet veri içinde en çok rastlanan sınıfa dahil edilir. k değeri eşitlik durumu ile karşılaşılmaması için tek sayı olarak seçilir [28].

$$\|a - b\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2.15)$$

Bu algoritmada izlenecek adımlar şu şekildedir;

- K parametresinin değeri belirlenir.
- Yeni verinin, mevcut verilere olan uzaklıkları tek tek hesaplanır,
- Hesaplanan uzaklıklar içerisinde en yakın k komşu ele alınır.
- Yeni veri bu komşulara göre etiketlenir

Yukarıda sıralanmış olan algoritma adımlarını aşağıdaki Şekil 2.9.görsel olarak ifade etmiştir.



Şekil2.9. KNN Algoritması

2.3.4. Naive Bayes Algoritması (NB)

Naive Bayes, verileri sınıflamak için Bayes teoremi kullanılarak oluşturulmuş olasılıklı bir makine öğrenimi algoritmasıdır. Sınıfın kısmi özelliklerinin varlığının veya yokluğunun, başka bir özelliğin varlığı (ya da yokluğu) ile ilişkili olup olmadığı ile ilgilenir [21]. Bayes teoremi bir değişken için geçerli olan olasılık dağılımları içinde koşullu olasılıklar ve marjinal olasılıklar arasındaki ilişkiyi göstermektedir. NB, veri kümesi içerisindeki bir eleman için, her durumun olasılığını hesaplamakta ve bu hesaplanan olasılıklar üzerinden değerlendirme yaparak en yüksek değerlikte olana göre sınıflandırma işlemini yapmaktadır [22].

Sınıfı belli olmayan $x = (x_1, x_2, \dots, x_d)$ d boyutlu veri örneğinin hangi sınıfa ait olduğunu bulmak için Bayes teoreminden yararlanır. Sınıfların $C = \{C_1, C_2, \dots, C_K\}$ veri kümesinde $P(C_k)$, C_k ($k=1, 2, \dots, K$) için öncü olasılıktır. Koşullu olasılık olan $P(x|C_k)$, x verisinin C_k sınıfında olduğunu kabul eder. Bayes teoremi Denklem 2.16' de verilmiştir. NB sınıflandırmasının formüle dökülmüş hali aşağıdaki gibi gösterilmiştir [32,34].

$$P(C_k/x) = \frac{P(x/C_k)*P(C_k)}{\sum_k P(x_{kk'})P(C_{k'})} \quad (2.16)$$

Yukarıdaki (2.16) denkleme göre;

$P(C|x)$: x durumu gerçekleştiği zaman C durumunun gerçekleşme ihtimalini göstermektedir.

$P(x|C)$: C durumu gerçekleştiği zaman x durumunun gerçekleşme ihtimalini göstermektedir.

$P(C)$: C durumunun gerçekleşme ihtimalini göstermektedir.

$P(x)$: x durumunun gerçekleşme ihtimalini göstermektedir.

Naïve Bayes algoritması bir sınıfa ait olan özneliğin değerinin başka bir öznelikten farklı kabul etmesinden dolayı koşullu olasılık Denklem 2.17' de verilmiştir.

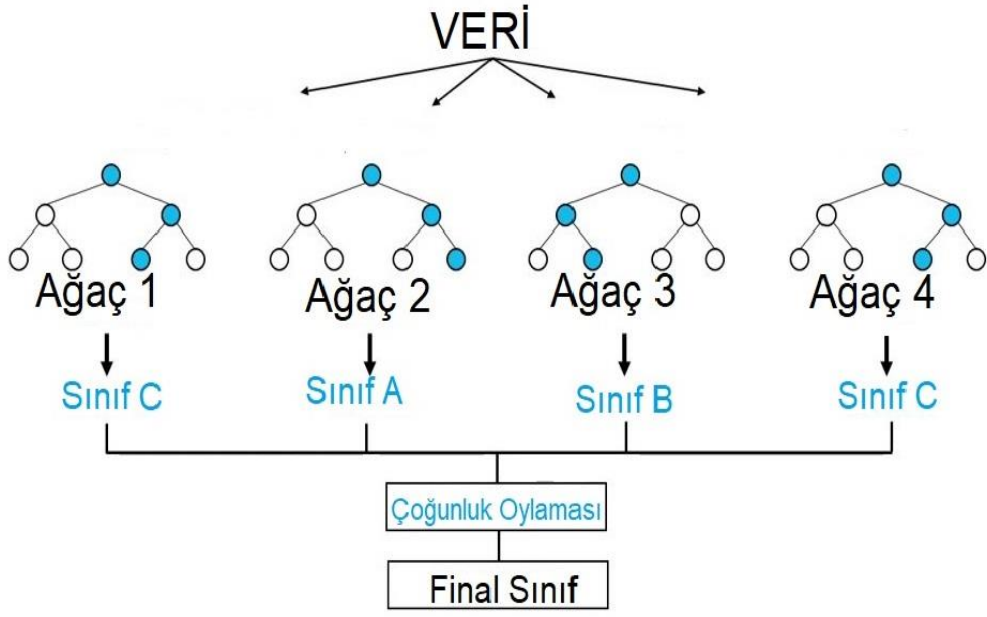
$$P(x|C_k) = \prod_{j=1}^d p(x^j | C_k) \quad (2.17)$$

Naive Bayes Sınıflandırmasını özetlemek gerekirse;

- Veri kümesi bir frekans tablosuna çevrilir.
- En olası sınıf sınıf tahmini olarak kabul edilecektir [24].
- Olasılık sınıflandırması için her değişkenin olasılığı belirlenir. Her sınıfın olasılığını hesaplarken, Naive Bayes denklemi kullanılır [25].

2.3.5.Rastsal Orman Algoritması (RF)

Rastsal Orman Algoritması, içerisinde birden fazla karar ağacını kullanarak sınıflandırma sonucunu iyileştirme hedefi güden bir algoritmadır. Algoritma içerisinde kullanılacak ağaç sayısı parametrikdir [26]. Bu parametre kapsamında oluşan karar ağaçları veri seti içerisinde Rastsal olarak seçilmiş veri kümeleridir. Rastsal olarak seçilen alt veri kümeleri üzerinde eğitim gerçekleşir ve her bir karar ağacı üzerinde tahmin etme yapılır. Yapılan bu tahminler sonucunda en yüksek başarı oranı olan karar ağacı ve tahmini sonuç olarak seçilmektedir. Aşağıdaki Şekil 2.10 üzerinde RF algoritmasının çalışma yöntemi gösterilmiştir

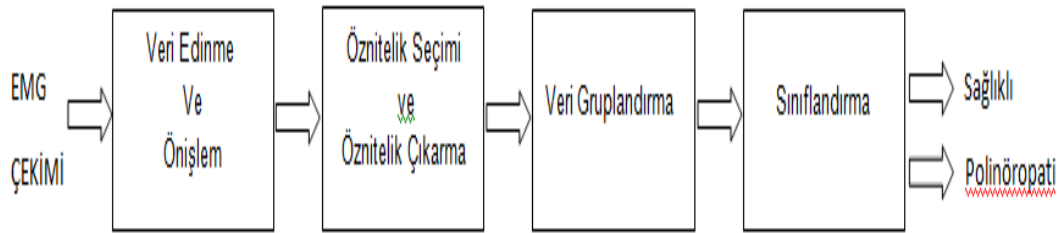


Şekil 2.10. Rastsal Orman Çalışma Yöntemi

2.4. EMG Verilerinin Sınıflandırılması

Bu çalışmada EMG verilerinden özellik çıkarılması, anlamlandırılması ve sınıflandırılması için hangi algoritmalarından faydalandığı açıklanacaktır. Daha güvenilir sonuçlar vermesi bakımından bilgisayar destekli karar verme mekanizmalarına ihtiyaç duyulduğu öngörülmüştür [25,26].

Bu amaçla, yapılan çalışmalarda destek vektör makineleri, yapay sinir ağları ve karar ağaçları gibi makine öğrenme yöntemleri kullanan bilgisayar destekli karar verme mekanizmaları geliştirilmesi düşünülmüştür. Çalışmaya ait akış şeması Şekil 2.11’de verilmiştir.



Şekil 2.11: EMG Verilerinin Sınıflandırılmasında Kullanılan Akış Şeması

➤ Veri Edinme ve Önışlem

Bu alıřmada kullanılan EMG verileri 72 sađlıklı insana ve 28 polinöropati hastasına aittir ve ayrıca her bir veri grubunda 22'şer tane alt veri bulunmaktadır.

➤ Öznitelik Seçimi ve Öznitelik ıkarma

EMG verisinden faydalı bilgileri ön plana ıkarmak için öznitelik seçimi yapılmış ve verinin içinden istenmeyen bölümleri, girişimleri kaldırmak adına EMG verilerinde özellik ıkarımı önemli bir aşamadır. EMG verisinin sınıflandırılmasıyla ilgili birçok alıřmaya bakıldığında gereksiz özellik vektörlerinin kullanıldığı görülmektedir. EMG verilerinin başarılı sınıflandırması, özellik vektörünün dođru seçilmesine bađlıdır. Veriye uygun özellik seçimi sınıflandırıcının yüksek dođruluđa yaklaşması açısından dođrudan ilgilidir. [29] Veri analizi aşamasında alıřmaya en yüksek deđer katan veri gruplarına yoğunlaşmıştır.

➤ Veri Gruplandırma

Veriler belli ölçütlere göre, veriler arasındaki aykırı verilerin işlemler sırasında bozulmalara yol açmasını engellemek amacıyla ayırıcı özelliklere göre gruplandırma yapılmıştır. Veriler 0 ile 1 arasında gruplandırılmıştır. Gruplandırma işlemi yöntem alanında Normalizasyon işlemi altında detaylı olarak anlatılmıştır.

➤ Sınıflandırma

Bu alıřmada EMG analizi için toplam 100 kişiye (72'u sađlıklı, 28 polinöropati hastası) ait toplam 2200 veri sınıflandırılmıştır. Özellik arz eden veriler Veri madenciliğinde yaygın olarak kullanılan Makine öğrenmesi algoritmalarına uygun olacak şekilde tekrar yapılandırılmıştır. Tüm modeller bir test veri seti kullanılarak deđerlendirildi. Sınıflandırma sonuçlarını deđerlendirmek için, makine öğrenimi modellerinde ortak ölçütler kullanıldı.

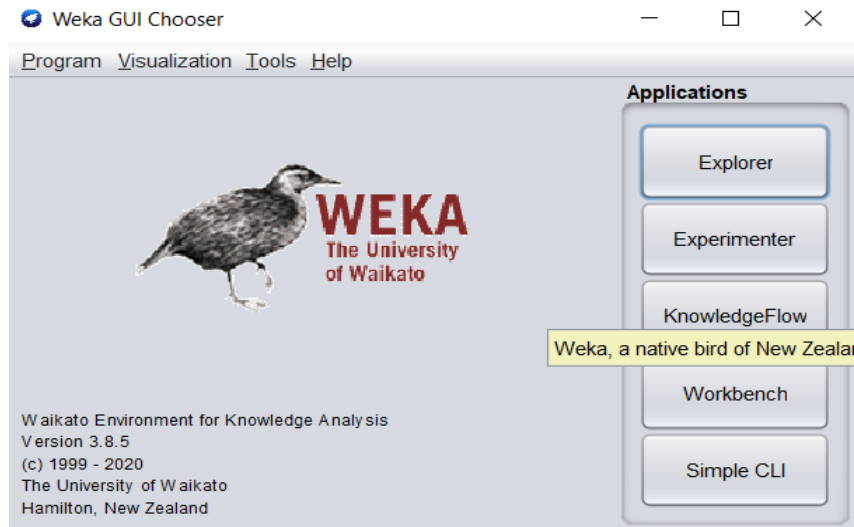
Makine Öğrenmesi Algoritmaları, Karar Ađaçları, Naive Bayes, K-NN Algoritması, Support Vector Machine (SVM) Algoritması ve Rastsal Orman sınıflandırma algoritmalarının başarımlarını testleri, WEKA yazılımı aracılığıyla karşılaştırılarak en iyi sonucu veren algoritma tespit edilmeye alışılmıştır.

3. DENEYSEL ÇALIŞMALAR

3.1 Problem Çözümünde Kullanılan Program

Veri Madenciliği uygulamaları için birçok program geliştirilmiştir. Bu programlara örnek olarak Rapid Miner (YALE), KoNstanz Information MinEr (KNIME), Knowledge Extractionbased on Evolutionary Learning (KEEL), Orange, Nefclass-J ve Waikato Environment for Knowledge Analysis (WEKA), gibi açık kaynak programlar kullanılmaktadır. Bu bölümde sadece bu çalışmamda yer alan veri madenciliği programlarından Waikato Environment for Knowledge Analysis (WEKA) programı anlatılmıştır [27,28].

WEKA “Waikato Environment for Knowledge Analysis: WEKA, Yeni Zelanda’da bulunan Waikato Üniversitesi’nde 1993 yılında C diliyle yazılmaya başlanan bir veri madenciliği yazılımıdır. 1996’da ilk 29 resmi sürümü yayınlanmıştır. 1997 yılında ise geliştirilerek Java diliyle güncellenmiştir. Sonraki sürümleri de Java dili ile geliştirilmeye devam etmektedir. Açık kaynak kodlu (open source) bir yazılımdır [29].



Şekil 3.1. WEKA aracının arayüzü

Programın veri dosya türü olan “.arff” (Attribute Relationship File Format) WEKA’ya özel olarak tasarlanmış ve bu sadece weka üzerinde çalışır. Bunun dışında metin tabanlı csv, c45, libsvm, svmlight, Xarff formatlarını da weka programı desteklemektedir. Ayrıca “jdbc” (Java Derby Client) sürücüsü bulunan veri tabanlarına doğrudan bağlantı yapabilmektedir. İnternet üzerinden http protokolünü kullanarak bu formatlara uygun dosyaları okuyabilme yeteneğine sahip olması da bir diğer özelliğidir. WEKA, makine öğrenmesi yazılımı olup temel olarak sınıflandırma(classification), kümeleme(clustering), denetleme, birliktelik analizi (association analysis), veri ön işleme(pre-processing) gibi temel veri madenciliği işlemlerini yapılabilmektedir. Bu anlamda veri madenciliği çalışmaları için oldukça kullanışlıdır.

WEKA beş farklı modülden oluşmaktadır. Bunlar:

- 1) Explorer: Ana işlemlerin sürdürüldüğü ara yüz alanıdır. Bu alanda veri kümesi görüntülenebilir, veriye ön işleme yapılabilir, sınıflandırma, kümeleme ve birliktelik işlemleri gerçekleştirilebilir.
- 2) Experimenter: Deneilerin gerçekleştirildiği ve öğrenme şemaları arasında istatistiksel testlerin yürütüldüğü ortamdır.
- 3) Knowledge Flow: Explorer ortamında var olan özelliklerin sürükle bırak tekniğiyle gerçekleştirilebildiği bölümdür. Explorerdan farklı olarak veri seti üzerinde zincirleme işlemler yapılan ve veri seti üzerinde modüler tasarım yapılabilen bölümdür.
- 4) Workbench: WEKA’nın en gelişmiş makine öğrenim koleksiyonudur. Kullanıcıların yeni veri setleri üzerinde mevcut makine öğrenme yöntemlerini hızlı bir şekilde uygulayabilmesini mümkün hale getirmek için tasarlanmıştır. Tüm standart veri madenciliği yöntemlerini barındırmaktadır. Birçok veri görselleştirme olanağı ve veri ön işleme araçları sunmaktadır.
- 5) Simple CLI (Command LineInterface): WEKA yazılımı komut satır ara yüzü ile kullanıcıların kendi algoritmalarını yazmalarına ve eklentiler yapmalarına olanak tanımaktadır.

3.2. Uygulama Veri Seti: EMG Veri Seti Analiz Bilgisi

Bu çalışmada kullanılan EMG data veri seti, Devlet Hastanesine polinöropati hastalığı şüphesi ile gelen gerçek hastaların verileri ile oluşturulmuştur. Bu veri seti 100 tane hastaya çekilen EMG verilerinden oluşmaktadır. Bu hastalardan 28 tanesi polinöropati hastası olup 72 tanesi normal hastalardan oluşmaktadır. Veri kümesinde toplam 21 tane alt özellik bulunmaktadır. 22. özellik ise hastalık durumunun olup olmadığının gösterildiği ve tıbbi uzmanlar tarafından oluşturulan alandır. Bu özellikler ve sahip oldukları veri tipleri Tablo 3.1’ de detaylı olarak verilmiştir.

Tablo 3.1. EMG veri setinin özellik bilgileri

Özellikler	Açıklaması	Veri Tipi
M.U.LAT	Motor sinir Üst Ekstremitte Ulnaris Latans Değeri	Numeric
M.U.AMP	Motor sinir Üst Ekstremitte Ulnaris Amplitud Değeri	Numeric
M.U.VEL	Motor sinir Üst Ekstremitte Ulnaris İleti Hızının Değeri	Numeric
M.M.LAT	Motor sinir Üst Ekstremitte Medianus Latans Değeri	Numeric
M.M.AMP	Motor sinir Üst Ekstremitte Medianus Amplitud Değeri	Numeric
M.M.VEL	Motor sinir Üst Ekstremitte Medianus İleti Hızının Değeri	Numeric
M.T.LAT	Motor sinir Alt Ekstremitte Tibial Latans Değeri	Numeric
M.T.AMP	Motor sinir Alt Ekstremitte Tibial Amplitud Değeri	Numeric
M.T.VEL	Motor sinir Alt Ekstremitte Tibial İleti Hızının Değeri	Numeric
M.P.LAT	Motor sinir Alt Ekstremitte Peroneal Latans Değeri	Numeric
M.P.AMP	Motor sinir Alt Ekstremitte Peroneal Amplitud Değeri	Numeric
M.P.VEL	Motor sinir Alt Ekstremitte Peroneal İleti Hızının Değeri	Numeric
D.U.LAT	Duyu sinir Üst Ekstremitte Ulnaris Latans Değeri	Numeric
D.U.AMP	Duyu sinir Üst Ekstremitte Ulnaris Amplitud Değeri	Numeric
D.U.VEL	Duyu sinir Üst Ekstremitte Ulnaris İleti Hızının Değeri	Numeric
D.M.LAT	Duyu sinir Üst Ekstremitte Medianus Latans Değeri	Numeric
D.M.AMP	Duyu sinir Üst Ekstremitte Medianus Amplitud Değeri	Numeric
D.M.VEL	Duyu sinir Üst Ekstremitte Medianus İleti Hızının Değeri	Numeric
D.P.LAT	Duyu sinir Alt Ekstremitte Peroneal Latans Değeri	Numeric
D.P.AMP	Duyu sinir Alt Ekstremitte Peroneal Amplitud Değeri	Numeric
D.P.VEL	Duyu sinir Alt Ekstremitte Peroneal İleti Hızının Değeri	Numeric
CLASS	Sınıf Bilgisi	Nominal

Polinöropati Hastalığına ait çekilen EMG hasta veri setindeki alanların ve bu alanlara ait bazı kayıtlar Şekil 3.1’ de verilmiştir.

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	CLASS
M.U.LAT	M.U.AMP	M.U.VEL	M.M.LAT	M.M.AMP	M.M.VEL	M.T.LAT	M.T.AMP	M.T.VEL	M.P.LAT	M.P.AMP	M.P.VEL	D.U.LAT	D.U.AMP	D.U.VEL	D.M.LAT	D.M.AMP	D.M.VEL	D.P.LAT	D.P.AMP	D.P.VEL	
5,88	9,74	52,3	6,68	5,59	52,5	3,4	5,59	53	3,28	6,19	58	2,24	22,1	53,6	1,92	25,2	72,9				NORMAL
4,68	5,74	77,6	6,16	7,77	67,3							2,17	15,7	55,3	2,08	37,7	67,3				NORMAL
						5,78	7,34	64	11	2,98	39,5										NORMAL
						14,4	4,78	40,7	11	3,12	43,4										NORMAL
						4,14	4,61		10,2	4,99	42,1										NORMAL
5,16	10,2	53,5	6,04	9,85	53							1,68	25,9	71,4	1,8	25,4	77,8				NORMAL
6,77	7,85	57,5	7,51	6,61	50,6	4,64	15,1	55	14,4	4,66	30,2	2,33	64,5	51,5	1,68	24,5	83,3				NORMAL
6,32	8,25	45,9	6,36	7,34	42,9				7,58	7,97	52,8	1,8	11,9	66,7	2,68	19,1	52,2				POLINOROPATI
5,48	9,4	52,3				9,72	5,94	53	6,64	8,51	51,5	1,84	27,1	65,2	1,72	22,8	81,4				NORMAL
									8,26	11,6	52,3							3,28	17,4	42,7	NORMAL
						3,04	4,27	38	3,96	0,265	30										POLINOROPATI
7,03	7,54	50,1	11	3,93	45,3							2,33	29,6	51,5	6,3	13,4	52,2				NORMAL
6,02	8,68	52,5	6,63	9,13	50,6	12	9,74	58,3	11	3,25	40,1	2,24	20,5	53,6	2,12	16	66				NORMAL
			12,8	13,3	52,9	8,68	3,56	48,7													NORMAL
6,2	9,02	53,8				12,4	9,61	44,1	14,4	1,54	41,7										POLINOROPATI
						3,39	8,09	42	11,8	4,32	37,5										NORMAL
5,36	8,5	54,2				10,6	6,19	62,1	7,83	5,24	52,2	1,8	17,4	66,7				2,52	16,1	55,6	NORMAL
						12,5	6,79	45	9,95	1,06	42										POLINOROPATI
5,4	10,4	53,6				9,6	11,2	61,6	9,08	4,3	48,1										NORMAL
10,6	3,21	28,7	10,4	4,05	32										2,96	18,7	47,2				POLINOROPATI
						10,8	5,82	63,2	6,8	1,99	82,4							1,84	4,4	65,2	NORMAL
						2,76	7,7	68,4	9	2,48	51,7							1,56	9,4	76,9	NORMAL
5,6	6,55	50,6	7,41	8,81	51,3	10,9	4,11	57,1	11,8	1,77	37,3	2,12	20	56,7	2,2	16,8	63,6				NORMAL
6,45	7,8	50,3	7,66	4,36	50,7	13,6	3,18	40,8	13,8	1,18	29,9	2,04	30,4	58,8	2,56	28,9	54,7				POLINOROPATI
5,4	11,1	53,6	8,26	5,43	43,3	3,36	1,06	45	9,64	0,411	47,5	1,8	23,8	66,7	2,96	15,5	47,3				NORMAL
6,46	10,7	50,9	7,94	7,63	52,3	14,6	5,28	38,6	12	1,51	36,8	2,22	24,9	54	2,75	26,1	50,9				POLINOROPATI
						9,52	7,52	64,1	8,56	5,71	66,4										NORMAL
6	9,06	47,9	6,76	4,86	53							2,28	21,6	52,6	2,12	28,2	66				NORMAL
10,6	8,65	46,1	9,24	6,51	54,9							2,8	44,7	52,8	4,07	16,6	34,4				NORMAL
8,68	5,16	33,6	8,92	5,73	44,1	17,6	2,07	36,1	17,4	1,55	30,1	2,22	6,8	54	2,91	8,7	48,1				POLINOROPATI
						12,4	10,4	62,9	9,52	3,46	53,2										NORMAL
5,61	7,32	50,9	7,4	8,86	50,9							1,92	18,9	62,5	2,32	16,7	60,3				NORMAL
7,49	7,63	50,3	8,48	10,6	51,2	13,6	9,97	40,8	12,6	3,74	43,8	2,52	30,5	51,6	2,36	35,8	59,3	2,28	16,5	52,6	NORMAL

Şekil 3.2: Kullanılan EMG veri setinden örnek kayıtlar

3.3. Yöntem

Bu tez çalışmasında yapılan işlemler sıralı bir şekilde gerçekleşen adımlardan meydana gelmektedir. İlk olarak verilerin elde edilmesi sağlanmıştır. Bu çalışma boyunca kullanılan verilerin elde edilmesi ve bunların detayları bir önceki bölüm olan “Veri Seti” bölümünde anlatılmıştır.

Veri seti bulunduktan sonraki işlem, veri setini makine öğrenmesi algoritmalarının çalışabileceği formata dönüştürmek olmuştur. Veri setinin hazırlanması aşaması veriler normalizasyon işlemine alınmıştır. Burada normalizasyon yapılmasının amacı parametreler arasındaki genel bütünlüğün sağlanmasıdır.

Normalizasyon işlemi sonrasında, tez çalışması için kullanılacak olan makine öğrenimi algoritmalarının uygulamaları yapılmıştır. Makine öğrenimi uygulamalarının başarılı bir şekilde bitmesi sonrasında, doğruluk, kesinlik ve duyarlılık puanları not edilmiştir. Not edilen değerler En son aşamada birbirleri ile karşılaştırılarak başarı oranları gözlemlenmiştir.

3.3.1. Normalizasyon İşlemi

Normalizasyon işlemi, veriler arasındaki aykırı verilerin işlemler sırasında bozulmalara yol açmasını engellemek amacıyla yapılan bir işlemdir. Normalizasyon işleminin yapılabilmesi için farklı yollar bulunmaktadır. Bu çalışma içerisinde bu yollardan min-max normalizasyonu yöntemi kullanılmıştır. Min-max normalizasyonu yöntemi kullanılan en yaygın normalizasyon yöntemlerinden biridir. [30,31]

Bu normalizasyon yöntemi ile en düşük değer 0 en yüksek değer 1 olacak şekilde bütün veriler bu aralığa yerleştirilmekte, sonrasındaki işlemler hesaplamalarına bu normalize edilmiş veriler üzerinden devam etmektedirler.

Min-max normalizasyonunun denklemi aşağıdaki gibidir;

$$X_y = (X_0 - X_{\min}) / (X_{\max} - X_{\min}) \quad (3.1)$$

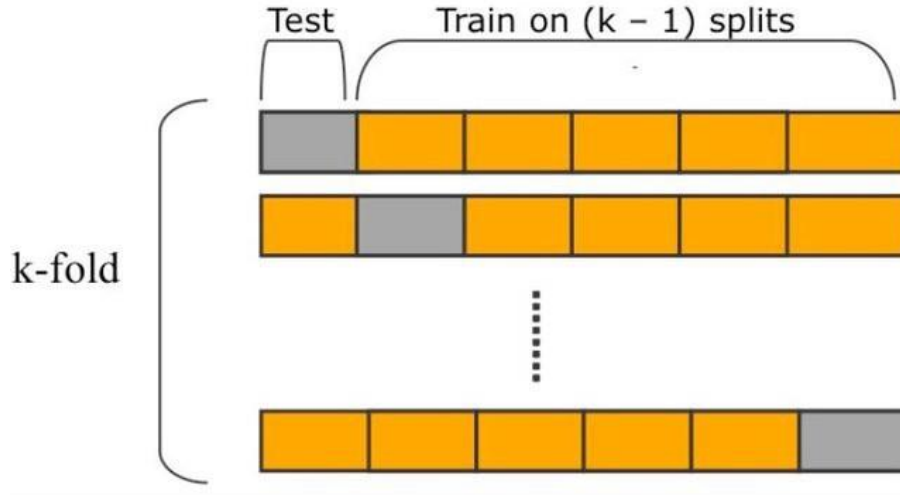
Formül içerisinde gösterilen ' x_0 ' değeri veri seti içerisindeki işleme alınacak her bir özelliğin var olan değerini temsil etmektedir. ' x_{min} ' değeri, veri seti içerisinde ilgili özellik ile alakalı en küçük değere sahip olan verinin değerini temsil ederken, ' x_{max} ' değeri, veri seti içerisinde ilgili özellik ile alakalı en büyük değere sahip olan veriyi temsil etmektedir.

Formülde yapılan işlem, mevcuttaki özellik değerinden minimum özellik değeri çıkartılıp, maksimum değer ile minimum değer farkına bölünmesi gerektiğini göstermektedir. Bu işlem sonucunda çıkan değer, normalize edilmiş olan yeni özellik değeri olarak kaydedilmektedir.

3.3.2. Çapraz Doğrulama Modeli (Cross Validation)

Bu çalışma içerisinde veri setini eğitim ve test verilerine ayırmak için Çapraz doğrulama tekniği kullanılmıştır. Bunun sebebi başka bir alternatif olan ayırma yönteminin büyük sayıda veri içeren veri setleri için kullanışlı olması ve test ve eğitim oranlarının manuel olarak belirlenmesidir. Bu manuel belirlemelerde modelin eğitim ve testlerinde bazı sapmalar meydana gelebilmektedir. Bu çalışmadaki veri setinin büyüklüğü göz önüne alındığında K-Katlamalı Çapraz Doğrulama yöntemi en doğru yöntem olarak karşımıza çıkmaktadır. K-Katlamalı Çapraz doğrulama bir veri seti içerisindeki verilerin rastgele bir şekilde K adet eşit parçaya bölünerek, bölünmüş olan bu gruplar üzerinden işlemlere devam edilmesidir. Burada bölünen gruplardan bir tanesi test için kullanılırken geriye kalan K-1 tanesi ise eğitim için kullanılmaktadır. Her bir grup için bu durum kaydırılarak tekrardan yapılır ve her bir grup için farklı sonuçlar elde edilerek, en son aşamada her grubun sonuçlarının aritmetik ortalaması alınarak bir başarı oranı ortaya koyar [35-37].

Makine öğrenimi uygulamalarının başarılı bir şekilde bitmesi sonrasında, K-Katlamalı Çapraz Doğrulama yöntemi kullanılarak, tez içerisinde kullanılacak olan makine öğrenmesi algoritmaları ile birlikte eğitime alınmıştır. Burada kullanılan 'K' dinamik değeri 10 olarak belirlenmiştir



Şekil 3.3 K-Katlamalı Çapraz Doğrulama modeli

3.3.3. Karışıklık Matrisi (Confusion Matrix)

Karışıklık matrisi, makine öğrenimi sınıflandırmaları için kullanılan bir performans ölçüm yöntemidir. Karışıklık matrisi Şekil 3.3 üzerinde görülebileceği gibi tahmin değerlerinin ve gerçek değerlerin dört farklı kombinasyonunu içeren bir tablo olarak gösterilmektedir.

		Öngörülen Sınıf	
		Sınıf=1	Sınıf=0
Doğru Sınıf	Sınıf=1	a	b
	Sınıf=0	c	d

Şekil 3.4 - Karışıklık Matrisi

a: TP(True Pozitif)

b: FN(False Negatif)

c: FP(False Pozitif)

d: TN(True Negatif)

Gerçek Pozitif (TP): Algoritma sonucunda alınan değer, orijinal değer ile aynıdır. Sınıflandırma işlemi doğru bir şekilde yapılmıştır.

Yanlış Negatif (FN): Algoritma sonucunda alınan değer, orijinal değerden farklıdır. Sınıflandırma işlemi yanlış bir şekilde yapılmıştır.

Yanlış Pozitif (FP): Algoritma sonucunda alınan pozitif değer, orijinal değerden farklıdır. Burada yanlış bir sınıflandırma işlemi yapılmıştır.

Gerçek Negatif (TN): Algoritma sonucunda elde edilen negatif değer, orijinal değer ile aynıdır. Sınıflandırma işlemi doğru bir şekilde yapılmıştır.

Karışıklık matrisi üzerinden elde edilen sonuçlar ile kesinlik, duyarlılık, doğruluk ve puanı hesaplamaları yapılabilmektedir. Bu değerlerin nasıl hesaplandıkları, aşağıdaki formüller ile gösterilmiştir.

Doğruluk: Doğruluk değeri, toplamda doğru olarak tahmin edilen verilerin, veri seti üzerindeki toplam veri sayısına oranı ile bulunmaktadır.

$$\text{Doğruluk} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (3.2)$$

Kesinlik: Kesinlik değeri, pozitif olarak tahmin edilen değerlerin gerçekten ne kadarının pozitif olduğunun oranını gösteren metriktir.

$$\text{Kesinlik} = \frac{TP}{(TP + FP)} \quad (3.3)$$

Duyarlılık: Duyarlılık değeri, gerçekte pozitif olan değerlerin tahminler sonucunda ne kadarının doğru bir şekilde pozitif olarak tahmin edildiğinin oranını gösteren metriktir.

$$\text{Duyarlılık} = \frac{TP}{(TP + FN)} \quad (3.4)$$

F1-Ölçüm: F1-Ölçüm değeri, duyarlılık ve kesinlik metriklerinin harmonik ortalamasını gösteren metriktir. Bu metrik dengesiz veri setleri için kullanışlıdır.

$$F1\text{-Ölçüm} = 2 * \frac{(\text{Kesinlik} * \text{Duyarlılık})}{(\text{Kesinlik} + \text{Duyarlılık})} \quad (3.5)$$

3.4. Araştırma Bulguları

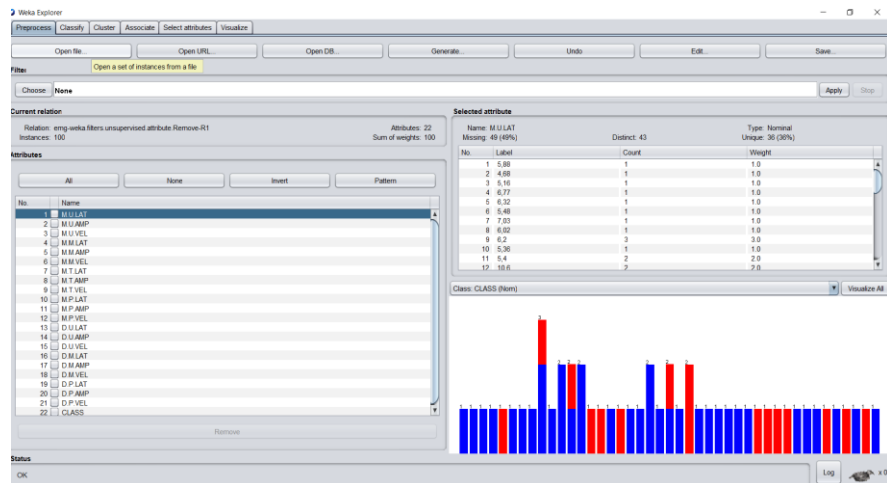
Bu bölümde J48, Random Forest, Naive Bayes, IBK algoritması ve SMO algoritmaları hakkında sırasıyla bilgi verilmiş ve bu algoritmalar Polinöropati hastalarına ait EMG Veri Seti kullanılarak elde edilen sonuçlar değerlendirilmiştir.

3.4.1. J48 algoritması

J48 algoritması, bir tür karar ağacı algoritmasıdır ve veriyi yukarıdan aşağıya doğru bir yapı ile sınıflandırır. [33]

John Ross Quinlan'ın geliştirdiği ve oldukça popüler olan C4.5 algoritması temeline dayanan bir karar ağacı algoritmasıdır. Karar ağaçları veri yapılarını anlamlandırmakta oldukça güçlü ve hızlı yollar sunmaktadır. Makine öğrenmede bilgiyi temsil etmek için kullanılan klasik bir yöntemdir. [39]

WEKA üzerinde uygulama yapılmak için kullanılacak olan EMG veri seti, emg.arff dosyası adıyla kaydedilmiştir. Bu dosyayı WEKA analiz aracında açmak için WEKA->Explorer->Preprocess->Open File->data>emg.arff adımları izlenir.



Şekil 3.5. WEKA preprocess ekranı

Şekil 3.4 'de görüldüğü üzere emg.arff dosyası içerisinde 21 farklı özellik (attributes) ve bu özelliklere bağlı olan polinöropati hastası olup olmama durumunu ifade eden seçici sınıf bilgisi (selector field) bulunmaktadır. Veri seti üzerinde karar ağacı sınıflandırma algoritmasını uygulamak için Classify->choose->trees->J48 algoritması seçilir ve “start” butonuna tıklanarak çalıştırılır.

Karışıklık matrisi, algoritmanın gerçekleştirdiği tahminlerin doğruluğu hakkında gerekli bilgiyi veren bir ölçüm aracı olarak kullanılmaktadır. Şekil 3.3'de ifade edilen karışıklık matrisinde sütunlar gerçek değerleri, satırlar ise tahmin edilen değerleri temsil etmektedir. Bu tez çalışmasında EMG verisine uygulanan Karar Ağacı algoritmalarından J48 algoritmasında elde edilen karışıklık matrisi Şekil 3.5 te verilmiştir.

Confusion Matrix	
20	0
6	0

Şekil 3.6. J48 Algoritması sonucundaki Karışıklık Matrisi

Yukarıdaki karışıklık matrisine göre değerlendirildiği zaman aşağıdaki hesaplamalar yapılarak doğruluk, kesinlik, duyarlılık, F1- ölçüm değerleri hesaplanmıştır.

$$\text{Doğruluk} = \frac{(TP + TN)}{(TP + FP + TN + FN)} = \frac{20+0}{20+0+6+0} = 0,769231$$

$$\text{Kesinlik} = \frac{TP}{(TP + FP)} = \frac{20}{20+6} = 0,769231$$

$$\text{Duyarlılık} = \frac{TP}{(TP + FN)} = \frac{20}{20+0} = 1$$

$$\text{F1-Ölçüm} = 2 * \frac{(\text{Kesinlik} * \text{Duyarlılık})}{(\text{Kesinlik} + \text{Duyarlılık})} = 2 * \frac{(0,769231 * 1)}{(0,769231 + 1)} = 0,8636$$

J48 algoritması çalıştırıldıktan sonra sınıflandırıcı sonuçları “Classifier Output” penceresinde gösterilmektedir. Bu algoritma örneklerinden daha önce anlattığımız Karışıklık Matrisi hesaplamaları aşağıdaki Şekil 3.6’de gösterilmiştir.

```

Classifier output

Size of the tree :      1

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      20          76.9231 %
Kappa statistic                    0
Mean absolute error                0.4379
Root mean squared error            0.4485
Relative absolute error            98.9967 %
Root relative squared error        99.3591 %
Total Number of Instances         26

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          1,000   1,000   0,769     1,000   0,870     ?       0,500    0,769    NORMAL
          0,000   0,000   ?         0,000   ?         ?       0,500    0,231    POLINOROPATI
Weighted Avg.   0,769   0,769   ?         0,769   ?         ?       0,500    0,645

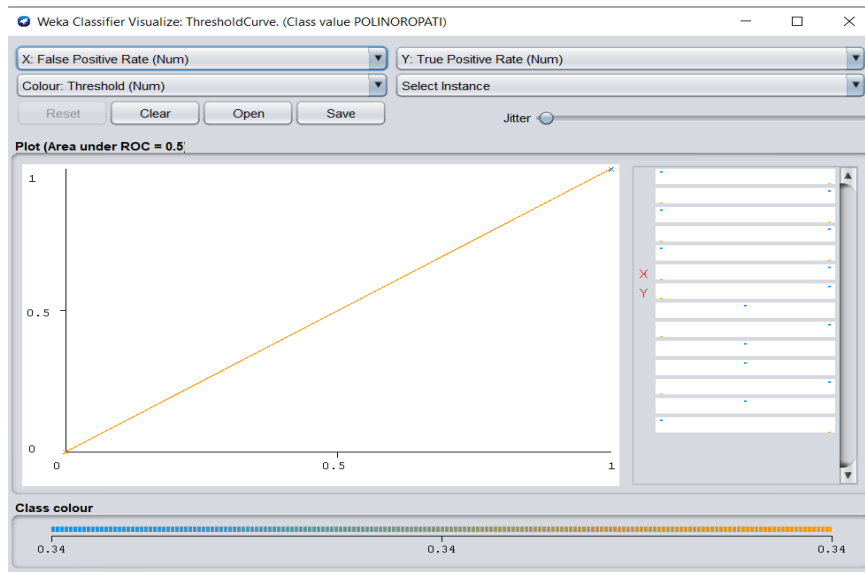
=== Confusion Matrix ===

 a  b  <-- classified as
20  0  | a = NORMAL
 6  0  | b = POLINOROPATI

```

Şekil 3.7. J48 Algoritmasının sınıflandırma sonuçları

J48 algoritmasının ROC eğrisi Şekil 3.7’deki gibidir. Bu şekle göre J48 algoritmasının AUC değeri 0,5 olduğu görülmektedir.



Şekil 3.8. J48 Algoritmasının ROC Eğrisi sonuçları

3.4.2. Random Forest algoritması

Random Forest Algoritması 2001 yılında Breiman tarafından geliştirilmiş bir algoritmadır. Bu algoritma tek bir karar ağacı oluşturmak yerine çok sayıda karar ağacının kararlarını birleştirmektedir. Random Forest (RF), karar ağacı oluşturmak için CART (Classification and Regression Tree) algoritmasını kullanır. Karar ağacındaki düğüm ve dallar bu algoritmanın özelliklerine göre oluşturulur. [38]

Tez çalışmasında kullanılan polinöropati hastaları için çekilen EMG veri seti, Random Forest algoritması ile çalışılmış ve Karışıklık Matrisinde aşağıdaki sonuçlar elde edilmiştir. Sonuçlara göre;

Confusion Matrix	
72	0
28	0

Şekil 3.9 RF sonucunda çıkan karışıklık matrisi

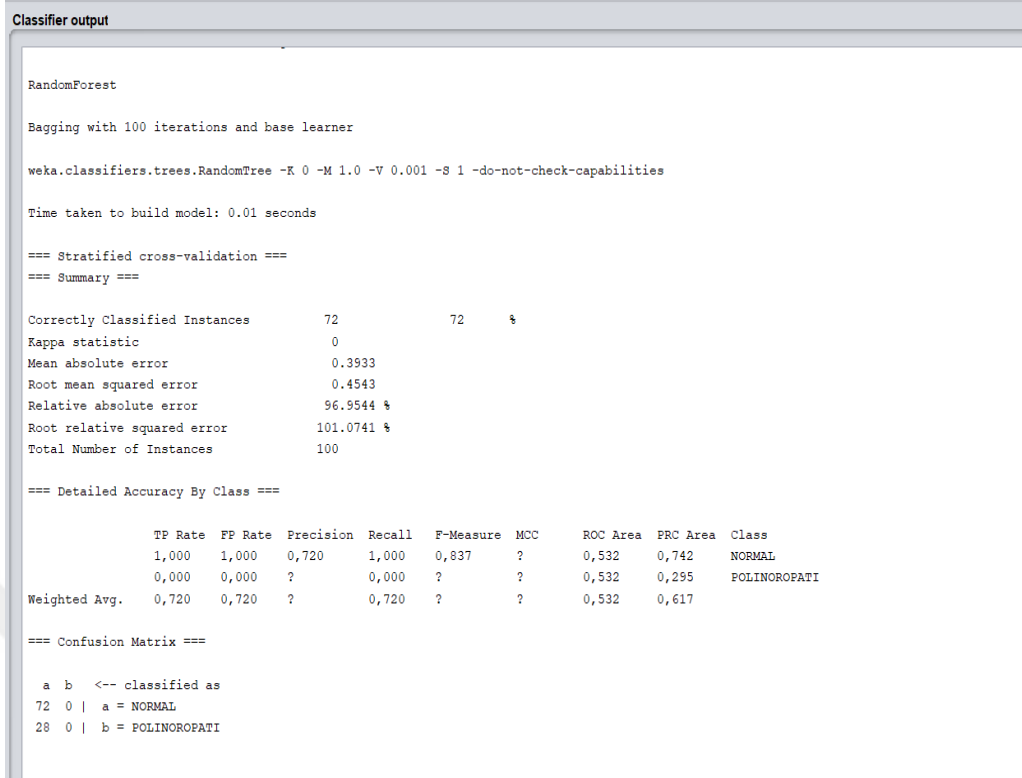
$$\text{Doğruluk} = \frac{(TP + TN)}{(TP + FP + TN + FN)} = \frac{72+0}{72+0+28+0} = 0,72$$

$$\text{Kesinlik} = \frac{TP}{(TP + FP)} = \frac{72}{72+28} = 0,72$$

$$\text{Duyarlılık} = \frac{TP}{(TP + FN)} = \frac{72}{72+0} = 1$$

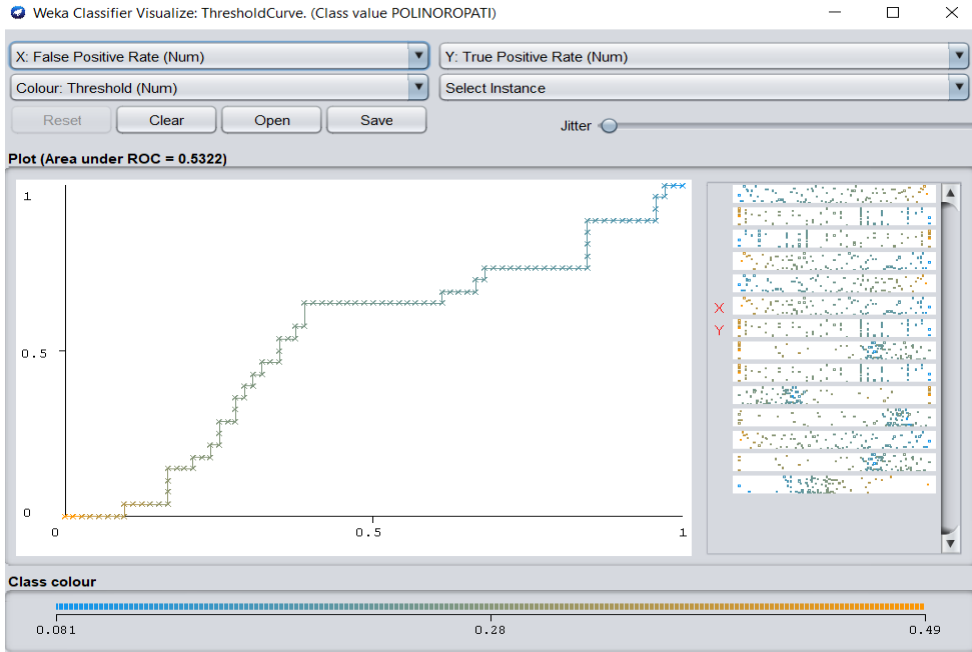
$$\text{F1-Ölçüm} = 2 * \frac{(\text{Kesinlik} * \text{Duyarlılık})}{(\text{Kesinlik} + \text{Duyarlılık})} = 2 * \frac{(0,72 * 1)}{(0,72 + 1)} = 0,837$$

Random Forest algoritmasının doğruluk oranı %72, kesinlik %72, duyarlılık %100, F1-Ölçüm %83,7 olarak sınıflandırmıştır. Random Forest algoritması çalıştırıldıktan sonra sınıflandırıcı sonuçları Şekil 3.9'de "Classifier Output" penceresinde gösterilmektedir.



Şekil 3.10. Random Forest algoritması sınıflandırma sonucu

Random Forest algoritmasının ROC eğrisi Şekil 3.10'deki gibidir. Bu şekle göre Random Forest algoritmasının AUC değeri 0,5322 olduğu görülmektedir.



Şekil 3.11. Random Forest algoritması ROC eğrisi

3.4.3. Naive Bayes algoritması

Bayes Teoremi, istatistiksel prensipleri kullanarak, önceki bilgilerden yola çıkıp yeni sınıf etiketleri oluşturmaktadır. Uygulamada kullanılan Naive Bayes sınıflandırma teknikleri ise araştırılan özelliklerin bağımsızlığı varsayımına dayanarak her sınıf etiketini şartlı olasılıkla tahmin etmeye çalışmaktadır [32]. Naive Bayes sınıflandırma basit olmasına rağmen bazı yöntemlerle karşılaştırıldığında iyi bir performans göstermektedir. Naive Bayes sınıflandırma algoritması bilinen veriler setinin belli bir gruba ait üyelik olasılıklarını tahmin eden istatistiksel sınıflandırıcılar olarak tanımlanabilir [40,41].

Tez çalışmasında kullanılan EMG veri seti, Naive Bayes algoritması ile ve aşağıdaki karışıklık matrisi sonuçları elde edilmiştir.

Confusion Matrix	
32	0
8	0

Şekil 3.12 NB sonucunda çıkan karışıklık matrisi

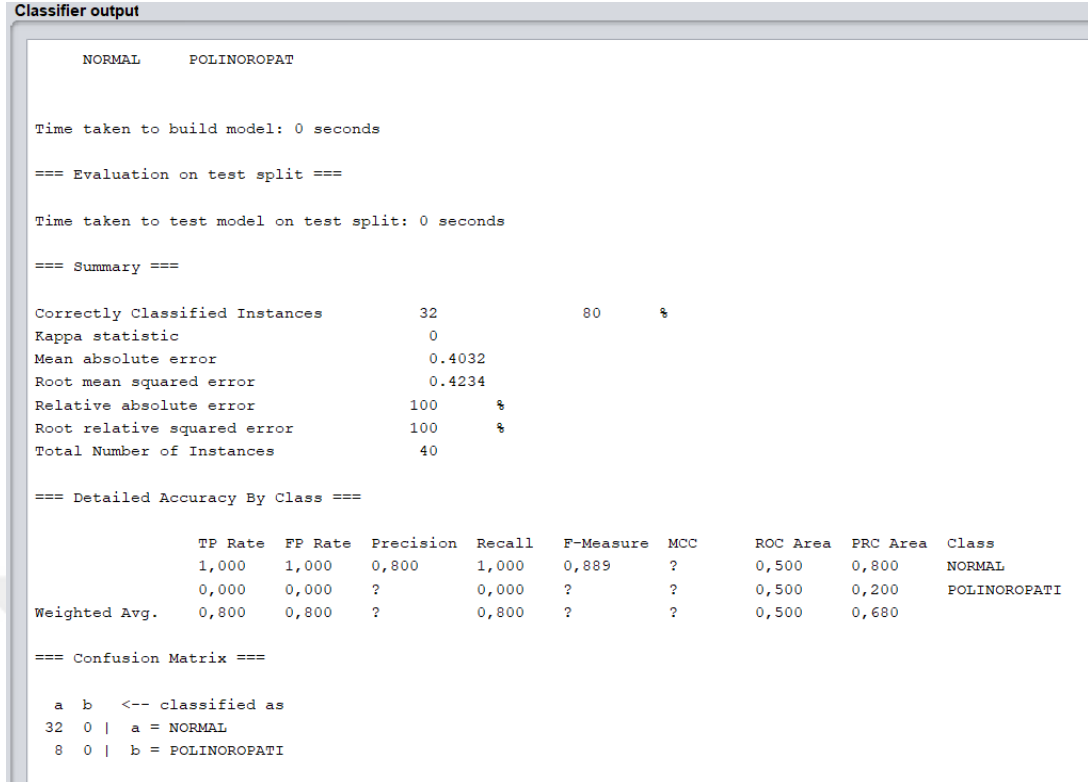
$$\text{Doğruluk} = \frac{(TP + TN)}{(TP + FP + TN + FN)} = \frac{32+0}{32+0+8+0} = 0,80$$

$$\text{Kesinlik} = \frac{TP}{(TP + FP)} = \frac{32}{32+8} = 0,80$$

$$\text{Duyarlılık} = \frac{TP}{(TP + FN)} = \frac{32}{32+0} = 1$$

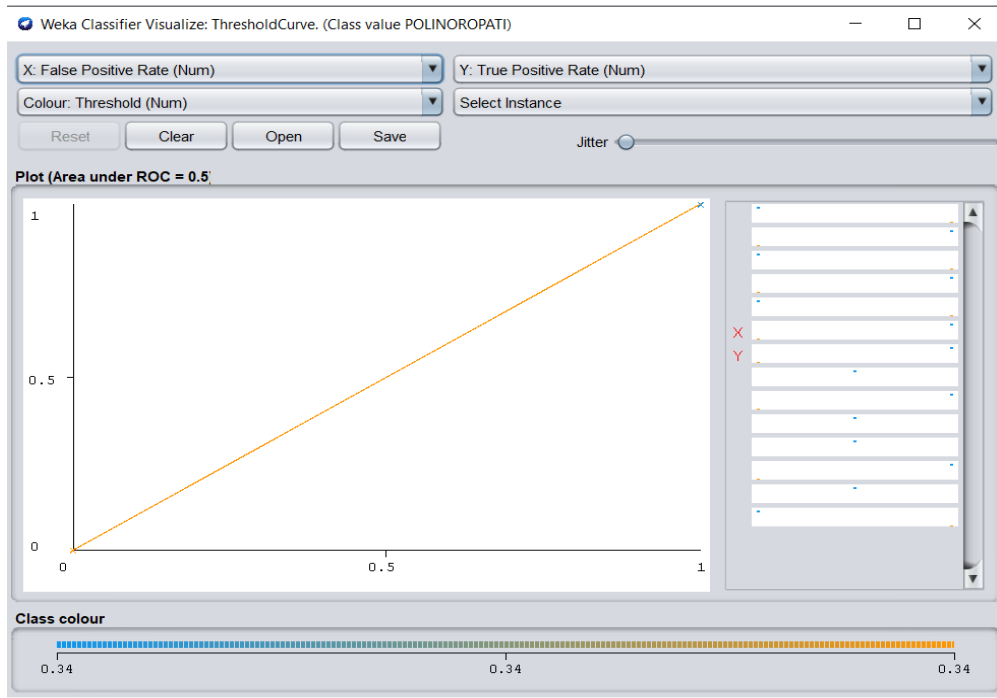
$$\text{F1-Ölçüm} = 2 * \frac{(\text{Kesinlik} * \text{Duyarlılık})}{(\text{Kesinlik} + \text{Duyarlılık})} = 2 * \frac{(0,80 * 1)}{(0,80 + 1)} = 0,889$$

Sonuçlara göre Naive Bayes algoritmasının doğru sınıflandırma oranı %80 kesinlik %80, duyarlılık %100, F1-Ölçün %88,9 olarak sınıflandırmıştır. Bu da Naive Bayes algoritmasının polinöropati hastalığının teşhisini de en fazla kestirim gücüne sahip olduğunu göstermektedir.



Şekil 3.13. Naive Bayes algoritması sınıflandırma sonucu

Naive Bayes algoritmasının ROC eğrisi Şekil 3.13'deki gibidir. Buna göre Naive Bayes algoritmasının AUC değeri 0,5 olduğu görülmektedir.



Şekil 3.14. Naive Bayes algoritması ROC eğrisi

3.4.4. IBK algoritması

K-En yakın komşu algoritması diğer bir adıyla “tek bağlantı kümeleme yöntemi” algoritması olan IBK (Instance Based Learner) algoritması, EMG verileri WEKA’ya yüklendikten sonra Classify->choose->lazy->IBk yolu takip edilerek çalıştırılır. Bu çalışmada k değeri 3 olarak seçilmiştir. Elde edilen karışıklık matrisi sonuçları aşağıdaki Şekil 3.14.de gösterilmiştir.

Confusion Matrix	
28	4
7	1

Şekil 3.15 IBK sonucunda çıkan karışıklık matrisi

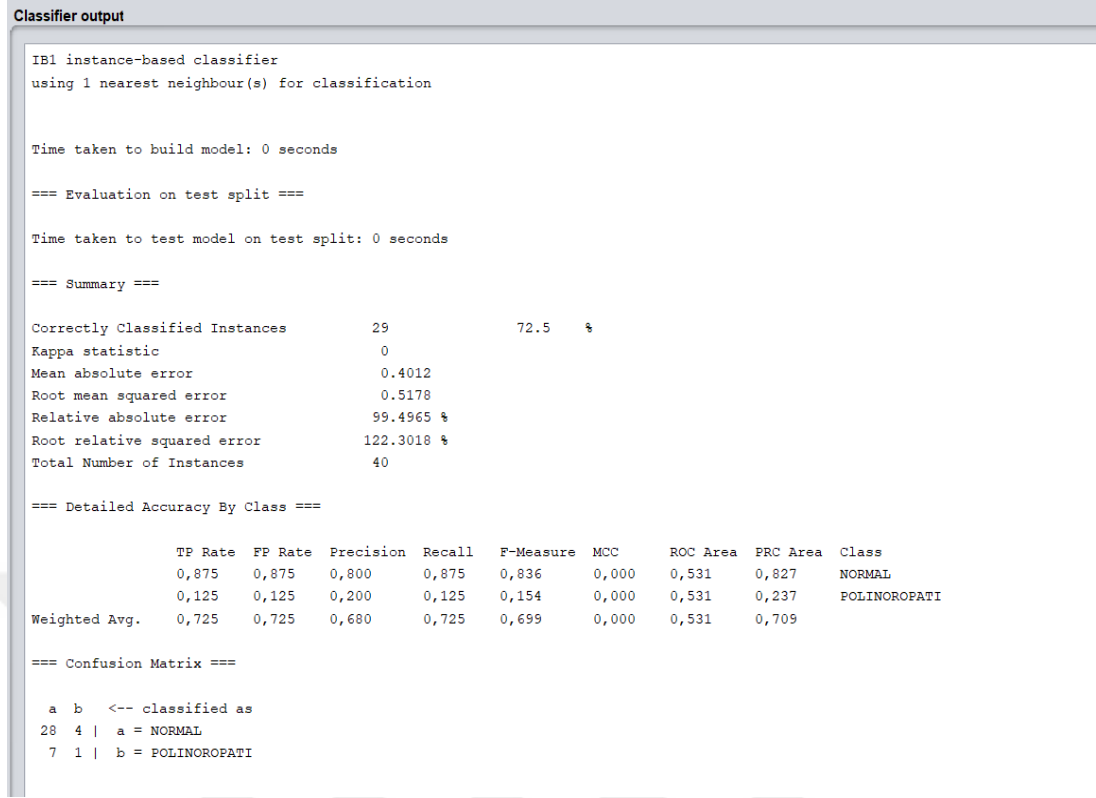
$$\text{Doğruluk} = \frac{(TP + TN)}{(TP + FP + TN + FN)} = \frac{28+1}{28+4+7+1} = 0,725$$

$$\text{Kesinlik} = \frac{TP}{(TP + FP)} = \frac{28}{28+7} = 0,80$$

$$\text{Duyarlılık} = \frac{TP}{(TP + FN)} = \frac{28}{28+4} = 0,875$$

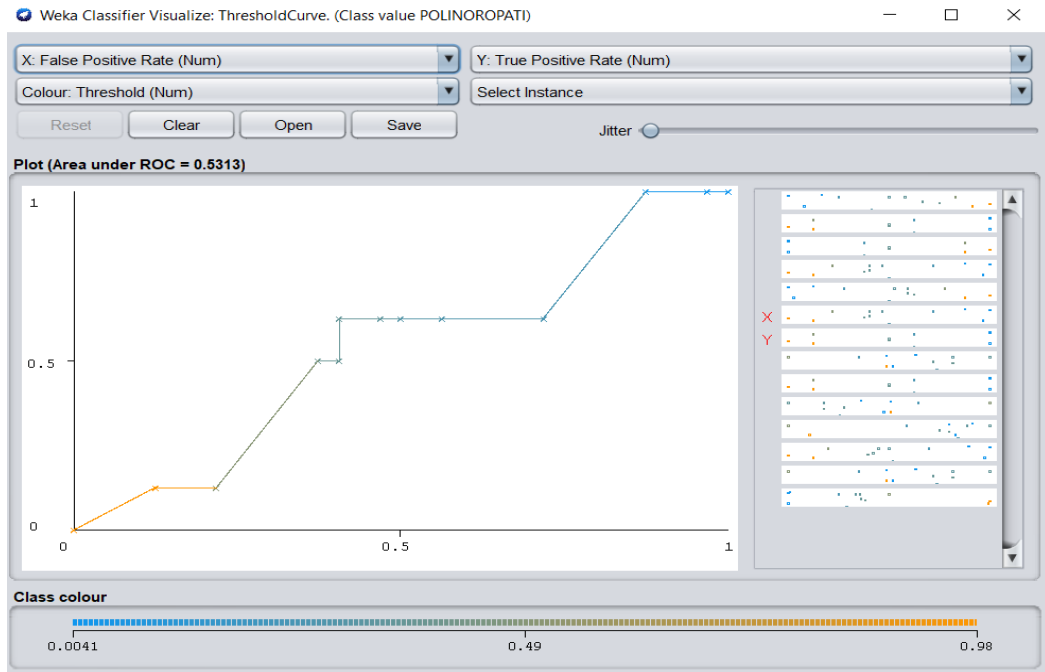
$$\text{F1-Ölçüm} = 2 * \frac{(\text{Kesinlik} * \text{Duyarlılık})}{(\text{Kesinlik} + \text{Duyarlılık})} = 2 * \frac{(0,80 * 0,875)}{(0,80 + 0,875)} = 0,836$$

Bu sonuçlara göre IBK algoritmasının doğruluk oranı %72.5, kesinlik %80, duyarlılık %87.5, F1-Ölçün %83.6 olarak sınıflandırmıştır. Buna göre IBK Algoritmasının Weka Programına ait ekran görüntüsü aşağıdaki Şekil 3.15 ‘de gösterilmiştir.



Şekil 3.16. IBK algoritması sınıflandırma sonucu

IBK algoritmasının ROC eğrisi Şekil 3.16'deki gibidir. Buna göre bu algoritmanın AUC değeri 0,5313 olduğu görülmektedir.



Şekil 3.17. IBK algoritması ROC eğrisi

3.4.5. SMO Algoritması

SMO (Sequential Minimal Optimization) algoritması Destek vektör makinesi yani Support Vektör Machine (SVM) algoritması grubundadır. SMO, her adımda mümkün olan en küçük problemi çözmeyi seçer. Bu çalışmada EMG verileri WEKA'ya yüklendikten sonra Classify->choose->functions->SMO yolu takip edilerek çalıştırılır. Elde edilen karışıklık matrisi sonuçları aşağıdaki Şekil 3.17 de gösterilmiştir.

Confusion Matrix	
68	4
25	3

Şekil 3.18 SMO sonucunda çıkan karışıklık matrisi

$$\text{Doğruluk} = \frac{(TP + TN)}{(TP + FP + TN + FN)} = \frac{68+3}{68+4+25+3} = 0,71$$

$$\text{Kesinlik} = \frac{TP}{(TP + FP)} = \frac{68}{68+25} = 0,731$$

$$\text{Duyarlılık} = \frac{TP}{(TP + FN)} = \frac{68}{68+4} = 0,944$$

$$\text{F1- Ölçüm} = 2 * \frac{(\text{Kesinlik} * \text{Duyarlılık})}{(\text{Kesinlik} + \text{Duyarlılık})} = 2 * \frac{(0,731 * 0,944)}{(0,731 + 0,944)} = 0,824$$

Bu sonuçlara göre SMO algoritmasının doğru sınıflandırma oranı %71, kesinlik %73.1, duyarlılık %94.4, F1-Ölçün % 82.4 olarak sınıflandırmıştır. Sonuç ekran kısmı aşağıdaki Şekil 3.18 de gösterilmiştir

```
Classifier output
- 0.6944

Number of kernel evaluations: 4935 (95.543% cached)

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      71          71    %
Kappa statistic                    0.0669
Mean absolute error                 0.29
Root mean squared error             0.5385
Relative absolute error             71.4898 %
Root relative squared error        119.8214 %
Total Number of Instances          100

=== Detailed Accuracy By Class ===

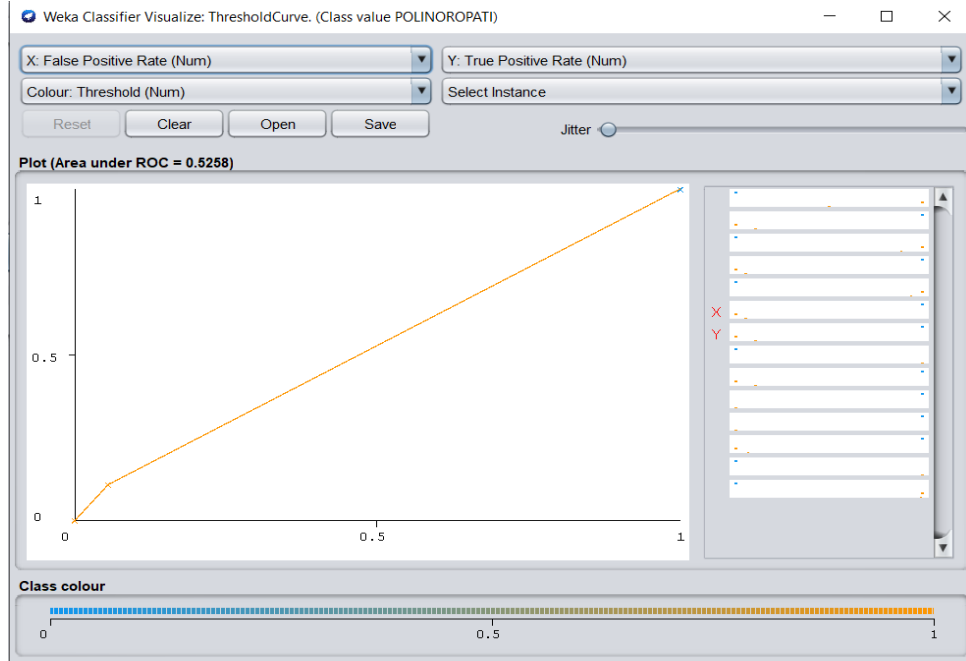
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
          0,944   0,893   0,731     0,944   0,824     0,091   0,526    0,731    NORMAL
          0,107   0,056   0,429     0,107   0,171     0,091   0,526    0,296    POLINOROPATI
Weighted Avg.   0,710   0,658   0,646     0,710   0,641     0,091   0,526    0,609

=== Confusion Matrix ===

 a b  <-- classified as
68 4 | a = NORMAL
25 3 | b = POLINOROPATI
```

Şekil 3.19. SMO algoritması sınıflandırma sonucu

SMO algoritmasının ROC eğrisi Şekil 3.19'deki gibidir. Bu şekle göre bu algoritmanın AUC değeri 0,5258 olduğu görülmektedir.



Şekil 3.20. SMO algoritması ROC eğrisi

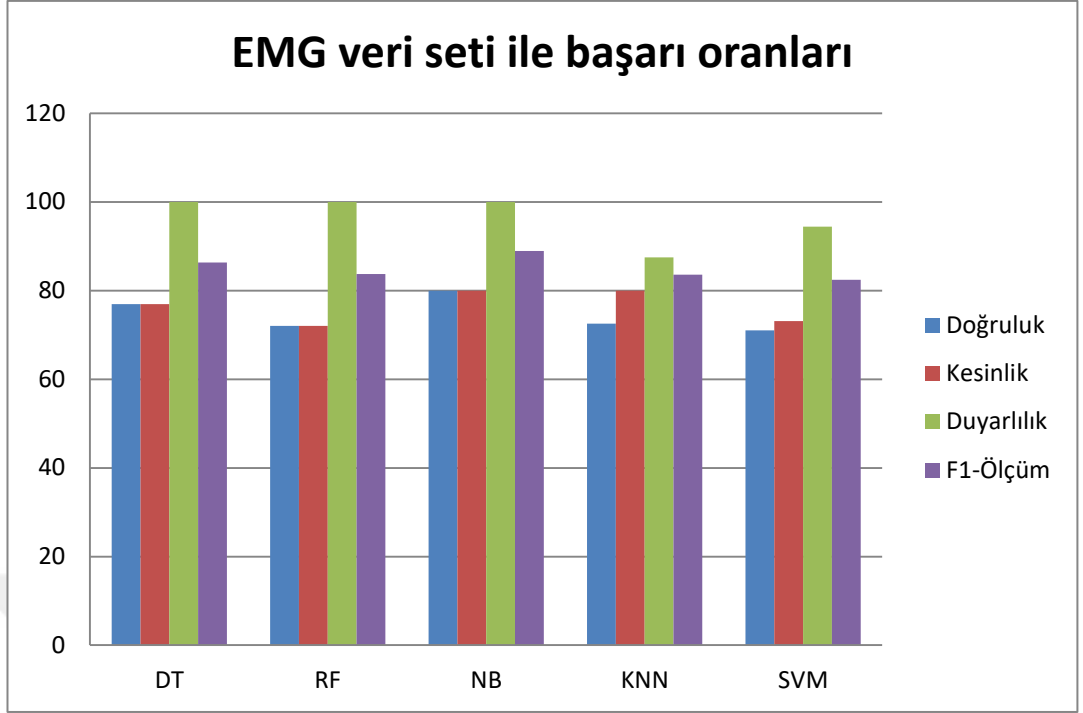
3.5 Araştırma Bulgularını Değerlendirme

Bu çalışmada EMG veri seti ile makine öğrenmesi algoritmaları kullanılarak Polinöropati hastalığı teşhisinin tespiti tahmin edilmeye çalışılmıştır. EMG veri setine gerçek hastalardan alınan verilere göre incelenmiştir. İncelenen makine öğrenmesi ile algoritmaları ile EMG veri seti kullanılarak K-NN algoritması ile veri setinin başarı oranı %72,5 olarak sınıflandırılmıştır. Naive Bayes algoritması ile %80 olarak başarı oranı bulunmuştur. Support Vector Machine (SVM) Algoritması olarak %71 olarak sınıflandırılmıştır. Karar Ağaçları (Decision Tree) algoritması ile başarı oranı %76,9231 olarak bulunmuştur. Random Forest algoritması ile başarı oranı %72 olarak sınıflandırılmıştır. Bu veriler sonucunda ortaya çıkan başarı oranları aşağıdaki Tablo3.2 de detaylı olarak gösterilmektedir.

Tablo 3.2. EMG veri seti ile makine öğrenmesi başarı oranları

Performans Ölçütleri %	Karar Ağacı (DT)	Random Forest (RF)	Naive Bayes (NB)	K-EN Yakın Komşu (K-NN)	Destek Vektör Makinesi (SVM)
Doğruluk ()	76,92	72	80	72,5	71
Duyarlılık	100	100	100	87,5	94,4
Kesinlik	76,92	72	80	80	73,1
F1-Ölçüm	86,36	83,72	88,89	83,6	82,4

Tablo 3.2 incelendiğinde Naive Bayes algoritması doğruluk değeri %80 ile en yüksek algoritma olduğu görülmektedir. Bu veriler sonucunda ortaya çıkan başarı oranları aşağıdaki Şekil 3.20 de detaylı olarak gösterilmektedir.



Şekil 3.21. EMG veri seti ile makine öğrenmesi başarı oranları grafiği

4.SONUÇ VE DEĞERLENDİRME

Makine öğrenme algoritmaları birçok alanda kullanıldığı gibi sağlık alanında da başarılı bir şekilde sıklıkla kullanılmaktadır. Günümüzde makine öğrenmesi ve veri madenciliği yöntemleriyle sağlık hizmeti sunucuları için doğru karar almalarına yardımcı olacak veya kararlarını destekleyecek pek çok hastalığın teşhisinde, sınıflandırılmasında, tahmin edilmesinde kullanılacak akıllı uygulamalar sağlık alanında geliştirilebilmektedir.

Bu çalışmada, Sağlık Bilgi Sistemlerinde EMG verileri ile makine öğrenmesi teknikleri kullanılarak yaşam kalitesini oldukça düşüren polinöropati hastalığının teşhisi tahmin edilmeye çalışılmıştır. Dünya genelinde yaygın olarak görülmekte olan polinöropati hastalıkları günümüzde birçok araştırmaya konu olmuştur. Fakat makine öğrenmesi yöntemleriyle Polinöropati hastalığını teşhisi ile ilgili bir çalışmaya konu olmamıştır. Polinöropati hastalığına sebep olan en önemli nedenler arasında diyabetik polinöropati gelmektedir ve genellikle hastalar diyabet hastası olduğunun farkında değildirler. Polinöropati teşhisi ilgili uzman tarafından tespit edildikten sonra yapılması gereken hastanın gerekli tahlillerinin istenmesi gerekmektedir ve çıkan sonuca göre tedavi planlaması yapılması gerekmektedir.

Bu çalışmada EMG verileri kullanılarak oluşturulan EMG veri seti kullanılmıştır. Bu veri seti Polinöropati hastalığının teşhisinde kullanılan değişkenleri içeren hastalara ait kayıtlardan oluşmaktadır. WEKA veri madenciliği programında yer alan J48, Random Forest, SMO, Naive Bayes ve IBk makine öğrenme algoritmaları ile EMG veri seti üzerinde çalışılmıştır. Bu algoritmaların sınıflama performansları yani teşhis güçleri tıpta kullanılan değerlendirme ölçülerinden faydalanılarak birer birer hesap edilmiş ve karşılaştırılmıştır. Tezde kullanılan 5 farklı makine öğrenme algoritması EMG veri setine uygulanmadan önce bir takım veri ön işleme adımları gerçekleştirilmiştir.

Gelecekteki çalışmalarımızda, makine öğrenmesi teknikleri kullanılarak daha doğru sonuçlar elde etmek için sisteme daha fazla veri seti tanımlanarak web tabanlı bir uygulama platformu geliştirilmesi düşünülmektedir. Böylelikle EMG çekimi yapıldıktan sonra verilerin platforma aktarılması sonucunda polinöropati hastalığının teşhisi tahmin edilerek hekimlerimizin iş yüklerinin azaltılması planlanmaktadır.



KAYNAKLAR

- [1] Basmajian V., De Luca C.J., *Muscles Alive: Their Functions Revealed by Electromyogram*, 5th edition, William & Wilkins, Baltimore, 1985.
- [2] Foster K.R., Koprowsk R., Skufca J.D.. Machine learning, medical diagnosis and biomedical engineering research-commentary. *Biomedical Engineering Online*, 2014, **13**(1), 94.
- [3] Bhardwaj, R., Nambiar,A.R.,Dutta D., Study of machine learning in healthcare. Annual Computer Software and Applications Conference (COMPSAC), IEEE, 2017, **41**(2), 236-241,.
- [4] Duch W., Swaminathan K., Meller,J., Artificial intelligence approaches for rational drug design and discovery, *Current Pharmaceutical Design*, 2007, **13**(14), 1497-1508.
- [5] Shiraishi J., Li Q., Appelbaum D., Doi K., Compute raided diagnosis and artificial intelligence in clinical imaging. *In Seminars in Nuclear Medicine*, WB Saunders, 2011, **41**(6), 449-462.
- [6] Taylor R.H., Mencias A., Fichtinger G., Fiorini P., Dario P., Medical robotics and computer integrated surgery, In Springer Handbook of Robotics, Springer, Cham, 2016, (pp.16571684).
- [7] Reaz M. B. I., Hussain M. S. and Mohd Y.F., Techniques of EMG signal analysis: Detection, processing, classification and applications, *Biol. Proced. Online*, 2006, **8**(1), 11–35
- [8] Ayaz F., Emg Sinyallerinin Sınıflandırılması, İnönü Üniversitesi Fen Bilimleri Enstitüsü (Yüksek Lisans Tezi), Malatya, 2018, 154848.
- [9] <https://www.medikalakademi.com.tr/polinoropati-nedir-neden-olur-belirtileri-ve-tedavisi/> (Ziyaret Tarihi:01 Nisan 2021)
- [10] https://edu.kpfu.ru/pluginfile.php/278552/mod_resource/content/1/MachineLearningR__Brett_Lantz.pdf (Ziyaret Tarihi:15 Nisan 2021)
- [11] Ozcift A, Gulden A., Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Compute Methods Programs Biomed*, 2011, **104**(3), 443–51.

- [12] Öztemel,E., Yapay Sinir Ağları, *Papatya Yayıncılık Eğitim AŞ.*, İstanbul, 2012, **11**(3), 232s.
- [13] Güven, Z.B. ve Bilgin, T.T. Zaman Serileri Madenciliği Kullanılarak Nüfus Artışı Tahmin Uygulaması. Akademik Bilişim'14- XVI. Akademik Bilişim Konferansı, Mersin Üniversitesi, Mersin, 5-7 Şubat 2014.
- [14] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I, Machine learning and data mining methods in diabetes research. *Comput Strt Biotechnol J.*,2017;**15**(1):104–116.
- [15] ATALAY, M.; ÇELİK, E. (2017). BüyükVeriAnalizinde Yapay Zekâ VeMakineÖğrenmesi Uygulamaları - Artificial Intelligence and Machine Learning Applications in Big Data Analysis, *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 155–172. doi:10.20875/makusobed.309727
- [16] Keleş, M. B., Keleş A. ve Keleş A., “Makine öğrenmesi yöntemleri ile uçuş fiyatlarının tahmini”, *Eurosia Journal of Mathematics*, 2020, **7**(11)0, 72-78.
- [17] Kaya Y., Motokaravan Sigortacılığı Tahmin Modellemesi Ve Uygulanan Yöntemlerin Karşılaştırılması, Beykent Üniversitesi Fen Bilimleri Enstitüsü (Yüksek Lisans Tezi), İstanbul, 2017, 459240.
- [18] Akı M.O., Sürücü Uykululuğunun Gerçek Zamanlı Görüntü İşleme Ve Makine Öğrenmesi Teknikleri İle Tespitine Yönelik Bir Sistem Tasarımı Ve Uygulaması, Trakya Üniversitesi Fen Bilimleri Enstitüsü (Doktora Tezi), Edirne, 2017, 469084.
- [19] Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H. Predicting diabetes mellitus with machine learning techniques. *Front Genet.* 2018, **9**,515.
- [20] Zhang X. Support vector machines. In: *Encyclopedia of machine learning and data mining*. Boston, MA: Springer US, 2017. p. 1214–20.
- [21] Fasidi F. and Adebayo O., “Rule-based Naïve Bayes Classifier for Heart Disease Risk Prediction and,” *Int. J. Clin. Med. Informatics Rev.*, 2019, **2**(2), 51–59.
- [22] GÖK M., Makine Öğrenmesi Yöntemleri ile Akademik Başarının Tahmin Edilmesi, *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, 2017, Vol. 5, No. 3, 139–148
- [23] Bozkurt M.R., EMG İşaretlerinin Modern Yöntemlerle Ön işleme ve Sınıflandırılması, Sakarya Üni. , Fen Bilimleri Ens., Doktora Tezi, Sakarya, 2007, 216254.
- [24] Irmak, S. Veri Madenciliği Yöntemleri İle Sağlık Sektörü Veri Tabanlarında Bilgi Keşfi: Tanımlayıcı ve Kestirimci Model Uygulamaları, Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü (Doktora Tezi), Antalya, 2009, 250551.

- [25] Kiremitçi B., Veri Ambarlarında Veri Madenciliği ve Ulaştırma-Lojistik Sektöründe Bir Uygulama, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü (Yüksek Lisans Tezi), İstanbul, 2005, 215573.
- [26] Savaş,S., Topaloğlu, N.,Yılmaz,M.,2012. Veri Madenciliği Ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, **11**(21), 1-23.
- [27] Dinçer E., Veri Madenciliğinde K-Means Algoritması Ve Tıp Alanında Uygulanması, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü (Yüksek Lisans Tezi), Kocaeli, 2006, 197927.
- [28] Karşlı Ö.B., Makine Öğrenme Yöntemleri İle Karaciğer Hastalığının Teşhisi, Ağrı İbrahim Çeçen Üniversitesi Fen Bilimleri Enstitüsü (Yüksek Lisans Tezi), 2019, 573609.
- [29] Daş, B. ve Türkoğlu, İ. 2014. DNA Dizilimlerinin Sınıflandırılmasında Karar Ağacı Algoritmalarının Karşılaştırılması. Eleco 2014 Elektrik – Elektronik – Bilgisayar ve Biyomedikal Mühendisliği Sempozyumu, 27 – 29 Kasım 2014, Bursa.
- [30] Xie J. et al., “A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges,” *IEEE Commun. Surv. Tutorials*, vol. 21, no. 1, pp. 393–430, 2019, doi: 10.1109/COMST.2018.2866942.
- [31] Wibawa A. P. et al., “Naïve Bayes Classifier for Journal Quartile Classification,” *Int. J. Recent Contrib. from Eng. Sci. IT*, vol. 7, no. 2, p. 91, 2019, doi: 10.3991/ijes.v7i2.10659.
- [32] Tyagi, I., Shad, J., Sharma, S., Gaur, S., ve Kaur, G., A novel machine learning approach to detect phishing websites, 2018 5th, *International conference on signal processing and integrated networks (SPIN)*, 2018, p. 425-430.
- [33] Hassan M. M., Jones E., and Buck C. E., “A simple Bayesian approach to tree-ring dating,” *Archaeometry*, vol. 61, no. 4, pp. 991–1010, 2019, doi: 10.1111/arcm.12466
- [34] Öztürk, K. ve Şahin, M. E., “Yapay sinir ağları ve yapay zekâ'ya genel bir bakış”, *Takvim-i Vekayî*, 2018, **6**(2), 25-36.
- [35] Mienye I. D., Sun Y., and Wang Z., “Prediction performance of improved decision tree-based algorithms: A review,” in *Procedia Manufacturing*, 2019, vol. 35, pp. 698–703, doi: 10.1016/j.promfg.2019.06.011.
- [36] Li L. L., Zhao X., Tseng M. L., and Tan R. R., “Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm,” *J. Clean. Prod.*, vol. 242, p. 118447, 2020, doi: 10.1016/j.jclepro.2019.118447.

- [37] Demolli H., Dokuz A. S., Ecemis A., and Gokcek M., "Wind power forecasting based on daily wind speed data using machine learning algorithms," *Energy Convers, Manag.*, vol. 198, no. March, p. 111823, 2019, doi: 10.1016/j.enconman.2019.111823.
- [38] Muppalaneni N. B., Ma M., and Gurumoorthy S., "Soft Computing and Medical Bioinformatics," p. 139, 2019, doi: 10.1007/978-981-13-0059-2.
- [39] Akşehirli Ö., Tıbbi Araştırmalarda Destek Vektör Makinelerinin Kullanımı, Düzce Üniversitesi Sağlık Bilimleri Enstitüsü (Yüksek Lisans Tezi), Düzce, 2012, 340813.
- [40] Rahat A. M., Kahir A. and Masum A. K. M., "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset", 8th, International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2019.
- [41] Çığşar, B., Kredi Risklerinde Veri Madenciliği Sınıflandırma Algoritmaları, Çukurova Üniversitesi Fen Bilimleri Enstitüsü (Yüksek Lisans Tezi), Adana, 2017, 465358.

KİŞİSEL YAYINLAR VE ESERLER

- [1] **Tamk N.**, Kavak A., “Machine Learning Based Polyneuropathy Diagnosis Using Electromyogram Data”, *ARTIFICIAL INTELLIGENCE THEORY and APPLICATIONS*, 2021, **1(2)**:148-155.



ÖZGEÇMİŞ

Lise öğrenimini Balıkesir Cumhuriyet Lisesinde tamamladı. 2004 yılında Erzurum Üniversitesi Bilgisayar Programcılığı Bölümü'nden 2006 yılında bölüm 2. olarak mezun oldu. 2006-2011 yılları arasında Balıkesir'de İlköğretim Okulunda ücretli öğretmenlik yaptı. 2011 yılında girdiği İstanbul Kültür Üniversitesi Bilgisayar Mühendisliği Bölümü'nden 2015 yılında mezun oldu. 2019 yılında Kocaeli Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı'nda yüksek lisans eğitimine başladı. Balıkesir Devlet Hastanesi Bilgi İşlem biriminde çalışmaktadır.

