



T.C.
KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MR SPEKTROSKOPİ TEMELLİ BEYİN TÜMÖRÜ
TEŞHİSİNDE VERİ MADENCİLİĞİ
UYGULAMALARI

SİNAN ALTUN

YÜKSEK LİSANS TEZİ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

KAHRAMANMARAŞ 2018

T.C.
KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MR SPEKTROSKOPİ TEMELLİ BEYİN TÜMÖRÜ
TEŞHİSİNDE VERİ MADENCİLİĞİ
UYGULAMALARI

SİNAN ALTUN

Bu tez,
Elektrik-Elektronik Mühendisliği Anabilim Dalında
YÜKSEK LİSANS
derecesi için hazırlanmıştır.

KAHRAMANMARAŞ 2018

Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü öğrencisi Sinan ALTUN tarafından hazırlanan “MR SPEKTROSKOPİ TEMELLİ BEYİN TUMÖRÜ TEŞHİSİNDE VERİ MADENCİLİĞİ UYGULAMALARI” adlı bu tez, jürimiz tarafından 11/05/2018 tarihinde oy birliği ile Elektrik-Elektronik Mühendisliği Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Doç. Dr. Ahmet ALKAN (DANIŞMAN)

.....

Elektrik-Elektronik Müh.

Kahramanmaraş Sütçü İmam Üniversitesi

Dr. Öğr. Üyesi Eser SERT (ÜYE)

.....

Bilgisayar Müh.

Kahramanmaraş Sütçü İmam Üniversitesi

Dr. Öğr. Üyesi Halil EROL (ÜYE)

.....

Elektrik-Elektronik Müh.

Osmaniye Korkut Ata Üniversitesi

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Doç. Dr. Mustafa ŞEKKEİ

.....

Fen Bilimleri Enstitüsü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada, alıntı yapılan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

Sinan ALTUN



Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

MR SPEKTROSKOPİ TEMELLİ BEYİN TÜMÖRÜ TEŞHİSİNDE VERİ MADENCİLİĞİ UYGULAMALARI

(YÜKSEK LİSANS TEZİ)
SİNAN ALTUN
ÖZET

Çağımızda bilişim teknolojilerinin gelişmesine bağlı olarak büyük kapasiteli veri tabanları oluşmuştur. Veri tabanlarında bulunan bu büyük verilerin işlenip kullanılması her açıdan olumlu olacaktır, işlenmeyen veri ise herhangi bir önem arz etmeyecektir. Veri tabanındaki bilgileri kullanarak yararlı sonuçlar çıkarma işlemi, veri madenciliği olarak adlandırılır.

Tıp alanında kan ölçüm değerleri, tıbbi görüntüleme ve sinyal işleme ile büyük boyutlu tıbbi veri tabanları oluşmaktadır. Bu veri tabanlarında bulunan veriler, veri madenciliği ile işlenip yararlı sonuçlar elde edilirse özellikle teşhis konusunda doktora yardımcı bir sistemin oluşturulmasını sağlar. Oluşturulacak bu uzman sistemler teşhis sırasında olabilecek yanlışlıkları azaltabilir.

Beyin tümörü teşhisi ivedilik arz eder ve teşhis ciddi bir uzmanlık gerektirir. Beyin kanaması gibi farklı rahatsızlıklar ile karıştırılıp gerekli müdahale yapılmazsa veya gereksiz bir cerrahi müdahale yapılırsa hastanın yaşam kalitesi düşebilir hatta ölümle sonuçlanan bir durumla karşılaşılabilir. MR Spektroskopi (MRS) ilk başta laboratuvar ortamlarında deneysel bir yöntem olarak kullanılmaya başlanmıştır. Daha sonra MRS' in klasik MR görüntülemelemedeki su ve yağ metabolitlerden farklı metabolitlerin görüntüleyebildiği keşfedilmiştir. MR Spektroskopinin (MRS) beyin tümörü teşhisinde kullanılmaya başlaması, MR Spektroskopinin kolin (Cho), kreatin (Cr) ve N-Asetil Aspartat (Naa) metabolit değerlerini ölçebildiğinin keşfi ile başlamıştır. Tümöre teşhisinde özellikler kolin (Cho) metabolit değeri oldukça önemlidir.

Bu çalışmada Ocak 2009 ve Mayıs 2017 tarihleri arasında, Kahramanmaraş Sütçü İmam Üniversitesi Tıp Fakültesi Beyin ve Sinir Cerrahisi kliniğinde beyin tümörü teşhisi konulmuş ve teşhis konulmamış hastaların bilgileri kullanılmıştır. Bu bilgileri kullanmak için gerekli izinler alınmıştır. Kullandığımız verilere göre konulan teşhisler, Beyin ve Sinir Cerrahisi Uzmanına kontrol ettirilmiştir. Böylece veri setinde eksik ya da yanlış bilginin olması engellenmiştir.

Çalışmamızda WEKA programı kullanılmış olup yüksek oranda başarılar sağlanmıştır. Sınıflandırma teknikleri olan KNN’ de % 95.69, Lojistik Regresyonda %90.32, Destek Vektör Makinelerinde %82.79, Çok Katmanlı Yapay Sinir Ağında %90.32, Karar Ağaçlarında %87.09 ve Naive Bayes yönteminde %87.09 başarı elde edilmiştir. Kümeleme teknikleri olan Kohonen de %92.92, K Ortalamada ise %98.92 başarı elde edilmiştir. Derin Öğrenme tekniği ile veri setimiz işlendiğinde %93.54 başarı elde edilmiştir. Bu yüksek başarılar doktora yardımcı, uzman bir sistemin oluşturulabileceğini göstermektedir.

Anahtar Kelimeler: Beyin Tümörü, Veri Madenciliği, MR Spektroskopi.

Kahramanmaraş Sütçü İmam Üniversitesi

Fen Bilimleri Enstitüsü

Elektrik-Elektronik Mühendisliği Anabilim Dalı, Mayıs / 2018

Danışman: Doç. Dr. Ahmet ALKAN

Sayfa sayısı: 65

MR SPECTROSCOPY BASED BRAIN TUMOR DIAGNOSIS DATA MINING APPLICATIONS

(M.Sc. THESIS)

SİNAN ALTUN

SUMMARY

In our age, large-scale databases have been created depending on the development of information technologies. The processing and use of these large data in the databases will be positive in every respect, unprocessed data will not be of any importance. Using the information in the database to extract useful results, it is called data mining.

Blood measurement values in medicine, large-scale medical databases are formed with medical imaging and signal processing. The data contained in these databases, it provides a system to assist the doctor in diagnosis especially if it is processed with data mining and beneficial results are obtained. These expert systems to be created can reduce the mistakes that may occur during diagnosis.

Brain tumor diagnosis is urgent and diagnosis requires serious expertise. If the patient is confused with a variety of disorders such as brain hemorrhage and if necessary intervention is not performed or an unnecessary surgical intervention is performed, the patient's quality of life may be reduced or even fatal. MR Spectroscopy (MRS) was first used as an experimental method in laboratory environments. It was later discovered that MRS can display metabolites different from water and fat metabolites in the classifiable MR image. Beginning to use MR Spectroscopy (MRS) in brain tumor diagnosis, It begins with the discovery that the MR Spectroscopy can measure the metabolite values of choline (Cho), creatinine (Cr) and N-acetyl aspartate (Naa). Choline (Cho) metabolite value is very important in the identification of the tumor.

In this study, between January 2009 and May 2017, Kahramanmaraş Sütçü İmam University Faculty of Medicine Department of Brain and Nerve Surgery clinic has been diagnosed with brain tumors and the information of patients who have not been diagnosed. Permission has been obtained to use this information. According to the diagnosis we use, Brain and Nerve Surgery Specialist. This prevents the data set from being incomplete or incorrect.

The WEKA program has been used in our work and has been achieved at high rates. The Classification techniques were 95.69% in KNN, 90.32% in Logistic Regression, 82.79% in Support Vector Machines, 90.32% in Multi-layer Artificial Neural Networks, 87.09% in Decision Trees and 87.09% in Naive Bayes method. In clustering techniques, Kohonen achieved 92.92%, K Means was 98.92%. When our data set was processed with the Deep Learning technique, 93.54% success was achieved. These high achievements show that a specialist system can be created to help the doctor.

Keywords: Brain Tumor, Data Mining, MR Spectroscopy.

University of Kahramanmaraş Sütçü İmam

Graduate School of Natural And Applied Sciences

Department of Electrical-Electronic Engineering, May / 2018

Supervisor: Assoc. Prof. Dr. Ahmet ALKAN

Page number: 65

TEŞEKKÜR

Tez çalışmalarım sırasında beni destekleyerek, bilgi ve tecrübesini benden esirgemeyen ve bana yol gösteren değerli danışman hocam Sayın Doç. Dr. Ahmet ALKAN' a veri setini oluşturmamızda desteğini hiç esirgemeyen Sayın Doç. Dr. İdiris ALTUN'a teşekkürlerimi sunarım.

Her zaman maddi ve manevi destekleri ile yanımda olan ve beni bugünlere getiren anneme ve babama sonsuz saygı ve sevgilerimi sunarım.

Ayrıca yoğun çalışma esnasında her zaman bana destek olan eşime teşekkürlerimi sunarım.

XXXX

Sinan ALTUN

İÇİNDEKİLER

Sayfa No

ÖZET	i
SUMMARY	iii
TEŞEKKÜR	v
İÇİNDEKİLER.....	vi
ŞEKİLLER DİZİNİ	ix
ÇİZELGELER DİZİNİ.....	xi
SİMGELER VE KISALTMALAR DİZİNİ	xii
1.GİRİŞ.....	1
1.1. Veri Madenciliği Yazılımları.....	3
1.1.1. WEKA	3
1.1.2. Theano	4
1.1.3. TensorFlow	4
1.1.4. Keras	4
1.1.5. Python	4
2.ÖNCEKİ ÇALIŞMALAR	5
3. VERİ MADENCİLİĞİ	7
3.1. Veri Madenciliği Nedir?	7
3.2. Veri Madenciliği Tarihi	8
3.3. Veri Madenciliği Uygulama Alanları	8
3.3.1. Finans alanında veri madenciliği uygulamaları	8
3.3.2. Tıp alanında veri madenciliği uygulamaları	9
3.3.3. Perakende alanında veri madenciliği uygulamaları	9
3.3.4. İletişim alanında veri madenciliği uygulamaları	9
3.4. Veri Madenciliğinde Karşılaşılan Problemler	10
3.4.1. Gürültü	10
3.4.2. Boş değer	10
3.4.3. Eksik veri	10
3.4.4. Dinamik veri	10
3.5. Veri Madenciliği Faydaları	10
3.6. Veri Madenciliği Modelleri	11
3.6.1. Sınıflandırma ve regresyon.....	11
3.6.2. Kümeleme.....	12
3.6.2.1. Paylaştırma temelli yöntemler	13

	<u>Sayfa No</u>
3.6.2.2. Hiyerarşik temelli yöntemler	13
3.6.2.3. Yoğunluk tabanlı yöntemler	13
3.6.2.4. İzgara tabanlı yöntemler	13
3.6.2.5. Model tabanlı yöntemler	13
3.7. Veri Madenciliği İşlem Akışı	14
3.7.1. Problem nedir?.....	15
3.7.2. Veriyi anlamak	16
3.7.3. Veri hazırlanması.....	16
3.7.4. Modelleme	16
3.7.5. Değerlendirme	16
3.8. Çalışmada Kullanılan Veri Madenciliği Sınıflandırma Yöntemleri	17
3.8.1. KNN.....	17
3.8.2. Destek vektör makineleri.....	18
3.8.3. Lineer regresyon	19
3.8.4. Non-lineer regresyon	20
3.8.5. Lojistik regresyon	21
3.8.6. Naive bayes.....	22
3.8.7. Karar ağaçları	25
3.8.7.1. Özellik 1 için entropi hesabı	27
3.8.7.2. Özellik 2 için entropi hesabı	27
3.8.7.3. Özellik 3 için entropi hesabı	27
3.8.7.4. Özellik 4 için entropi hesabı	28
3.8.8. Çok katmanlı yapay sinir ağları.....	28
3.9. Çalışmada Kullanılan Veri Madenciliği Kümeleme Yöntemleri.....	29
3.9.1. K-Ortalamlar (K-means).....	29
3.9.2. Kohonen (Self organizing map)	30
3.10. Derin Öğrenme	31
3.10.1. Tam bağlantılı yapay sinir ağı (Dense fully connected).....	32
3.10.2. Konvolüsyon sinir ağı (CNN).....	32
3.10.3. Aktivasyon fonksiyonları	33
3.10.4. Derin öğrenme için hata fonksiyonları	35
3.11. Başarı Ölçütleri	35
3.11.1. ROC (Receiving operating characteristic) analizi	35
3.11.2. Duyarlılık (Sensitivity – Doğru pozitif oranı- Recall).....	36
3.11.3. Özgüllük (Specificity – Doğru negatif oranı).....	36
3.11.4. Yanlış pozitif oranı (1-Özgüllük)	36
3.11.5. Yanlış negatif oranı (1-Duyarlılık)	36
3.11.6. Gerçek doğruluk (Accuracy)	37
3.11.7. Artı yorum gücü (Positive predictive value-Duyarlılık (Precision):-Kesinlik)	37
.....	37
3.11.8. Eksi yorum gücü (Negative predictive value)	37
3.11.9. ROC eğrisi altında kalan alan (AUC).....	37
3.11.10. Ortalama mutlak hata (Mean absolute error).....	38
3.11.11. Ortalama hata kareleri toplamı kökü (Root mean square error).....	38
3.11.12. Bağlı mutlak hata (Relative absolute error).....	38
3.11.13. Ortalama bağlı mutlak hata kareleri toplamı (Root relative squared error)	39
.....	39

	<u>Sayfa No</u>
3.11.14. F-Ölçütü (F-Ölçütü):.....	39
3.11.15. Kappa istatistiği	39
3.12. “n” Kat Çapraz Doğrulama Eğitim Verisi	40
3.13. Yüzdelik Eğitim Verisi	42
4. MR SPEKTROSKOPİ İLE BEYİN TÜMÖRÜ	43
4.1. MR Spektroskopi Nedir?	43
4.1.1. Voksel	43
4.1.2. Manyetik alan homojenizasyonu	44
4.1.3. Su ve yağ baskılanması.....	44
4.1.4. MRS incelenmesi.....	45
4.2. Metabolitler.....	46
4.3. MR Spektroskopi İle Beyin Tümörü Teşhisi	46
5.DENEYSEL ÇALIŞMALAR	49
5.1. Sınıflandırma Yöntemleri İçin Deneysel Çalışmalar	49
5.2. Kümeleme Yöntemleri İçin Deneysel Çalışmalar	53
5.3. Derin Öğrenme İçin Deneysel Çalışmalar	55
6.SONUÇLAR.....	59
KAYNAKLAR.....	61
ÖZGEÇMİŞ.....	65

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.1 WEKA programı giriş ekranı	4
Şekil 3.1 Veri madenciliğinin bileşenleri	7
Şekil 3.2 Sınıflandırma örneği.....	12
Şekil 3.3 Kümeleme örneği	14
Şekil 3.4 Veri madenciliği işlem süreci.....	15
Şekil 3.5 KNN örneği	18
Şekil 3.6 (a) Lineer bir çizgi ile ayrılabilen sınıflar, (b) doğrusal olarak ayrılamayan	19
Şekil 3.7 İki boyutlu uzaya verilerin yerleştirilmesi	19
Şekil 3.8 İki boyutlu uzaydaki veriler için oluşturulan doğru	20
Şekil 3.9 WEKA’ da non-lineer regresyon ile sınıflandırma örnek ağaç yapısı	21
Şekil 3.10 Logaritmik olarak oluşturulan çizgi	22
Şekil 3.11 Naive Bayes hesap için veri seti özelliklerin gösterimi	24
Şekil 3.12 Karar ağaçları hesap için veri seti özelliklerin gösterimi	27
Şekil 3.13 Karar ağaçları hesaplama veri seti karar ağacı	28
Şekil 3.14 Yapay sinir ağları modeli	29
Şekil 3.15 K ortalama iş akış şeması	30
Şekil 3.16 Kohonen yöntemi nöron dizilimi	31
Şekil 3.17 Derin öğrenme.....	31
Şekil 3.18 LeNet mimarisi.....	32
Şekil 3.19 Alexnet mimarisi	33
Şekil 3.20 Tanh aktivasyon fonksiyonu (Anwer, 2017).....	33
Şekil 3.21 ReLU aktivasyon fonksiyonu (Anwer, 2017)	34
Şekil 3.22 Sigmoid aktivasyon fonksiyonu (Anwer, 2017)	34
Şekil 3.23 ROC eğrisi altında kalan alan AUC	38
Şekil 3.24 10 kat çapraz doğrulamaya göre veri seti.....	41

Sayfa No

Şekil 3.25 10 kat çapraz doğrulamaya göre işlem aşamaları.....	42
Şekil 4.1 MRS işlem aşamaları (Cüce, 2005’den değiştirilmiştir.)	43
Şekil 4.2 Veksol seçimi (Başak, 2005).....	44
Şekil 4.3 (a) TE süresi kısa seçilen MRS görüntüsü. (b) TE süresi uzun seçilen.....	45
Şekil 4.4 Su baskısı sonucu oluşan bozuk spektrum (Başak, 2005).....	45
Şekil 4.5 MR Spektroskopi ölçümü ekran görüntüsü.....	47
Şekil 4.6 (a) Tümör olmayan alan ölçümü. (b) Tümörü olan alanda ölçüm	48
Şekil 5.1 Sınıflandırma yöntemleri için yapılan çalışmanın blok diyagramı	49
Şekil 5.2 Karar ağaçları yöntemi için ağaç yapısı	52
Şekil 5.3 Kümeleme yöntemleri için yapılan çalışmanın blok diyagramı.....	53
Şekil 5.4 WEKA programında Kohonen (Self Organizing Map) yöntemi için elde edilen kümeleme ekranı.....	54
Şekil 5.5 WEKA programında K means yöntemi için elde edilen kümeleme ekranı	54
Şekil 5.6 Python’ da elde edilen giriş özelliklerinin grafiksel gösterimi.....	56
Şekil 5.7 50.000 iterasyon için konfüzyon matrisi	57
Şekil 5.8 100.000 iterasyon için konfüzyon matrisi	57
Şekil 5.9 200.000 iterasyon için konfüzyon matrisi	58

ÇİZELGELER DİZİNİ

	<u>Sayfa No</u>
Çizelge 1.1 Veri madenciliği yazılımlarının bazıları.....	2
Çizelge 3.1 Veri madenciliği gelişim kronolojisi.....	8
Çizelge 3.2 Naive Bayes hesabı için örnek veri seti	23
Çizelge 3.3 Değer 1 özelliğinin olasılıkları	24
Çizelge 3.4 Değer 2 özelliğinin olasılıkları	24
Çizelge 3.5 Değer 3 özelliğinin olasılıkları	24
Çizelge 3.6 Değer 4 özelliğinin olasılıkları	24
Çizelge 3.7 Karar ağaçları hesap için veri seti	26
Çizelge 3.8 ROC eğrisi elde etmede kullanılan adı.....	35
Çizelge 3.9 ROC eğrisi test sonucu ifadeleri.....	35
Çizelge 3.10 ROC eğrisi ölçütleri için kontenjans tablosu.....	36
Çizelge 3.11 Örnek Kappa katsayısı hesabı	40
Çizelge 5.1 Sınıflandırma yöntemleri istatistik sonuçları	50
Çizelge 5.2 Sınıflandırma yöntemleri doğruluk sonuçları	51
Çizelge 5.3 Kümeleme yöntemlerinin başarısı.....	55
Çizelge 5.4 Derin öğrenme yöntemi başarısı.....	57

SİMGELER VE KISALTMALAR DİZİNİ

p	:Tahmin edilen değer
\bar{a}	:Gerçek değerlerin ortalaması
n	:Veri sayısı
a	:Gerçekleşen değer
MRS	:Manyetik rezonans spektroskopisi
VM	:Veri madenciliği
VT	:Veri tabanı
BT	:Beyin tümörü
YY	:Yüzyıl
DVM	:Destek vektör makineleri
KNN	:K-En yakın komşu
MAE	:Mean Absolute Error
RMSE	:Ortalama hata kareleri toplamı kökü
RAE	:Relative absolute error
RRSE	: Ortalama bağıl mutlak hata kareleri toplamı
Cho	:Kolin
Cr	:Kreatin
Naa	:N-asetil aspartat
SVM	:Support vector machine
VA	:Veri ambarları
ROC	:Receiving operating characteristic
Sensitivity	:Duyarlılık
Specificity	:Özgüllük
1-Özgüllük	:Yanlış pozitif oranı
1-Duyarlılık	:Yanlış negatif oranı

Accuracy	:Genel doğruluk oranı
Positive Predictive Value	:Artı yorum gücü
Negative Predictive Value	:Eksi yorum gücü
AUC	:ROC eğrisi altında kalan alan
DP	:Gerçek durum pozitifken test sonucu pozitif
YP	:Gerçek durum negatifken test sonucu pozitif
DN	:Gerçek durum negatifken test sonucu negatif
YN	:Gerçek durum pozitifken test sonucu negatif
LM	:Lineer model
T	:Tesla
LM	:Lineer Model
CACHE	:Öne bellek
D	:Duyarlılık
Duyarlılık (Precision)	:Kesinlik
Recall	:Duyarlılık

1.GİRİŞ

21. Yüzyılda her alanda yazılımlar etkili şekilde kullanılmaktadır. Bilişim sistemlerinin bu kadar yoğun kullanılması beraberinde büyük verileri ve bu büyük verilerin depolanmasını beraberinde geliştirmiştir. Bu büyük verilerle oluşturulan veri tabanları ise insanoğlunu araştırmaya yöneltmiş ve mevcut bilgilerle faydalı bilgiler üretmeye yönlendirmiştir. Mevcut bilgiler kullanılarak istatistik ve yazılım temelli sistemler geliştirmek, bir meslek dalında karşılaşılan herhangi bir hataya tekrar tekrar düşme engellenebileceği gibi tıp alanında hızlı ve doğru teşhis, tedavi gibi birçok faydayı beraberinde getirecektir. Bir anlamda nasıl bir cevher madenden çıkarılıp işleniyor ve sonrasında insanoğluna yararlı bir ürün oluyorsa, veri madenciliği de mevcut veri tabanında yer alan verileri kullanıyor ve onları işleyip insanoğluna faydalı bir ürün sunuyor. Veri madenciliğinde bu işlem istatistik bilimi ve makine öğrenmesi gibi farklı teknikleri kullanarak yapılmaktadır (Kumdereli, 2012).

Beyin tümörü, ciddi bir hastalık olup gerekli müdahale yapılmazsa insanın temel ihtiyaçlarını karşıladığı yürüme, yemek yeme gibi fonksiyonları ortadan kaldıracabileceği gibi hasta yaşamının son bulmasına da neden olabilir. Ayrıca beyin tümörü teşhisi ciddi bir uzmanlık gerektirmektedir ve maalesef bu anlamda uzman kişi sayısı ülkemizde azdır. Beyin tümörü olmayıp beyin tümörü teşhisi konularak yapılacak bir cerrahi müdahale taşıdığı yüksek riskler nedeniyle çok farklı sorunlar oluşturabilir. Gerçekten beyin tümörü varsa ve buna teşhis konulup gerekli tedavi yapılmazsa da hasta olumsuz etkilenecektir.

Beyin tümörü teşhisinde, MR Spektroskopi kısa bir geçmişe sahip olsa da etkin bir teşhis yöntemi olarak kullanılmaktadır. Beyin tümörü gibi ciddi hastalıkların teşhisi tek bir yöntemle yapılmaz. Farklı tekniklerle teşhis doğrulanmaya çalışılır. MR Spektroskopide beyin tümörü teşhisi için tek başına bir yöntem değildir ancak teşhis koymada etkili bir yöntemdir.

MR Spektroskopinin teşhis koymada bu kadar etkili olmasını kolin (Cho) metabolitini ölçebilmesine bağlayabiliriz. Çünkü tümör olan bölgede hızlı bir hücre artışı söz konusudur. Hücre sayısının hızlı şekilde arttığı bölgede kolin (Cho) metabolit değeri yüksektir. Kreatin (Cr) ve N-Asetil Aspartat (Naa) metabolitleri de beyin tümörü teşhisinde önemlidir. Şöyle ki beyin kanaması gibi farklı hastalıklarda da bu metabolit değerleri değişiklik gösterebilir. Burada rahatsızlığın tespiti iyi bir uzmanlık gerektirmektedir.

Çünkü yukarıda bahsedildiği gibi metabolit değerleri doğru değerlendirilmez ise beyin kanaması tümör, tümörde beyin kanaması olarak teşhis edilebilir.

Çizelge 1.1 Veri madenciliği yazılımlarının bazıları

ÜRÜN	GÜÇLÜ NİTELİĞİ	ZAYIF NİTELİĞİ
WEKA	Görsel ara yüzü, kullanım kolaylığı, sonuçlarının kabul görmesi	Derin Öğrenme tekniğinin olmaması.
Clementine	Görsel ara yüzü, algoritma genişliği	Ölçeklenebilirlik
DataCruncher	Kullanım kolaylığı	Tek algoritma içermesi
Enterprise Miner	Algoritma derinliği, görsel ara yüzü	Kullanımı zor, kompleks ara yüzü
GainSmarts	Veri dönüştürme, algoritma seçeneği genişliği	Sınırlı görselleştirme
Intelligent Miner	Algoritma genişliği, grafiksel ağaç çıktısı	Algoritma seçeneği az
MineSet	Veri görselleştirme	Algoritma seçeneği az
ModelQuest	Algoritma genişliği	Sezgisel olmayan ara yüz seçenekleri
PRW	Kapsamlı algoritmalar, otomatik model seçimi	Sınırlı görselleştirme
CART	Ağaç opsiyonları derinliği	Dosya açılması zorluğu, sınırlı görselleştirme
Scenario	Kullanım kolaylığı	Dar analiz yolu
OLPARS	Çoklu istatistiksel algoritmalar, sınıf tabanlı görselleştirme	Eski ara yüz, dosya açılması zorluğu
See-5	Ağaç opsiyonları derinliği	Sınırlı görselleştirme, veri eksikliği
S-plus	Algoritma derinliği, görselleştirme, programlanabilme	Sınırlı çıkarım metodu, dik öğrenme eğrisi
WizWhy	Kullanım ve model kavrama kolaylığı	Sınırlı görselleştirme

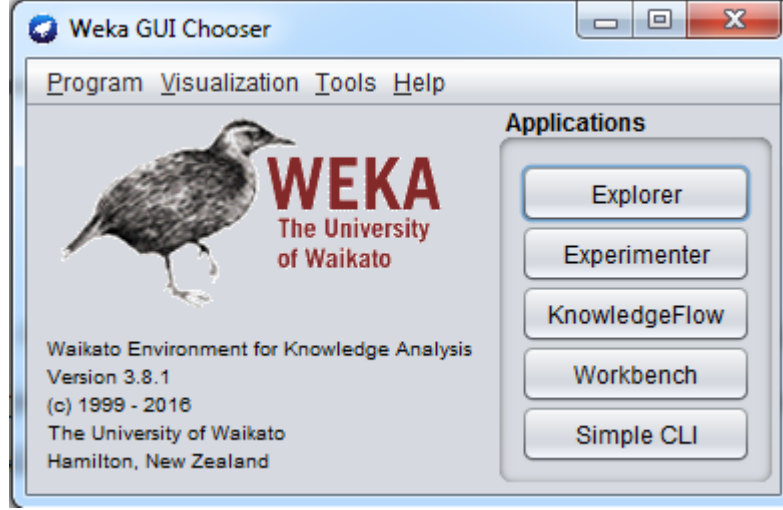
1.1. Veri Madenciliği Yazılımları

Veri madenciliğinin gelişimi ile beraber, veri madenciliği algoritmalarını kullanan yazılımlar geliştirilmiştir. Görsel içerikli işlem yapan yazılımlar kullanıcı tarafından daha rahat etkin şekilde kullanıldığından son zamanlarda geliştirilen yazılımlar görsel içeriklidir. Aynı zamanda veri madenciliği ile elde edilen sonuçlar grafik gibi görsel içerik taşırsa anlaşılması kolay olacaktır. Veri madenciliği yazılımları ile ilgili çeşitli bilgiler Çizelge 1.1’ de verilmiştir. Veri madenciliği ile uğraşan kişilerin hangi yazılımı seçeceği çok önemlidir ve yazılım seçerken şunlara dikkat etmeleri gereklidir:

- Yazılımın hangi teknikleri desteklediği.
- Yazılım kullanım kolaylığı.
- Yazılım sonuçları veya işlem süreci farklı yazılımlarla da kullanılabileceğinden diğer yazılımlar ile uyumu.
- İşlem sırasında karşılaşılabilecek bir sorun ile ilgili yardım desteğinin yeterli olup olmadığı.
- Sonuçların görsel olarak verilmesi ve anlaşılmasının kolay olması.
- Kullanılacak veri tabanı boyutu ile işlem yapıp yapamayacağı.

1.1.1. WEKA

Java tabanlı bir yazılımdır. Waikato Üniversitesi tarafından makine öğrenme projesi kapsamında hazırlanmıştır. Açık kaynak kodlu bir yazılım olup herhangi bir lisans ücreti ödmeden kullanılabilir. WEKA programı görsel ara yüzü sayesinde çok fazla talep görmektedir. Bunun yanı sıra komut satırı kullanarak programlama yapmaya da imkân tanımaktadır. WEKA programı ARFF, CSV ve C4.5 uzantılı veri dosyalara imkân tanımaktadır. WEKA programında kullanılacak verilerin ön işleme yapma imkânı bulunmaktadır. WEKA programı vasıtası ile sınıflandırma, kümeleme işlemleri görsel ara yüzler kullanılarak çok rahat yapılabilir (Girginer ve Cankuş, 2008).



Şekil 1.1 WEKA programı giriş ekranı

1.1.2. Theano

Yöntem, Montreal Üniversitesinde araştırma grubu tarafından geliştirilmiştir. Matematiksel ifadeleri yüksek başarılarla değerlendirmek üzere oluşturuldu. Theano' yu geliştirmek için çalışmalar hala sürmektedir (Anwer, 2017).

1.1.3. TensorFlow

Sayısal hesaplama için açık kaynaklı bir yazılım kütüphanesidir. Bir tek uygulama programlama ara yüzü ile mobil cihazındaki bir ya da daha fazla CPU'ları ya da GPU'ları uygulamamıza olanak tanır (Anwer, 2017).

1.1.4. Keras

Tensorflow ve Theano üzerinde çalışabilen bir uygulama programlama ara yüzüdür. Python açık kaynak kodlu yazılım dilinde yazılmıştır (Anwer, 2017).

Biz de çalışmamızda, python programlama dilinde “keras” derin öğrenme kütüphanesini kullandık.

1.1.5. Python

Açık kaynak kodlu bir yazılımdır. Çok fazla doküman bulunabilmesi önemli bir avantajıdır. Son yıllarda veri madenciliği ve derin öğrenme için yapılan akademik çalışmalarda sıklıkla kullanılmaktadır.

2.ÖNCEKİ ÇALIŞMALAR

Destek Vektör Makinelerini kullanarak çalışmada Nagori, M. Joshi, M., MRS görüntüsünü kullanarak kolin (Cho), kreatin (Cr) ve N-Asetil Aspartat (Naa) metabolit değerlerini okumuşlardır. Yapılan bu çalışmada kullanılan bilgilerin tamamı tümör teşhisi konulmuş ve farklı evredeki hasta bilgileridir. Bizim veri tabanımızda yer alan bilgiler ise tümör olan veya olmayan hastaların bilgileridir ve çalışmamız teşhis koymada doktora yardımcı bir sistem temeli olabilecek bir çalışma özelliği taşımaktadır. Nagori, M. Joshi, M. çalışmalarında 88 hasta bilgisini kullanmışlardır ve bu bilgiler farklı hastanelerden alınmıştır. Çalışmalarının temelini oluşturan MRS görüntüsüne göre değerleri okuyup, kendi oluşturdukları veri tabanına yazma başarıları %100 olarak gerçekleşmiştir. Okuyup veri tabanına yazdıkları verilere göre hastanın hangi evrede olduğunu anlamada aynı başarıyı gösteremeyip %66 başarı sağlamışlardır (Nagori ve Joshi, 2013).

Tümör olan ve tümörün hangi evrede olduğunu tahmin etmeye yönelik başka bir çalışma Lukasa, A ve arkadaşları tarafından yapılmıştır. Sınıflandırma için Doğrusal Ayırma Analizi, Destek Vektör Makineleri algoritmalarını kullanmışlar ve bu algoritmaların LS ve RBF çekirdeklerini seçmişlerdir. Bizim çalışmamız ise beyin tümörü teşhisi konusunda tahmine dayalı olup çalışmamızda beyin tümörü evre bilgileri yoktur. Doğrusal Ayırma Analizi (PCA/LDA) ile MRS tam spektrum görüntüsünü kullanarak yaptıkları çalışmada %64,6995 başarı, Destek Vektör Makineleri algoritmasının (SVM), LS çekirdeğini kullandıkları çalışmada %93,98, RBF çekirdeğinde ise %97,81 başarı sağlamışlardır (Lukas ve ark., 2004).

Veri madenciliği hesaplamaları için çok önemli olan ROC eğrisi ve AUC alanı hesabını anlatan bir çalışma Tomak, L., Bek, Y. yapmıştır. Çalışmada tanı testleri kullanılmıştır. Çalışma sonucunda ROC eğrisi ve AUC alanı hesaplama görsel nitelik taşıdığından anlaşılması kolay olduğunu vurgulamışlardır (Alan, 2014).

ROC analizini detaylı olarak hesaplamalar ile anlatan Dirican A. çalışmasını tanı testleri için yapmıştır. Çalışmada "Uygulama örneğimizde BERA yönteminin duyarlılığı "1" özgülüğü "0.77" ve doğruluğu "0.83", CT yönteminin ise duyarlılığı, özgülüğü ve doğruluğu "1" olarak tespit edilmiştir (İşler ve Narin, 2012).

MRS spektral görüntüsünü kullanarak bir çalışmayı Carlos A ve arkadaşları yapmıştır. Bu çalışmada tümörün hangi evrede olduğunu tahmin etmeye yönelik olup, Gaussian Ayırıştırması ve Yapay Sinir Ağlarını teknikleri kullanılmıştır. Bu çalışmada MRS ile elde

edilen Cho, Cr ve Naa metabolit deęerleri kullanılmayıp MRS spektral grnts kullanılmıřtır. İlk nce Gaussian Ayırıtması ile ilgisi az olan zellikler ihmal edilmiřtir. alıřma sonucunda en yksek bařarı %93.847 olmuřtur (Arizmendi ve ark., 2014).

Carlos A ve arkadařları alıřmalarında, beyin tmr hastalarının MRS spektral grntlerini alarak kullanmıřlardır. Gaussian Ayırıtması ve Yapay Sinir Aęlarını yntemleri ile tmrn hangi evrede olduęunu tahmine dayalı bir alıřmadır. Beyin tmr teřhisinde ve tmrn hangi evrede olduęunu belirlemede nemli yere sahip olan Cho/Cr, Cho/Naa, Naa/Cr deęerlerini kullanmamıřlardır. alıřmalarında ilk nce Gaussian Ayırıtması teknięi ile zniteliklerden daha az olanları ıkarmıřlardır. Yaptıkları alıřmada sınıflamadaki en yksek bařarı %93.847 olarak kayıt edilmiřtir (Carlos ve ark., 2014).

Firmaların pazar arařtırmalarına destek saęlayabilecek bir alıřma Tařkın, ., Emel, G.G. tarafından yapılmıřtır. alıřma bir firmanın mřterilerinin kmelenmesini Kohonen kmeleme teknięini iermektedir. Byklkleri %11 ile %20 arasında deęiřen 8 adet kme elde etmiřlerdir. alıřmalarında, firmanın kendi mřterilerinin bilgilerini kullanmıřlar ve mřteri bilgilerinin hangilerinin pazar elde etmede nem arz ettięi hangilerinin nem arz etmedięi sonucunu da ıkarmıřlardır (Tařkın ve Emel, 2010).

Derin ęrenme yntemleri zerine bir alıřmayı, Derin, A., Diri, B., Balık, H.,H. Yaptmıřtır. alıřmalarında derin ęrenme ve geliřim sreci hakkında bilgiler vermiřlerdir (Derin ve ark., 2017).

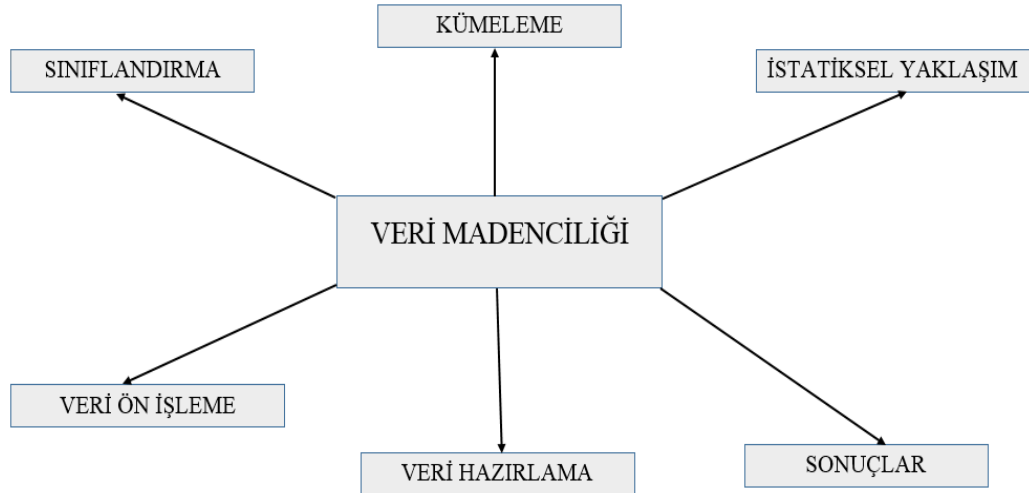
Anwer, O., M., A., UCI makine ęrenmesi veri havuzundan gęs kanseri ile ilgili veriler alarak alıřma yapmıřtır. alıřmada Python programlama dilini kullanılmıřtır. Konvolsyon yapay sinir aęları (CNN) ve makine ęrenmesi yntemlerini kullanarak %96 gibi yksek bařarı ile sınıflandırma yapılmıřtır (Anwer, 2017).

3. VERİ MADENCİLİĞİ

3.1. Veri Madenciliği Nedir?

Veri madenciliği (VM), günümüzde devasa boyutlara ulaşan verilerin özellikle istatistik alanında kullanılarak yararlı bilgiler elde etmeye denilebilir. Aslında VM kelime anlamına bakıldığında da faydalı bir ürün elde edilebilecek cevherin çıkartılması olduğunu görüyoruz. VM mevcut bilgilerden geleceğe yönelik karar vermek için kişiye tahmin etme aracı sağlar. İstenilen faydalı sonuç bilgileri, veri tabanında yer almayan ancak veri tabanından (VT) çıkarılabilecek potansiyel bilgilerdir. Çağımızda resmi ve özel kurumlar tarafından büyük miktarda veriler üretilmekte ve saklanmaktadır. Ancak bu veriler bu haliyle hiçbir anlam ifade etmemektedir. Bu veriler çok büyük olduğundan bu veri yığımına veri ambarları demek de mümkündür. Bu büyük verileri işlemek istatistik yöntemleri açısından oldukça zorlayıcıdır ve daha gelişmiş sistemler ile bu verilerin işlenerek faydalı hale gelmesi zorunlu hale gelmiştir (Yurdakul, 2015; Kumdereli, 2012; Hark, 2013).

Bilgisayarın gelişim sürecine bakıldığı zaman, ilk başta küçük boyutlu veriler bilgisayar aracılığı ile analiz edilebilmiştir. Ancak bilgisayar teknolojisi geliştikçe veri boyutları artmış ve veri tabanları oluşturulmuştur. Bu veri tabanlarındaki verilerin analizi SQL kullanılarak yapılmıştır. Ancak gelişim arttıkça verilerin boyutu da artmış ve veri ambarları oluşturulmuştur. Bu veri ambarlarındaki verilerin analizi ise yapay zeka kullanılarak yapılır hale gelmiştir (Yurdakul, 2015).



Şekil 3.1 Veri madenciliğinin bileşenleri

3.2. Veri Madenciliği Tarihi

Veri madenciliği yapay zekâyı kullanarak veri işleme amacıyla 1960'lı yıllarda matematik bilimi tarafından araştırılmaya başlanmıştır. Bu başlangıç aşamasında bilim insanları makine eğitime dayalı sistemler üzerinde durmuşlardır. 1960'larda ise istatistik bilimi ile uğraşan insanlar bu keşfe dâhil olmuş ve yeni algoritmalar keşfetmişlerdir. Daha sonraki süreçte bilgisayar biliminin gelişmeye başlaması ve buna paralel olarak veri boyutlarının artmasına bağlı olarak veri tabanları daha büyük veriler için veri ambarları oluşturulmaya başlanmıştır. Bu veri tabanları ve veri ambarları (VA) içerisindeki verilerin arasındaki ilişki araştırılmaya devam edilmiştir. Böylece veri madenciliği gelişimini hızlandırmıştır. Veri madenciliğinin gelişim kronolojisi Çizelge 3.1' de görülmektedir (Girginer ve Cankuş, 2008).

Çizelge 3.1 Veri madenciliği gelişim kronolojisi

Gelişim Adımları	Cevaplanan Karar Problemi	Kullanılabilen Teknolojiler
Veri Toplama (1960'lar)	İlimizde teşhis konmuş hastanın kan değerleri.	Bilgisayarlar
Veri Erişimi (1980'ler)	İlimizde Başlangıç, orta ve ileri seviyedeki hastasının kan değerleri nelerdir.	Veri tabanları
Veri Ambarları ve Karar Destek Sistemleri (1990'lar)	Ülkemizde geçen yıl teşhis konmuş hasta bilgileri satışları ne kadar?	Çok boyutlu veri tabanı sistemleri, veri ambarları
Veri Madenciliği (Bugün)	Kan değer ölçümü yaptıran hastanın teşhisi nedir ve hangi evrededir?	Çok büyük veri ambarları, çok hızlı işlem yapan bilgisayarları.

3.3. Veri Madenciliği Uygulama Alanları

3.3.1. Finans alanında veri madenciliği uygulamaları

Bankacılıkta müşterilerin kredi kartı taleplerinin karşılanıp karşılanamayacağının tespiti, müşteri kredi kartı limit artış talebinin karşılanıp karşılanamayacağı, kart hırsızlıkları tespiti gibi birçok sayıda bankacılık işleminde VM kullanılabilir. Ancak

burada önemli olan daha doğru bir sonuç almak için verilerin güvenilirliğidir (Yurdakul, 2015; Kumdereli, 2012; Hark, 2013).

3.3.2. Tıp alanında veri madenciliği uygulamaları

Kanser gibi ciddi hastalıkların tespitinde, kronik rahatsızlıkların teşhisinde, salgın bir hastalığın tespitinde, uygulanan bir tedavinin başarılı olup olmadığının tespiti gibi birçok karar teşhis işlemlerinde kullanılır. Ayrıca ilaç sanayinde ilaçların farmakolojik yapısı tespitinde VM yaygın olarak kullanılmaktadır (Yurdakul, 2015; Kumdereli, 2012; Hark, 2013).

Bizde çalışmamızda, beyin tümörü teşhisinde doktora yardımcı olmak için geliştirilecek olan sistemin temelini oluşturmak amacıyla eldeki verileri kullanan makine öğrenmeye dayalı bir çalışma yaptık.

3.3.3. Perakende alanında veri madenciliği uygulamaları

Günümüzde, rekabet şartları oldukça zorlaşmıştır ve firmalar doğru müşteri kitlesine doğru zamanda erişmesi çok büyük önem arz etmektedir. Doğru yönetilemeyen bir yapı ise zarara hatta iflasa firmaları götürmektedir. İşte perakende alanda veri madenciliği uygulamaları firmalara, müşteri alışveriş eğiliminin belirlenmesi, müşteri memnuniyetinin belirlenmesi hatta ürünün müşteri ulaşım aşamasında maliyetlerin düşürülmesi gibi birçok alan iyi sonuçlar verir (Hark, 2013).

3.3.4. İletişim alanında veri madenciliği uygulamaları

Günümüzde sosyal ağlar oldukça yaygın olarak kullanılmaktadır. Hatta sosyal ağlar çok fazla bir geçmişe sahip olmasa bile önceki nesillerde sosyal ağlara merak duymuştur. Sosyal ağların hangi zamanlarda daha çok kullanıldığı, hangi içeriklerin daha fazla kullanıldığı verileri kullanarak yapılan bir analiz firmaların sosyal ağları kullanarak reklam yapabilmelerine olanak sağlamaktadır. Sosyal ağlardaki verilerin büyüklüğü düşünüldüğünde analizi yapabilecek sistem gelişmiş bir sistem olan veri madenciliğidir. VM ile oluşturulan bir reklam yapısı ile firmalar doğru zamanda reklamını yapacak ve reklam maliyetlerini düşürecektir. Bu da firma açısından her açıdan kazanç demektir (Hark, 2013).

3.4. Veri Madenciliğinde Karşılaşılan Problemler

3.4.1. Gürültü

Veri tabanında yer alan veriler arasında yanlış veriler yer alabilir ve yanlış verilere gürültü diyebiliriz. Verilerdeki bu gürültünün nedeni verinin yanlış girilmesi veya eksik girilmesinden kaynaklı olabilir. Bazı veri tabanları eksik girilen veriyi tamamlasa veya yanlış girilen veriyi düzeltse de pek yaygın değildir. Gürültü veri madenciliği için çok büyük bir problemdir. Bu nedenden dolayı veri madenciliği sistemi işleme başlamadan önce gürültüyü tespit edip bunu yok saymalıdır (Çokluk, 2010).

3.4.2. Boş değer

Bir veri tabanında, bazı verilerin sahip oldukları özelliklere diğerleri sahip olmayabilir ya da o veri o özelliğe sahip değildir. Veri madenciliğinde, veri, tabanında yer alan boş veriler için 2 farklı yöntem uygulanmaktadır. Bunlardan ilki bu değer yok sayılması, ikincisi ise bu veri için diğerlerinden elde edilen değer bu boş değere atanmasıdır (Çokluk, 2010).

3.4.3. Eksik veri

Bir veri tabanında yalnız belirli bir gruba ait veriler yer alırsa, farklı bir gruptaki veriyi incelemek doğru sonuç vermez. Mesela herhangi bir hastalık için sadece kadınlara ait kan değerleri veri tabanında yer alırsa bu veri tabanını kullanarak erkek hastaya ait sonuç alınamaz, alınsa bile bu sonuç bağlayıcı kabul edilemez (Çokluk, 2010).

3.4.4. Dinamik veri

Dinamik yapıları veri tabanlarındaki veriler süreklilik arz etmez. Yani herhangi bir veriye göre yapılan makine eğitimi verinin değişmesi ile değişir ve sonuç farklılık gösterir. Dinamik sistemler, tetiklemeli bir veri tabanına sahip sistemlerdir ve bu tetiklemeyi kullanarak veriler veri madenciliğinde kullanılabilir (Çokluk, 2010).

3.5. Veri Madenciliği Faydaları

Veri madenciliği faydalarını şu şekilde sıralayabiliriz (Çokluk, 2010):

- Firmalar müşterilerini sınıflara ayırarak bu sınıfların ihtiyaçlarına göre doğru pazarlama stratejileri geliştirebilirler.
- Firmalar kampanyalarını veri madenciliğinin verdiği imkânlar doğrultusunda farklılaştırarak doğru müşteriye ulaşmasını sağlayabilirler.

- Firmalar müşterilerini net tanıyarak müşterisinin neye ihtiyacını olduğunu yorumlayabilir.
- Öğrenci bilgileri VM ile analiz edilerek başarı ve başarısızlık nedenleri ortaya konulabilir.
- Kronik rahatsızlığı olanları tedavi süreci takip edilerek etkin bir tedavi sağlanabilir.
- Beyin tümörü gibi teşhisi ciddi uzmanlık gerektiren rahatsızlıkların zaman kaybetmeden teşhisinde doktora yardımcı olabilir.
- Kalp krizi gibi beklenmedik anda gelen hastalıkların önceden tahmin edilmesi ve ona göre tedbir alınması sağlanabilir.
- Büyük verilerin yer aldığı veri ambarlarında (VA), benzer veriler ortaya çıkarılabilir.
- Sosyal ağlarda kişilerin ilgi alanı belirlenip firmalar ona göre reklam politikası geliştirebilir.
- Üniversite girişlerindeki okulların başarıları ölçülebilir.
- Laboratuvar ortamında elde edilen yüksek veriler daha kolay analiz edilebilir.
- Borsada portföy yönetimi daha kolay yapılabilir.
- Müşterilerin ödeme planları analiz edilerek kredi notu ayarlaması yapılabilir.

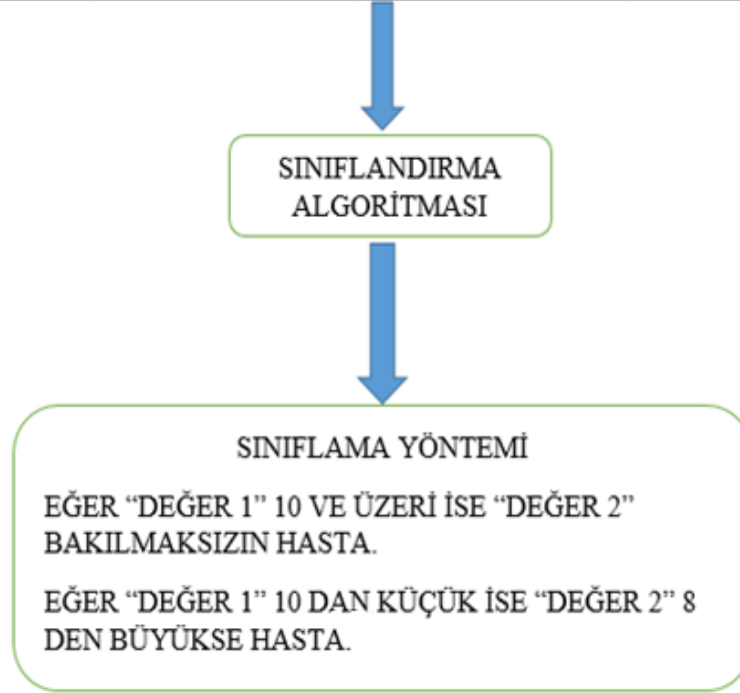
3.6. Veri Madenciliği Modelleri

3.6.1. Sınıflandırma ve regresyon

Sınıflandırma ve regresyon, eldeki eğitim verisine göre yeni gelen verinin hangi sınıfa dâhil edileceğinin belirlenmesidir. Yani sınıflandırmada işlem sonucunda elde edilecek sınıflar önceden belirlidir. Ayrıca hangi sınıftaki verinin hangi özelliklere sahip olduğu eğitim verisi olarak makine eğitiminde kullanılır. Sınıflandırma ve regresyon en yaygın kullanılan veri madenciliği tekniğidir. Sınıflandırma ve regresyondaki temel fark tahmin edilen ilişkili değişkenin kategorik veya süreklilik göstermesidir. Son zamanlardaki çalışmalarda sınıflandırma ve regresyon giderek birbirine yaklaşmıştır. Sınıflandırma teknikleri ve regresyon teknikleri için kullanılan algoritmalar benzerdir. Örneğin belirli bir hastalığa sahip kişilerin kan değerleri yaşı, cinsiyeti gibi bilgilerine göre bir eğitim veri seti oluşturulabilir ve yeni gelen hastanın bu hastalığa sahip olup olmadığı tespit edilebilir. Bir firma bir ürünü ile ilgili yaptığı reklamın ürün satışını nasıl etkilediğini, reklam yapılan saati, reklam süresi ve görsel içeriklere bakarak bir eğitim veri seti oluşturabilir. Bu eğitim veri setine göre, yeni bir ürün için yapılacak reklamın ne şekilde yapılacağına dair firma yetkililerine bilgi verecektir ve böylece gereksiz reklam masraflarından firmayı

kurtaracaktır. Başka bir örnekte ise finans kuruluşu kişilerin gelir seviyesi, harcama bilgileri gibi verilerden bir eğitim veri seti oluşturularak müşterilere kredi not sistemi geliştirebilir (Çokluk, 2010; Sattari, 2009; Özcan, 2014; Yurdakul, 2015).

HASTA ADI	DEĞER 1	DEĞER 2	SONUÇ
AHMET	10	2	HASTA
MEHMET	15	27	HASTA
YELİZ	8	14	HASTA
HAMİT	5	3	HASTA DEĞİL
AYNUR	3	4	HASTA DEĞİL



Şekil 3.2 Sınıflandırma örneği

Yaygın olarak kullanılan sınıflandırma ve regresyon veri madenciliği teknikleri: Yapay Sinir Ağları, Destek Vektör Makineleri, Karar Ağaçları, Genetik Algoritmalar, K-en yakın komşu, Naive Bayes. Sınıflandırma 2 aşamadan olmak üzere, eğitim setinin oluşturulması ve eğitim setine göre tahmin işlemi.

3.6.2. Kümeleme

Kümeleme tekniğinde, sınıflandırma yönteminin sahip olduğu eğitim veri seti yoktur. Belirli özellikleri olan veriler belirli özelliklerine göre kümelere ayrılırlar. Bu küme ayırma işlemi mesafe ölçümüne göre yapılır. Verilerin kaç kümeye ayrılacağı uzman kişi tarafından belirlenebileceği gibi, bilgisayar programları aracılığıyla da yapılabilir. Kümelemede eğitim veri seti olmadığından makine eğitimi yoktur. Kümeleme tekniğinde

farklı yöntemler vardır. Bunları şöyle sıralayabiliriz (Sattari, 2009; Özcan, 2014; Yurdakul, 2015).

3.6.2.1. Paylaştırma temelli yöntemler

Oluşturulan her kümede en az bir eleman yer alır. Büyük verilerin oluşturduğu veri ambarlarında pekiyi sonuç vermeyen bu yöntem küçük ve orta büyüklükteki veri tabanlarında kullanılır ve iyi sonuçlar verir. Bu yöntem verileri küre şeklindeki kümelere ayırır. Bu yöntem altına k-ortalama algoritması alınabilir (Hark, 2013).

3.6.2.2. Hiyerarşik temelli yöntemler

Bu yöntemde bütün veri seti aynı kümeye atılır daha sonra veriler farklı kümelere ayrılır. BIRCH ve ROCH algoritmaları bu yöntemin algoritmalarıdır (Hark, 2013).

3.6.2.3. Yoğunluk tabanlı yöntemler

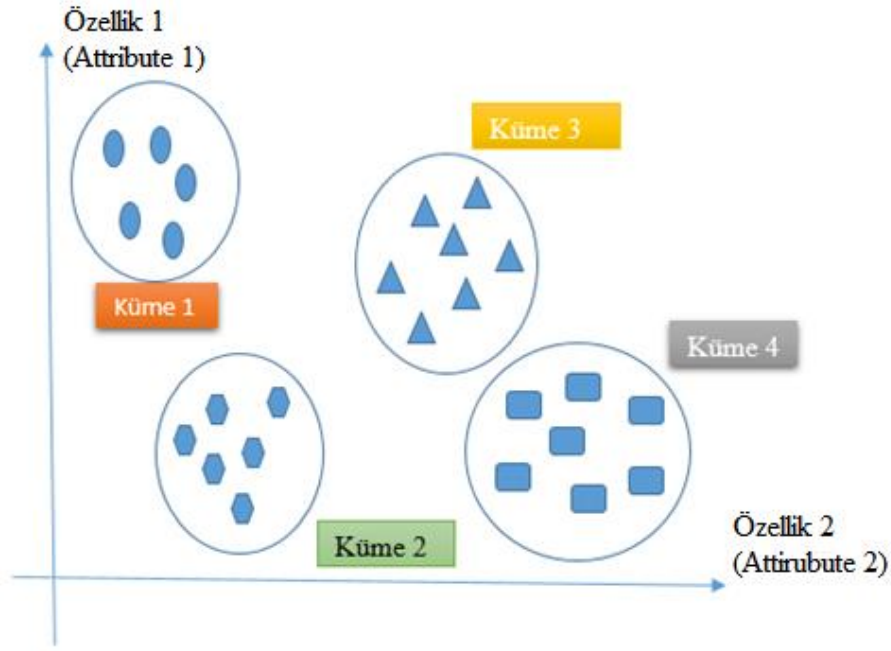
Verileri rastgele bulup kümelere ayırır. DBSCAN ve OPTICS bilinen algoritmalarıdır (Hark, 2013).

3.6.2.4. Izgara tabanlı yöntemler

Veriler ızgara yapısına sahip bir uzaya taşınır ve kümeleme işlemi bu ızgaralar üzerinde yapılır. STING algoritması ızgara tabanlı kümelemeye tabidir (Hark, 2013).

3.6.2.5. Model tabanlı yöntemler

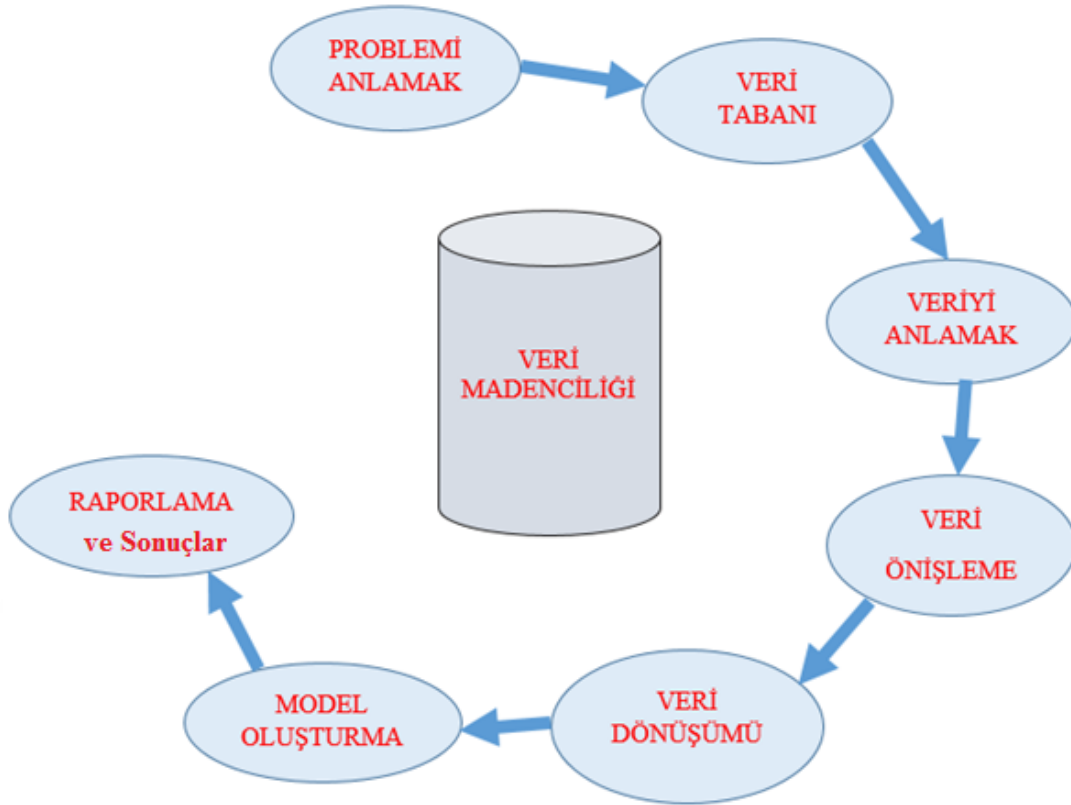
Oluşturulan her bir küme için bir model oluşturulur ve veriler bundan sonra kümelere ayrılır. EM ve COBWEB algoritmaları model tabanlı yöntemlerdir (Hark, 2013).



Şekil 3.3 Kümeleme örneği

3.7. Veri Madenciliği İşlem Akışı

Veri madenciliği tek başına bir işlem olarak görülüp ona göre işlem yapılırsa yanlış yapılmış olur. Çünkü veri madenciliği bir işlem süreci sonunda elde edilen anlamlı verilerdir. Bu işlem sürecinin adımlarını doğru tanımlamak ve işlem süreci adımlarını eksiksiz yapmak veri madenciliğinde daha doğru sonuç verecektir. Oluşturulan bilgisayar programları veri madenciliği işlem akışını net bir şekilde yapmaktadır. Özellikle son yıllarda oluşturulan programlar hem iş akış sürecinde kullanıcıya görsel olarak iş yapma imkânı sağlamaktadır. Bu da veri madenciliğine yeni başlayanlara büyük bir imkân tanımaktadır. Ayrıca bu kolay işlem süreci yeni başlayanların veri madenciliğine ilgisini arttırmaktadır. Veri madenciliği işlem sürecini oluşturan aşamalar aşağıda detaylı olarak anlatılmıştır (Çokluk, 2010; Özcan, 2014; Yurdakul, 2015).



Şekil 3.4 Veri madenciliği işlem süreci

3.7.1. Problem nedir?

Uygulama yapılacak alanda nasıl bir sonuç elde edilmek istendiği açık olarak bilinmelidir. Yani işlem sonunda nasıl bir sonuç alınırsa işlem başarılı sayılacaktır bunun kararı yapılan işin bir şeye yarayıp yaramadığını anlamak için oldukça önemlidir. Aslında veri madenciliğinin temeli bu aşama diyebiliriz. Çünkü her işte olduğu gibi ne yapılmak istendiği bilinmeyen bir işlem sonuç vermeyecektir. Tıp alanında, bankacılık alanında, ilaç sanayi alanında, sosyal ağlar aracılığı ile reklam alanı gibi her alanda ne yapılmak istendiği ve nasıl bir sonuca varılmak istendiği net olarak belirlenirse yapılan işlem bir işe yarayacaktır. Aksi halde verilerin toplanması, değerlendirilmesi uygun yöntemin seçilmesi gibi süreçlerde çok fazla zaman kaybedilecektir. Bu durum firmaların rekabet ortamını yakalayamaması, teşhis konulacak bir hastalığa tedavi açısından geç kalınması gibi birçok açıdan zararlı olacaktır. Ters açıdan bakıldığında ise veri madenciliği kullanılarak yapılacak bir işlemin doğru anlaşılması ve ne istenildiğinin net olarak bilinmesi alınan sonuçla firmanın rekabet gücünün artmasını sağlayacaktır. Veri madenciliğini kullanan sistemlerde bazı hastalığın teşhisi hızlanacaktır (Özcan, 2015).

3.7.2. Veriyi anlamak

Veri tabanında yer alan verilerin neleri içerdiği, verilerin özelliklerinin neler olduğunu anlamak oldukça önemlidir. Veri madenciliği işleminin 2. ayağı olan bu bölümde verilerin kalitesi değerlendirilir. Birden fazla ortamdan toplanan verilerin uyumluluğunu ve işlem için uygunluğu bu bölümde test edilir. Veriyi anlama safhası doğru yapılmazsa bu durum doğrudan sonucu olumsuz etkileyecektir. Ayrıca veriyi anlamadan işlem yapmak, ne yapılmak istendiği bilinse dahi sonuca varan yolda neleri kullandığını bilmemektir (Özcan, 2014).

3.7.3. Veri hazırlanması

Veri madenciliği işlem sürecinde en uzun zamanın ayrıldığı kısım bu aşamadır. Bu aşamada yapılacak yanlış bir işlem doğrudan sonucu olumsuz etkileyecektir. Ayrıca veri madenciliği işlem sürecinde yapılacak bir hatanın tespiti sırasında bakılacak ilk safhadır. Burada yapılacak işlerin kalitesi doğrudan sonucunu belirleyecektir. Veri hazırlanması işlemi aşamasında yapılan işlemler şu şekilde sıralanabilir (Özcan, 2014; Yurdakul, 2015).

- Veri seçimi.
- Veri temizleme.
- Veri kurma.
- Veri birleştirme.
- Veri formatlama

3.7.4. Modelleme

Veri madenciliği işlem sonucunda, istenilen sonucu alabilmek için modelleme kısmında farklı algoritmalar denenmelidir. Seçilen bir modelin doğruluğunu test etmek için, eldeki veri setinin %5 ile %33 arasındaki kısmı modelde test verisi olarak kullanılır. Kalan kısım ise test işlemi için kullanılır. Bu işlem makine eğitimi içeren sınıflandırma teknikleri için geçerlidir. Kümeleme tekniğinde ise, uzman kişi tarafından önce verilerin kaç kümeye ayrılacağı sisteme bildirilir veya bilgisayar verilerin özelliklerine göre kaç küme oluşturacağını kendisi belirler (Çokluk, 2010; Özcan, 2014; Yurdakul, 2015; Kumdereli, 2012).

3.7.5. Değerlendirme

Veri madenciliğinde son aşama olup, burada elde edilen sonucun istenilen sonuç olup olmadığı, veri madenciliği işlem aşamalarında olabilecek hataların kontrolü ve düzeltilmesi sağlanır. Değerlendirme ile sonuç istenilen şekilde ise bu sonucun daha geniş

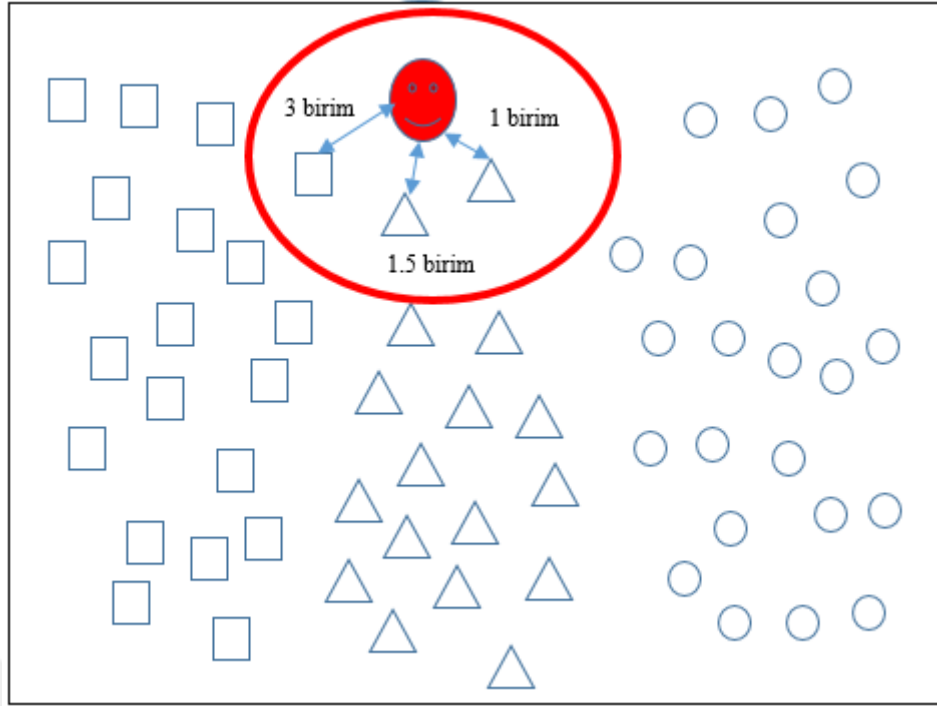
kitlelere veya ilgililere yayılması sağlanabilir. Bu da işlem süreci ve özellikle de sonuçların iyi anlatılması ile mümkündür. Birçok işlemde olduğu gibi veri madenciliği ile elde edilen sonuçlar da görsel olarak daha iyi anlatılabilir. Son yıllarda oluşturulan veri madenciliği programları görsel sonuç vermekte ve veri madenciliği işlem sonuçlarının kolay anlaşılmasını sağlamaktadır.

3.8. Çalışmada Kullanılan Veri Madenciliği Sınıflandırma Yöntemleri

3.8.1. KNN

Sınıflandırma için en yakın “k” kadar komşulara bakan algoritmadır. 1950’lerde literatüre kazandırılmış bir algoritmadır. Bu algoritmada boyut sınırı yoktur yani 2 boyutlu veya daha çok boyutlu uzayda sınıflandırma yapabilir. Genelde “k” değeri tek sayı alınır bunun sebebi, çift sayı seçilirse eşit sayıda farklı sınıf elemanı komşu olabileceğinden, sistem yeni verinin sınıflandırmasını yapamaz. Komşulara bakarken uzaklık hesaplamak için farklı teknikler vardır bunlar: Euclidean Distance, Manhattan ve Minkowski’dir. WEKA programında KNN algoritmasını kullanarak yaptığımız çalışmalarda “k” değerini kendimiz seçebiliriz. “k” değeri seçimi belirli değere kadar başarı oranını arttırsa da belirli bir değerden sonra başarıyı düşürdüğü gözlemlenmiştir (Kaya ve Pehlivan, 2014; Yücebaş, 2016).

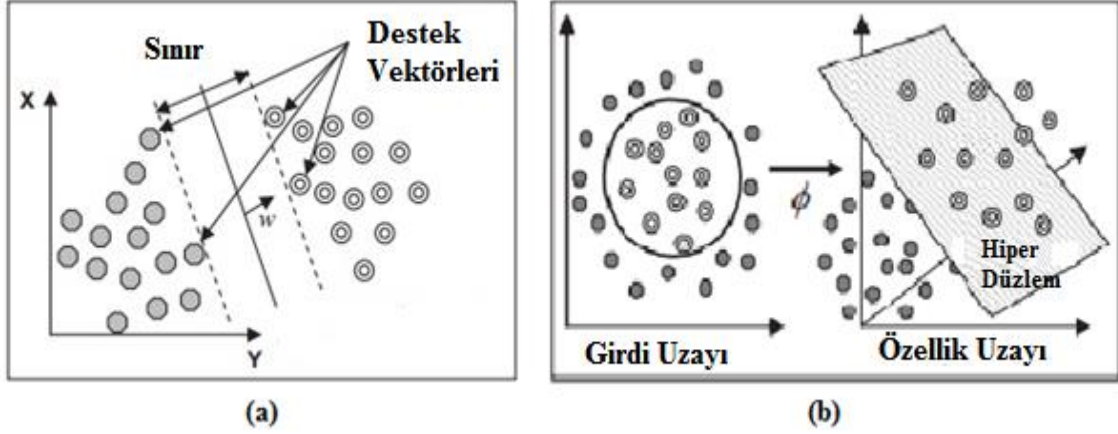
Şekil 3.5’ de görüleceği üzere bir veri seti kare, üçgen ve yuvarlak şekilleri olarak 3 sınıftan oluşsun. Veri setinde yer alan veriler sahip oldukları özelliklere (attribute) göre örnek uzaya yerleştirilsin. Sınıflandırılmak istenen “gülen yüz” verisi de sahip olduğu özelliklere göre yerleştirilir. KNN sınıflandırma tekniği için tanımlanan “k” değerine göre kendisine en yakın örneklerle bakılır. Şekil 3.5’ de “k” değeri 3 seçilmiş olsun ve buna göre en yakın 2 adet üçgen 1 adet kare yer almaktadır. Dolayısıyla “gülen yüz” verisi üçgen sınıfına dâhil olur.



Şekil 3.5 KNN örneği

3.8.2. Destek vektör makineleri

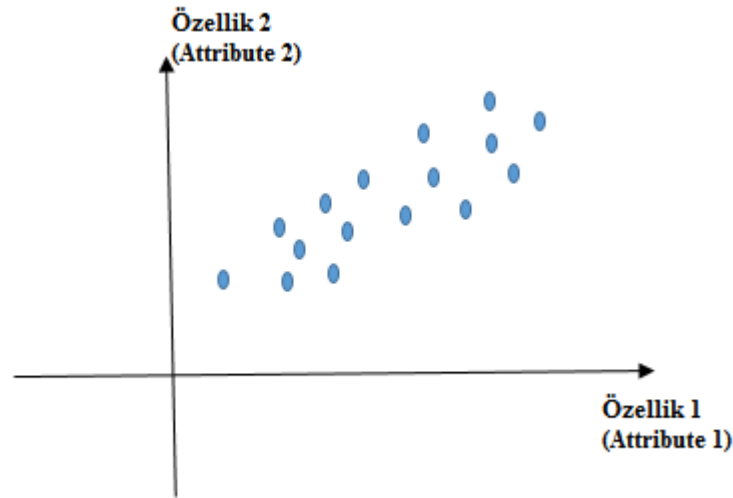
2 adet sınıf oluşturulup sınıflandırılma yapılacağı zaman kullanılan destek vektör makineleri zamanla geliştirilmiş ve 2'den daha fazla küme oluşturacak hale getirilmiştir. 2 adet sınıf oluşturulup yapılacak işlemde destek vektör makineleri çok iyi sonuçlar verir. Destek vektör makinelerinin çalışma yapısı 2 farklı sınıf arasına lineer bir çizgi çekilir ve her bir küme elemanına bu lineer çizginin uzaklığı maksimumdur. Şekil 3.6(a).’da lineer bir çizgi ile ayrıştırılmış 2 küme gösterilmektedir. Lineer bir çizgi kullanılarak ayrılamayan sınıflarda ise durum biraz daha farklıdır. Şekil 3.6(b)’de lineer bir çizgi ile ayrılamayan 2 farklı sınıfın çok boyutlu bir uzaya taşınarak ayrıştırılması gösterilmektedir (Kaya ve Pehlivan, 2014; Yücebaş, 2016; Kavzoğlu ve Çölkesen, 2010).



Şekil 3.6 (a) Lineer bir çizgi ile ayrılabilen sınıflar, (b) doğrusal olarak ayrılamayan verilerin çok boyutlu düzleme taşınarak ayrılması

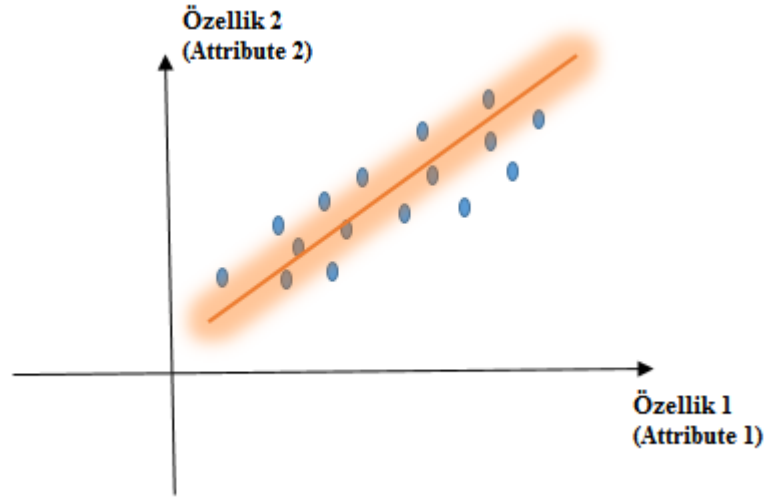
3.8.3. Lineer regresyon

Lineer regresyon algoritması verilerin “n” boyutlu bir uzaya aktarılmasından sonra sınıflandırılmasıdır. Boyut sayısını veri setinin özellik sayısı belirler. Örneğin yaş ve kilo bilgisi olan bir veri seti 2 boyutlu iken yaş, kilo, cinsiyet özellikleri olan bir veri seti 3 boyutlu uzaya taşınarak işlem yapılır (Şeker, 2017).



Şekil 3.7 İki boyutlu uzaya verilerin yerleştirilmesi

Veriler “n” boyutlu uzaya taşındıktan sonra bu verilere en yakın uzaklıktan geçen bir doğru çizilir. Bu doğru için formülü $y=ax+b$ şeklinde genellemek mümkündür. Özellik sayısının 2’ den fazla olduğunu düşünürsek eğer işte bu algoritma tamda karmaşık olan özellikleri ve bu özelliklerin matematiksel karşılığını basite indirgeyerek bir doğruya çeviriyor.



Şekil 3.8 İki boyutlu uzaydaki veriler için oluşturulan doğru

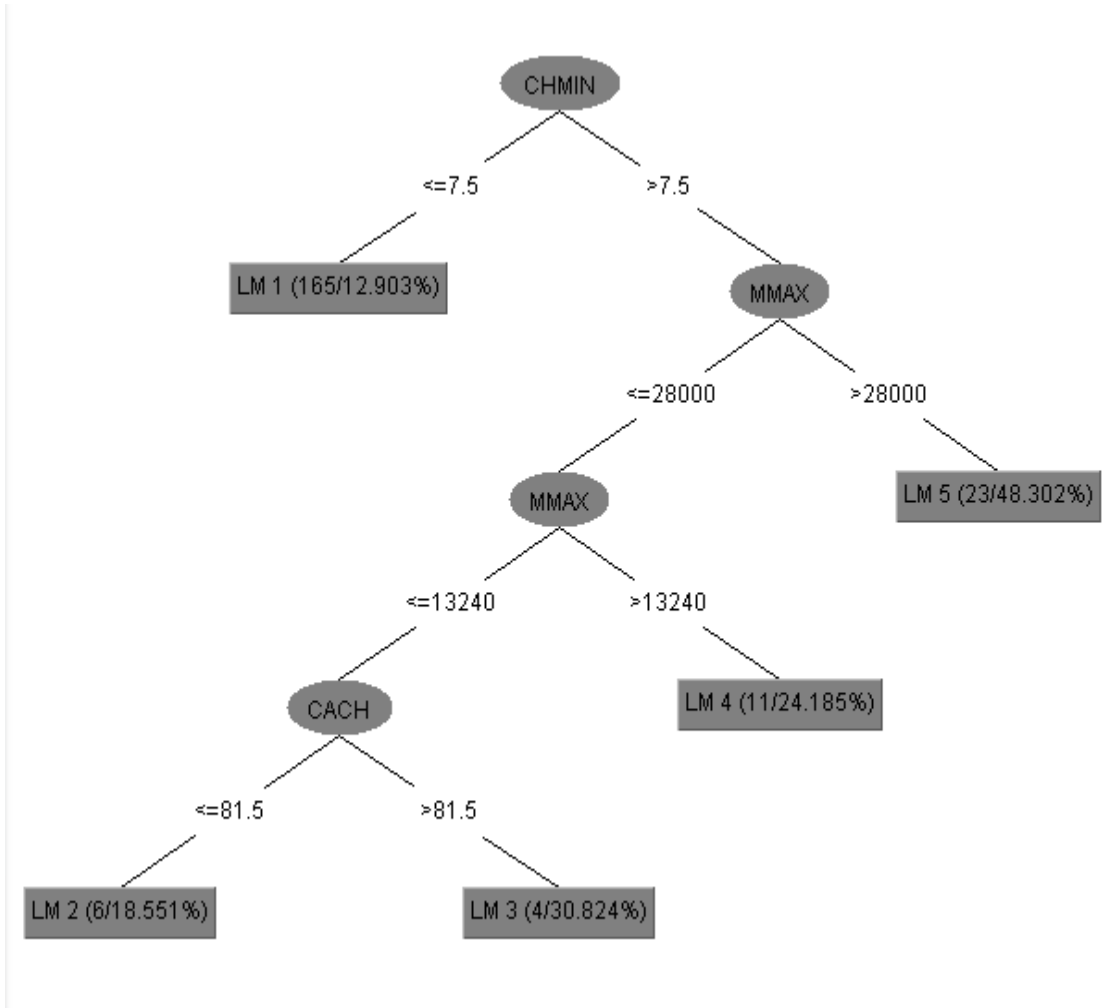
$$\text{asg } H(a, b) \text{ için } H(a, b) = \sum_{i=1}^n h_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad (3.1)$$

Çizilecek olan doğrunun bütün noktalara en kısa mesafede olması gerektiği daha önce belirtmiştir. İşte bunun için doğru denkleminde bu hata oranları en az olmalıdır ve bu toplam hata oranını H olarak isimlendirirsek Denklem (3.1) bize oluşturulacak doğrunun genel denklemini verir (Şeker, 2017).

3.8.4. Non-lineer regresyon

Regresyon veriler arasındaki ilişkiye dayanarak geleceği tahmin etmeye dayalı bir yöntemdir. Veriler serpmeye yöntemi ile özellik sayısının belirlenecek boyuttaki bir uzaya taşınması doğrusal bir doğru çizilemeyebilir. Veri setindeki özellik sayısı artınca bu durum daha belirgin hale gelir. Doğrusal çizgi çekebilecek hale getirmek için WEKA programı karar ağaçlarındaki dallanmaya benzer bir dallanma yapar ve işlemin en sonunda Lineer Model'ler oluşturur (Şeker, 2017).

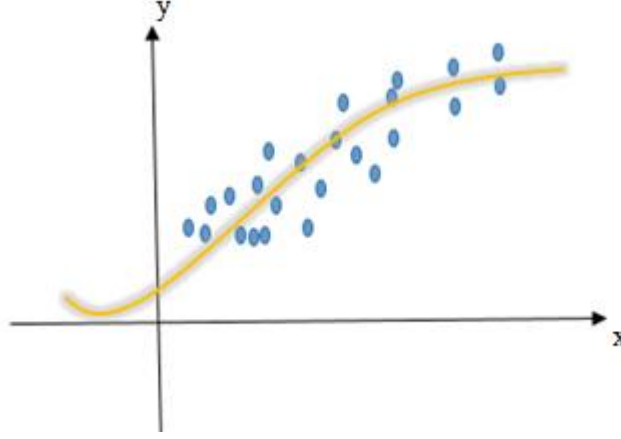
WEKA programının belleğinde bulunan bir veri seti için non-lineer regresyon modeli çalıştırılmıştır ve Şekil 3.9' da yer alan ağaç yapısı elde edilmiştir. Burada yer alan veri seti bilgisayar işlemcisi ile ilgili olup ön belleğin (CACHE) kapasitesi ve hafıza bilgileri yer almaktadır. Ön belleğin (CACHE) kapasitesi ve hafızaya bağlı olarak 2'ye dallanmıştır ve ikinci kısım daha fazla dallanmadan lineer model oluşturulmuştur. İlk kısım ise 2 adım daha dallanmıştır ve toplamda 5 adet lineer model oluşturulmuştur.



Şekil 3.9 WEKA’ da non-lineer regresyon ile sınıflandırma örnek ağaç yapısı

3.8.5. Lojistik regresyon

Bu yöntem genelde sosyal araştırmalarda sıklıkla kullanılsa da tıp alanının da kullanımı yaygındır. Lineer regresyonda oluşturulacak doğru $y=ax+b$ şeklinde düz olarak oluşturulmaya çalışılırken burada logaritmik bir doğru elde etme amacı vardır. Burada yapılmak istenen verilerin taşındığı uzayda oluşturulacak doğrunun verilere en yakın şekilde geçmesini sağlamaktır. Yöntem 2 aşamalı veya çok aşamalı verilerden olması durumunda, bağımlı ve bağımsız değişkenler arasındaki ilişkilerin incelenmesinde önemli bir yere sahiptir (Şeker, 2017; Girginer ve Cankuş, 2008).



Şekil 3.10 Logaritmik olarak oluşturulan çizgi

$$\sum_{i=1}^n (1 - x^{(i)}) \log \left(1 - \Pr \left[1 \mid a_1^{(1)}, a_2^{(2)}, \dots, a_k^{(k)} \right] \right) + x^{(i)} \log \left(\Pr \left[1 \mid a_1^{(1)}, a_2^{(2)}, \dots, a_k^{(k)} \right] \right) \quad (3.2)$$

Denklem (3.2) lojistik regresyon modelini nasıl oluşturduğumuzun fonksiyonudur. Denklem (3.2)' de "a" olarak gösterilen değer veri setimiz de yer alan özellikler (attribute) ve bunların sahip oldukları ağırlıkların çarpımıdır. "Pr" ise koşullu olasılığı göstermektedir. Yani Denklem (3.2) veri setinde yer alan özelliklere göre, verilere en yakın noktadan geçecek şekilde logaritmik bir çizgi oluşturmayı hedefler. Bunu da " $x^{(i)}$ " gerçek değer ile tahmin edilen değer arasındaki hatayı hesaplayarak yapar (Şeker, 2017).

3.8.6. Naive bayes

Bu yöntem koşullu olasılığa dayanan bir yöntemdir. Makine eğitimi, eldeki veriler üzerinden olasılık hesabı yaparak ona göre hesaplıyor. Olasılık tabanlı olduğu için bu yöntemde her bir giriş verisinin önemi vardır. Yani herhangi bir hastalık teşhisi için yapılan bazı yöntemler cinsiyet veya yaş gibi özelliklerin sisteme etkisi az veya hiç olmadığından veri madenciliğinde bu özelliği yok sayabilirler. Ancak Naive Bayes yönteminde böyle bir durum söz konusu olmayıp her bir özellik için sınıflandırma sonucuna göre olasılık hesabı yapılır.

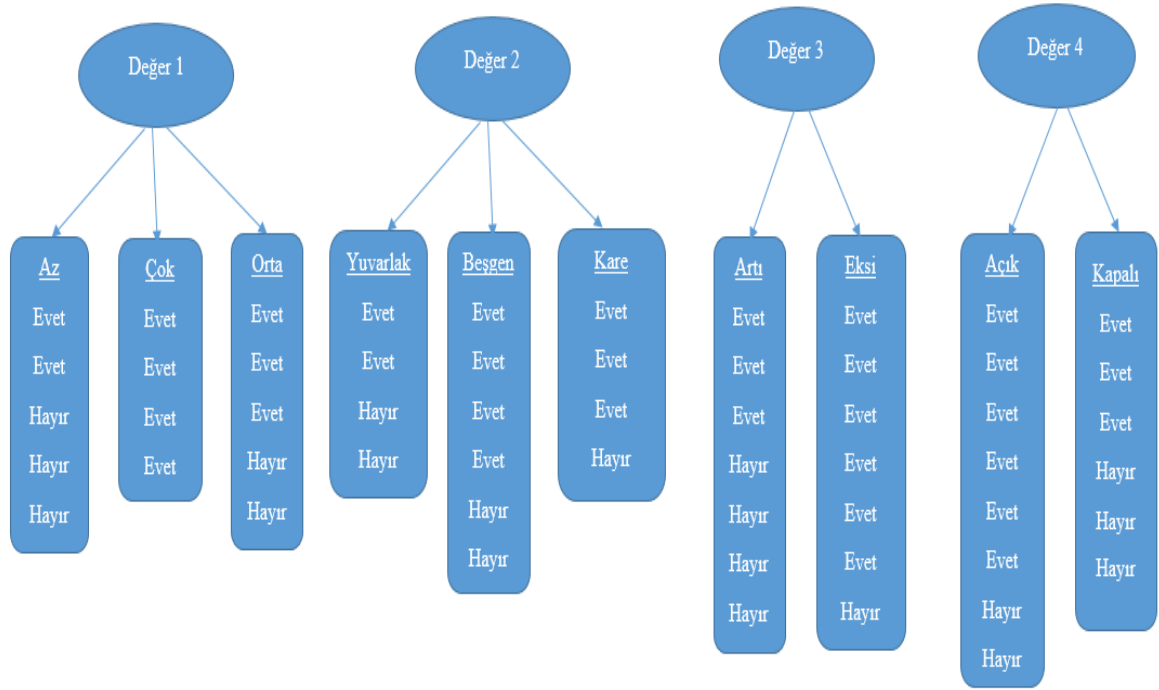
Denklem (3.3)' de Naive Bayes yöntemi temelini oluşturan bayes teoremi olasılık hesabı formülü yer almaktadır. Buradaki $P(c|x)$, "x" in olması durumunda "c" nin olması olasılığı olup koşullu olasılık olarak adlandırılır. $P(x|c)$ olayların her birinin ihtimallerini çarpımını gösterir. $P(c)$ sınıfın ihtimalidir (Haltaş ve Alkan, 2014; Kaya ve Pehlivan, 2014; Alkuşak ve Gök, 2014).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (3.3)$$

Çizelge 3.2’ de yer alan veri setine göre Naive Bayes yöntemi için örnek bir hesaplama yapabiliriz. Her bir özelliğin sahip olduğu durumlara göre olasılıklar hesaplanmalıdır. Çizelge 3.2’ de 2 adet sınıf yer almaktadır. Toplam veri sayımız 14’ dür. “Evet” sınıfından 9, “Hayır” sınıfından 5 verimiz vardır. Hesaplama yaparken “Evet” sınıfı için oranımız $\frac{9}{14}$ “Hayır” sınıfı için oranımız $\frac{5}{14}$ olacaktır.

Çizelge 3.2 Naive Bayes hesabı için örnek veri seti

DEĞER 1	DEĞER 2	DEĞER 3	DEĞER 4	SINIF
Az	Yuvarlak	Artı	Açık	Hayır
Az	Yuvarlak	Artı	Kapalı	Hayır
Çok	Yuvarlak	Artı	Açık	Evet
Orta	Beşgen	Artı	Açık	Evet
Orta	Kare	Eksi	Açık	Evet
Orta	Kare	Eksi	Kapalı	Hayır
Çok	Kare	Eksi	Kapalı	Evet
Az	Beşgen	Artı	Açık	Hayır
Az	Kare	Eksi	Açık	Evet
Orta	Beşgen	Eksi	Açık	Evet
Az	Beşgen	Eksi	Kapalı	Evet
Çok	Beşgen	Artı	Kapalı	Evet
Çok	Yuvarlak	Eksi	Açık	Evet
Orta	Beşgen	Artı	Kapalı	Hayır



Şekil 3.11 Naive Bayes hesap için veri seti özelliklerin gösterimi

Çizelge 3.3 Değer 1 özelliğinin olasılıkları

Değer 1			
	Evet	Hayır	Toplam
Az	4/9	0/5	4/14
Çok	3/9	2/5	5/14
Orta	2/9	3/5	5/14

Çizelge 3.4 Değer 2 özelliğinin olasılıkları

Değer 2			
	Evet	Hayır	Toplam
Yuvarlak	3/9	1/5	4/14
Kare	2/9	2/5	4/14
Beşgen	4/9	2/5	6/14

Çizelge 3.5 Değer 3 özelliğinin olasılıkları

Değer 3			
	Evet	Hayır	Toplam
Artı	3/9	4/5	7/14
Eksi	6/9	1/5	7/14

Çizelge 3.6 Değer 4 özelliğinin olasılıkları

Değer 4			
	Evet	Hayır	Toplam
Kapalı	3/9	3/5	6/14
Açık	6/9	2/5	8/14

Veri setimizde yer alan Değer 1: Orta, Değer 2: Beşgen, Değer 3: Artı, Değer 4: Kapalı için hesaplama yapalım. Denklem (3.4)' de örnek için verilen durumunun sınıflandırma sonucu “evet” olma olasılığını hesaplama formülüdür. Denklem (3.5) ise sınıflandırmanın “hayır” olarak hesaplanmasının formülüdür.

$$P(\text{evet}|x) = \frac{(P(\text{orta}|\text{evet}) \times P(\text{beşgen}|\text{evet}) \times P(\text{artı}|\text{evet}) \times P(\text{kapalı}|\text{evet}) \times P(\text{evet}))}{P(x)} \quad (3.4)$$

$$P(\text{hayır}|x) = \frac{(P(\text{orta}|\text{hayır}) \times P(\text{beşgen}|\text{hayır}) \times P(\text{artı}|\text{hayır}) \times P(\text{kapalı}|\text{hayır}) \times P(\text{hayır}))}{P(x)} \quad (3.5)$$

Sınıflandırma sonucunun “evet” için değerlerin yerine konulması ile Denklem (3.6) elde edilir. Sınıflandırma sonucu “hayır” için değerlerin yerine konulması ile Denklem (3.7) elde edilir.

$$P(\text{evet}|x) = \frac{\left(\frac{3}{9} \times \frac{4}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}\right)}{9} \quad (3.6)$$

$$P(\text{hayır}|x) = \frac{\left(\frac{2}{5} \times \frac{2}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}\right)}{5} \quad (3.7)$$

$$P(\text{evet}|x) = 0.0011 \quad (3.8)$$

$$P(\text{hayır}|x) = 0.0054 \quad (3.9)$$

Sınıflandırma sonucu “evet” için 0.0011, hayır için 0.0054 çıkmıştır. Bu da bize örnekteki durumun sınıfının “hayır” olacağını göstermektedir.

3.8.7. Karar ağaçları

Bu yöntemde, sistemin çalışması kararlar üzerindedir. Yani oluşturulacak karar ağacının kökünde hangi özellik yer alacak, alt dallanmalar nasıl olacak bu kararlar üzerine belirlenecektir. Bu belirlemeleri ise entropi hesabı ile yapılması gerekiyor. Denklem (3.10)' da bu denklem yer almaktadır (Kaya ve Pehlivan, 2014).

$$P_1 \times \log_2(P_1) + P_2 \times \log_2(P_2) + \dots + P_n \times \log_2(P_n) \quad (3.10)$$

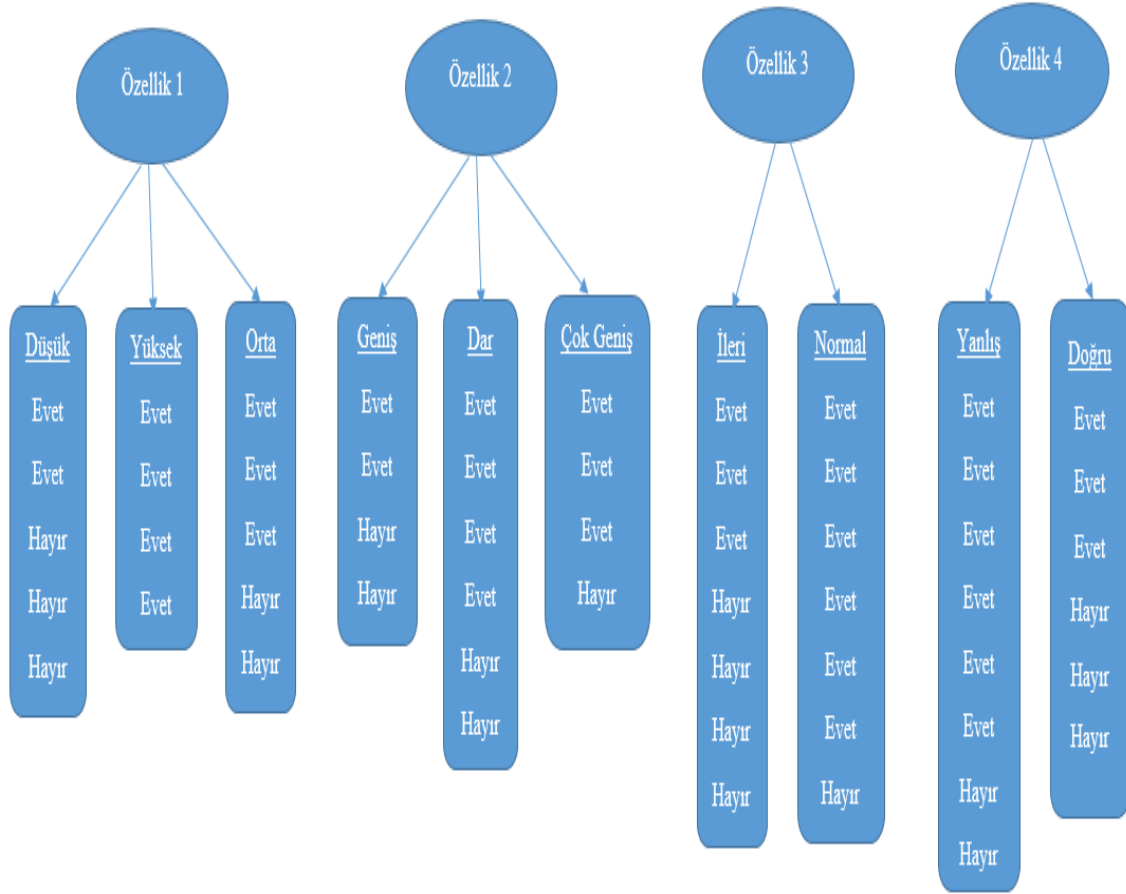
P_1 özellik altında yer alan her bir değerin toplam sınıf sayısına bölünmesi ile elde edilir. Çizelge 3.7’ de yer alan veri setine göre hesaplama yapabiliriz. Bu veri setinde 4 adet özellik bulunmaktadır.

Çizelge 3.7 Karar ağaçları hesap için veri seti

ÖZELLİK 1	ÖZELLİK 2	ÖZELLİK 3	ÖZELLİK 4	SINIF
Düşük	Geniş	İleri	Yanlış	Hayır
Düşük	Geniş	İleri	Doğru	Hayır
Yüksek	Geniş	İleri	Yanlış	Evet
Orta	Dar	İleri	Yanlış	Evet
Orta	Çok Geniş	Normal	Yanlış	Evet
Orta	Çok Geniş	Normal	Doğru	Hayır
Yüksek	Çok Geniş	Normal	Doğru	Evet
Düşük	Dar	İleri	Yanlış	Hayır
Düşük	Çok Geniş	Normal	Yanlış	Evet
Orta	Dar	Normal	Yanlış	Evet
Düşük	Dar	Normal	Doğru	Evet
Yüksek	Dar	İleri	Doğru	Evet
Yüksek	Geniş	Normal	Yanlış	Evet
Orta	Dar	İleri	Doğru	Hayır

Entropi hesabını yapacağımız özelliklerin gösterimi Şekil 3.11’ de yer almaktadır. Yapılan hesaplama sonucunda ağırlık en fazla Özellik 1 verisinde çıkıyor ve ağacın temelini bu özellik oluşturuyor. Ağacın aşağı doğru diğer dallanması ise kalan özelliklere göre entropi hesabı ile gerçekleşir.

Şekil 3.12’ de ise bu veri setinin karar ağaçları yöntemi ile işlenmesi sonucu elde edilen karar ağacı yer almaktadır.



Şekil 3.12 Karar ağaçları hesap için veri seti özelliklerin gösterimi

3.8.7.1. Özellik 1 için entropi hesabı

Şekil 3.12’ de görüldüğü üzere Özellik 1 özelliği altında 3 seçenek yer almaktadır. Toplam sonuç sınıflandırması ise 14 tanedir. Özelliğin altında yer alan her bir seçenek için hesaplama yapılıyor.

$$\frac{5}{14} \times \log_2 \left(\frac{5}{14} \right) + \frac{4}{14} \times \log_2 \left(\frac{4}{14} \right) + \frac{5}{14} \times \log_2 \left(\frac{5}{14} \right) = 0.544 \quad (3.11)$$

3.8.7.2. Özellik 2 için entropi hesabı

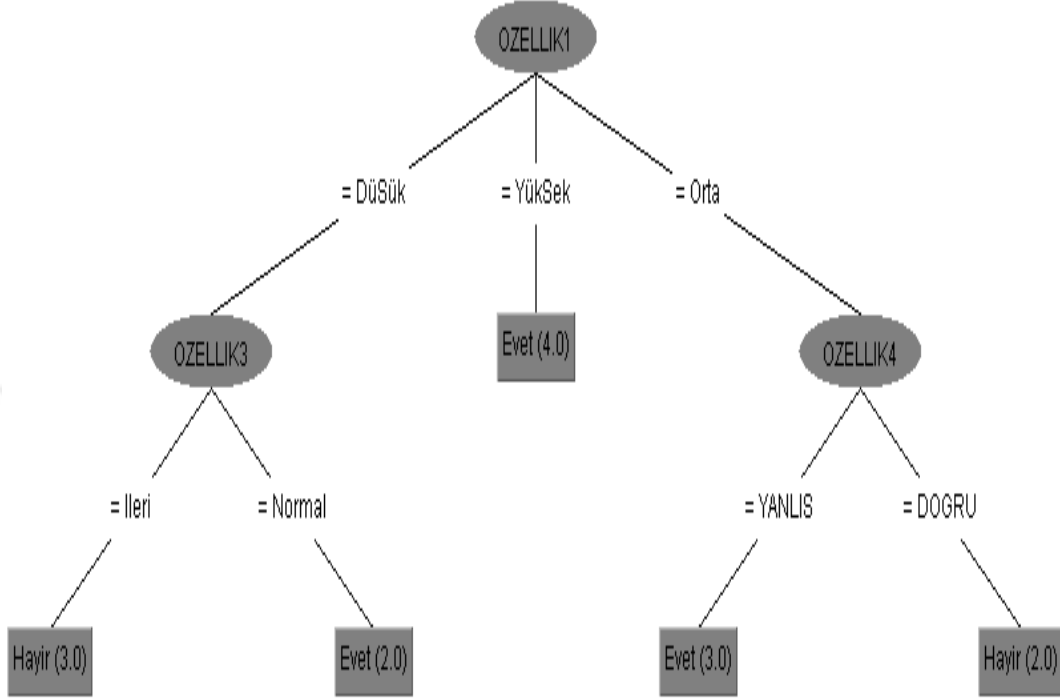
$$\frac{4}{14} \times \log_2 \left(\frac{4}{14} \right) + \frac{6}{14} \times \log_2 \left(\frac{6}{14} \right) + \frac{4}{14} \times \log_2 \left(\frac{4}{14} \right) = -0.1902 \quad (3.12)$$

3.8.7.3. Özellik 3 için entropi hesabı

$$\frac{7}{14} \times \log_2 \left(\frac{7}{14} \right) + \frac{7}{14} \times \log_2 \left(\frac{7}{14} \right) = 0.152 \quad (3.13)$$

3.8.7.4. Özellik 4 için entropi hesabı

$$\frac{8}{14} \times \log_2 \left(\frac{8}{14} \right) + \frac{6}{14} \times \log_2 \left(\frac{6}{14} \right) = 0.2713 \quad (3.14)$$

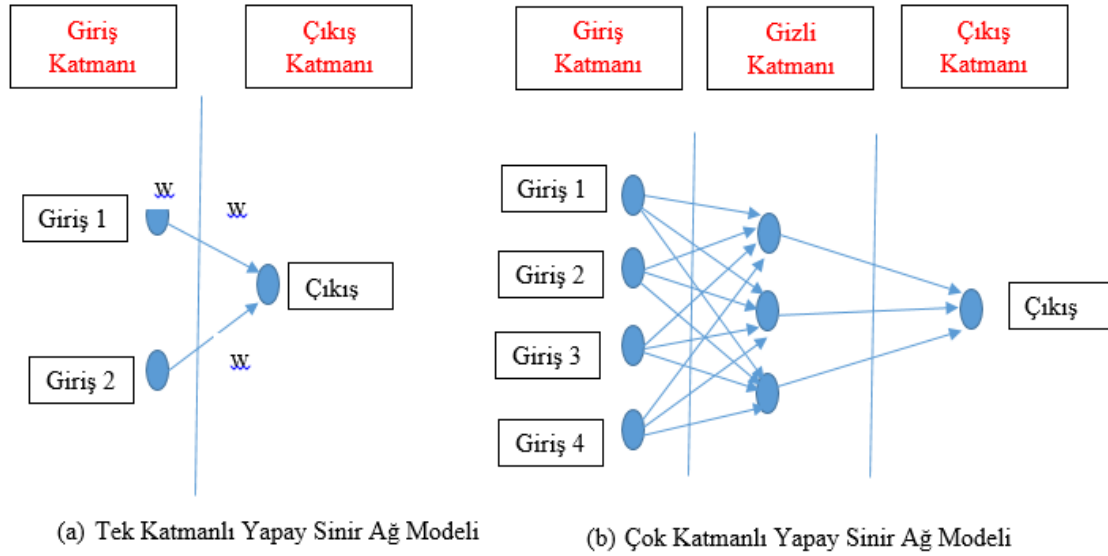


Şekil 3.13 Karar ağaçları hesaplama veri seti karar ağacı

3.8.8. Çok katmanlı yapay sinir ağları

Yapay sinir ağları insan öğrenmesinden etkilenilerek oluşturulmuş bir yöntemdir. İnsan öğrenmesinde sinir hücreleri olan nöronların önemi büyüktür. Yapay sinir ağlarında da nöronlar bulunmaktadır ve bu nöronlar aslında transfer fonksiyonlarıdır. İnsan öğrenmesinde nöronlar arası sinyal iletiminde sinyal kayıpları söz konusu iken yapay sinir ağlarında bu kayıplar yoktur; ancak daha iyi sonuç alabilmek için yapay sinir ağlarında, nöronlar arası geri beslemeler söz konusudur. Yapay sinir ağlarındaki nöronlar arasında iletimde her birinin bir ağırlığı söz konusudur (Kaya ve Pehlivan, 2014; Arora ve Suman, 2012).

Yapay sinir ağlarında; giriş katmanları, çıkış katmanları ve gizli katmanlar bulunmaktadır. Normal bir yapay sinir ağı modelinde giriş ve çıkış katmanları bulunur. Bizim kullanacağımız çok katmanlı yapay sinir ağı modelinde giriş katmanları arasında gizli katman bulunmaktadır (Kaya ve Pehlivan, 2014; Arora ve Suman, 2012).



Şekil 3.14 Yapay sinir ağları modeli

3.9. Çalışmada Kullanılan Veri Madenciliği Kümeleme Yöntemleri

3.9.1. K-Ortalamalar (K-means)

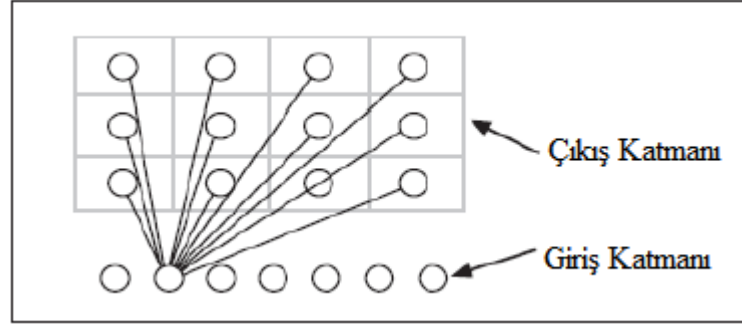
K-ortalamalar (K means) yöntemi kümeleme yöntemlerinde yaygın olarak kullanılan bir yöntemdir. K-ortalamalar kümeleme yönteminde, verinin kaç kümeye ayrılacağı önceden bildirilir ve bu yöntemin dezavantajıdır. Bu yöntemde öncelikle sistem bir veriyi seçer ve merkeze yerleştirir. Kümedeki diğer verilerin merkezdeki veriye olan uzaklıkların ortalaması küme benzerliğidir. Bu yöntemde kümede yer alan veriler arası mesafe minimum, kümeler arası mesafe ise maksimum olmak zorundadır (Alan, 2014; İşler ve Narin, 2012).



Şekil 3.15 K ortalama iş akış şeması

3.9.2. Kohonen (Self organizing map)

Yapay sinir ağları temelli bir yöntemdir. Yapay sinir ağlarında yer alan giriş çıkış katmanları mevcuttur. Giriş katmanlarını ve sayısını veri setinde yer alan özellikler belirler. Çıkış katmanı ise yöntem kaç kümeye göre kümeleme yapacaksa ona göre belirlenir. Burada yalnız klasik bir yapay sinir ağları modellemesinden farklı bir dizilim vardır. Bu dizilim küp, doğrusal veya dikdörtgen şeklinde olabilir. Şekil 3.16' da yaygın olarak kullanılan dikdörtgen dizilimin şeması yer almaktadır (Taşkın ve Emel, 2010; Coşkun ve Karaağaoğlu, 2011).

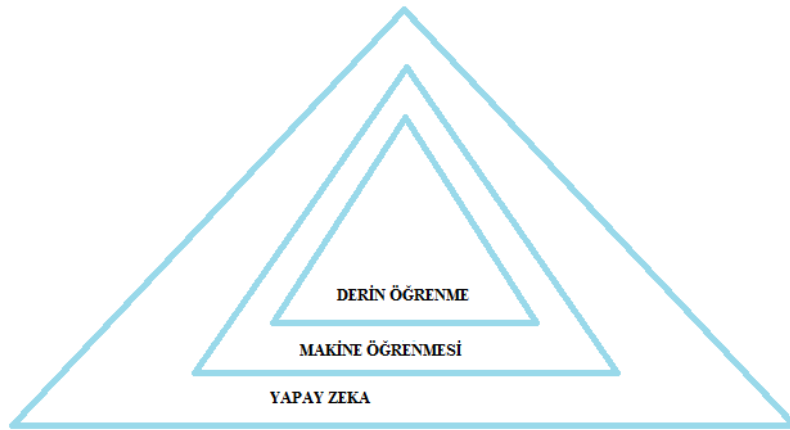


Şekil 3.16 Kohonen yöntemi nöron dizilimi

K boyutlu bir veri, denetimsiz bir öğrenme algoritması ile daha az boyuta dönüştürülür. İndirgenmiş boyutta yer alan veriler benzer özelliklerine göre birbirlerine daha yakın konumlarda bulunan bir uzaya taşınır. Görsel olarak incelendiğinde uzay, veri setini de görsel olarak yorumlamak daha kolaydır, ayrıca burada veriler ilgi derecesine göre bir kümeleme de gerçekleştirilebilir.

3.10. Derin Öğrenme

Derin öğrenmeye çok katmanlı yapay sinir ağıdır denebilir. Ancak derin öğrenme klasik çok katmanlı yapay sinir ağına göre oldukça gelişmiştir ve hata oranları daha da düşmüştür. Derin öğrenme Şekil 3.17’de görüleceği üzere yapay zekâ ve makine öğrenmesinin altında yer alır (Anwer, 2017).



Şekil 3.17 Derin öğrenme

Son yıllarda yapılan akademik çalışmalar ile özellikle görüntü işleme alanında derin öğrenme sayesinde yüksek başarılar elde edilmiştir. Ancak buradaki en büyük sorun verilerin fazla olması ile beraber daha yetenekli, yani matris hesaplarını daha hızlı yapan

cihazlara gereksinimin artmasıdır. Tabi bu sorun da çeşitli firmaların ürettiği GPU (Grafik İşlemci Ünitesi) sayesinde aşılmaktadır.

Derin öğrenme hayatımıza hızlı bir şekilde girmiş olup, her gün tarama yaptığımız ve video izlediğimiz dev firmalar derin öğrenme tekniğini kullanarak bize hitap eden içerikleri hizmetimize sunmaktadırlar.

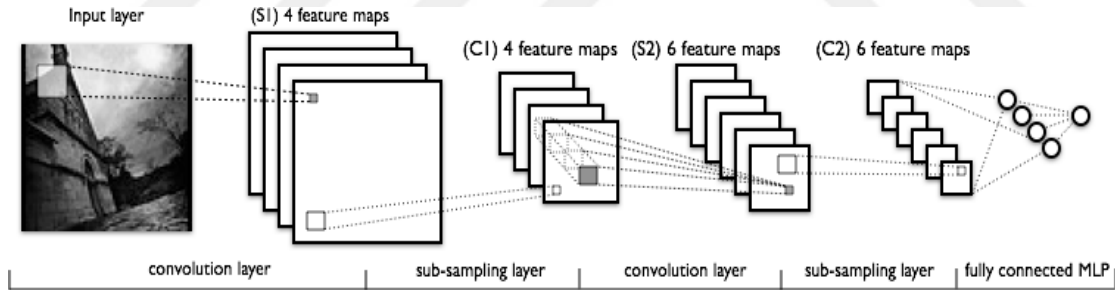
3.10.1. Tam bağlantılı yapay sinir ağı (Dense fully connected)

Kısaca bu yapay sinir ağını, her bir nöronun kendinden önceki ve sonraki nörona bağlı olması olarak adlandırabiliriz. Dense olarak adlandırılan bu yapay sinir ağının en basit yöntem olduğu söylenebilir (Anwer, 2017; Aygün, 2017).

Veri setimiz sayısal verilerden oluştuğundan yani herhangi bir görüntü yer almadığından, çalışmamızda Dense yöntemini kullanacağız.

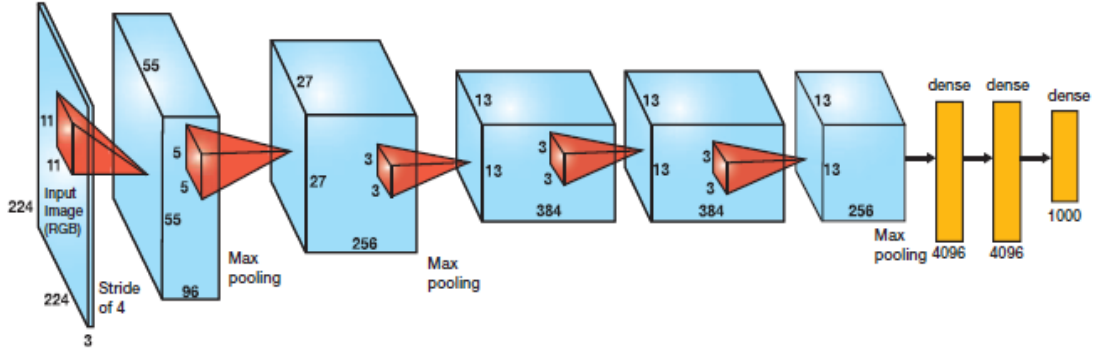
3.10.2. Konvolüsyon sinir ağı (CNN)

Bu yöntem, görüntü işleme temeline dayanır. Yöntemin temel mantığı veri setinde yer alan görüntülerden çeşitli özellikler çıkarmaktır. CNN ağını 1988 yılında Yann LeCun ortaya atmıştır ve 1998 yılına kadar geliştirilmeye devam edilmiştir. Şekil 3.18’ de bu ağın mimarisi yer almaktadır (Şeker ve ark, 2017)



Şekil 3.18 LeNet mimarisi

Ancak 2012 yılında, Alex Krizhevsky ve arkadaşları devrim niteliğinde bir çalışma yaparak hata oranını %15 civarına düşürmüşlerdir.



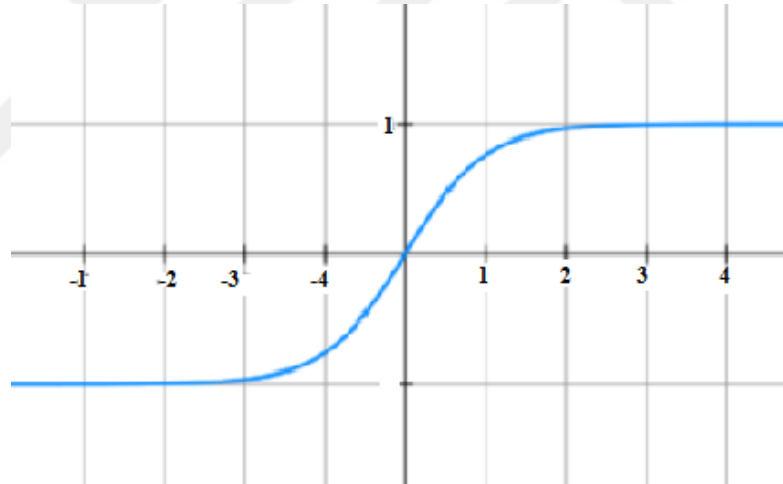
Şekil 3.19 Alexnet mimarisi

3.10.3. Aktivasyon fonksiyonları

Herhangi bir katmandaki nörondaki veri bir sonraki nörona aktivasyon fonksiyonları sayesinde iletilir. Çeşitli aktivasyon fonksiyonları vardır (Anwer, 2017).

Aktivasyon fonksiyonlarının yaptığı işleme çok büyük hesaplamalar sonucu oluşan sayısal değerleri -1 ile 1 arasına taşımakta diyebiliriz.

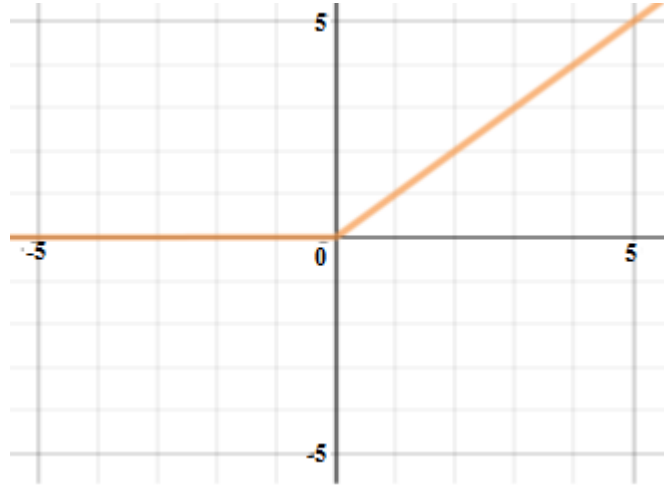
Tanh aktivasyon fonksiyonu:



Şekil 3.20 Tanh aktivasyon fonksiyonu (Anwer, 2017)

Tanh aktivasyon fonksiyonunun normalize edilmiş hali Şekil 3.20’ de görüleceği üzere -1 ve 1 arasındadır. Bu fonksiyon hiperbolik bir fonksiyondur (Anwer, 2017).

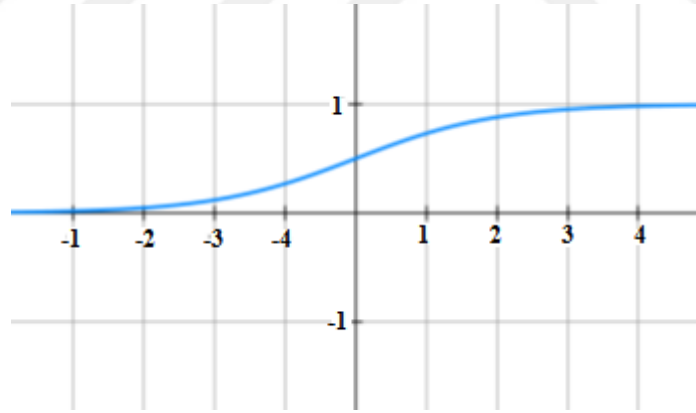
Rectified linear aktivasyon fonksiyonu (ReLU):



Şekil 3.21 ReLU aktivasyon fonksiyonu (Anwer, 2017)

Giriş değeri 0 ise bu fonksiyonda çıkış değeri de 0 olur. Giriş değeri arttıkça çıkış değeri de belirli bir orantıda artış gösterir. Bu fonksiyon başarılı sonuçlar verdiği için kabul gören bir fonksiyondur ve biz de çalışmamızda ReLU aktivasyon fonksiyonunu kullandık (Anwer, 2017).

Sigmoid aktivasyon fonksiyonu:



Şekil 3.22 Sigmoid aktivasyon fonksiyonu (Anwer, 2017)

3.10.4. Derin öğrenme için hata fonksiyonları

Hata fonksiyonunun amacı hatayı azaltmaktır. Eğitimi iyi olan bir yapay sinir ağı minimum hataya sahiptir. Hata ölçümü hata fonksiyonları ile ölçülür. Derin öğrenme veya yapay sinir ağı tasarlayan uzmana göre hata fonksiyonları farklılık gösterir. Hata fonksiyonlarının bazıları şunlardır:

1. Çapraz entropi hata fonksiyonu.
2. Log-likelihood hata fonksiyonu.

Çalışmada çapraz entropi hata fonksiyonunu kullanıldı.

3.11. Başarı Ölçütleri

3.11.1. ROC (Receiving operating characteristic) analizi

Duyarlılık ve seçicilik arasındaki ilişki grafiksel olarak ROC eğrisi ile gösterilir. Böylece en uygun eşik değeri belirlenerek örneğin sınıflandırmaya buna göre karar verilir. ROC eğrisinde tanı testleri karar verme yöntemleridir. Tanı testlerinde pozitif veya negatif kararların doğruluk derecesi önem arz etmektedir. ROC eğrisi elde etmede kullanılan malzemeler Çizelge 3.8’ de gösterilmiştir (Alan, 2014; İşler ve Narin, 2012).

ROC başarı ölçütü sınıflandırma tekniklerinde ve derin öğrenme tekniğinde kullanılacaktır.

Çizelge 3.8 ROC eğrisi elde etmede kullanılan adı

Sıra:	ROC Eğrisi Elde Etmede Kullanılan Malzeme Adı:
1.	Duyarlılık (Sensitivity)
2.	Özgüllük (Specificity)
3.	Yanlış pozitif oranı (1-Özgüllük)
4.	Yanlış negatif oranı (1-Duyarlılık)
5.	Genel doğruluk oranı (Accuracy)
6.	Artı yorum gücü (Positive Predictive Value)
7.	Eksi yorum gücü (Negative Predictive Value)
8.	ROC eğrisi altında kalan alan (AUC)

Çizelge 3.10’ da gerçek durum ile test sonucu elde edilen değerlerin ne anlama geldiği gösterilmektedir.

Çizelge 3.9 ROC eğrisi test sonucu ifadeleri

Sıra:	Adı:	Açıklama:
1.	DP	Gerçek durum pozitifken test sonucu pozitif
2.	YP	Gerçek durum negatifken test sonucu pozitif
3.	DN	Gerçek durum negatifken test sonucu negatif
4.	YN	Gerçek durum pozitifken test sonucu negatif

ROC eğrisi elde etmek için kullanılan malzemeleri hesaplamak için kullanılan Çizelge 3.8’ deki malzemeleri Çizelge 3.9’ da yer alan değerlere göre formülüle etmek mümkündür. Bu formülüle için Çizelge 3.10’ da yer alan kontenjans tablosundan yararlanılabilir (Alan, 2014; İşler ve Narin, 2012).

Çizelge 3.10 ROC eğrisi ölçütleri için kontenjans tablosu

Gerçek Durum:					
Test Sonucu:	i \ j		Pozitif	Negatif	Toplam:
			j=1	j=2	
	Pozitif:	i=1	n_{11} (DP)	n_{12} (YP)	$n_{1.}$ (DP+YP)
	Negatif:	i=2	n_{21} (YN)	n_{22} (DN)	$n_{2.}$ (YN+DN)
	Toplam:		$n_{.1}$ (DP+YN)	$n_{.2}$ (YP+DN)	n (DP+YP+YN+DN)

3.11.2. Duyarlılık (Sensitivity – Doğru pozitif oranı- Recall)

Hastalığı gerçek olanları, testin hasta olarak ayırabilmesidir.

$$\text{Duyarlık} = \text{Doğru Pozitif Oranı} = P(y_i > t | y_1 < t < y_m) = \frac{n_{11}}{n_{.1}} = \frac{DP}{(DP + YN)} \quad (3.15)$$

3.11.3. Özgüllük (Specificity – Doğru negatif oranı)

Hastalığı olmayanları, testin hasta olmayarak ayırabilmesidir.

$$\text{Özgüllük} = \text{Doğru Negatif Oranı} = P(y_i \leq t | y_{m+1} < t < y_n) = \frac{n_{22}}{n_{.2}} = \frac{DN}{(DN + YP)} \quad (3.16)$$

3.11.4. Yanlış pozitif oranı (1-Özgüllük)

Hasta olmayanlar, test tarafından yanlış olarak hasta olarak seçilenlerdir.

$$YPO = (1 - \text{Özgüllük}) = P(y_i > t | y_{m+1} < t < y_n) = \frac{n_{12}}{n_{.2}} = \frac{YP}{(YP + DN)} \quad (3.17)$$

3.11.5. Yanlış negatif oranı (1-Duyarlılık)

Hasta olanlar, test tarafından yanlış olarak hasta değil olarak seçilenlerdir.

$$YNO = (1 - \text{Duyarlılık}) = P(y_i \leq t | y_{m+1} < t < y_n) = \frac{n_{21}}{n_{.1}} = \frac{YN}{(DP + YN)} \quad (3.18)$$

3.11.6. Gerçek doğruluk (Accuracy)

Testin doğru olarak seçtiği hasta ve hasta olmayanların, testin toplan seçimine oranıdır.

$$\text{Accuracy} = \frac{(n_{11} + n_{22})}{n} = \frac{(DP + DN)}{(DP + DN + YP + YN)} \quad (3.19)$$

3.11.7. Artı yorum gücü (Positive predictive value-Duyarlılık (Precision):-Kesinlik)

Test hasta sonucu verdiğiinde, gerçek hasta olma olasılığıdır.

$$\text{PKD} = P(H + |T+) = \frac{n_{11}}{n_{11} + n_{12}} = \frac{DP}{(DP + YP)} \quad (3.20)$$

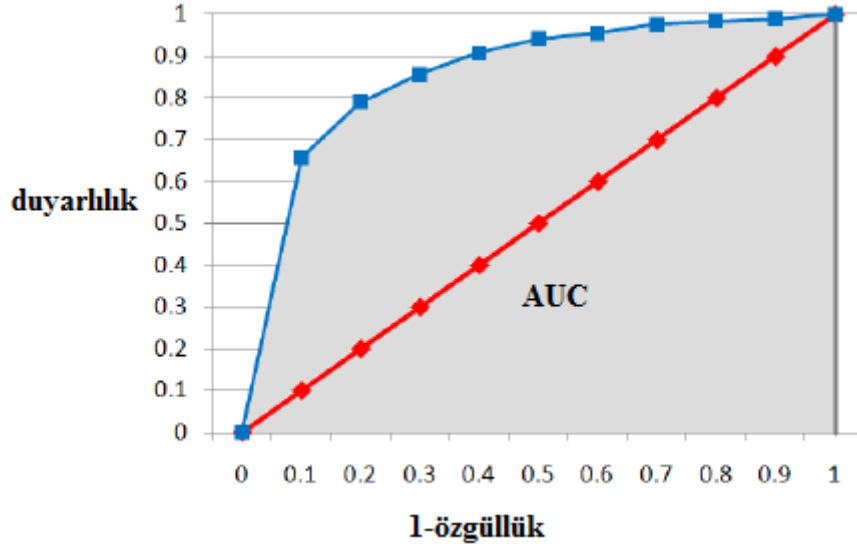
3.11.8. Eksi yorum gücü (Negative predictive value)

Test hasta değil sonucu verdiğiinde, gerçek hasta değil olma olasılığıdır.

$$\text{NKD} = P(H - |T -) = \frac{n_{22}}{n_{21} + n_{22}} = \frac{DN}{(DN + YN)} \quad (3.21)$$

3.11.9. ROC eğrisi altında kalan alan (AUC)

Koordinat sisteminde X eksenini 1-özgüllük, Y eksenini duyarlılık yer alır ve kesim noktalarına karşılık gelen noktalar birleştirilerek ROC eğrisi çizilir. ROC eğrisi altında kalan alan AUC olarak isimlendirilir. Pozitif ve negatifler birbirinden tamamen ayrılırsa AUC değeri 1 olur. AUC değeri azaldıkça başarı oranı azalır (Alan, 2014, İşler ve Narin, 2012).



Şekil 3.23 ROC eğrisi altında kalan alan AUC

3.11.10. Ortalama mutlak hata (Mean absolute error)

Tahmin edilen değer ile gerçekleşen değerlerin farkının mutlak değerini aldıktan sonra bu değerlerin ortalamasıdır. Denklem 3.22’ deki “p” değeri tahmin edilen değeri “a” ise gerçekleşen değeri ifade etmektedir (Şeker, 2017).

$$\frac{|p_1 - a_1| + |p_2 - a_2| + \dots + |p_n - a_n|}{n} \quad (3.22)$$

3.11.11. Ortalama hata kareleri toplamı kökü (Root mean square error)

MAE (Mean Absolute Error)’ dan pek farklı değildir ancak daha iyi sonuç verir. Burada, tahmin edilen değer ile gerçekleşen değerlerin ortalamasının karekökü alınıyor (Şeker, 2017).

$$\sqrt{\frac{(p_1 - a_1)^2 + (p_2 - a_2)^2 + \dots + (p_n - a_n)^2}{n}} \quad (3.23)$$

3.11.12. Bağlı mutlak hata (Relative absolute error)

Burada, gerçekleşen değer ile gerçek değerlerin ortalamasının farkı mutlak değer ile alınır ve toplanır. Formülde bu payda kısmını oluşturur. Pay kısmında yapılan ise tahmin edilen değer ile gerçekleşen değerlerin farklarının toplamı yer alır. Aslında RAE sapmayı belirler. Yani tahmin edilen değerler ile gerçekleşen değerlerin sapmasını oranlayarak veri madenciliği algoritmasının sapmasını bulur. RAE “0” olarak hesaplanırsa mükemmel bir

sınıflama yapılmıştır. Eğer bu değer “1” olarak hesaplanırsa sınıflandırma başarısız olmuştur (Şeker, 2017).

$$\frac{|p_1 - a_1| + |p_2 - a_2| + \dots + |p_n - a_n|}{|a_1 - \bar{a}| + |a_2 - \bar{a}| + \dots + |a_n - \bar{a}|} \quad (3.24)$$

3.11.13. Ortalama bağıl mutlak hata kareleri toplamı (Root relative squared error)

Burada yapılan ise RAE ile hesaplanan değer karekökünü almaktır.

$$\sqrt{\frac{(p_1 - a_1)^2 + (p_2 - a_2)^2 + \dots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + (a_2 - \bar{a})^2 + \dots + (a_n - \bar{a})^2}} \quad (3.25)$$

3.11.14. F-Ölçütü (F-Ölçütü:)

Daha önce anlatılan kesinlik ve duyarlılık doğrudan sistemin başarısı hakkında bilgi vermez. Bu iki sonuç işlendiğinde ise anlamlı bir sonuç elde edilir. Bu anlamlı sonuç F-Ölçütü olarak isimlendirilir ve duyarlılık ve kesinliğin harmonik ortalaması alınarak elde edilir (Nizam ve Akın, 2014).

$$F = \frac{2xDxK}{(D + K)} \quad (3.26)$$

3.11.15. Kappa istatistiği

Kappa katsayısı iki veya daha fazla gözlemcinin bir durumla ilgili uyumunu ortaya koyar. Kappa katsayısı uyuma göre -1 ile +1 arasında değer alır. +1 çıkan sonuçta iki gözlemci arasındaki uyum tamdır. -1 çıkan durumda ise iki gözlemci birbiri ile zıt değerlendirmede bulunmuşlardır. 0 çıkan sonuç ise gözlemciler arası uyum şansa bağlıdır şeklinde yorumlanabilir. Denklem (3.27)’deki Pr(a) iki gözlemcinin ortak kararını Pr(e) ise ortak karar verme ihtimallerini gösterir (Kılıç, 2015).

$$K = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (3.27)$$

Kappa katsayısını bir örnekle açıklayalım. Örneğin bir hastanenin beyin ve sinir cerrahisi kliniğinde iki beyin cerrahı 100 tane tümör ya da tümör değil teşhisi koysunlar. Bunu Çizelge 3.11’de görebiliriz.

Çizelge 3.11 Örnek Kappa katsayısı hesabı

	Gözlemci 2 Tümör	Gözlemci 2 Tümör değil.	Toplam:
Gözlemci 1 Tümör	54	14	68(%68)
Gözlemci 1 Tümör değil.	16	16	32(%32)
Toplam:	70(%70)	30(%30)	100(%100)

Çizelge 3.11’ de görüleceği üzere gözlemci 1 olan cerrah, 54 hastaya tümör 16 hastaya da tümör değil olarak teşhis koymuştur. Yine burada görüleceği üzere gözlemci 2 cerrah 54 hastaya tümör 14 hastayı tümör değil olarak teşhis koymuştur. İki cerrahın da tümör teşhis sayısı 54, tümör değil teşhisi 16’ dır. Böylece $Pr(a)=0.7$ olur. Tümör teşhisine göre uyum olasılığı $0.7 \times 0.68 = 0.476$ olarak hesaplanır. Tümör değil için olasılık $0.3 \times 0.32 = 0.096$ olarak hesaplanır. Toplamı ise $0.096 + 0.476 = 0.572$ ’dir. Denklem 3.28’deki gibi değerleri yerine koyarsak sonuç 0.299 olarak elde edilir.

$$K = \frac{0.7 - 0.572}{1 - 0.572} = \frac{0.128}{0.428} = 0.299 \quad (3.28)$$

3.12. “n” Kat Çapraz Doğrulama Eğitim Verisi

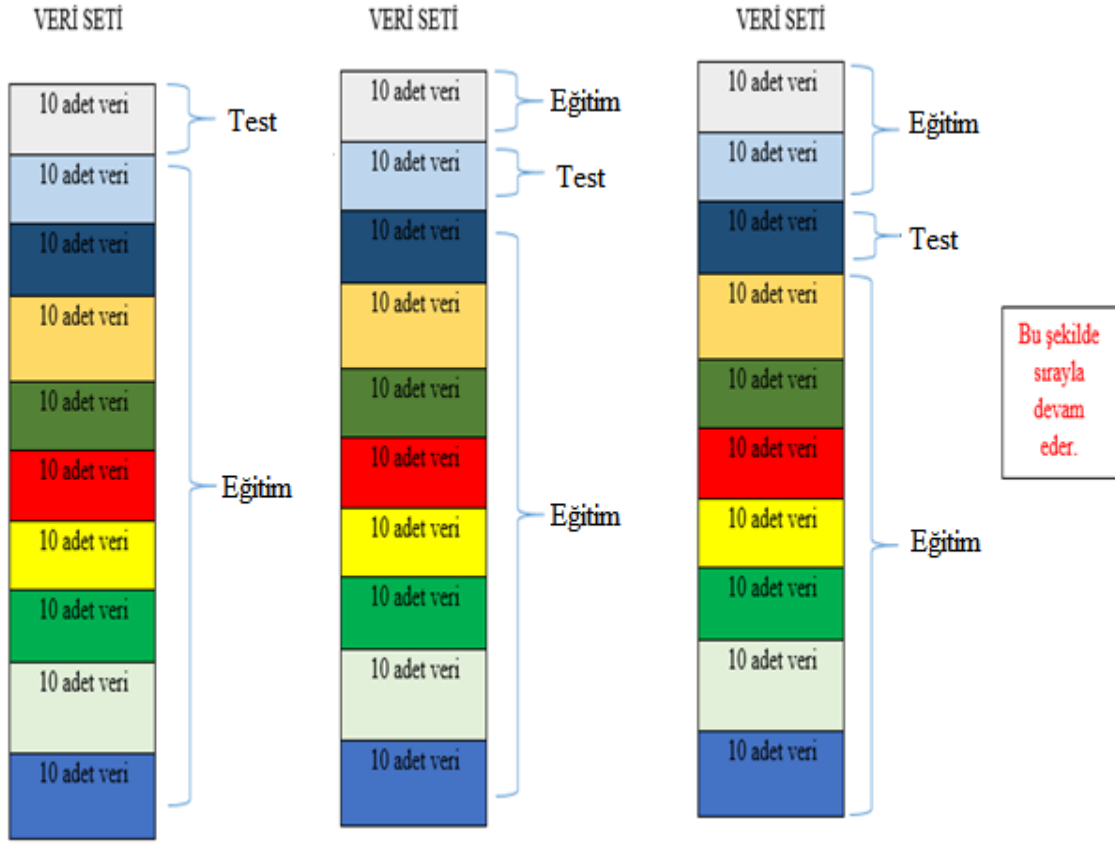
Veri madenciliği sınıflandırma yöntemleri makine öğrenmesine dayalı yöntemlerdir. Basit doğrulama tekniğinde, makine eğitimini eldeki veri setinin belirli bir kısmı eğitim kısmı kullanılarak yapılır. Basit doğrulama tekniğinde genelde, eldeki verilerin %66’sı eğitim, %33 test için kullanılır. “n” katlı çapraz doğrulamada veri seti “n” kadar bölüme ayrılır. Veri madenciliği genel uygulaması olarak “n” 10 seçilir. Örnek bir uygulama yapacak olursak veri setimizde 100 adet veri olsun. 10 kat çapraz doğrulamada bu veriler Şekil 3.24’ de görüldüğü üzere 10’ar eşit parçaya bölünür (Şeker, 2017).

VERİ SETİ

10 adet veri
10 adet veri
10 adet veri
10 adet veri
10 adet veri
10 adet veri
10 adet veri
10 adet veri
10 adet veri
10 adet veri
10 adet veri

10 KAT ÇAPRAZ
DOĞRULAMAYA GÖRE
VERİ SETİ 10 EŞİT PARÇAYA
BÖLÜNÜYOR

Şekil 3.24 10 kat çapraz doğrulamaya göre veri seti



Şekil 3.25 10 kat çapraz doğrulamaya göre işlem aşamaları

Şekil 3.25’ de görüldüğü gibi, eğitim seti her bir bölme için tekrarlanıyor. 10 kat çapraz yöntemi ile 10 defa eğitim ve test işlemi yapılıyor. İşlem sonucunda oluşan 10 sonuca göre, en başarılı sonucun olduğu sistem için sonuç olarak alınır.

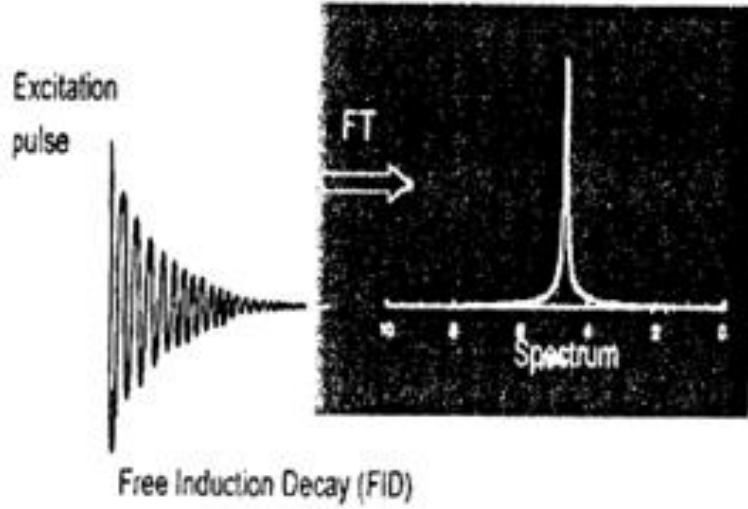
3.13. Yüzdelik Eğitim Verisi

Bu teknikte makine eğitimi veri setinin belirlenen kısmını eğitim verisi diğer kısmını ise test için kullanır. Veri setinin ne kadarının eğitim için kullanılabileceği WEKA programında seçilebilir.

4. MR SPEKTROSKOPİ İLE BEYİN TÜMÖRÜ

4.1. MR Spektroskopi Nedir?

MR Spektroskopi molekülün kimyasal yapısını belirlemede kullanılan yöntemlerdendir. MR Spektroskopi patolojik veya anatomik dokunun biyokimyasal yapısını ve karakterini spektrumda gösteren bir tekniktir. Beyin, MR Spektroskopi (MRS) kullanılarak görüntüleme yapılan en belirgin organdır. MRS' in beyin hakkında verdiği sonuçlar invazif olmayan teknikle elde edilen sonuç bağlamındadır. MRS klinik kullanımda 1.5 T ve daha yüksek değerlerde kısa sürede sonuçlar vermektedir. MRS ölçümü öncelikle ölçüm yapılacak bölgenin seçilmesiyle yapılır. Bundan sonra sadece seçilen bölgede ölçülen sinyaller spektrumda yerini alır. MRS 1.5 T gibi çok büyük manyetik alana sahip cihazlar aracılığıyla gerçekleştirilir. Bu büyük manyetik alan metabolik farklılıkları daha belirgin hale getirmektedir. MRS Şekil 4.1' de görüldüğü gibi 4 aşamadan meydana gelir (Cüce, 2005). İlk önce çekirdek uyarılır (Excitation Pulse). 2. aşamada serbest indüksiyon azalması (Free induction decay) gerçekleşir. 3. aşamada Fourier dönüşümü (Fourier Transform) gerçekleşir. Son aşamada spektrum gösterilir.

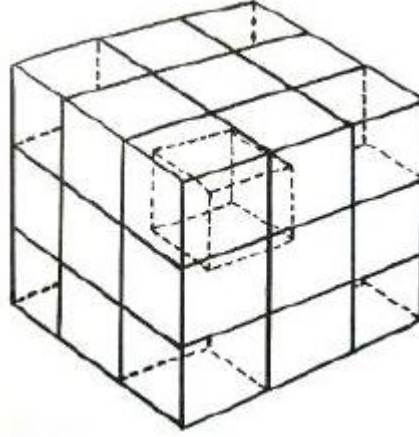


Şekil 4.1 MRS işlem aşamaları (Cüce, 2005'dan değiştirilmiştir.)

4.1.1. Voksel

İncelenecek bölgeden alınan örnektir. Klinik olarak voksel hacmi genel olarak 2-8 cm³ seçilir. Kullanılan MRS cihazının kalitesine bağlı olarak daha küçük hacimdeki voksel seçilebilir. Hacim küçüldüğü için sinyal gürültü oranı azalmaktadır ve daha iyi sonuç almak için sinyal gürültü oranının artırılması gerekir. Bu da küçük vokselde sinyal

averaj sayısını (NSA) arttırmakla mümkündür. Uygulamaya bakıldığı zaman genelde tümör alanı ve az miktarda da olsa tümörün etrafındaki alan seçilerek işlem yapılır. Ayrıca genel uygulamada tek vöksel seçilir ve MRS sonuçları seçilen vökselin konumuna bağlıdır. Yani seçilen vöksel içerisinde canlı tümör hücre sayısı fazla olmalı ve bunların kenarındaki hücreler bulunmalıdır. İyi vöksel seçimi ile daha iyi sonuç alınabilir (Cüce, 2005).



Şekil 4.2 Vöksel seçimi (Başak, 2005)

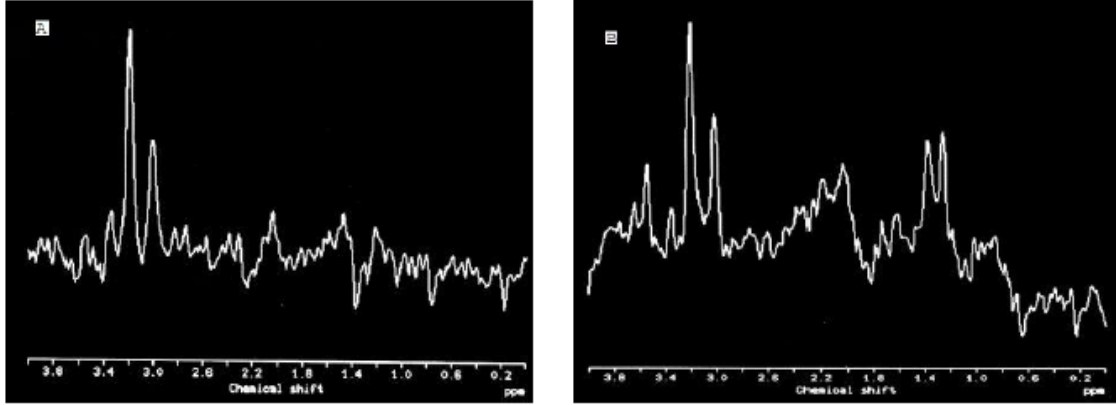
4.1.2. Manyetik alan homojenizasyonu

Spektrumda yer alan pikler arasındaki mesafe birbirine eşit uzaklıkta olmalıdır ve bunlar birbirinden net olarak ayrılmalıdır. MRS cihazlarının Tesla değeri düştükçe salınım oranı artmaktadır. Bundan dolayı 1.5 T altındaki cihazlar MRS ölçümünde pek kullanılmaz. MRS ile elde edilen spektrum kalitesi manyetik alanın homojenliğine doğrudan bağlı olduğundan su metabolitinin baskılanması gerekir (Cüce, 2005).

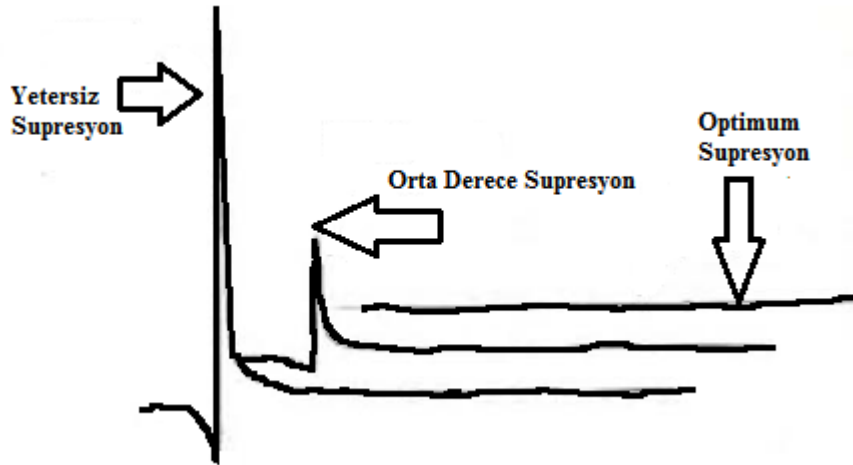
4.1.3. Su ve yağ baskılanması

MRS görüntülemeye seçilen vökseldeki su baskılanmazsa ve yağ ortadan kaldırılmazsa daha küçük metabolitler ölçülemez. Suyu baskılama yönteminin adı CHESS' dir. CHESS tekniği frekans seçen 90^0 işaretler kullanır. Bu işaretler suyun baskılanmasında çok iyi sonuçlar verir. Yağ baskılanması diğer metabolitlerin ölçülmesini engellemesi daha kolaydır. Yağ metabolitleri skalp denilen bir bölgede olduğundan vöksel seçimi bu bölge haricinde yapılarak bunun önüne geçilebilir. Diğer bir yöntem ise uzun TE süresi seçilerek yağın spektruma etkisi azaltılmaktadır. Sudan elde edilen değerin diğer metabolitlerden elde edilen değere göre çok daha büyük olduğundan sudan elde edilen sinyalleri ölçümün sonuna kadar baskılamak gerekir. Su baskılanması dengeli olarak yapılmalıdır. Çünkü aşırı su baskılanması spektrumdaki su sinyalinin spektrumda aşırı şekilde ortaya çıkmasına neden

olmaktadır. Suyun az baskılanması ise spektrum çizgisinin aşağı yönlü eğilmesine neden olmaktadır (Cüce, 2005).



Şekil 4.3 (a) TE süresi kısa seçilen MRS görüntüsü. (b) TE süresi uzun seçilen MRS görüntüsü (Başak, 2005)



Şekil 4.4 Su baskısı sonucu oluşan bozuk spektrum (Başak, 2005)

4.1.4. MRS incelenmesi

MRS incelemek için kullanılan metotlar şunlardır:

- DRESS (depth resolved surface coil spectroscopy)
- PRESS (point resolved surface coil spectroscopy)
- SPARS (spatially resolved spectroscopy),
- STEAM (stimulated echo acquisition method)

Bunlardan en yaygın kullanılan teknikler PRESS ve STEAM' dir. PRESS tekniği daha uzun TE süresi kullandığı için STEAM tekniğine göre daha avantajlıdır. PRESS tekniğinde elde edilen gürültülü sinyal oranı daha fazladır. Ayrıca su baskılaması için

gerekli işaret miktarı daha azdır. Bu üstünlüğü ayrıca PRESS' in daha büyük vokselde ölçüm yapılmasını sağlar (Başak, 2005).

4.2. Metabolitler

MRS görüntüleme 1.5 T ve üzerinde yapılır ve MRS ile Cr (kreatin), Cho (kolin), Naa (N-Asetil Aspartat) metabolitlerin ölçümü yapılır. Normal bir beyin bölgesinde çekilen MRS' te en büyük piki Naa oluşturur. Naa 2.0-2.02 ppm' de rezonans yapar. Tümör olan bölgede Naa metaboliti azalma gösterir. Cr, beyin hücrelerinin enerji sistemlerinin korunmasında yer alır. Cr piki 3.0-3.02 ppm' de spektrumda ortaya çıkar. Cr, metaboliti normal bir beyin bölgesinde farklı oranlarda pik yapabilir. Ayrıca tümör olan bölgede Cr metabolitinde azalma görülür. Cho metaboliti 3.22 ppm' de ortaya çıkar. Cho metabolitinin artışı hücrenin hızla çoğalmasına bağlıdır. Tümör olan bölgede de hücreler hızla artış gösterdiğinden, bu bölgede ölçülen Cho değeri yüksektir. Lipitler, alanin, glutamat ve glutamin, miyoinositol ve laktat MRS ile ölçülüp görüntülenebilen diğer metabolitlerdir ancak bu metabolitler bizim çalışmamızda kullanılan MRS değerlerinde yoktur (Başak, 2005; Cüce, 2005).

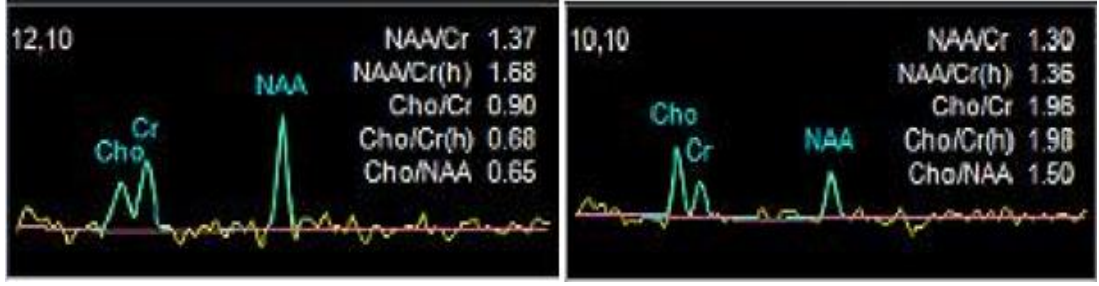
4.3. MR Spektroskopi İle Beyin Tümörü Teşhisi

Beyin tümörü teşhisinde, yıllarca laboratuvar çalışmalarında kullanılan MR Spektroskopi önemli bir yere sahiptir. MR Spektroskopi MR görüntülemeden farklı olarak su ve yağ haricindeki metabolitlerin görüntülenmesini sağlar. Tümör tespitinde kolin (Cho), N-Asetil Aspartat (Naa) ve kreatin (Cr) metabolitlerin değerleri önemlidir. Tümör bölgesinde ölçülen kolin (Cho) değeri yüksektir. Ancak beyin kanaması gibi farklı rahatsızlık geçiren hastaların şüphelenilen bölgelerde ölçülen bu metabolit değerleri de normale göre farklıdır ve bu değerlerin doğru yorumlanması gereklidir. MRS ilgili uzman doktora ölçüm yapılan bölgenin Cr, Cho ve Naa metabolit değerleri ile bunların grafiksel gösterimini verir (Esen ve ark., 2015; Başak, 2005).



Şekil 4.5 MR Spektroskopi ölçümü ekran görüntüsü

Şekil 4.5’ de görüleceği üzere uzman tarafından, MRS’ te şüphelenilen alan seçiliyor ve bu alan 25 farklı alana ayrılıyor. Bu 25 alanın her biri için Cho/Naa, Cho/Cr ve Naa/Cr oranları veriliyor ve metabolit değerleri grafiklerle çiziliyor. Beyin tümörü olan bölgede Cho değeri yüksek olduğunda Cho/Cr ve Cho/Naa oranları yükseliyor. Daha önce bahsedildiği gibi beyin kanaması ve beyin tümörü karıştırılma durumu yüksek olduğundan grafiklerde incelendiğinde tümör olan bölgede Cho değeri yüksektir, Şekil 4.6’ da MRS’ te alınan bir görüntü ile bu durum desteklenmektedir.

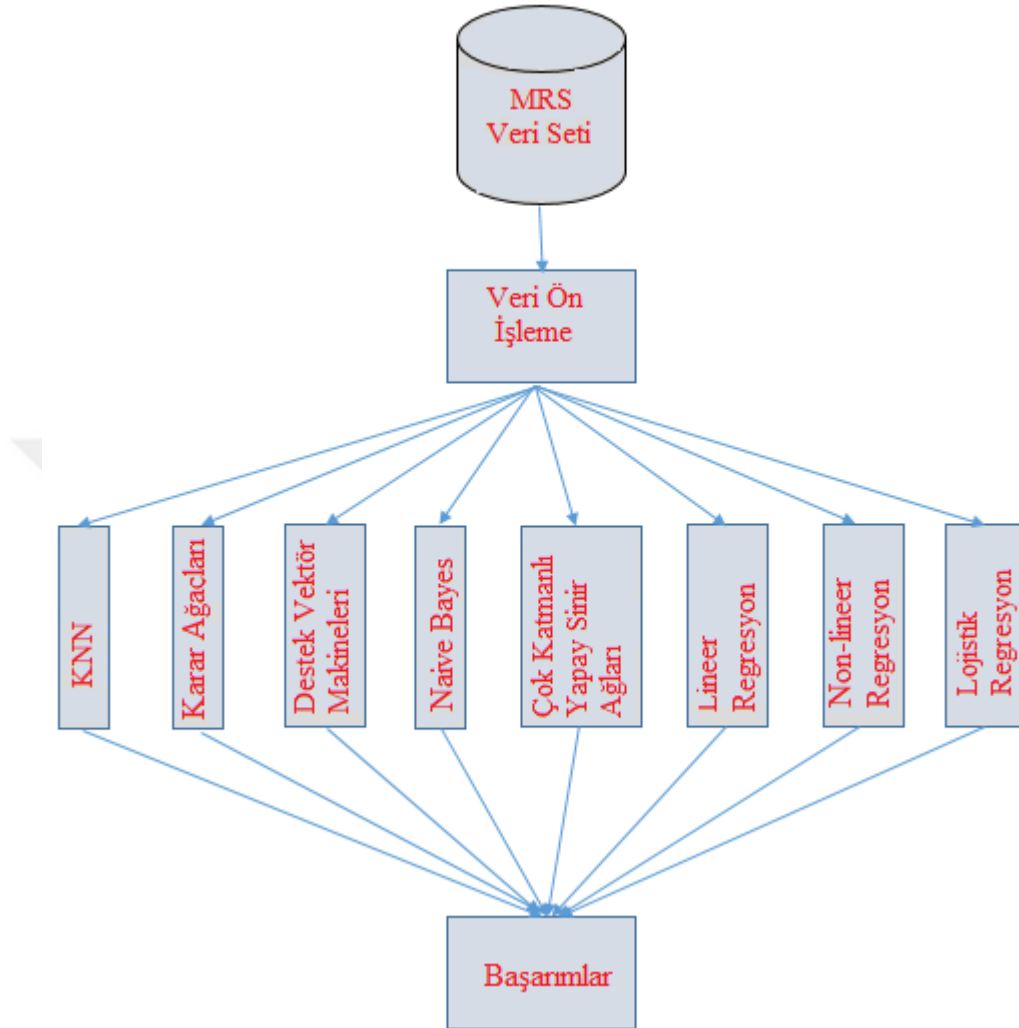


Şekil 4.6 (a) Tümör olmayan alan ölçümü. (b) Tümörü olan alanda ölçüm

Çalışmamızda Ocak 2009 ve Mayıs 2017 tarihleri arasında, Kahramanmaraş Sütçü İmam Üniversitesi Tıp Fakültesi Beyin ve Sinir Cerrahisi kliniğinde beyin tümörü teşhisi konulmuş ve teşhis konulmamış hastaların çeşitli bilgileri kullanılmıştır. Bu verilerin kullanılması için gerekli izinler alınmıştır ve kesinlikle başkaları ile paylaşılmayacaktır. Veri tabanında bulunan verilerin tamamı, Beyin ve Sinir Cerrahisi uzmanına kontrol edilmiştir. Bu sayede yanlış verinin sonucu olumsuz etkilemesinin önüne geçilmiştir.

5.DENEYSEL ÇALIŞMALAR

5.1. Sınıflandırma Yöntemleri İçin Deneysel Çalışmalar



Şekil 5.1 Sınıflandırma yöntemleri için yapılan çalışmanın blok diyagramı

Çalışma WEKA programı kullanılarak yapılmıştır. Yapılan 5 adet yöntemin hepsi 10 kat çapraz doğrulama kullanılarak yapılmıştır.

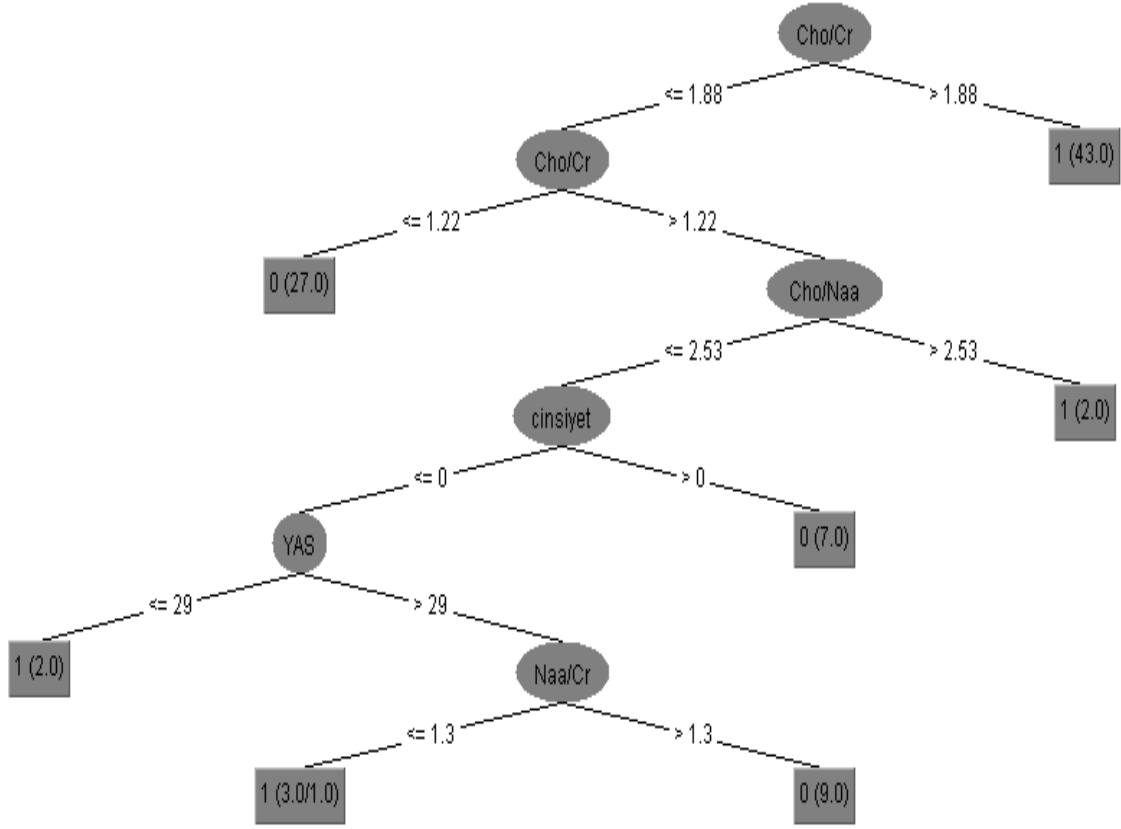
Çizelge 5.1’ de KNN, Lojistik Regresyon, Destek Vektör Makineleri, Çok Katmanlı Yapay Sinir Ağları, Karar Ağaçları ve Naive Bayes yöntemleri kullanılarak elde edilen hata oranları ve başarıları sıralanmaktadır. Çizelge 5.2’ de ise KNN, Lojistik Regresyon, Destek Vektör Makineleri, Çok Katmanlı Yapay Sinir Ağları, Karar Ağaçları ve Naive Bayes yöntemleriyle elde edilen doğru pozitif oranı, yanlış pozitif oranı, hassasiyet gibi bilgiler yer almaktadır.

Çizelge 5.1 Sınıflandırma yöntemleri istatistik sonuçları

Başarım Ölçütü:	KNN	Lojistik Regresyon	Destek Vektör Makinele ri	Çok Katmanlı Yapay Sinir Ağları	Karar Ağaçları	Naive Bayes
Doğru Olarak Sınıflandırılan Örnek Sayısı:	89	84	77	84	81	81
Yanlış Olarak Sınıflandırılan Örnek Sayısı:	4	9	16	9	12	12
Kappa İstatistik Değeri:	0.9139	0.8061	0.6597	0.8057	0.7412	0.743
Ortalama Mutlak Hata (Mean Absolute Error):	0.0547	0.1156	0.172	0.1715	0.1381	0.1399
Ortalama Hata Kareleri Toplamı Kökü (Root Mean Square Error):	0.1936	0.2781	0.4148	0.3211	0.3453	0.3212
Bağıl Mutlak Hata (Relative Absolute Error):	10.96 %	23.16 %	34.48 %	34.37 %	27.67 %	28.04%
Ortalama Bağıl Mutlak Hata Kareleri Toplamı (Root relative squared error):	38.75 %	55.66 %	83.02 %	64.27 %	69.11 %	64.28%
Toplam Örnek Sayısı:	93	93	93	93	93	93

Çizelge 5.2 Sınıflandırma yöntemleri doğruluk sonuçları

Sınıf Adı:	Sınıf:	TP Oranı:	FP Oranı:	Kesinlik (Precision):	Duyarlılık (Recall):	F- Ölçütü:	ROC Alanı:
KNN	1	0,939	0,023	0,979	0,939	0,958	0,985
	0	0,977	0,061	0,935	0,977	0,956	0,985
	Ortalama:	0,957	0,041	0,958	0,957	0,957	0,985
Lojistik Regresyon	1	0,898	0,091	0,917	0,898	0,907	0,977
	0	0,909	0,102	0,889	0,909	0,899	0,977
	Ortalama:	0,903	0,096	0,904	0,903	0,903	0,977
Destek Vektör Makineleri	1	0,714	0,045	0,946	0,714	0,814	0,834
	0	0,955	0,286	0,750	0,955	0,840	0,834
	Ortalama:	0,828	0,159	0,853	0,828	0,826	0,834
Çok Katmanlı Yapay Sinir Ağı	1	0,918	0,114	0,900	0,918	0,909	0,930
	0	0,886	0,082	0,907	0,886	0,897	0,930
	Ortalama:	0,903	0,098	0,903	0,903	0,903	0,930
Karar Ağaçları	1	0,878	0,136	0,878	0,878	0,878	0,900
	0	0,864	0,122	0,864	0,864	0,864	0,900
	Ortalama:	0,871	0,130	0,871	0,871	0,871	0,900
Naive Bayes	1	0,816	0,068	0,930	0,816	0,870	0,912
	0	0,932	0,184	0,820	0,932	0,872	0,912
	Ortalama:	0,871	0,123	0,878	0,871	0,871	0,912

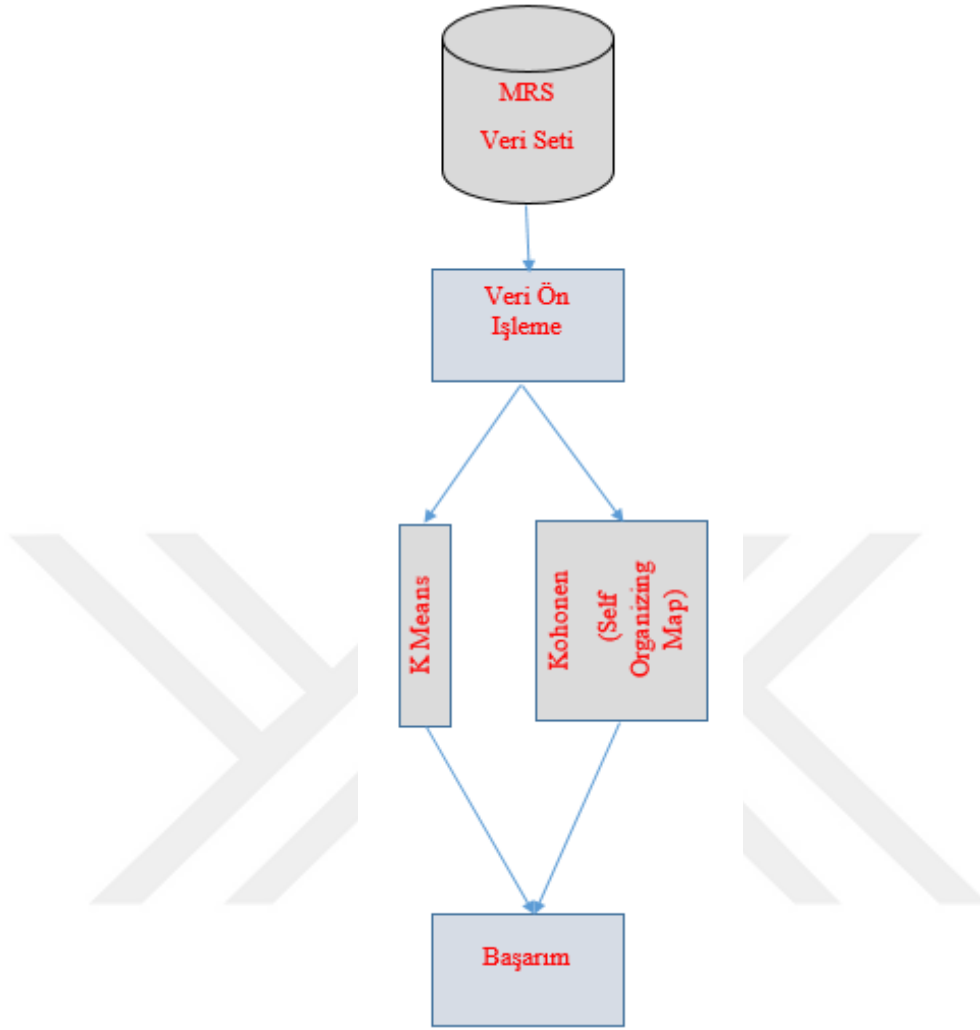


Şekil 5.2 Karar ağaçları yöntemi için ağaç yapısı

WEKA programında Naive Bayes, Çok Katmanlı Yapay Sinir Ağları, Destek Vektör Makineleri, KNN, Karar Ağaçları, Lojistik Regresyon sınıflandırma yöntemlerinde çalışılmıştır. Çalışmada 10 kat çapraz doğrulama tekniği kullanılmıştır. En yüksek başarı %95.69 ile KNN yönteminde en düşük başarı ise %82.79 ile Destek Vektör Makineleri yönteminde elde edilmiştir. Destek Vektör Makineleri için elde edilen başarı da düşük bir başarı olmayıp bu yöntem için de uzman karar destek sistemi oluşturulabilir. Şekil 5.2' de WEKA programında, Karar Ağaçları yöntemi için elde edilen ağaç yapısı yer almaktadır.

Lojistik Regresyon yöntemi için yapılan çalışma da lojistik regresyon için %95.69 başarı elde edilmiştir. Ortalama mutlak hata (mean absolute error) 0.11 ve ROC Area 0.977 olarak gerçekleşmiştir. Bu değerler lojistik regresyonda da yüksek başarı ile sınıflandırma yapılabileceğini göstermiştir. Lineer ve Non-lineer regresyonlarda yüksek başarı elde edilememesi bu yöntemlerin daha basit, yani özellik (attribute) sayısı az olan veri setlerinde iyi sonuçlar vermesine bağlayabiliriz.

5.2. Kümeleme Yöntemleri İçin DeneySEL Çalışmalar

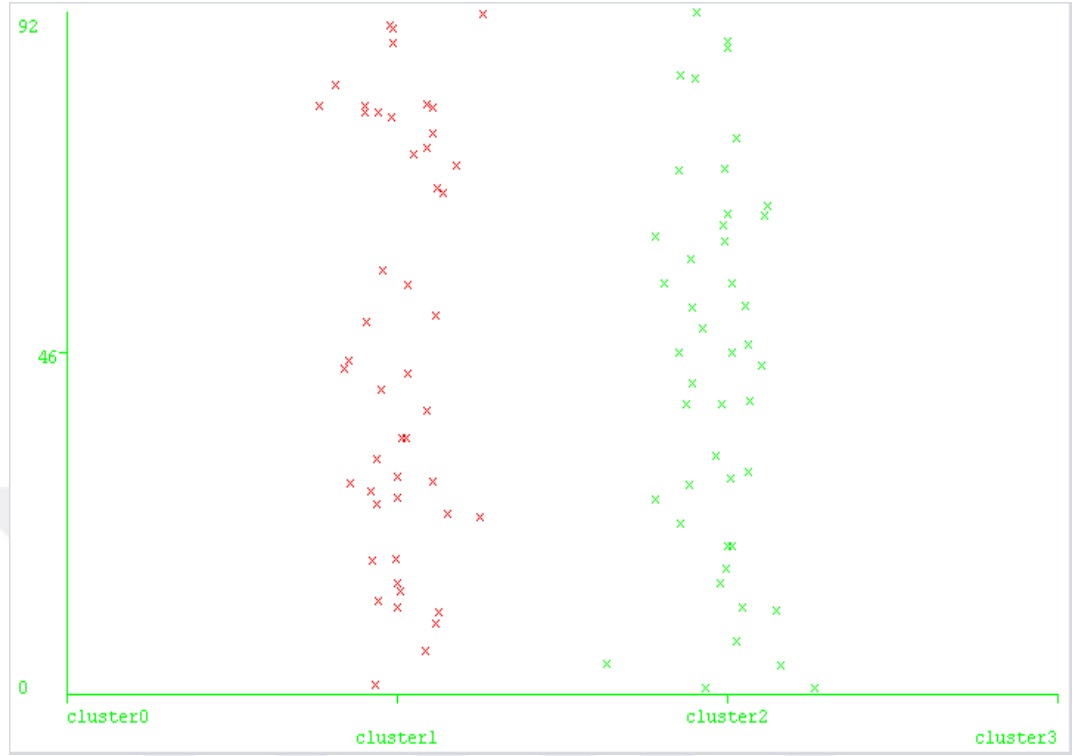


Şekil 5.3 Kümeleme yöntemleri için yapılan çalışmanın blok diyagramı

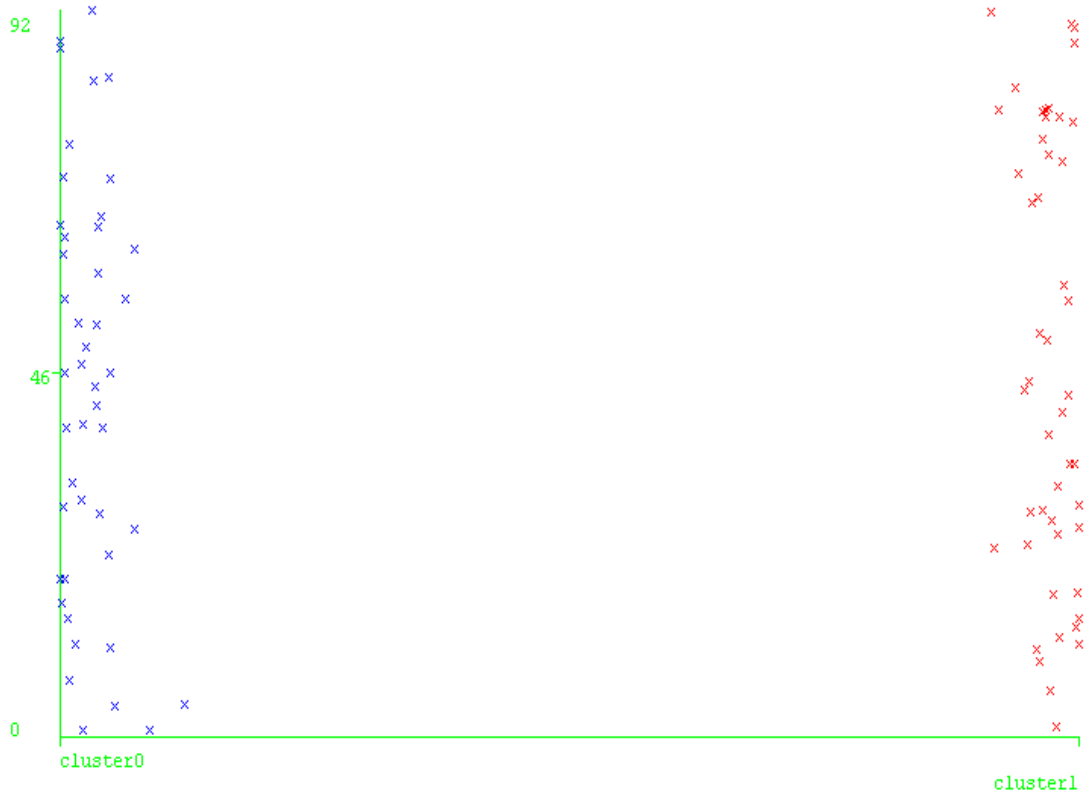
Self Organizing Map (SOM) yönteminde, veri setimizde yer alan 93 adet veri kümelenmiştir. SOM 93 adet veriden 45 tanesini bir kümeye 48 tanesini diğer kümeye toplamıştır. 93 veriden 49 tanesi beyin tümörü teşhisi konulmuş diğer 44 tanesine beyin tümörü teşhisi konulmamıştır, yani yöntem 1 adet veri için yanlış kümeleme yapmıştır. Çizelge 5.3'te görüleceği gibi %98,92 başarımlı sağlamıştır. Şekil 5.4' te WEKA programında Self Organizing Map (SOM) yönteminden elde edilen kümeleme grafiği yer almaktadır.

K-ortalama yöntemi için başlangıçta 2 küme seçilmiştir. Yöntem 93 adet veriden 45 tanesini bir kümeye 48 tanesini diğer kümeye yerleştirmiştir. Veri setimizde yer alan 93 veriden 49 tanesi beyin tümörü teşhisi konulmuş diğer 44 tanesine beyin tümörü teşhisi konulmamış hastalardır. Sonuçta yanlış kümelenen veri sayısı 1' dir. Çizelge 5.3'te

görülebileceği gibi %98,92 başarımlı sağlamıştır. Şekil 5.5’de WEKA programında K-ortalama yönteminden elde edilen kümeleme grafiği yer almaktadır.



Şekil 5.4 WEKA programında Kohonen (Self Organizing Map) yöntemi için elde edilen kümeleme ekranı



Şekil 5.5 WEKA programında K means yöntemi için elde edilen kümeleme ekranı

Çizelge 5.3 Kümeleme yöntemlerinin başarısı

Yöntem Adı:	Başarı:
Kohonen (Self organizing Map)	%92.92
K means	%98.92

5.3. Derin Öğrenme İçin DeneySEL Çalışmalar

Veri setimizi derin öğrenme yöntemini kullanarak işledik. Sınıflandırma ve kümeleme yöntemlerini çalıştığımız WEKA programında derin öğrenme ara yüzü bulunmadığından, Python yazılımını kullandık.

Çalışmamız veri seti 93 adet olduğundan 1 adet gizli katman oluşturuldu. Giriş katmandaki nöron sayısı, giriş özellik sayımız 5 olduğundan 5 olarak belirlendi. Çıkış katmandaki nöron sayısı ise çıkışı ya Tümör (1) ya da Tümör Değil (0) olduğundan 2 olarak belirlendi. Gizli katmandaki nöron sayısı ise giriş ve çıkış nöron sayısının ortalaması olan 4 olarak alındı.

```
model.add(Dense(4, input_dim=5))
model.add(Activation('relu'))
model.add(Dense(2))
model.add(Activation('softmax'))
```

Veri sayısı az olduğundan %33 test %66 eğitim için ayrıldı. % 33, 31 adet veriye denk gelip bu yazılım tarafından rastgele seçilmiştir. Ayrıca bu 31 adet veri hiçbir eğitime tabi tutulmadan test edilmiştir. Elde edilen doğruluk, ROC Alanı:, f1 ölçütü, Duyarlılık:, konfüzyon matrisi bu 31 adet veri test edilerek elde edilmiştir. Doğruluk-İterasyon grafiği ise hem eğitim hem de test için elde edilmiştir.

```
X_train, X_test, y_train, y_test = train_test_split(giris, cikis,
test_size=0.33)
```

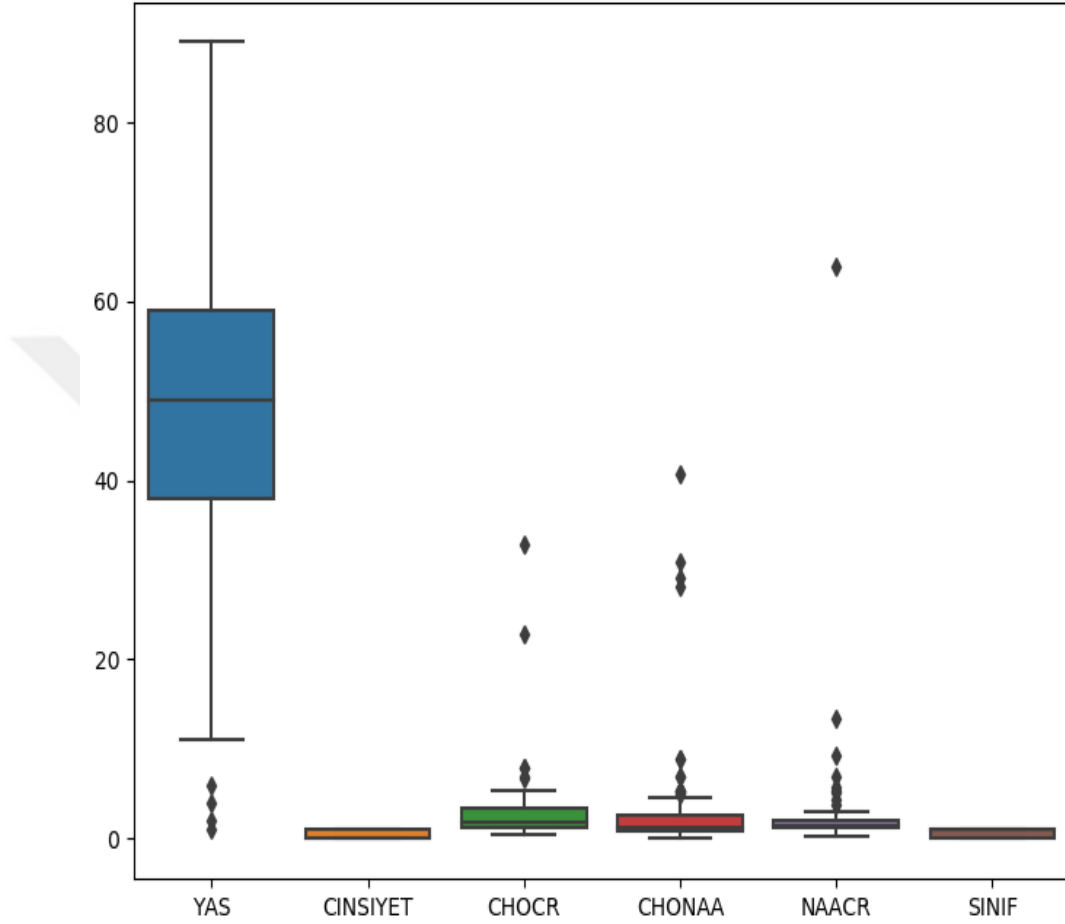
Sonuçları daha doğru görebilmek için 50.000, 100.000 ve 200.000 iterasyonlar için çalışılmıştır.

Python programlama dilinde kullandığımız kütüphaneler şunlardır:

```
from keras.models import Sequential
from keras.layers import Dense, Activation, Embedding, LSTM, Dropout,
advanced_activations
from keras.callbacks import ModelCheckpoint, TensorBoard
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.utils import shuffle
import seaborn as sns
from sklearn.metrics import confusion_matrix, average_Duyarlılık:_score,
recall_score
```



```
from sklearn import metrics
from sklearn.metrics import roc_curve, auc, recall_score
```

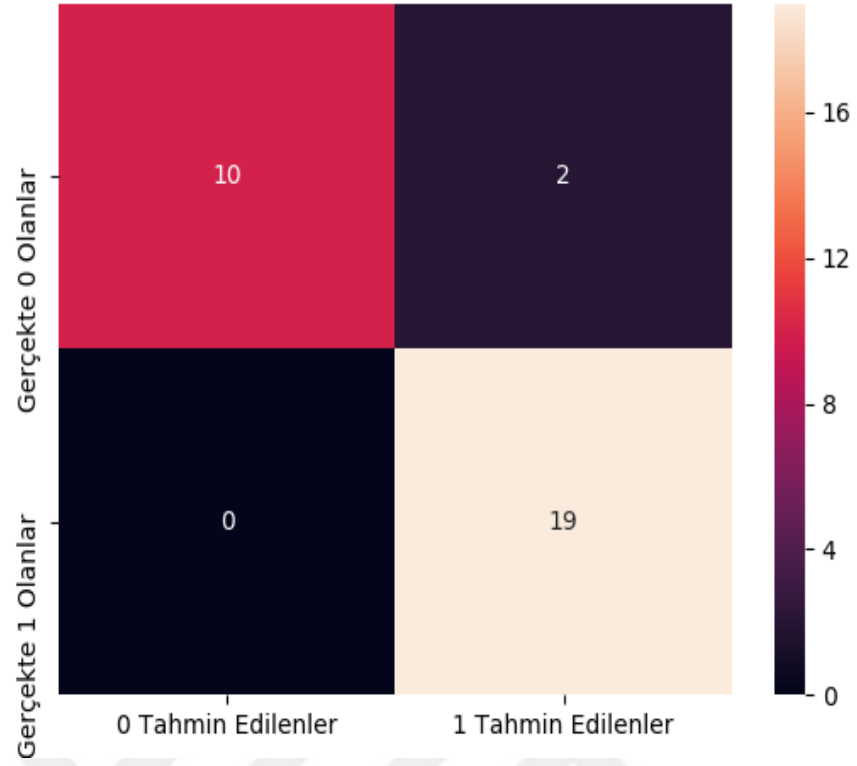


Şekil 5.6 Python’ da elde edilen giriş özelliklerinin grafiksel gösterimi

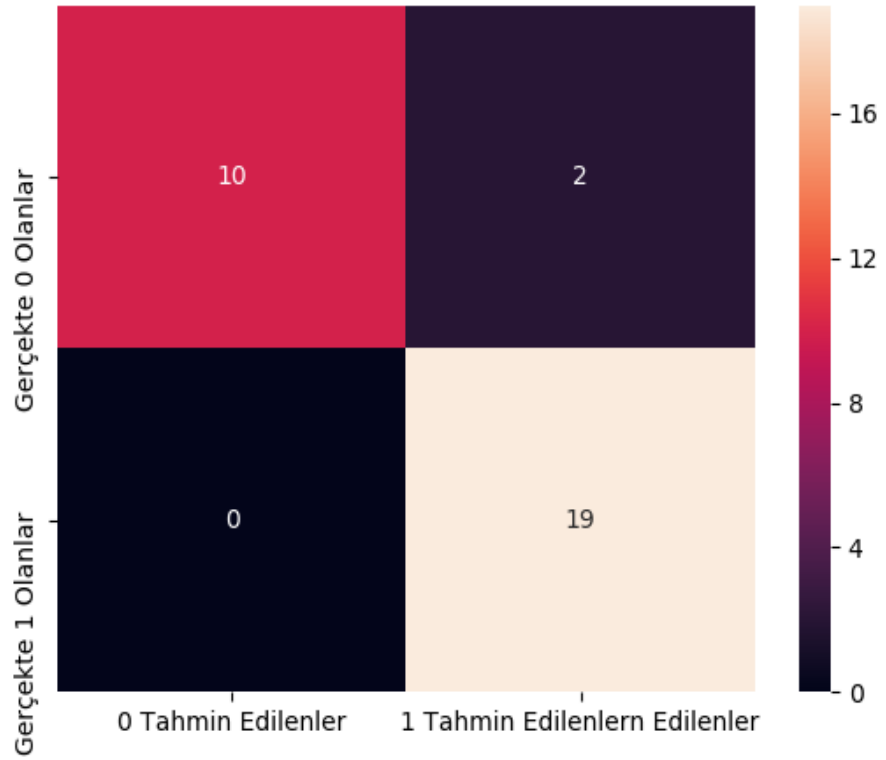
Şekil 5.6’ da Python’ dan elde ettiğimi veri setimizde yer alan 5 adet özelliğin grafiği yer almaktadır. Kullanılan veri setindeki hastalar 35- 60 yaş aralığındadır.

Çizelge 5.4’ de 50.000, 100.000 ve 200.000 iterasyonların da elde edilen başarı ölçütleri yer almaktadır. Buradan anlaşılacağı üzere iterasyon sayısı artmasına rağmen elde edilen başarı sabit kalmıştır.

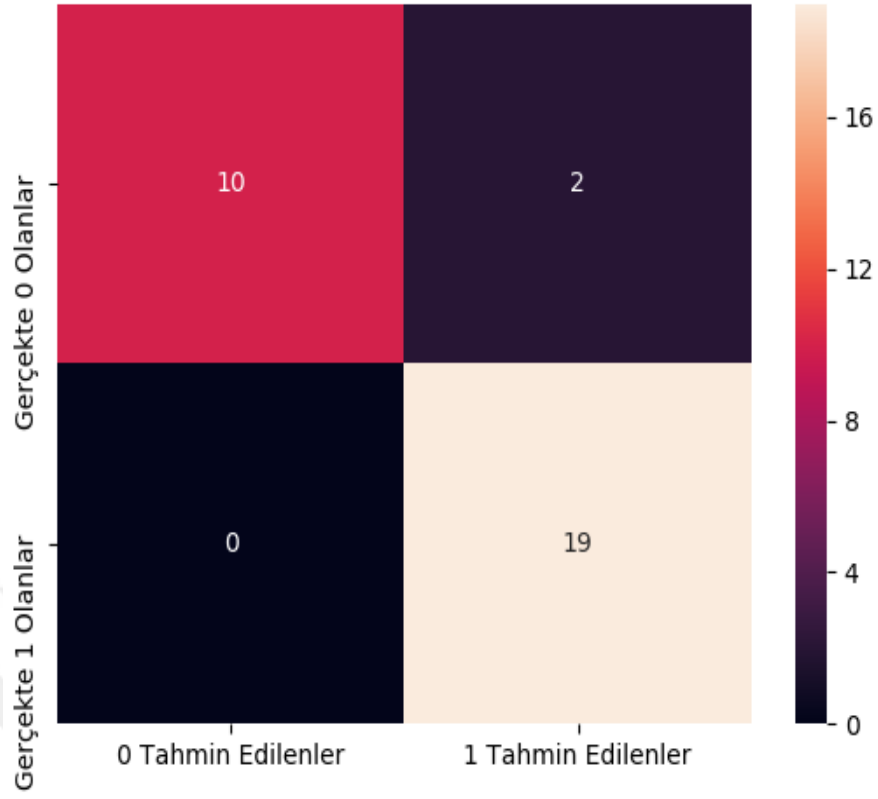
Şekil 5.7 50.000 iterasyon, Şekil 5.8 100.000 iterasyon ve Şekil 5.9 200.000 iterasyonlar da elde edilen konfüzyon matrisini göstermektedir.



Şekil 5.7 50.000 iterasyon için konfüzyon matrisi



Şekil 5.8 100.000 iterasyon için konfüzyon matrisi



Şekil 5.9 200.000 iterasyon için konfüzyon matrisi

Çizelge 5.4 Derin öğrenme yöntemi başarıları

İterasyon Sayısı:	Sınıf:	Kesinlik (Precision):	Duyarlılık (Recall):	F-Ölçütü:	AUC:	Doğruluk Başarısı:
50.000	1	1	0,83	0,91	0,916	0,9354
	0	0,9	1	0,95		
	Ortalama:	0,94	0,94	0,93		
100.000	1	1	0,83	0,91	0,916	0,9354
	0	0,9	1	0,95		
	Ortalama:	0,94	0,94	0,93		
200.000	1	1	0,83	0,91	0,916	0,9354
	0	0,9	1	0,95		
	Ortalama:	0,94	0,94	0,93		

6.SONUÇLAR

Beyin tümörü, insan hayatını olumsuz etkileyen, hatta ölümle sonuçlanabilen ciddi bir rahatsızlıktır. Bu rahatsızlığın ivedi olarak teşhis edilmesi ve uygun tedavi yönteminin kullanılması oldukça önemlidir. Basından takip edebildiğimiz kadarıyla yanlış tedavi sonucu hastaların yaşam kaliteleri düşmekte hatta yaşamları son bulabilmektedir. Söz konusu rahatsızlığın beyinde olduğu düşünülürse ciddi bir uzmanlık gerektirdiği ortadadır.

MRS beyin tümörü teşhisinde tek ve bağlayıcı yöntem özelliği olmasa da beyin tümörü teşhisi ve tedavi yönteminin başarısını tespitinde önemli bir yöntemdir. Çalışmamızda bize yardımcı olan beyin ve sinir cerrahisi uzmanı Sayın Doç. Dr. İdris ALTUN' un da belirttiği üzere maalesef MRS görüntüleri ve elde edilen değerlerin yorumlanması pek de iyi bilinen bir konu değildir. MRS ile elde edilen verilerin çok doğru okunması ve yorumlanması gerekmektedir. Çünkü elde edilen değerler beyin kanaması veya farklı bir rahatsızlığın sonucu ortaya çıkan değerlerde olabilir. Ama beyin tümörü kaynaklı değer değişimi söz konusu ise bunun da ivedi olarak tedavi edilmesi gerekir.

Çalışmamızda 93 adet tümör teşhisi konulmuş veya konulmamış hastaların çeşitli bilgileri kullanılmıştır. Veri setimiz Kahramanmaraş Sütçü İmam Üniversitesi Tıp Fakültesi'nde 2009 ve 2017 yılları arasında beyin tümörü tespiti için çekilen hastaların MRS görüntü ve raporlarından elde edilen bilgiler kullanılmıştır. Veri setimiz arttırılmak istenmiştir, ancak uzun çalışmalar sonucu bu kadar veri toplanabilmiştir. Veri setinde yer alan verilerin sayısı arttıkça sınıflandırma başarısının daha da artacağı düşünülmektedir.

Çalışmamızda beyin tümörü teşhisinde yaygın olarak kullanılmaya başlanılan MRS verilerini kullanarak doktora yardımcı bir sistemin temelini oluşturan çeşitli sınıflandırma yöntemlerinin başarıları incelenmiştir.

Sınıflandırma için kullanılan KNN, Destek Vektör Makineleri, Çok Katmanlı Yapay Sinir Ağları, Naive Bayes, Karar Ağaçları, Lojistik Regresyon, Lineer Regresyon, Non-lineer Regresyon yöntemlerinin başarımları %82 ile %95'ler arasında değişmektedir. Bu konuda en başarılı sınıflandırma %95 ile KNN yönteminde elde edilmiştir.

Kümeleme yöntemlerinden K-ortalama ve Kohonen (Self Organizing Map) kullanılmıştır. Bunların başarımları da %92.92 ile %98.92 aralığında olup, oldukça iyi sonuçlar elde edilmiştir. Bu konuda en başarılı sonuç %98.92 ile K-ortalama yöntemiyle elde edilmiştir.

Elde edilen sonuçlar böyle bir sistemin oluşturulabileceğini göstermiştir.



KAYNAKLAR

- Alan, M.A. 2014. “Sivas Erzincan Kalkınma Projesi (SEKP) Verilerinin Veri Madenciliği ile Sınıflandırılması ve Kümelenmesi” *Manas Sosyal Araştırmalar Dergisi* 2014; Cilt: 03 Sayı: 10 2014: 129-144.s.
- Al-Azzawi, M.I.M., 2017. Heart Disease System Prediction Using Data Mining Techniques. *The University Of Turkish Aeronautical Association Institute of Science and Technology*. 70s.
- Alkan, A., Altun, S., “Analysis of MRS Signals by Using Clustering Techniques” International Advanced Researches & Engineering Congress(IAREC’2017), 16-18 Kasım 2017, Osmaniye, Türkiye.
- Alkan, A., Altun, S., Altun, İ., “Classification of MR Spectroscopy Data in Brain Tumor Diagnosis” International Advanced Researches & Engineering Congress (IAREC’2017), 16-18 Kasım 2017, Osmaniye, Türkiye.
- Alkan, A., Günay, M., “Identification of EMG Signals Using Discriminant Analysis and SVM Classifier”, *Expert Systems with Applications* 39, (2012) s.44–47.
- Alkuşak, E., Gök, M., “Karaciğer Yetmezliğinin Teşhisinde Makine Öğrenmesi Algoritmalarının Kullanımı”, ISITES 2014 Sempozyumu (ISITES’2014), 18-20 June 2014, Karabük, Türkiye.
- Anwer, O, M, A., 2017. Derin Öğrenme Yöntemleri İle Göğüs Kanseri Teşhisi. Yüksek Lisans Tezi. Türk Hava Kurumu Üniversitesi Fen Bilimleri Enstitüsü. 92s.
- Arora, R., Suman. 2012. “Comparative Analysis of Classification Algorithms on Different Datasets using WEKA” *International Journal of Computer Applications* (0975–887) Volume 54– No.13, September 2012, 21-25s.
- Aygün, C, R., 2017. Derin Öğrenme Yöntemleri İle Bilgisayar Ağlarında Güvenliğe Yönelik Anormallik Tespiti. Yüksek Lisans Tezi. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü. 69s.
- Carlos A., Daniel A., S., Alfredo V., Enrique R. 2014. “Automated Classification of Brain Tumours from Short Echo Time in Vivo MRS data using Gaussian Decomposition and Bayesian Neural Networks” *Expert Systems with Applications* 41 (2014) 5296–5307s.
- Çokluk Ö. Kuram ve Uygulamada Eğitim Bilimleri / Educational Sciences: *Theory & Practice* 10 (3) • Yaz / Summer 2010 • 1357-1407.
- Coşgun, E., Karaağaoğlu, E. 2011. “Veri Madenciliği Yöntemleriyle Mikro Dizilim Gen İfade Analizi” *Hacettepe Tıp Dergisi* 2011; 42:180-189
- Cüce, A.M. 2005. “Beyin Kitlelerinin Tanısında MR Spektroskopinin Yeri İle Kısa ve Uzun TE MR Spektroskopinin Karşılaştırılması”. G.Ü. Uzmanlık Tezi, İstanbul, 98s.

- Dirican A. Evaluation of the Diagnostic Test's Performance and Their Comparisons. *Cerrahpaşa J Med* 2001; 32 (1): 25-30.
- Eroğlu K., Palabaş, T. http://www.emo.org.tr/ekler/a4e4f37b054dd7c_ek.pdf (Erişim Tarihi:01.10.2017)
- Esen, S.Ö., Bozkurt, M., Adıbelli, H.Z., Aykut, E., Canverenler, S. 2014. "Proton MR Spektroskopinin Beyin Tümörlerinde Tanısal Değeri." *ITÜ Tepecik Eğit Hast. Derg.* 2014; 24 (2): 93-98.s.
- Girginer, G., Cankuş, B. "Tramvay Yolcu Memnuniyetinin Lojistik Regresyon Analiziyle Ölçülmesi: Estra Örneği" YÖNETİM VE EKONOMİ Yıl:2008 Cilt:15 Sayı:1 *Celal Bayar Üniversitesi İ.İ.B.F. MANİSA*.
- Göker, H., Şahin, İ., Tekedere, H. 2015. "Erken Çocukluk Döneminde Otizm Teşhisine Yönelik Dinamik Uzman Sistem Tasarımı" *Bilişim Teknolojileri Dergisi*, Cilt: 8, Sayı: 3, 167–172.s.
- Haltaş, A., Alkan, A. İmmunohistokimyasal Boyalar ile Tiroid Tümörü Teşhisinde Naive Bayes Algoritması Kullanılması, 5-7 Şubat, 2014, Mersin, s.931-935.
- Hark, C., 2013. Öğrencilerin Akıllı Tahtaya İlişkin Tutumlarının İncelenmesine Yönelik Bir Veri Madenciliği Uygulaması. Yüksek Lisans Tezi. Fırat Üniversitesi Eğitim Bilimleri Enstitüsü. Elazığ. 51s.
- İşler, Y., Narin, A. 2012. "WEKA Yazılımında k-Ortalama Algoritması Kullanılarak Konjestif Kalp Yetmezliği Hastalarının Teşhisi" *SDU Teknik Bilimler Dergisi* 2012; Cilt: 2 Sayı: 4 2012: 21-29.s
- Jin X, Bie R. Random Forest and PCA for Self-Organizing Maps Based Automatic Music Genre Discrimination, Conference on Data Mining, 2006: 414-7.
- Karatağ, O. 2005. "İntrakranyal Yer Kaplayıcı Lezyonların Ayırıcı Tanısında MRSpektroskopinin Yeri". Radyoloji Uzmanlık Tezi. Sağlık Bakanlığı Şişli Etfal Eğitim ve Araştırma Hastanesi. İstanbul, 50s.
- Karlıdağ, O., 2016. Electricity Tariff Usage Prediction via Data Mining. The Republic of Turkey Bahçeşehir University Graduate School of Natural and Applied Sciences Information Technologies. İstanbul. 45s.
- Kavzoğlu, T., Çölkesen, İ. 2010. "Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi" *Harita Dergisi* Temmuz 2010 Sayı 144: 73-82.s.
- Kaya, Y., Pehlivan, H., "KNN, NN, BAYES, DT ve SVM Kullanılarak EKG Vurularının Sınıflandırılması", ISITES 2014 Sempozyumu (ISITES'2014), 18-20 June 2014, Karabük, Türkiye.
- Kılıç, S. 2015. Kappa testi. *Journal of Mood Disorders* Volume: 5, Number: 3, 2015: 142-144.s.

- Kumdereli, C., Ü. 2012. EEG Sinyallerindeki Epileptiform Aktiviteye Veri Madenciliği Yöntemlerinin Uygulanması. Yüksek Lisans Tezi. Trakya Üniversitesi Fen Bilimleri Enstitüsü. Edirne. 130s.
- Lukasa, A. Devosa, J.A.K. Suykensa, L. Vanhammea, F.A. Howeb, C. Majo'sc, A. Moreno-Torresd, M. Van Der Graafe, A.R. Tateb, C. Aru'sf, S. Van Huffela. 2004. "Brain tumor Classification based on long echo proton MRS signals" Artificial Intelligence in Medicine (2004) 31, 73—89s.
- Nagori, M. Joshi, M. 2013. "Methods and Algorithms for Extracting Values from MRS Graph for Brain Tumor Detection" IERI Procedia 4 (2013) 331 – 336s.
- Nizam, H., Akın, S., S, "Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması", 19. Türkiye' de İnternet Konferansı(inet-tr'14), 27-29 Kasım 2014, İzmir, Türkiye.
- Özcan, C., 2014. Veri Madenciliğinin Güvenlik Uygulama Alanları ve Veri Madenciliği ile Sahtekârlık Analizi. Yüksek Lisans Tezi. İstanbul Bilgi Üniversitesi Sosyal Bilimler Enstitüsü. İstanbul. 89s.
- Sattari, T.M., 2009. Eleviyan Sulama Rezervuarında Veri Madenciliği Yaklaşımları İle İşletme Kurallarının Belirlenmesi. Doktora Tezi. Ankara Üniversitesi Fen Bilimleri Enstitüsü. 142s.
- Şeker E. Ş. Bilgisayar Kavramları. <http://bilgisayarkavramlari.sadievrenseker.com> (Erişim Tarihi= 10.10.2017).
- Şekeri, A., Diri, B., Balık, H, H., 2017. "Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme" *Gazi Mühendislik Bilimleri Dergisi* 3 (3). (2017) s.47-64.
- Shokry, A. 2012. "MRS of brain tumors: Diagrammatic representations and diagnostic approach" *The Egyptian Journal of Radiology and Nuclear Medicine* (2012)43, 603–612.s
- Solmaz, R., Günay, M., Alkan, A., "Uzman Sistemlerin Tiroit Teşhisinde Kullanılması", Akademik Bilişim 2013 Sempozyumu(ab'2013), 23-25 Ocak 2013, Antalya, Türkiye.
- Taşkın, Ç., Emel, G.G. 2010. "Veri Madenciliğinde Kümeleme Yaklaşımları ve Kohonen Ağları İle Perakendecilik Sektöründe Bir Uygulama" *Süleyman Demirel Üniversitesi İktisadim ve İdari Bilimler Fakültesi Dergisi* Y.2010, C.15, S.3 s.395-409.
- Tomak, L., Bek, Y. 2009. " İşlem Karakteristik Eğrisi Analizi ve Eğri Altında Kalan Alanların Karşılaştırılması" *Journal of Experimental and Clinical Medicine* 27 (2010) 58 – 65.
- Wehrens R, Buydens LMC. Self and Super-Organizing Maps in R: The Kohonen Package. *J Stat Soft* 2007; 21:1-9.
- Weka: <http://www.cs.waikato.ac.nz/~ml/weka/>. (Erişim Tarihi: 12.12.2017)

Yücebaşı, C., S.,“ Prostat Kanseri Teşhisinde Veri Madenciliği Yöntemlerinin Başarım Karşılaştırması”, Akademik Bilişim 2016 Sempozyumu(ab’2016), 30 Ocak 5 Şubat 2016, Aydın, Türkiye.

Yurdakul, S., 2015.Veri Madenciliği ile Lise Öğrenci Performanslarının Değerlendirilmesi. Yüksek Lisans Tezi Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü. Kırıkkale.105s.



ÖZGEÇMİŞ

Kişisel Bilgiler

Adı, soyadı : Sinan ALTUN
Uyruğu : T.C.
Doğum tarihi ve yeri : 22.10.1986, Kahramanmaraş
Medeni hali : Evli
e-posta : s.altun86@gmail.com

Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Lisans	KSÜ/ Elektrik-Elektronik Müh. Bölümü	2009
Lise	Kadriye Çalık Anadolu Lisesi	2005

İş Deneyimi

Yıl	Yer	Görev
2014-	Kahramanmaraş 112 Acil Çağrı Merkezi	Sistem Yöneticisi

Yabancı Dil

İngilizce

Yayınlar

1. Alkan, A., Altun, S., “Analysis of MRS Signals by Using Clustering Techniques” International Advanced Researches & Engineering Congress (IAREC’2017) 16-18 Kasım 2017, Osmaniye, Türkiye.

2. Alkan, A., Altun, S., Altun, İ., “Classification of MR Spectroscopy Data in Brain Tumor Diagnosis” International Advanced Researches & Engineering Congress (IAREC’2017), 6-18 Kasım 2017, Osmaniye, Türkiye.

Hobiler

Basketbol, Masa tenisi, Kitap okuma