

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**VERİ MADENCİLİĞİ YÖNTEMLERİNİ
KULLANARAK DİYABETİK RETİNOPATİ
HASTALIĞININ TEŞHİSİ**

YÜKSEK LİSANS TEZİ

Menduh ÇERKEZİ

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ**

Tez Danışmanı : Yrd. Doç. Dr. Hayrettin EVİRGEN

Haziran 2013

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

VERİ MADENCİLİĞİ YÖNTEMLERİNİ
KULLANARAK DİYABETİK RETİNOPATİ
HASTALIĞININ TEŞHİSİ

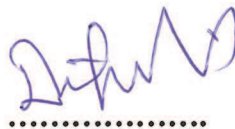
YÜKSEK LİSANS TEZİ

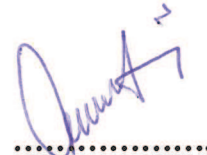
Bil. Müh. Menduh Çerkezi

Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ

Bu tez ^{26 06}16 / 01 /2013 tarihinde aşağıdaki jüri tarafından Oybirliği ile kabul edilmiştir.


.....
Jüri Başkanı


.....
Üye


.....
Üye

Yrd. Doç. Dr. Hayrettin Evirgen

Prof. Dr. Emin GÜNDOĞAR

Doç. Dr. Cemil ÖZ

TEŐEKKÜR

Bu tez alıőmasında, bana rehberlik eden, kısıtlı zaman ierisinde yer ve mekan tanımadan okul dıőında da desteęini esirgemeyen ve olumlu yaklaőımları ile srekli teővik eden tez danıőmanım Sayın Yrd. Do. Dr. Hayrettin Evirgen'e itenlikle teőekkr ederim.

Ayrıca yurt dıőı eęitimim sresince her trl desteklerini esirgemeyen deęerli aileme sonsuz Őkranlarımı sunuyorum.

İÇİNDEKİLER

TEŞEKKÜR.....	ii
İÇİNDEKİLER	iii
ŞEKİLLER LİSTESİ	vi
TABLolar LİSTESİ	vii
ÖZET.....	viii
SUMMARY	ix
BÖLÜM.1.	
GİRİŞ	1
1.1. İlgili Çalışmalar	2
BÖLÜM.2.	
SAĞLIK BİLGİ SİSTEMLERİ	4
2.1. Bilişimin Tanımı	4
2.2. Tıp Bilişimi.....	5
2.3. Tıp Bilişimin Uygulama Alanları	6
2.3.1. Hastane bilgi sistemleri.....	7
2.3.1.1. Karar destek sistemleri	7
2.3.1.2. Hastane bilgi sistemlerinin gelişimi	10
BÖLÜM.3.	
VERİ AMBARLARI.....	12
3.1. Veri Ambarının Tanımı	12
3.2. Konuya Yöneliktir	12
3.3. Bütünleşiktir	13
3.4. Zaman Boyutu Vardır	13
3.5. Sadece Okunabilir.....	14

3.6. Veri Ambarının Temel Özellikleri.....	14
3.7. Veri Ambarı Mimarisi	14
3.7.1. Veri dönüştürülmesi	15
3.7.2. Veri ambarının oluşturulması.....	15
3.7.3. Kullanıcının veri ambarına erişimi	16
BÖLÜM 4.	
VERİ MADENCİLİĞİ.....	17
4.1. Veri Madenciliği Tanımı	17
4.2. Uygulama Alanları.....	18
4.2.1. Sağlık sektörü	18
4.2.2. Devlet uygulamaları.....	19
4.2.3. Haberleşme sektörü	20
4.2.4. Finans sektörü	20
4.3. Veri Madenciliği Süreci	21
4.3.1. Veri temizleme	21
4.3.2. Veri bütünleştirme.....	22
4.3.3. Veri indirgeme.....	22
4.3.4. Veri dönüştürme.....	23
4.3.4.1 Min-Max normalleştirme.....	23
4.3.4.2. Z-score standartlaşma	24
4.3.5. Veri madenciliği algoritmasını uygulama.....	24
4.3.6. Sonuç ve değerlendirme	24
BÖLÜM 5.	
VERİ MADENCİLİĞİ YÖNTEMLERİ	25
5.1. Öngörü Yöntemleri	25
5.1.2. Sınıflandırma ve eğri uydurma yöntemleri.....	26
5.1.2.1. K-en yakın komşu.....	26
5.1.2.2. Ağırlıklı oylama	27
5.1.2.3. Karar ağaçları.....	28
5.1.2.4. Yapay sinir ağları	29
5.1.2.5. Naive Bayes	31

5.2. Tanımlayıcı Yöntemler	32
5.2.1. Kümeleme yöntemleri ve birliktelik kuralları	32
5.2.1.1. K-ortalamlar yöntemi.....	33
5.2.1.2. Apriori algoritması.....	33
BÖLÜM 6.	
DİYABET VE GÖZ HASTALIKLARI	35
6.1. Diyabet.....	35
6.2. Diyabette Göz	35
6.3. Retina.....	36
6.4. Diyabetik Retinopati Hastalığı	36
6.5. Diyabetik Retinopati Sınıflandırılması	37
6.6. Diyabetik Retinopati Seyrini Etkileyen Faktörler	37
BÖLÜM 7.	
UYGULAMA	39
7.1. Yöntemler ve Mesafe Ölçümleri	39
7.2. Giriş Değerler	41
7.3. Sonuç Bölümü	42
7.3.1 Sınıflandırma yöntemlerinin sonuçları	42
7.3.2 Çapraz geçerlilik ve karışım matrisi	42
7.3.2.1. Çapraz geçerlilik.....	42
7.3.2.2. Karışım matrisi	43
7.3.2.3. Çapraz geçerlilikten bir örnek	44
7.3.2.4. Sonuçların değerlendirilmesi.....	45
BÖLÜM 8.	
TARTIŞMA VE SONUÇ	46
KAYNAKLAR.....	47
ÖZGEÇMİŞ	50

ŞEKİLLER LİSTESİ

Şekil 3.1. Veri ambarı mimarisi	15
Şekil 4.1. Veri madenciliği süreci	21
Şekil 5.1. Veri madenciliği yöntemleri	25
Şekil 5.2. K-en yakın komşu algoritması	27
Şekil 5.3. Karar ağacı algoritmasının yapısı	29
Şekil 6.3. Retina	36
Şekil 7.1. Uygulama arayüzü	39
Şekil 7.2. İki nokta arasındaki d uzaklığı	40
Şekil 7.3. n Katlamalı Çapraz Geçerlilik	40

TABLolar LİSTESİ

Tablo 7.1. Giriş deęerlerinin kategorik deęerleri	41
Tablo 7.2. Karışım matrisinin tablosu	43
Tablo 7.3. Çapraz Geçerlilik ve Karışım Matrisinin uygulamada bir örneęi.....	44
Tablo 7.4. Sınıflandırıcıların teşhis sonuçları	45

ÖZET

Anahtar kelimeler: Diyabetik Retinopati, Hastalık Teşhisi, Veri Madenciliği

Son yıllarda hastane bilgi sistemlerinde hastalara ait büyük miktarda tıbbi bilgilere sahip olunuyor. Dijital ortamlarda bilgilere sahip olma amacı hastalara daha kaliteli sağlık hizmeti sunmaktır. Bu tür verilerin arasında muhakkak dikkate değer, saklı kalmış bilgiler mevcuttur. Amaç veri madenciliği yöntemlerini kullanarak elimizdeki saklı kalmış bilgileri ortaya çıkarmaktır.

Bu tez çalışmasında, veri madenciliği yöntemlerini kullanarak diyabetik retinopati hastalıklarının teşhisi yapılmıştır.

Uygulama bölümünde, gerçek veriler kullanılarak K-en yakın komşu, ağırlıklı oylama KNN (Weighted K-nearest neighbor) ve Bayes algoritmaları ile sınıflandırma yapılmış ve elde edilen sonuçlar tartışılmıştır. Çalışmada kullanılan veriler, Sakarya Üniversitesi Eğitim ve Araştırma Hastanesi Göz Polikliniği bölümünden alınmıştır. Veri setinde bulunan bütün parametreler doktorlar tarafından seçilmiş ve bu parametreler baz alınarak hastalığın teşhisi yapılmıştır.

DIAGNOSIS OF DIABETIC RETINOPATHY USING DATA MINING TECHNIQUES

SUMMARY

Key Words: Diabetic Retinopathy, Disease Diagnosis, Data Mining

In recent years, large amount of medical information of patients are stored in hospital information systems. The purpose of having the information in digital environments is to provide better health care for patients. Of course there are remarkable and hidden data among patient's data. The aim is to discover these hidden data by using data mining techniques.

In this study, diagnosis of diabetic retinopathy has been made using data mining techniques.

In the application part, the classification has been made with K-Nearest Neighbor (KNN), Weighted KNN and Naive Bayes algorithms in real life data set. And finally, the results were discussed.

The data set was obtained at the Eye Clinic of the Sakarya University Educational and Research Hospital. All the parameters in the data set were selected by doctors and evaluation was made based on these parameters.

BÖLÜM 1. GİRİŞ

Diyabet ya da tıptaki adıyla Diabetes Mellitus, insülin hormonunun, vücutta bulunmaması ya da az miktarda bulunmasından kaynaklanan bir hastalıktır. Bu hastalık vücudun besinlerden yararlanmasını sağlayan normal süreçleri bozar ve kandaki glikoz düzeylerinin yükselmesine neden olur. Diyabetten en çok etkilenen organlar, küçük damarlardan zengin olan göz, böbrek ve beyindir. [1].

Diyabette gözün bütün bölümleri etkilenir. Görme kaybının nedeni ise, retina (ağsı tabaka) damarlarda ortaya çıkan değişikliklerdir. Bu durum "diyabetik retinopati" olarak tanımlanır. [1].

Şeker hastası olan insanlarda hastalığın seyri ve evresi nedeniyle çeşitli organlarının zarar gördüğü bilinmektedir. Bu organlardan biri de gözdür. Şeker hastalığı olanlarda gözdeki ağ tabakası (retina) ve sarı nokta (makula)'nın etkilenme sıklığı 40% 'tır. Bu nedenle Türkiye'de iki milyon kişide şeker hastalığına bağlı ağ tabakası rahatsızlığı (diyabetik retinopati) olduğu düşünülmektedir. Bu oranların yüksekliği, şeker hastalığının çok ciddi bir toplum sağlığı sorununa yol açtığına göstergesidir. [2].

Bu çalışmada, diyabetik retinopati hastalığının teşhisi yapılarak doktorlara yardımcı olacak bir yazılım geliştirilmiştir. Bu yazılım, geçmişteki gerçek hastalara uzman doktorlar tarafından uygulanmış teşhislere dayanmaktadır. Doktorlar, bu gerçek bilgiler ile oluşturulan yazılımın kullanıcı arayüzüne dokuz parametreyi girerek programı çalıştırmakta ve yine kullanıcı arayüzünde yazılımın önerdiği teşhisi görebilmektedir.

Bu tez, diyabetik retinopati hastalarının gerçek verilerini kullanarak uzman ve pratisyen doktorla için hastane bilgi sistemlerinde bir karar destek sistemi oluşturup

hastalığı teşhis etmeyi amaçlamaktadır. Bu amaçla hazırlanan tez aşağıdaki bölümlerden oluşmuştur:

- İkinci bölümde, yapılan çalışma tıp bilimiyle ilgili olduğu için sağlık bilgi sistemleri ele alınmıştır.
- Üçüncü, dördüncü ve beşinci bölümlerde ise veri ambarları, genel olarak veri madenciliği ve veri madenciliği yöntemleri anlatılmıştır.
- Altıncı bölümde ise hastalıkla ilgili genel bilgiler verilmiştir.
- Yedinci bölümde, diyabetik retinopatide uzman ve pratisyen doktorlara yardımcı olabilecek karar destek sistemi sunulmuştur.
- Son bölümde ise, elde edilen sonuçların geçerliliği tartışılmış olup, ayrıca ileride yapılabilecek benzeri çalışmalar ve uygulama alanları için öneriler tartışılmıştır.

1.1. İlgili Çalışmalar

Veri madenciliğinde tıp alanında pek çok çalışma mevcut olup, bunlardan en çok kanser teşhisi çalışmaları yapılmıştır. Örnek bir tez çalışmasında meme kanserinde tedavi yöntemlerinin belirlenmesi için bir karar destek sistemi tasarlanmıştır, bu çalışmada gerçek veri seti üzerinde sınıflandırma yöntemleri uygulanmış olup, araç olarak Weka¹ yazılımı kullanılmıştır. [3]. Veri madenciliği yöntemlerini kullanarak başka bir karar destek sistemi ise biyokimya alanı ile ilgilidir, bu çalışmada karar ağacı yöntemini kullanarak biyokimyada dört tip hastalığının teşhisi yapılmıştır. [4].

Soni vd. (2011) tarafından yapılan çalışmada kalp hastalıkları alanında yapılmıştır. Konuyla ilgili yazılmış makalede, veri seti 15 parametreden oluşmaktadır. Veri seti üzerinde pek çok algoritma uygulanmış olup, sınıflandırma algoritmalarından k-en yakın komşu, bayes sınıflandırıcı, karar ağacı ve kümeleme yöntemlerine göre sınıflandırma yapılmıştır. Bunlardan en etkili karar ağacı ve belli aralıklarda bayes sınıflandırma yöntemleri olup, istenilen sonuçları vermişlerdir. [5]. Çataloluk vd. (2012) tarafından deri hastalıkları ile ilgili bir çalışma yapılmış olup, bu çalışmada 34 parametrelili bir veri setinde K-en yakın komşu ve ağırlıklı oylama KNN sınıflandırma

¹ Weka bir veri madenciliği uygulama geliştirme programıdır.

yöntemleri kullanılmıştır. Çalışmada iki yöntem arasında karşılaştırma yapılmış olup, ayrıca sınıflandırma yöntemlerinde kullanılan mesafe ölçmesi için Öklid ve Manhattan ölçümlerin arasında karşılaştırma yapılmıştır. [6].

Diyabetik hastalıklarla ilgili ilk çalışmalar arasında diyabet hastalarının klinik verileri ve ölüm raporları incelenmiştir. Bu çalışmada hastaların hastaneye ilk başvurduğu zamanki gözlem sonuçları ile erken ölümleri arasındaki ilişkiler bulunmuştur. [7].

Konu ile ilgili yapılan araştırma sırasında görünmüştür ki konu üzerinde yapılan çalışmalar pek azdır, bu alana yönelik bir çalışmada 16 parametrelilik gerçek veri seti üzerinden %85 bir doğruluk oranıyla sonuçlanmıştır. Çalışmada Veri Madenciliği ve Durum Tabanlı Çıkarsama (Case-Based Reasoning) yöntemleri kullanılmıştır. [8].

Nabiev vd. (2007) tarafından yapılan çalışmada ise görüntü işleme yöntemlerini kullanarak hastalığın teşhisi yapılmış olup, gözdeki damarların eşiklenmesi üzerinde durulmuş ve ardıl eşikleme yöntemi uygulanmıştır. [2].

BÖLÜM 2. SAĞLIK BİLGİ SİSTEMLERİ

2.1. Bilişimin Tanımı

Bilişim, insan hayatının hemen hemen her alanında çok etkin bir şekilde yer almaktadır. Akademik, endüstriyel, teknolojik ve buna benzer tüm alanlarda bilişim bileşenini görmek mümkündür. [9]. Bilişim biliminin kökleri matematik, fizik ve mühendislik gibi dallara dayanır. Bir mühendislik alanı olarak bilişim, verileri aktarabilen, depolayabilen ve algoritmalar yardımıyla verileri işleyebilen matematiksel makineler tasarlar. Böylelikle bilişim özellikle gerçek süreçlerin benzetimini (simülasyon) mümkün kılar. Bir "yardımcı bilim" olarak düşünüldüğünde bilişim diğer bilimlerdeki olguları soyutlaştırır ve algoritmalar yardımıyla işler. [10].

Bilişim sözcüğü son on yılda çok büyük gelişim gösteren internet ve mobil iletişim sistemlerinin bir yan ürünü olarak ortaya çıkan ve yabancı dillerde informatik, information, data processing gibi sözcüklerin karşılığı olarak teknolojik gelişimin dilimize kazandırdığı bir sözcüktür. [10].

Veri (İng. data), işlenmemiş bilgi veya bilginin ham hâlidir. Bilgi ise, en basit anlamda verinin işlenmiş şeklidir. Bilgi dendiğinde iki farklı bilgiyi de bir birinden ayırmakta yarar vardır. 'Knowledge' kelimesinin karşılığı olan bilgi statik bir varlığı ifade ederken 'information' kelimesinin karşılığı olan bilişim (Çoğu kez bilgi olarak çevrilmektedir) ise dinamik bir varlığı ifade etmektedir. Statik bir bilgi (knowledge), gönderici tarafından belli bir ortamda karşı taraftaki bir alıcıya iletilirse bilişim adını alır. Bilişim (İnformation) sözcüğü içinde, bilgi (info) ve iletişim (communication) sözcükleri gizlidir. [10].

Veri işleme ve bununla ilgili iş alanları için genel bir kavram olarak İngilizce "information technology" (IT) yerine Türkçede bilişim teknolojisi (BT) kavramı kullanılmaktadır. [10].

2.2. Tıp Bilişimi

Tıpta yürütülen bilimsel çalışmaların fazlalığı ve bu literatürün pratiğe uygulanmasındaki güçlükler ve gecikmeler göz önüne alındığında, bunların aşılmasında yararlı bir yöntem olarak bilişim teknolojilerinin önemi kaçınılmaz. Modern dünyada bilginin en yoğun olarak kullanıldığı alanlardan birisi insan sağlığı ve tıp uygulamaları. Kullanılan ölçüm ve görüntüleme yöntemlerinin, test, analiz ve izleme cihazlarının hızla gelişmesi ve yaygınlaşması sonucunda tıp bilgisi gün geçtikçe zenginleşmekte ve hastalar için toplanan tıbbi veri ve bilgi miktarı da büyük bir hızla artmakta. Tıp bilişimi, bu bilgi ve verilerin oluşturulması, biçimlendirilmesi, paylaşılması ve sonuçta hastaların bakımlarının ve tedavilerinin belirlenmesi, seçilmesi ve geliştirilmesini hedef alır. [11].

Tıp bilişimi, araç olarak bilişim teknolojilerini kullanarak Tıp pratiğinde ortaya çıkan bilgi yönetimi gereksinimine cevap vermeye çalışır. Temel olarak ilgilendiği konular arasında aşağıdaki başlıklar sıralanabilir:

- Veri, bilgi (elde edilmesi, saklanması vb.)
- Kodlama sistemleri
- Veri işleme
- Veri tabanı yönetim sistemleri
- Telekomünikasyon sistemleri (teletıp uygulamaları)
- Tıbbi sınıflandırma sistemleri (snomed, ICD-10 vb.)
- Hasta kayıt sistemleri, elektronik hasta kayıtları
- Biyosinyal analizleri (EKG yorumlayan yazılımlar vb.)
- Tıbbi görüntüleme sistemleri (USG, MRG vb.)
- Görüntü işleme ve analiz yöntemleri
- Klinik bilgi sistemleri

- Toplum hekimliđi bilgi sistemleri
- Hemřirelik bilgi sistemleri
- Karar destek sistemleri
- Sađlık bilgi sistemleri (halk sađlıđı, birinci basamak bilgi sistemleri)
- Hastane bilgi sistemleri (idari ve finansal uygulamalar)
- Bilgi sistemleri gvenliđi. [11].

2.3. Tıp Biliřimin Uygulama Alanları

Gnmzdeki tm teknolojik ilerlemelere rađmen, sađlık bilimleri ve teknolojiye en ileri gitmiř lkelerde bile bilgilerin toplanamaması, verilerin yeteri kadar gvenli olmaması, gzlemlerdeki sbjektiflik ve benzeri nedenlerle byk miktarlarda iř gc, para ve zaman kaybı sregelmektedir. Bu kayıplar dolayısıyla bilimin ilerlemesi ister istemez yavařlamakta ve byk bir kaynak israfı ortaya çıkmaktadır. [11].

Hemen hemen tamamı gzlemlerle, elde edilen bulguların sınıflandırılıp deđerlendirilmesine ve istatistiki alıřmalara dayalı olan tıp biliminde, verilerin kresel bazda homojen olarak toplanabilmesi, sratle iřlenip tasnif edilebilmesi, standardizasyonu ve daha sonra kolayca ulařılabilmesi belirgin nem tařımaktadır. Tıbbi biliřim uygulamaları bu aıdan kresel veri tabanları ve bilgilerin tasnif edilip sratle eriřilebilmesinde sađladıđı kolaylıklarla modern tıp biliminin vazgeilemez bir alt bilim dalıdır. [11].

Dnya cođrafyasına yayılmıř en kk yerleřim birimlerine geleneksel yntemlerle eksiksiz bir sađlık hizmeti gtrmek ok pahalı, ok zor ve hatta imknsızdır. Ancak uzaktan tıp (telemedicine) uygulamalarıyla binlerce kilometre uzaktan hastaların kalp ritmini belirlemek, kan biyokimyasını đrenmek kısaca tanı koymak ve tedaviyi ynlendirmek mmkndr. Hatta bu konuda son yıllarda atılan dev adımlar sonucu ameliyat robotları sayesinde bir bařka lkedeki bir cerrahın kilometrelerce uzaklıktaki hastasına kalp ameliyatı yapması bile mmkn olmuřtur. [11].

2.3.1. Hastane bilgi sistemleri

Sağlık hizmetleri diğer alanlardan daha fazla bilgiye duyarlı bir alandır ve neticesinde etkin bir yönetim için sistematik olarak bilginin elde edilmesini gerektirir. Kaliteli bir sağlık hizmet sunumu, geniş kapsamlı ve iyi planlanmış bir bilgi sürecine bağlıdır.

Hastane bilgi sistemi hastanenin idari ve tıbbi bilgilerinin yönetimini kolaylaştırmak ve sağlık hizmetlerinin kalitesini yükseltmek için düzenlenmiş bir bilgi sistemi olarak tanımlanabilir. Bilgisayarların kullanıldığı sistemlere bilgisayara dayalı hastane bilgi sistemi denir. İdari ve tıbbi bilgileri iç içe, bir arada tutabilen sistemlere bütünlük hastane bilgi sistemleri denilmektedir. Bütünlük bir hastane bilgi sisteminin fonksiyonları şunlardır:

- Hastanın tedavisi ve bu tedaviyle ilgili yönetsel görevleri desteklemek amacıyla hasta hakkında doğru, kalıcı ve güncel bilginin doğru kişiye, doğru yerde ve kullanılabilir formatta sağlanması. Bu da bilginin doğru olarak toplanması, saklanması, işlenmesi ve belgelenmesi ile gerçekleşir.
- Hastalık hakkında gelişmiş bilgi desteği sağlama. Örneğin ilaçların, teşhis ve tedavilerin ters etkileri hakkında bilgi sağlamak
- Hastanın tedavisinin kalitesi, hastanenin performansı ve maliyetleri hakkında bilgi sağlamak
- Gelişmiş hastane enformasyon sistemlerinde farklı hastaneler ve hastanelerin ilişkide bulunduğu kurumlar arasında bilgi aktarımına olanak veren yüksek kaliteli bir iletişim sağlamak.
- Eğitim hastanelerinde hasta tedavisine ilave olarak araştırma ve eğitim amacıyla hastanın tedavisinde özel deneyimlerle ortaya çıkan verileri dikkatli bir şekilde toplayarak yeni bilgiler elde edilmesini sağlamaktır. [11].

2.3.1.1. Karar destek sistemleri

Sağlık kuruluşları günümüzde bilişim sistemlerinden, yönetim hizmetleri, hastalıkların teşhis edilmesi, hekimlerin hastayla ilgili vereceği kararların

desteklenesi, hemşire ve hekimlerin yapacağı işlerde rehberlik, sinyal yorumlama, laboratuvar hizmetleri ve hasta yönetimi gibi çok çeşitli alanlarda faydalanılmaktadır. Bu amaçla kullanılan sistemlerin başında ise Klinik Karar Destek Sistemleri gelmektedir. [12].

Klinik karar destek sistemleri, hekimlere veya diğer sağlık personeline alacağı klinik kararlarda destek sağlayan bilgisayar programlarıdır. Bu sistemler bir bakıma karar desteği sağlamak için klinik veri yada medikal bilgiyle ilgilenen bilgisayar sistemleridir. Bugün hekimler tıbbi bilginin miktarında meydana gelen artış nedeniyle bu bilginin yönetimiyle başa çıkmak ve uzman yokluğunda uygun seçenekler arasında karar verebilmek için karar desteğine ihtiyaç duymakta ve bu amaçla klinik konularda akıl yürütme özelliğine sahip karmaşık bilgisayar programları olan klinik karar destek sistemler kullanmaktadır. [12].

Bu programlardan bazıları, klinisyenler tarafından girilen temel klinik bilgileri dikkate alarak teşhise yönelik gayretleri arttırmakta, hastalara özel değişkenlere bağlı olarak özel ilaç tavsiyesinde bulunabilmekte ve hastalara ait özel bilgileri uzman bilgi tabanı ile mukayese ederek hasta yönetimi ve konsültasyon işlevini gerçekleştirmektedir. [12].

Klinik Karar Destek sistemlerini üç başlık altında sınıflandırabiliriz: bilgi yönetimi için araçlar, uyarı ve dikkatin bir alana odaklanması için sistemler ve Uzman sistemleri. [12].

- **Bilgi yönetimi için araçlar**, Sağlık kuruluşlarında bilgi yönetimi, gelişmiş iş istasyonları yardımıyla gerçekleştirilir. Bu istasyonlar bilginin depolanması ve geri çağırılması için gerekli olan altyapıyı sağlar. Bilgi yönetim araçları, hekimlerin ve sağlık personelinin ihtiyaç duyacağı bilgi ve verileri sağlamakla birlikte, özel bir kararın alınmasına yardım etmezler. Böyle bir durumda problemin çözümü için gereksinim duyulan bilgiler temin edildikten sonra, karar verme işlemi hekime bırakılır. [12].

- **Uyarı ve dikkatin bir alana odaklanması için sistemler;** Bu tür programlar, kullanıcıların problemleri ve teşhis yöntemlerini hatırlamaları için dizayn edilir. Bu sistemler tipik olarak potansiyel anormalliklerin gösterilmesini yada belirli bir standardı içene alan yanıtların listesini ihtiva eden basit bir mantık kullanır. Örneğin eczacılık karar destek sistemleri; muhtemel ilaç etkileşimleri ilgili uyarılar vererek kullanıcıların dikkatinin bu alana çekilmesini sağlarken, laboratuvar karar destek sistemleri, anormal değerleri belirler veya bu anormal değerlerle ilgili muhtemel açıklamaların listesini verir. [12].

Klinik anımsatma ve uyarı sistemleri hekim ve hemşireleri, yanıt vermedeki gecikmelerin kritik olabileceği olası ciddi durumlara karşı uyarılmak için geliştirilmiş olan sistemlerdir. Hekim ve diğer sağlık personeline tedavi sürecinin muhtemel sonuçları hakkında uyarılarda bulunarak arzulanmayan tehlikeli sonuçların ortaya çıkması önlenmiş olur. Örneğin hekim hastaya kullanmaması gereken bir ilacı tavsiye ederse sistem devreye girerek hekimi uyarmaktadır. [12].

- **Uzman sistemler;** Bu programlar hastaların özel verilerine dayalı tavsiyeler ve değerlendirmeler sunar. Temel olarak karar teorisine ya da maliyet-fayda yaklaşımına dayanan bir mantıkla hareket eder. Bazen de problemlerin çözümü için sayısal yaklaşımlardan faydalanabilir. Bazı teşhis yardımcıları (örneğin DxPlain gibi) dar alandaki etiyolojik ihtimaller durumunda, farklı teşhisleri veya ek bilgileri önerebilir. İnternist gibi programlar hastanın semptomlarını değerlendirerek en olası teşhisi koyabilir. Bazı sistemler ise klinik bağlamda hasta kayıtlarını özetler ve yorumlar. [12].

Uzman sistemlerin üç ana ögesi bulunmaktadır. Birinci öge; tıp uzmanları tarafından geliştirilen bilimsel tabandır (Knowledge base). Bilimsel taban, belirli hastalık grubuyla ilgili kararların nasıl alınacağını ortaya koymaktadır. İkinci öge; hastadan alınan bilgilerdir. Üçüncü ögesi olan kurallara dayalı çıkarımlar/sonuçlar üreten motorlar (rule-based inference engines) ise; hastadan elde edilen bilgileri, bilimsel tabanda bulunan bilgileri referans kabul ederek işler ve hekimin kullanımına sunar. [12]

2.3.1.2. Hastane bilgi sistemlerin geliřimi

Hastane bilgi sistemleri öncelikle faturalama, maliyet ve ödeme kontrol etme gibi idari amaçlarla kullanılmaya başlanmıřtır. Doktorlar istemlerini geleneksel olarak elle yazmıřlar ve sekreterler tarafından sisteme girilmiřtir.

1970'lerde finansal bilgi sistemleri temelli birok hastane bilgi sistemine klinik laboratuvar sistemleri, radyoloji, eczane gibi hastane bölümleri için yazılım paketleri geliřtirdiler. 1970'leri tanımlayan üç özellik olarak; donanım ve yazılımın güvenilirliđinin geliřmesi, teknolojik maliyetlerin azalması, arařtırma ve geliřtirme faaliyetlerinde önemli yatırımlar sayılabilir. [10].

1980'lerde sađlık kurumlarındaki uygulamaları destekleyebilecek iletiřim teknolojileri geliřti. Hastanelere kabul edilen hastaların verileri otomatik olarak paylařılabilmekteydi. Hasta istemleri diyet, ev idaresi, laboratuvar, radyoloji ve benzeri birimlere elektronik olarak iletilebiliyor, hekim istemleri hemřire istasyonundan dođrudan laboratuvar, eczane, radyoloji ve benzeri birimlere yapılabilmekteydi. Bilgisayar ađ yapılarında önemli kazanımlar elde edilmiřti. [10].

1980'lerden 1990'lara dođru hastane enformasyon sistemlerinde idari uygulamalardan hekimler, hemřireler ve diđer sađlık alıřanları tarafından kullanılan klinik enformasyon sistemlerine dođru bir geliřim söz konusudur. [10].

1990'larda pek ok iřlevsel düzeyde iřlemler için hastane bilgi sistemleri, birok fonksiyonel alanı kaplayan geniř bir yapıya kavuřmuřtur. Bu bölümler ađ iletiřimi yoluyla birbirine bađlanarak bütünlük bir bilgi yapısı oluřturulmuřtur. Kapsamlı bir hastane bilgi sistemi altı öđeyi içermektedir:

- Hastaların randevusu, kabulü, taburcu iřlemleri ve nakli ile ilgili iřlemler
- Hasta hesapları, fatura ve ücret bordrosu gibi iřletme ve finansal sistemler
- İletiřim ve ađ uygulamaları ile hemřirelik hizmetleri, yardımcı sađlık hizmetleri alanları gibi bölümler arasında mesajların iletilmesi ve bu bölümler

arasında hasta kabulü, istemlerin izlenmesi ve bunlara yanıt verilmesinin sağlanması

- Eczane, radyoloji ve laboratuvar gibi bölüme özgü sistemler, bölümlerin işletme fonksiyonlarının idare edilmesi ve verilerin kurumsal veri tabanlarına bağlanması
 - Hasta verilerini toplamak, saklamak ve yeniden elde etmek için dokümantasyon sistemlerinin oluşturulması ve bu kategorideki uygulamaların hemşirelik dokümantasyonu için bakım noktasındaki hasta başı terminallerinden klinik raporları tutan ve depolayan kopyalama modüllerine kadar kapsanması
 - Hekimlere önemli test sonuçları için uyarı mesajları, kullanım ölçütleri ve ilaç / ilaç etkileşimi konusunda veriler içeren; tedavinin klinik yollar (clinical pathway) ile uyum sağlayıp sağlanmadığını kontrol etme olanağı veren, hastanın tıbbi bakımını planlamada yardım eden alarm ve uyarı sistemleri.
- [10].

BÖLÜM 3. VERİ AMBARLARI

3.1. Veri Ambarının Tanımı

Veri ambarı, bir işletmenin ya da kuruluşun değişik birimleri tarafından toplanan bilgilerden değerli olanlarının, gelecekte analiz işlemlerinde kullanılması amacıyla işletimsel sistem veritabanından farklı bir ortamda birleştirilmesinden oluşan büyük çaplı bir veri deposudur. [13].

Veri ambarı, bir ürün değil ortamdır. Aynı zamanda, çoklu veri tabanı veya diğer bilgi sistemlerinden ilgili veriyi elde etmek için gerekli olan algoritmaları, araçları içeren mimari topluluğudur. İlgili veriler, veri sorgulama ve raporlama amaçlı elde edilir.

Veri ambarı, birbirleriyle bütünleşik olmayan uygulamaların bütünleştirilmesi açısından bir olanak sağlar. Veri ambarı, bir zaman boyutu içinde analitik işlemlerin yapılması için ihtiyaç duyulan bilgi temelini sağlar. Veri ambarı, karar verme sürecinde destek vermek üzere hazırlanmış,

- konuya yönelik
- bütünleşik
- zaman boyutu olan
- sadece okunabilen

veri topluluğudur. [14].

3.2. Konuya Yöneliktir

Veri ambarı birinci özelliği, işletmenin belli başlı amaçlarına ya da konularına yönelik olmasıdır. Konuya yönelik olmasının anlamı, veri ambarının işletmedeki

yüksek seviyeli varlıklar üzerine odaklanmış olmasıdır. Bu varlıklar bir okul ortamı için öğrenciler, dersler, notlar vb. olabilir. [14].

3.3. Bütünleşiktir

Veri ambarı ortamdaki verinin en belirgin görünümü, bütünleşik durumda olmasıdır. Veri ambarı içindeki veri mutlaka bütünleşik olmalıdır. Bunun istisnası olamaz.

İşlemsel sistemlerde farklı uygulamalarda farklı tasarımcılar tarafından verinin tanımlanması açısından değişik yollar tercih edilmiş olabilir. Kodlamalardaki farklılık, anahtar yapılar da farklılıklar, fiziksel karakteristiklerdeki farklılıklar, isimlendirmelerdeki farklılıklar vb. birçok farklılık olabilir. [14].

Örneğin, cinsiyet ile ilgili bir alan uygulamada sadece “E” ve “K” kodlarıyla belirtilmiş olabilir. “E” kodu erkek, “K” kodu ise kadınları simgelemektedir. Bir başka uygulamada ise söz konusu cinsiyet ile ilgili alan 1 veya 0 değerleri ile ifade edilmiş olabilir. Eğer veri ambarında sadece “E” ve “K” kodlarının kullanılması söz konusu ise, işlemsel veritabanından veri ambarına veriyi taşıırken, farklı tanımların tümünü buna göre dönüştürmek gerekecektir. [14].

3.4. Zaman Boyutu Vardır

Veri ambarındaki tüm veri zamanın belirli bir anına aittir. Veri ambarındaki verinin bu temel karakteristiği, işlemsel sistemlerdeki veriden oldukça farklıdır. İşlemsel sistemlerde veri o anda var olan değerdir. İşlemsel sistemlerde bir veriye ulaşıldığında, çoğunlukla onun şu andaki değeriyle ilgilenir. İşlemsel veri de zaman dağılmış olabilir. Ancak bu süre örneğin 80 gün olabilir. [14].

Veri ambarındaki veri ise, sadece o andaki değerlerle değil, geçmişteki değerlerle de ilgilidir. Veri zaman içinde bir noktayla birleştirilerek değerlendirilir. Örneğin sömestre, mali yıl, ödeme dönemi gibi unsurlar göz önüne alınır. Veri ambarı verinin en az 5 yıllık değerlerini içermelidir. [14].

3.5. Sadece Okunabilir

Veri ambarının diğeri bir özelliđi, veri ambarında yer alan verinin sadece okunabilir bir yapıda olmasıdır. Veri ambarındaki veri yönetiminin gereksinimlerine yanıt vermek üzere tasarlandıđı işlemlere tabi tutulmaz, yani silinmez veya güncelleştirilmez. [14].

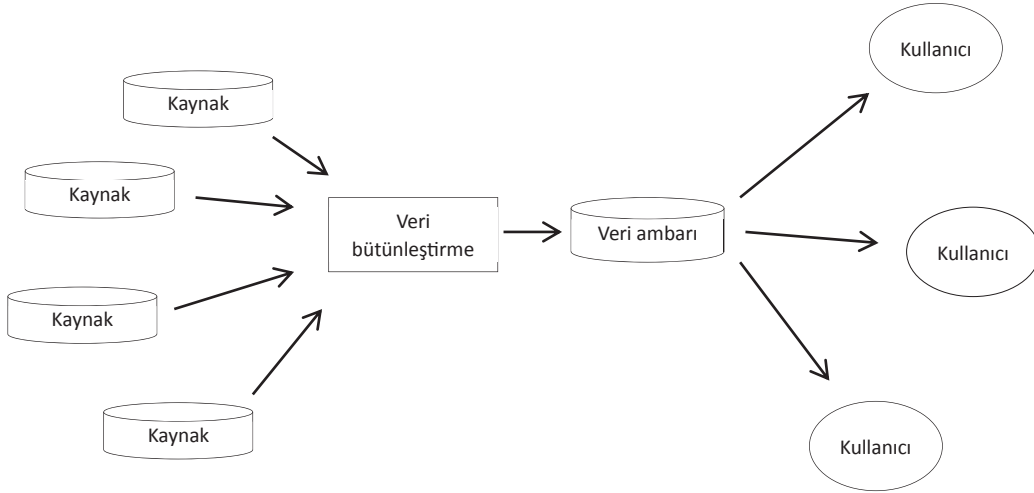
3.6. Veri Ambarının Temel Özellikleri

- Çok boyutlu bir modellemede depolama için işlem gören, çoklu kaynaklardan gelen entegre verinin deposudur.
- Zaman serileri ve eğilim analizi gibi tarihsel bilgi gerektiren yapıları destekler.
- İşlemsel veri tabanları gibi uçucu değildir. (Bilgi çok daha az sıklıkta deđişir ve periyodik güncelleme yapılıır.)
- Karar mekanizmasındaki kişinin daha iyi ve daha hızlı kararlar vermesini sađlayan karar destek teknolojileri topluluđu olarak da tanımlanabilir.

3.7. Veri Ambarı Mimarisi

Veri ambarı belli bir mimariye sahip olduđunu söyleyebiliriz. Veri ambarı mimarisi genel karakteristikleri řu şekilde sıralanabilir:

- Kaynaklardan alınan veri dönüştürülür
- Veri ambar oluşturulur
- Kullanıcının veri ambarına erişimi sađlanır. [14].



Şekil 3.1. Veri ambarı mimarisi [14]

3.7.1.1. Verinin dönüştürülmesi

Üretildiği kaynaklardan veri temin edilmesi gerekmektedir. Veri bu kaynaklarda farklı biçimde yer alabilir. Örneğin farklı veritabanlarında saklanıyor olabilir. Ayrıca aynı veri farklı biçimlerde temsil ediliyor olabilir. Ayrıca veri üzerinde bozukluklar söz konusu olabilir. Bu gibi veriyi veritabanına doğrudan kaydetmek sorunlara neden olabilir. Söz konusu verinin öncelikle düzenlenmesi gerekir. [14].

3.7.1.2. Veri ambarının oluşturulması

Veri dönüştürüldükten sonra veri ambarına aktarılır. Veri ambarının istenen amaçlara uygun biçimde çalışabilmesi için özel bir tasarıma gereksinim duyulur.

Veri ambarında veri modeli olarak “çok boyutlu model” tercih edilir. Bu veri modeli, adından da anlaşılacağı gibi veriye bazı boyutlar kazandırarak sorgulamaları bu boyutlar yardımıyla oluşturmak amacıyla düzenlenir. Bir işletmenin satış verilerini içeren bir veri ambarında yıl, ürün grubu ve mağazalar birer boyut olarak düşünülebilir. [14].

3.7.1.3. Kullanıcının veri ambarına erişimi

Veri ambarı oluşturduktan sonra kullanıcılar farklı biçimde erişerek kullanabilirler. Kullanıcı doğrudan veri ambarına erişerek sorgulama yapabilir.

BÖLÜM 4. VERİ MADENCİLİĞİ

4.1. Veri Madenciliği Tanımı

Günümüzde kurumlarda büyük veriler üretilmektedir, bu büyük verilerin genelde gizli ya da çok net olmayan ve önceden bilinmeyen kullanışlı bilgilere sahiptirler. Birikilen bu verileri kurum için yararlı olanları çıkarma işlemine veri madenciliği adı verilmiştir. Veri madenciliğinin genel amacı bir veri kümesinden bilgileri ayıklamak ve kullanım için anlaşılır bir yapı haline dönüştürmektir.

Veri Madenciliği çok büyük veri yığınlarından kritik bilgileri elde etmeyi sağlar. Böylelikle normal şartlar altında uzun zaman süren araştırmalarla doğruluğu kesin olmayacak şekilde elde edilen bilgi veri madenciliği ile kısa sürede ve kesin olarak elde edilir. Elde edilen bu bilgi objektif değerlendirmeler yapılmasında ya da stratejik kararlar almada kullanılır. Bu bilgiler kurumsal veri kaynaklarının iyi analiz edilmesine ve iş dünyasındaki yaklaşımlara ilişkin tahminlerde bulunulmasına yardımcı olur. Kısaca veri madenciliği sayesinde şirketler stratejik adımlar atarken çok büyük veri yığınları arasından kendilerine yol gösterecek kritik verileri ayıklayarak analiz edebilir. [15].

Günümüzün ekonomik koşulları ve yaşanan hızlı değişim ortamlarında, iş deneyimi ve önseziyle dayanarak alınan kararlarda yanlış karar alma riski çok yüksektir. Riski azaltmanın tek yolu bilgiye dayalı yönetimi öngören karar destek çözümleridir. Veri madenciliği teknikleri gerçek anlamda bir karar destek sistemi oluşturmada olmazsa olmaz araçlardır. Bu noktada bilgi teknolojilerinden yararlanmak kaçınılmaz olmuştur. [16].

Veri madenciliği büyük veri kümelerinde gizli kalan verileri, istatistik, makine bilgisi, veri tabanları ve yüksek performanslı işlem gibi temelleri kullanarak yeni ve yararlı bilgiler keşfeder.

4.2. Uygulama Alanları

Veri madenciliğinin günümüzde yaygın bir kullanım alanı bulunmaktadır. Örneğin pazarlama bankacılık ve sigortacılık gibi alanlarda ve elektronik ticaret ile ilgili alanlarda yaygın şekilde kullanılmaktadır.

4.2.1.1. Sağlık Sektörü

Sağlık alanında yapılan birçok veri madenciliği araştırmalarında hastaların elektronik tıbbi kayıtları ve idari işleri belgeleyen veriler kullanılmaktadır. Bu verilerden yararlanılarak farklı tahminler yapılabilir. Örneğin bunlardan bazıları şunlardır:

- Belirli bir hastalığa sahip kişilerin ortak özelliklerinin tahmin edilmesi
- Tıbbi tedaviden sonra hastaların durumlarının tahmin edilmesi
- Hastane maliyetlerinin tahmin edilmesi
- Ölüm oranları ve salgın hastalıkların tahmin edilmesi. [17].

Hastalıkların yönetimi ile ilgili veri madenciliği çalışmaları hastalıkların ve durumlarının tanımlanmasını ve maliyetlerin modellenmesi gibi araştırmaları içerir. Bu çalışmalarda amaç pozitif sonuç elde etmektir. Örneğin Harleen Kaur ve arkadaşları hastaların yaş ve cinsiyet gibi verilerini karar ağacı yöntemleri ile analiz ederek göğüs kanseri olup olmadığını tahmin etmeye çalışmışlardır. [17].

Hastane bilgi sistemlerindeki verilerle yapılmış diğer bir çalışmada, hastaların sık sık farklı doktorları ziyaret etmeleri araştırılmış ve hasta demografik bilgileri ve işlemsel veriler analiz edilmiştir. Göşkisel kural analizi (Association rules analysis) kullanılarak yapılan veri madenciliği çalışmasında, yaşın, cinsiyetin, hastanelerin özelliklerinin, kronik ve akıl hastalıklarının sürekli doktor ziyaret etme davranışında etkili oldukları ortaya çıkartılmıştır. [17].

Hastanelerde maliyetleri etkileyen en önemli konulardan birisi de hastaların kalış süreleridir. Kalış sürelerinin etkileyen faktörler de günümüzde veri madenciliği çalışmalarının araştırma konusudur ve birçok çalışma yapılmıştır. Örneğin, yapılan bir çalışmada hastaların demografik ve çevresel bilgileri, sinir ağları ile analiz edilmiş ve bazı önemli bilgiler elde edilmiştir. Bu bilgilere göre 40 yaşından büyük hastalar, şehirlerde yaşayan hastalar, alkol ve sigara bağımlılığı olan hastalar daha uzun süre hastanede kalmaktadırlar. Ayrıca özel hastanelerdeki kalış süreleri devlet hastanelerinden daha kısadır. [17].

İlaçlar da tıbbın önemli araştırma konularından birisidir. Amerika Birleşik Devletleri'nde yeni bir ilaç geliştirildiğinde, klinik denemelerden sonra FDA (Food and Drug Administrators) kurumu tarafından onaylanarak piyasaya sürülür. Onaylanmadan önce ilacın faydalarının risklerinden daha çok olması göz önünde bulundurulur. Bazı ilaçlar piyasaya sürüldükten sonra risklerinin çok fazla görülmesi nedeniyle kaldırılmışlardır. İlaçların önceden tanımlanmamış yan etkilerinin bulunabileceği olasılığı, web üzerinden tıbbi yayınlar analiz edilerek veri madenciliği çalışmaları da yapılmaktadır. [17].

4.2.1.2. Devlet uygulamaları

Kamu yöneticileri günümüzde verinin ve bilginin önemini kavramışlardır. Müşteriye özel hizmet sunan ticari kuruluşlarda olduğu gibi devlet kurumları da vatandaşlarının ihtiyaçlarına özel hizmet sunabilmenin önemini kavramışlardır. Kamu yöneticileri için en önemli uygulamalar kaynakların doğru olarak kullanılması ve planlanmasıdır. Kamu güvenliğini sağlama amacı ile güvenlik problemlerini önceden tahmin etmek, rastlantısal olaylardaki sorunların çözümüne dair izleri keşfetme ve olası güvenlik sorunlarını eş zamanlı olarak tespit edebilme ve çözüm üretebilme, vergi ile ilgili yolsuzlukları ve izlerini belirleme, yolsuzlukları eş zamanlı olarak belirleme, sağlık ödemeleri gibi uygulamalar uygulanmaktadır. Kamuda enformasyon ve bilgi ihtiyacı sonsuzdur. Emniyet birimleri için suç istatistiklerine dair online raporlama, hangi profildeki insanların ne tür suçlara meyilli olduklarını belirleme, eş zamanlı suç engelleme politikaları oluşturmak ancak ileri analitik uygulamalar ile mümkündür. Günümüzde e-devlet kavramı oldukça kritiktir. E-devlet uzmanlarının en önemli

hedefi bilgiye eş zamanlı olarak ulaşmak ve daha iyi hizmet vermektir. E-devlet uygulaması gerçekleştirilen ülkelerde kamu kuruluşları ziyaretçilerin sayfalarını nasıl kullandığı, ihtiyaç duyulan formlara kolayca ulaşıp ulaşılamadığı, web sayfa tasarımının nasıl en iyi kullanılabilir hale getirilebileceği, hangi sayfaların hangi sıra ile ziyaret edildiğinin anlaşılması, geçmişteki ziyaretçi davranışlarına göre kurumun web sayfasını vatandaşın ihtiyacına daha iyi yanıt verecek şekilde yeniden düzenlemek mümkündür. [18]. Yukarıda belirttirilen uygulamalarının temelinde veri madenciliği çalışmaları bulunmaktadır.

4.2.1.3. Haberleşme sektörü

Telekom sektöründe en önemli sorun müşteri kaybıdır. Kuruluşlar hangi müşterilerini kaybedebileceklerini önceden belirleyebildikleri takdirde bu müşterilerini elde tutma amaçlı stratejiler geliştirebilir, düşük maliyetli ve etkili kampanyalar düzenleyebilirler. Kaybetme olasılığı olmayan bir müşteriye kalıcılığını sağlama amaçlı bir mesaj göndermek hem müşterinin kendisine verilmek istenen mesajın ne olduğunu algılamasını zorlaştıracak hem de maliyetleri artıracaktır. Örneğin Amerika'nın en büyük kablosuz iletişim sağlayıcısı olan Verizon kaybetme olasılığı yüksek olan müşterilerini ve müşteri kaybına neden olan faktörleri belirleme amaçlı bir Veri Madenciliği çalışması yapmıştır. [18].

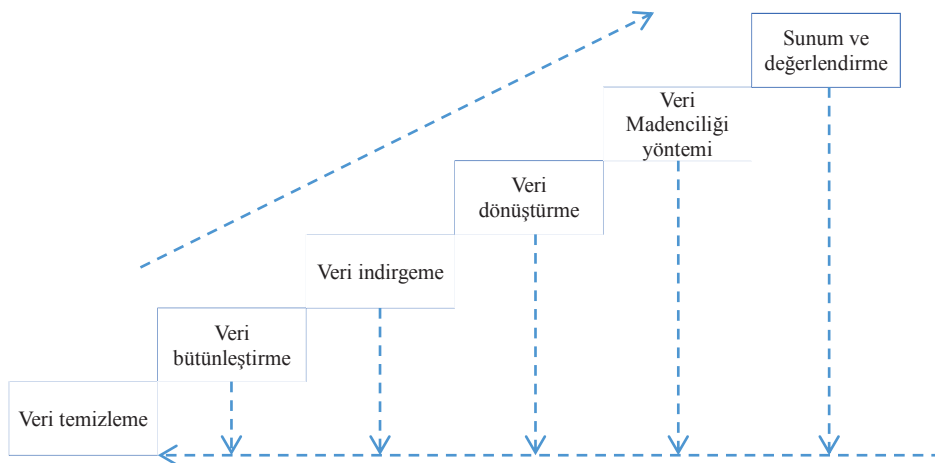
4.2.1.4. Finans sektörü

Finans sektöründe en temel uygulamalar çapraz satış, risk derecelendirme, mevcut müşteriyi elde tutma, yeni müşteriler kazanma, maliyetleri azaltma, kayıp ve kaçakları engelleme, alternatif kanallar oluşturma, müşteri memnuniyetini sağlama olarak özetlenebilir. Hangi müşteri profilinin neyi, ne zaman ve neden tercih ettiğini anlayabilen bir kuruluş hem talep yaratma, hem de doğru zamanda doğru talebi karşılama ve sunma avantajına sahip olacaktır. Bankalar, mevcut müşterilerden rakip bankaya geçme ihtimali olan müşterileri, profillerini ve kaybettikleri müşterilerin hangi sebepler yüzünden sistemden ayrıldıklarını tespit etmek istemektedir. [18].

4.3. Veri Madenciliği Süreci

Veri madenciliğini bir süreç olarak değerlendirilir, söz konusu süreç aşağıda belirtilen adımları içermektedir:

- Veri temizleme
- Veri bütünleştirme
- Veri indirgeme
- Veri dönüştürme
- Veri madenciliği algoritmasını uygulama
- Sonuçları sunum ve değerlendirme.



Şekil 4.1. Veri madenciliği süreci [14]

4.3.1.1. Veri temizleme

Veri temizleme işleminin amacı, veriler içindeki uygun olmayan veya hatalı girilmiş verileri ayıklamaktır. Bu işlemde eksik veriler uygun değerler ile doldurulur. Eğer eksik veri çok ise bu kaydın silinmesi gerekir. Eksik verilerin yerine yenileri belirlenerek konulmalıdır. Bunun için aşağıda belirtilen yöntemlerden biri kullanılabilir.

- a) Eksik değer içeren kayıtlar veri kümesinden atılabilir.

- b) Kayıp değerlerin yerine bir genel sabit kullanılabilir. Bütün kayıp değerler için aynı sabit kullanılabilir.
- c) Değişkenlerin tüm verileri kullanarak ortalaması hesaplanır ve eksik değer yerine bu değer kullanılabilir.
- d) Değişkenin tüm verileri yerine, sadece bir sınıfa ait örneklerin değişken ortalaması hesaplanarak eksik değer yerine kullanılabilir.
- e) Verilere uygun bir tahmin yapılarak, örneğin regresyon ya da karar ağacı modeli kurularak eksik tahmin edilebilir ve eksik değer yerine kullanılabilir. [14].

4.3.1.2. Veri bütünleştirme

Farklı veri tabanlarından ya da veri kaynaklarından elde edilen verilerin birlikte değerlendirmeye alınabilmesi için farklı türdeki verilerin tek türe dönüştürülmesi yani bütünleştirilmesi söz konusu olacaktır. Bunun en yaygın örneği cinsiyette görülmektedir. Çok fazla tipte tutulabilen bir veri olup, bir veri tabanında 0/1 olarak tutulurken diğer veri tabanında E/K veya Erkek/Kadın şeklinde tutulabilir. Bilginin keşfinde başarı verinin uyumuna da bağlı olmaktadır.

4.3.1.3. Veri indirgeme

Veri indirgeme teknikleri, daha küçük hacimli olarak ve veri kümesinin indirgenmiş bir örneğinin elde edilmesi amacıyla uygulanır. Bu sayede elde edilen indirgenmiş veri kümesine veri madenciliği teknikleri uygulanarak daha etkin sonuçlar elde edilebilir.

Veri indirgeme yöntemleri aşağıdaki biçimde özetlenebilir:

- a. Veri Birleştirme veya Veri Küpü
- b. Boyut indirgeme
- c. Veri Sıkıştırma
- d. Örnekleme
- e. Genelleme.

Veri küpleri aynı özelliklere sahip olan verileri aynı küpte gösterilir. Örneğin bir firmanın satış tutarları belli bir zaman içinde aynı küpte gösterilir, satışı yapılan ürünler ve firmanın farklı satış yerleri için aynı küp üzerinde gösterilebilir. Veri küpleri özet bilgiye herhangi bir hesaplama yapmadan hızlı bir biçimde erişilmesini sağlarlar. Veri sıkıştırma aşamasında büyük veri kümelerinin sıkıştırılarak daha az yer işgal etmeleri sağlanır. Örnekleme aşamasında ise, büyük veri topluluğu yerine onu temsil eden daha küçük veri kümelerinin oluşturulması amaçlanır. Genelleme verilerin tek tek değil genel kavramlarla ifade edilmesini sağlar.

4.3.1.4. Veri dönüştürme

Veri dönüşümünün amacı ise, kaynak veriyi farklı formatlara veya değerlere dönüştürmektir. Örneğin, Veritabanındaki mantıksal (boolean) bir alanı integer bir tipe dönüştürülebilir. Bunun sebebi ise kullanılan bazı veri madenciliği algoritmalarının integer veri tipiyle Boolean veri tipine göre daha başarılı sonuçlar üretmesidir. Bazı veri setlerinde değişkenlerin sahip olduğu çok büyük ve çok küçük değerler sağlıklı bir çözüm almasını engeller, bunun için bu değişkenleri normalleştirme veya standartlaşmaya uygun bir yol olacaktır. [19]

4.3.1.1. Min-Max normalleştirilmesi

Verileri 0 ile 1 arasındaki sayısal değerlere dönüştürmek için min-max normalleştirme yöntemi uygulanır. Bu yöntem ile veri setinde bir alandaki değişkenlerden en büyük ve en küçük değerini alarak buna uygun bir şekilde her değişken dönüştürülür. Min-max normalleştirilmesi şu şekilde ifade edilmektedir:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4.1)$$

X' dönüştürülmüş değer, X gözlem değeri, X_{min} en küçük gözlem değeri, X_{max} en büyük gözlem değeri.

4.3.4.2. Z-score standartlaşma

İstatistik çözümlerde sıkça kullanılan bir değer dönüşüm yöntemidir. Bu yöntemle yeni değerlere dönüştürülmesi, verilerin ortalaması ve standart hatası alınarak hesaplanır. Z-score standartlaşma şu şekilde ifade edilmektedir:

$$X' = \frac{X - \bar{X}}{\sigma_x} \quad (4.2)$$

X' dönüştürülmüş değer, X gözlem değeri, \bar{X} verilerin aritmetik ortalaması ve σ_x değerlerin standart sapması.

4.3.4.3. Veri madenciliği algoritmasını uygulama

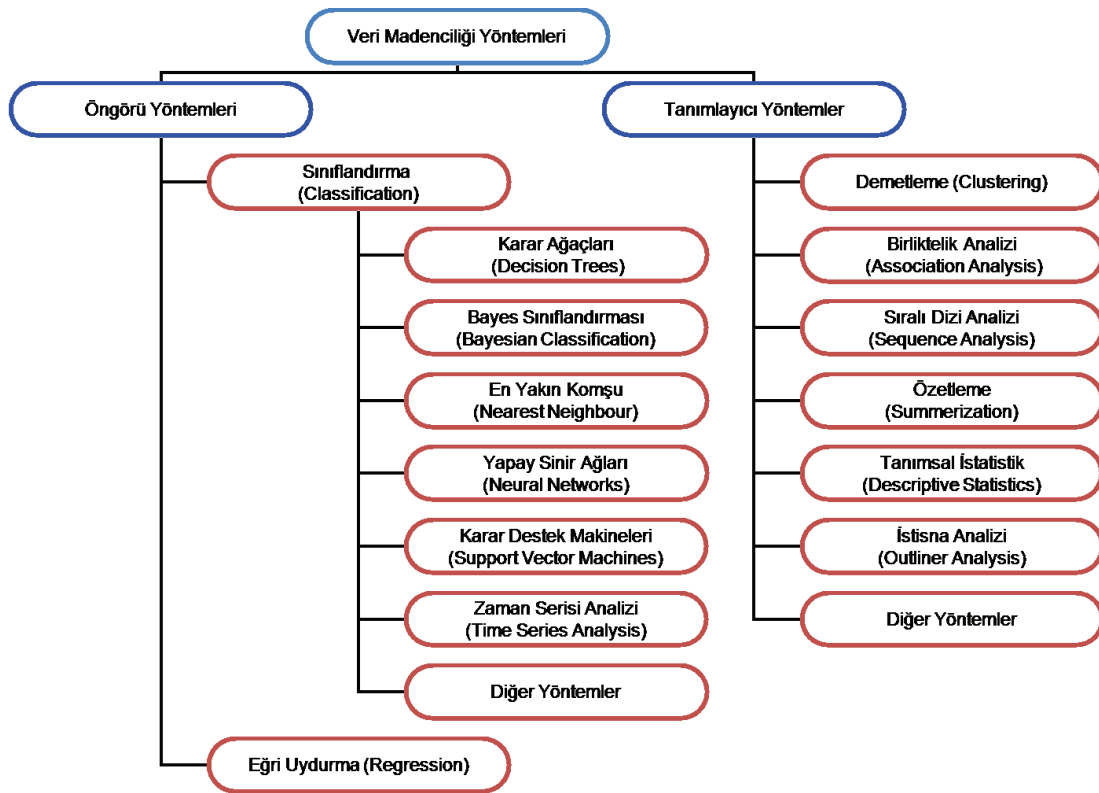
Veri madenciliği algoritmasını uygulamak için öncelikle veriler düzenlenir ve hazır bir hale gelir. Verilere göre algoritma seçimi olur, eğer veriler sınıflandırılabilir durumda ise sınıflandırma algoritmaları kullanılır, eğer veriler kümelenebilir ise kümeleme algoritmaları kullanılır. Bu aşamada amaç eldeki verilere uygun algoritmaları kullanarak en iyi sonuç veren algoritmayı kullanmaktır.

4.3.4.4. Sonuç ve değerlendirme

Uygulamanın tam olarak gerçekleştirdiğinden emin olmak için modelin eksiksiz bir şekilde değerlendirilmesi ve modeli gerçekleştirmek için oluşturulan adımların gözden geçirilmesi önemli bir adımdır. Temel amaç, yeteri derecede doğruluk oranı olup olmadığını belirlemektir. Bu aşamanın sonunda veri madenciliği sonuçlarının bir neticeye varıp varmadığını öğrenilir. Algoritmanın performansını değerlendirmek için mevcut yöntemler vardır bunlardan en yaygını çapraz geçerlilik (cross validation) yöntemidir.

BÖLÜM 5. VERİ MADENCİLİĞİ YÖNTEMLERİ

Veri madenciliğinde kullanılan yöntemler, öngörü (Prediction) ve tanımlayıcı (Description) olmak üzere ikiye ayrılır.



Şekil 5.1. Veri madenciliği yöntemleri [31]

5.1. Öngörü Yöntemleri

Öngörü yöntemlerinde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Örneğin bir banka önceki dönemlerde vermiş olduğu kredilere ilişkin gerekli tüm verilere sahip olabilir. Bu verilerde bağımsız değişkenler kredi alan müşterinin özellikleri, bağımlı değişken

değeri ise kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceğinin tahmininde kullanılmaktadır. [20].

5.1.2. Sınıflandırma ve eğri uydurma yöntemleri

Tahmin etmede faydalanılan ve veri madenciliği teknikleri içerisinde en yaygın kullanıma sahip olan sınıflama ve regresyon yöntemleridir. Sınıflamada tahmin edilen bağımlı değişken kategorik, Regresyonda ise süreklilik gösteren bir değere sahiptir.

Sınıflandırma, veri nesnesini daha önceden belirlenen sınıflardan biriyle eşleştirme sürecidir. Verileri ve karşı gelen sınıfları içeren eğitim kümesi ile eğitilen sistem, sonraki aşamalarda sınıf bilgisine sahip olunmayan verilerin ait olduğu sınıfların bulunması için kullanılır. Sınıflama algoritması bir sınıfı diğerinden ayıran örüntüleri keşfeder.

Regresyon, sürekli sayısal bir değişkenin, aralarında doğrusal ya da doğrusal olmayan bir ilişki bulunduğu varsayılan diğer değişkenler yardımıyla tahmin edilmesi yöntemidir. [21].

Regresyon modeli, sayısal değerleri tahmin etmeye yönelik olması dışında sınıflandırma yöntemine benzetilebilir. Sınıflama gruplanacak verileri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır. [21].

Sınıflama ve regresyon modellerinde karar ağaçları, yapay sinir ağları, genetik algoritmalar, K-en yakın komşu ve Naive Bayes gibi teknikler kullanılmaktadır.

5.1.2.1. K-en yakın komşu

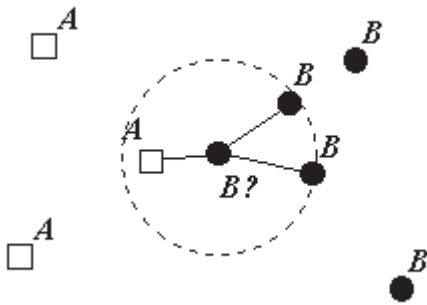
Bu teknikte tüm örnekler bir örüntü uzayında saklanır. Algoritma, bilinmeyen bir örneğin hangi sınıfa dahil olduğunu belirlemek için örüntü uzayını araştırarak bilinmeyen örneğe en yakın olan k örneklerini bulur. Yakınlık Öklid, Manhattan ve

Minkowski uzaklıkları ile tanımlanır. Daha sonra, bilinmeyen örnekler, k en yakın komşu içinden en çok benzediği sınıfa atanır.

K-en yakın komşu aşağıdaki adımlarla uygulanır:

- K parametresi belirlenir. Bu parametre verilen bir noktaya en yakın komşuların sayısıdır.
- Bu algoritma verilen bir noktaya en yakın komşuları belirleyeceği için, söz konusu nokta ile diğer tüm noktalar arasındaki uzaklıklar tek tek hesaplanır.
- Yukarıda hesaplanan uzaklıklara göre satırlar sıralanır ve bunlar arasında en küçük olan k tanesi seçilir.
- Seçilen satırların hangi kategoriye ait oldukları belirlenir ve en çok tekrarlanan kategori değeri seçilir.
- Seçilen kategori, tahmin edilmesi beklenen gözlem değerinin kategorisi olarak kabul edilir. [14].

Şekil 5.2.'de K en yakın komşu yapısı genel anlamda gösterilmiştir.



Şekil 5.2. K-en yakın komşu algoritması

5.1.2.2. Ağırlıklı oylama

K komşu algoritması verilen bir gözleme en yakın k komşunun belirlenmesi ve sınıfı bilinmeyen yeni bir gözlem değeri için, bu k gözlem içindeki en fazla tekrar eden sınıfın seçilmesi esasına dayanıyordu. Ancak seçilen bu sınıf, sadece k komşunun

gözönüne alınması nedeniyle her zaman uygun olmayabilir. [14]. Tahmin edilen noktanın sınıfı, en yakın k noktasına bakılarak karar vermek yerine, ağırlıklı oylama (weighted voting) yöntemi de kullanılabilir.

Ağırlık oylama yöntemi, gözlem değerlerine göre aşağıdaki eşitliğe göre ağırlıklı uzaklıkların hesaplanmasına dayanır.

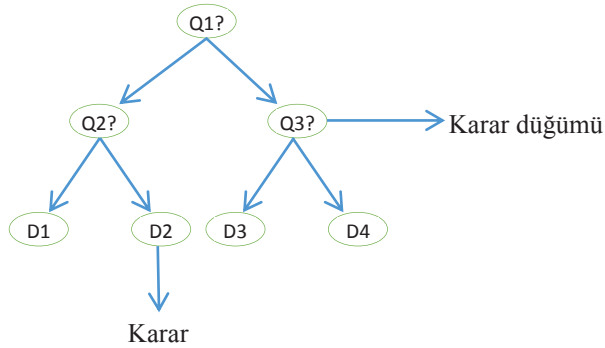
$$d(i,j) = \frac{1}{d(i,j)^2} \quad (5.1)$$

Burada $d(i,j)$ gözlemler arası öklid uzaklığıdır. Her bir sınıf değeri için bu uzaklıkların toplamı hesaplanarak ağırlıklı oylama değeri elde edilir. En büyük ağırlıklı oylama değerine sahip olan sınıf, yeni gözlemin sınıfı kabul edilir.

5.1.2.3. Karar ağaçları

Karar ağacı, adından da anlaşılacağı gibi bir ağaç görünümünde, tahmin edici bir tekniktir. Ağaç yapısı ile, kolay anlaşılabilen kurallar yaratabilen, bilgi teknolojileri işlemleri ile kolay entegre olabilen en popüler sınıflama tekniğidir. Karar ağacı karar düğümleri, dallar ve yapraklardan oluşur. Her düğümde test ve dallara ayrılma işlemleri ardışık olarak gerçekleşir ve bu ayrılma işlemi üst seviyedeki ayrımlara bağımlıdır. Ağacın her bir dalı sınıflama işlemi tamamlamaya adaydır. Eğer bir dalın ucunda sınıflama işlemi gerçekleşemiyorsa, o dalın sonucunda bir karar düğümü oluşur. Ancak dalın sonunda belirli bir sınıf oluşuyorsa, o dalın sonucunda yaprak vardır. Bu yaprak, veri üzerinde belirlenmek istenen sınıflardan biridir. Karar ağacı işlemi kök düğümünden başlar ve yukarıdan aşağıya doğru yaprağa ulaşana dek ardışık düğümleri takip ederek gerçekleşir. [22].

Karar ağacında en önemli sorunlardan birisi herhangi bir kökten itibaren dallanmanın hangi kıstasa göre yapılacağıdır. Genelde dallanmada kullanılan en sık algoritmalar entropiye dayalı algoritmalar, örnek olarak ID3 ve onun gelişmiş biçimi olan C4.5 algoritması verilebilir. Şekil 5.3.'de Karar ağacı yapısı genel anlamda gösterilmiştir.



Şekil 5.3. Karar ağacı algoritmasının yapısı

5.1.2.4. Yapay sinir ağları

Yapay sinir ağları, insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacı ile geliştirilen bilgisayar sistemleridir. [23].

İnsan beyni fonksiyonel özelliklerine benzer şekilde öğrenme, ilişkilendirme, sınıflandırma, genelleme, özellik belirleme ve optimizasyon gibi konularda başarılı bir şekilde uygulanmaktadırlar. [23].

Yapay sinir ağları, kaynak (girdi), çıktı ve iç (gizli) düğümlerle yönetilen bir grafik olarak görülebilir. Girdi düğümü girdi katmanında, çıktı düğümü ise çıktı katmanında bulunur. Gizli düğümler, bir yada daha çok gizli katmanda bulunur. Veri madenciliğinde, çıktı düğümü tahmini belirler. Tek bir girdi düğümünün olduğu (ağacın kökü) karar ağaçlarından farklı olarak yapay sinir ağlarında, her öznitelik değeri için bir girdi düğümü vardır. Yapay sinir ağları karmaşık sorunları çözebilir, ayrıca temel uygulamalardan “öğrenebilir”. Yani ağ, soruna kötü bir çözüm bulduysa, bu soruna bir dahaki sefer daha iyi bir çözüm bulacak biçimde değiştirilir. [21].

Yapay sinir ağıları üç bölümden oluşur:

- Yapay sinir ağının veri yapısını tanımlayan yapay sinir ağı grafiği.
- Öğrenmenin nasıl gerçekleşeceğini belirten öğrenme algoritması.
- Bilginin ağdan nasıl elde edileceğini belirleyen teknikler.

Bir problemin yapay sinir ağı ile çözülmesi için şu şartlardan birini sağlanması gerekir:

- Sadece yapay sinir ağıları ile problemlere pratik çözümler üretebilme durumunun söz konusu olması gerekir.
- Başka çözüm yolları olmasına rağmen yapay sinir ağlarının daha kolay ve daha etkin çözümler üretebilmesinin sağlanması gerekir. [21].

Yapay Sinir Ağları, bağlantı ve öğrenme türlerine göre sınıflandırılabilir.

- İleri beslemeli bağlantıda bağlantılar yalnızca yapıdan daha sonraki katmanlardır.
- Geri beslemeli bağlantıda ise bazı bağlantılar daha önceki katmanlardır.

Yapay Sinir Ağları öğrenme türleri ise

- Denetimli (supervised) öğrenme,
- Denetimsiz (unsupervised) öğrenmedir

Denetimli öğrenme, temel olarak iki aşamalı bir işlemdir:

- Yapay sinir ağını, örnek dizileri göstererek verideki farklı sınıfları tanıyacak biçimde eğitmek.
- Önceden görmediği bir veri grubu sağlayarak yapay sinir ağının bu örneklerden ne kadar öğrendiğini denemektir.

Denetimsiz öğrenmede ise sinir ağına, sunulan verinin doğru olarak sınıflandırılmasına ilişkin hiçbir ön bilgi verilmez. Sinir ağı, denetimsiz öğrenmeyi, o veride doğal olarak var olan kümeleri ve altkümeleri bulmak amacıyla çok boyutlu bir veri grubunu çözümlmek için kullanır. Sinir ağları denetimsiz öğrenme tekniği, sağlanan verinin yapısını temel alarak kendi sınıflandırma şemalarını tanımlamak için kullanır. [21].

5.1.2.5. Naive Bayes

Naive Bayes sınıflandırıcısı adını İngiliz matematikçi Thomas Bayes'ten alır. Bayes sınıflandırıcıları istatistiksel sınıflandırma teknikleri arasında yer alır. İstatistiksel Bayes teoremine dayanır, tahmin edici modeldir ve basit uygulanabilir bir yöntemdir. Naive Bayes, hedef değişkenle bağımsız değişkenler arasındaki ilişkiyi analiz eden tahminci ve tanımlayıcı bir sınıflama algoritmasıdır. [21].

Naive Bayes, sürekli veri ile çalışmaz. Bu nedenle sürekli değerleri içeren bağımlı yada bağımsız değişkenler kategorik hale getirilmelidir. Örneğin; bağımsız değişkenlerden biri yaş ise, sürekli değerler “<20” “21-30”, “31-40” gibi yaş aralıklarına dönüştürülmelidir. [21].

Yukarıda anlatılan Naive Bayes sınıflandırıcısını aşağıdaki örnekle açıklanmıştır.

X sınıf üyeliği bilinmeyen veri örneği olsun. Örnek $X = \{x_1, x_2, x_3 \dots x_n\}$ nitelik değerlerinden olsun. Bu örnek sınıfta m sınıf olduğunu varsayalım $C_1, C_2, \dots C_m$ sınıf değerleri olsun. Sınıfı belirlenecek olan örneğe ilişkin olarak,

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (5.2)$$

olasılıkları hesaplanır. Her bir sınıf için hesaplanan bu olasılıkların, olasılığı en büyük veren sınıf, veri örneği o sınıfa aittir. [14].

5.2. Tanımlayıcı Yöntemler

Tanımlayıcı yöntemlerde ise karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu aileler ile, çocuğu olmayan ve geliri X/Y aralığından düşük olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir. [20].

5.2.1. Kümeleme yöntemleri ve birliktelik kuralları

Kümeleme verilerin kendi aralarındaki benzerliklerin göz önüne alınarak gruplandırılması işlemidir. Kümeleme analizinde veritabanındaki kayıtların hangi kümeler ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı konunun uzmanı olan bir kişi tarafından belirtilebileceği gibi veritabanındaki kayıtların hangi kümeler ayrılacağını geliştirilen bilgisayar programları da yapabilmektedir. [21].

Kümeleme yöntemleri istatistik, biyoloji ve makine öğrenmesi gibi pek çok alanda kullanılır. Kümeleme tekniğinde, sınıflama tekniğinde olan veri sınıfları yoktur. Sınıflama tekniğinde, verilerin sınıfları bilinmekte ve yeni bir veri geldiğinde bu verinin hangi sınıftan olabileceği tahmin edilmektedir. [21]. Kümeleme yöntemi, danışmansız sınıflama modeli olarak da bilinir. [21].

Birliktelik kuralları, veritabanı içinde yer alan kayıtların birbirleriyle olan ilişkilerini inceleyerek, hangi olayların eş zamanlı olarak birlikte gerçekleşebileceklerini ortaya koymaya çalışan bir veri madenciliği yöntemidir.

Birliktelik kurallarının kullanıldığı en tipik örnek Market Sepeti Analizidir (Market Basket Analyses). Bu işlem, müşterilerin yaptıkları alışverişlerdeki ürünle arasındaki birliktelikleri bularak müşterilerin satın alma alışkanlıklarını analiz eder. Bu tip birlikteliklerin keşfedilmesi, müşterilerin hangi ürünleri bir arada aldıklarını bilgisini ortaya çıkarır ve market yöneticileri de bu bilgi ışığında daha etkili satış stratejileri geliştirebilirler. [24].

5.2.1.1. K-ortalamlar yöntemi

1967 yılında Mac Queen tarafından bulunan k-means algoritması, kümeleme problemini çözen en basit denetimsiz öğrenme (herhangi bir öğrenme olmaksızın) algoritmalarından biridir. Bölümleyici kümeleme tekniklerinden olan k-means, bilimsel ve endüstriyel uygulamalarda en yaygın olarak kullanılan kümeleme algoritmaları arasında yer almaktadır. Kümeleme, benzer özellik gösteren veri elemanlarının kendi aralarında gruplara ayrılmasıdır.

K-ortalama algoritmasına başlamadan önce k küme sayısının belirlenmesi gerekir. Söz konusu k değeri belirlendikten sonra her bir kümeye gözlem değeri atanır ve böylece C_1, C_2, \dots, C_m kümeleri belirlenmiş olur. Ardından aşağıdaki işlemler gerçekleşir:

- a) Her bir kümenin merkezi belirlenir. Bu merkezler M_1, M_2, \dots, M_m biçimindedir.
- b) e_1, e_2, \dots, e_m küme içi değişimler hesaplanır. Bu değişimlerin toplamı olan E_k^2 değeri bulunur
- c) M_k merkez değerleri ile gözlem değerleri arasındaki uzaklıklar hesaplanır. Bir gözlem değeri hangi merkeze yakın ise, o merkez ile ilgili küme içine dahil edilir.
- d) Yukarıdaki b ve c adımları, kümelerde herhangi bir değişiklik olmayınca dek sürdürülür. [14].

5.2.1.2. Apriori algoritması

Birliktelik kuralların en yaygın yöntemi Apriori Algoritmasıdır, bu algoritma Agrawal ve Srikant tarafından 1994 yılında geliştirilmiştir. Algoritmanın ismi, yaygın nesnelere önsel bilgilerini kullanmasından yani bilgileri önceki adımdan almasından “önceki (prior)” anlamında aprioridir. Bu algoritma aşağıda belirtilen aşamalara sahiptir:

- a) Birliktelik çözümlemesinin yapılabilmesi için öncelikle destek ve güven ölçütlerini karşılaştırmak üzere eşik değerler belirlenir. Uygulamadan elde edilen sonuçların bu eşik değerlere eşit ya da büyük olması beklenir.
- b) Veritabanı taranarak çözümlemeye dahil edilecek her bir ürün için tekrar sayıları, yanı destek sayıları hesaplanır. Bu destek sayıları eşik destek sayısı ile karşılaştırılır. Eşik destek sayısından küçük değerlere sahip satırlar çözümlmeden çıkarılır ve koşula uygun kayıtlar göz önüne alınır.
- c) Yukarıdaki adımda seçilen ürünler bu kez ikişerli gruplandırılacak, bu grupların tekrar sayılar, yanı destek sayıları elde edilir. Bu sayılar eşik destek sayıları ile karşılaştırılır. Eşik değerden küçük değerlere sahip satırlar çözümlmeden çıkarılır.
- d) Bu kez üçerli, dörderli ve benzeri gruplandırmalar yapılarak bu grupların destek sayıları elde edilir ve eşik değer ile karşılaştırılır, eşik değerlere uygun olduğu sürece işlemlere devam edilir.
- e) Ürün grubu belirlendikten sonra kural destek ölçütüne bakılarak birliktelik kuralları türetilir ve bu kuralların her birisiyle ilgili olarak güven ölçütleri hesaplanır. [14].

BÖLÜM 6. DİYABET VE GÖZ HASTALIKLARI

6.1. Diyabet

Diyabet vücutta glikoz metabolizmasının bozulması sonucu ortaya çıkan bir hastalık tablosudur. Glikozu vücut tarafından kullanılabilir hale getirmeye yarayan insülin hormonu, pankreasta bulunan beta hücrelerinin tahribatı sonucunda daha az salgılanmaya başlar. İnsülin hücrelerde kullanımı bozulduğundan kanda dolaşan glikoz hücrelerde kullanılamaz duruma gelir. Hücreler glikoz emilimini yapamaz ve kanda glikoz seviyesi yükselir. [25].

Bunun sonucunda vücut damarlarında bozukluklar başlar. Başta göz, böbrekler, kalp ve damar sistemi olmak üzere birçok organda hasarlar gelişir. İki tip Diyabet vardır: insüline bağımlı (tip 1) ve insüline bağımlı olmayan (tip 2). Her iki tip diyabette de hasar, kan glikoz seviyelerinin kontrolüne ve diyabetin süresine bağlıdır. [25].

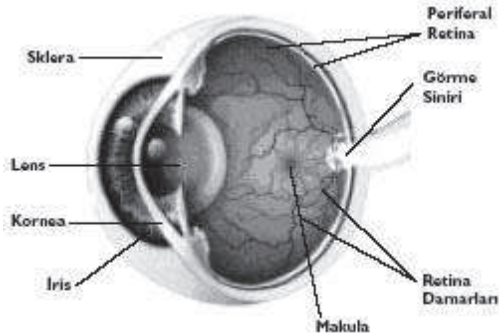
6.2. Diyabette Göz

Diyabet hastalarında gözde katarakt, glokom ve en önemlisi diyabetik retina hastalığına sebep olarak görme azalmasına yol açabilir. Diyabetlerde görme kaybı gelişme ihtimali normalden 25 kez daha fazladır. 20-65 yas arasındaki görme kaybının en sık sebeplerinden biri olarak ortaya çıkar. Diyabetik hastalarda göz hasarı gelişme ihtimali 10 yıllık diyabetiklerde %20, 30 yıllık diyabetiklerde %80 civarındadır. Fakat hastalığın teşhis ve tedavisindeki gelişmeler sayesinde, zamanında yapılan girişimlerde görmeyi ciddi şekilde etkileyen hasarlara günümüzde daha düşük oranlarda rastlanmaktadır. [26].

Katarakt, diyabeti olanlarda olmayanlara oranla daha sıklıkla ve daha genç yaşlarda oluşmaktadır. Ayrıca, görme sinirini etkileyebilir ve göz kaslarında felçlere yol açarak paralitik göz kaymasına (şaşıklık) yol açabilmektedir. Diyabetin en önemli komplikasyonu retina tutulumudur. Retinada ödem (su tutulması), kanama odakları ve yeni damarlanmaya yol açabilmektedir. Bu yeni oluşan hassas damarlar da kanayarak göz içi kanama ve retina dekolmanı oluşturabilmektedir. Hastalığın daha ileri evrelerinde de glokom (göz içi basıncının artması) oluşabilmektedir. Hastalık ve komplikasyonları tedavi edilmezse körlükle sonuçlanabilmektedir. [27].

6.3. Retina

Gözün arka tarafında yer alan ve görüntüyü elektrik uyarısına dönüştüren tabakadır. Elektrik uyarısı haline gelen uyarılar beyine görme siniriyle taşınır. Retina iki bölümden oluşur. Birincisi makula olarak isimlendirilen görme merkezi, diğeri de periferik retinadır. Makula merkezi görmeyi ve renkli görmeyi sağlar. Periferik retinaysa çevre görmeyi ve gece görmeyi sağlar. [29].



Şekil 6.1. Gözün yapısı

6.4. Diyabetik Retinopati Hastalığı

Diyabetik retinopati retina olarak adlandırılan, görme hücrelerinin bulunduğu tabakayı besleyen küçük damarların hastalığıdır. Diyabette bu damarlar yüksek şeker nedeniyle hasar görür. Bu hasar nedeniyle hem damar dışına sızıntı olur hem de retinanın beslenmesi bozulur. Tedavi edilmezse ciddi görme kayıpları olur.

Diyabetin komplikasyonlarından biri olan retinopati gelişmiş ülkelerde 20-64 yaş arasında en sık körlük sebebidir. Hastalığın prevalansı hasta yaşı ve hastalık süresi ile ilişkilidir. Tip 1 Diabetes Mellitus (D.M.)' da 20. yılda %99, Tip 2 D.M.' da 20. yılda %60 oranında değişik derecelerde retinopati bulunmuştur. Bunlardan Tip 1 D.M.'lu hastaların %3.6'sında Tip 2 D.M.'luların %1.6'sında yasal körlük görüldüğü bildirilmiştir. [28]

6.5. Diyabetik Retinopati Sınıflandırılması

Diyabetik retinopati, temel olarak iki gruba ayrılır:

- a) Non-Proliferatif diyabetik retinopati
- b) Proliferatif diyabetik retinopati

Nonproliferatif Diyabetik Retinopatide, hasar oluşmuş damarlardan kan ve sıvı sızar. Bunun sonucunda ödem olur. Retinada beslenme bozulur ve tedavi edilmezse yavaş yavaş ilerler. Eğer sıvı ve kan sızmaları periferal retinada oluyorsa hasta bu ilerlemenin farkına varmaz. Eğer görme merkezinde oluyorsa görme azalır. Bu noktada tedavinin zamanlaması çok önemlidir. Çünkü tedavi kaybolan görmeyi geri getirmez, daha fazla hasar oluşmasını engeller. [29].

Proliferatif diyabetik retinopati hastalığın daha da ileri evreleridir. Doku beslenmesi çok bozulunca yeni damar oluşumları meydana gelir. Bu yeni damar oluşumları çok incedir, kendi kendine göz içine kanama yapabilir, kanama çok yoğun olursa retina tabakasında çekintilere ve dekolmana neden olabilir. Proliferatif diyabetik retinopati acil tedavi gerektirir. Tedavi edilmediğinde görme kaybı kaçınılmazdır. [29].

6.6. Diyabetik Retinopati Seyrini Etkileyen Faktörler

1. **Hastalığın süresi:** Hastalığın süresi uzadıkça DR bulgularına rastlama olasılığı artmaktadır. Diabetes mellitus tanısı konulduktan 20 yıl sonra tip 1 DM hastalarının hemen hemen tamamında, tip 2 DM hastalarının ise

%60'ından fazlasında herhangi bir seviyede DR bulgularına rastlanmaktadır. [30].

2. **Kan glikoz düzeyinin kontrolü:** *Diabetes Control and Complications Trial (DCCT)* çalışma grubunun yaptığı geniş, randomize klinik çalışmada yoğun insülin tedavisi ve DM monitorizasyonu ile konvansiyonel tedavi karşılaştırılmıştır. Yoğun insülin tedavisi ile DR başlangıcının veya ilerleyişinin azaldığı gösterilmiştir. [30].
3. **Böbrek hastalığı:** Proteinüri, kan üre nitrojeni ve kreatininin yüksekliği ile karakterize kronik böbrek hastalığı retinopati gelişimini hızlandıran bir faktördür. [30].
4. **Sistemik hipertansiyon:** Kan basıncı yüksekliğinin DR ortaya çıkmasında risk faktörü olduğu gösterilmiştir. [30].
5. **Gebelik:** Diyabetik hastalarda gebelikle birlikte retinopati gelişme riski artmakta, aynı zamanda gebelik öncesi retinopati varsa gebelik döneminde seyri hızlanabilmektedir. [30].
6. **Hiperlipidemi:** Hiperlipideminin diyabetik retinopati seyrinde olumsuz etkileri olduğu ve okside lipoproteinlerin perisit kaybına yol açtığı gösterilmiştir. [30].
7. **Oküler faktörler:** Glokomlu gözlerde ve miyopik gözlerde DR gelişme olasılığının daha az olduğu gösterilmiştir. [30].

BÖLÜM 7. UYGULAMA

Bu tez çalışmasında diyabetik retinopati hastalığının teşhisi için uygulamada veritabanı boyutundaki işlemler için MySQL Veri Tabanı Yönetim Sistemi ve XAMPP sunucusu kullanılmıştır. Ayrıca, görsel ara yüzler PHP ve JavaScript dilleri ile yazılmıştır.

Veri Madenciliği Yöntemlerini Kullanarak Diyabetik Retinopati Hastalığının Tahmin ve Taahhisi

Uygulama Hastalık

Yöntem

K-En Yakın Komşu
 Ağırlıklı Oylama KNN
 Bayes Sınıflayıcısı

Mesafe Ölçümü

Öklid Uzaklığı
 Manhattan Uzaklığı
 Minkovski Uzaklığı

Yeni Hasta Kayıtları:

HGB: 13.5 DM (vii): 10 Glukoz: 133
ÜRE: 59 Trigliserid: 155 HDL: 44
LDL: 170 Kreatin: 0.9 HbA1c: 5.5

Sonuçlar

7-En Yakın Komşu/Bayes

En Yakın mesafeler:

0.1428000
0.1429000
0.1485000
0.1497000
0.1678000
0.1702000
0.1714000

En Yakın Komşu sonuçları:

0
1
0
0
0
0
1
0

Teşhis: 0

10-katlamalı çapraz geçerliliğin sonuçları

1.Katman

4	0
1	34

Diyagonal Toplamı: 38
Satır Sayısı: 39
Doğruluk Oranı: 97.44%

2.Katman

2	3
0	34

Diyagonal Toplamı: 36
Satır Sayısı: 39
Doğruluk Oranı: 92.31%

3.Katman

7	2
---	---

Copyright @menduhcerkezi 2013

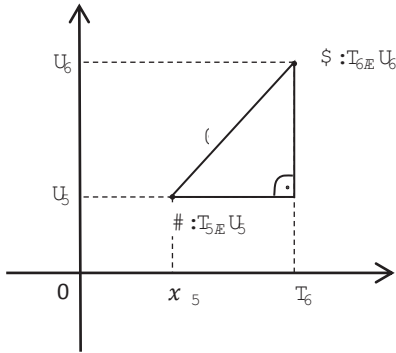
Şekil 7.1. Uygulama Arayüzü

7.1. Yöntemler ve Mesafe Ölçümleri

Yöntemler kısmında şekilde görüldüğü gibi arayüzde üç seçenek mevcuttur, K-en Yakın Komşu, Ağırlıklı Oylama KNN ve Naive Bayes sınıflandırıcıları kullanılabilir. Bu yöntemler ayrıntılı bir şekilde 5. Bölümde anlatılmıştır.

K-en Yakın Komşu ve Ağırlıklı Oylama KNN yöntemleri bilindiği gibi mesafe ölçümleri kullanmış olup, şekilde görüldüğü gibi mesafe ölçümleri için üç seçenek mevcuttur, bunlar aşağıda ayrıntılı bir şekilde anlatılmaktadır.

Öklid uzaklığı: Uygulamalarda en çok kullanılan uzaklık ölçüsü Öklid uzaklık bağıntısı adıyla bilinmektedir. Bu uzaklık, aşağıdaki şekil üzerinde görüldüğü gibi, iki boyutlu uzayda Pisagor teoreminin bir uygulaması olarak karşımıza çıkmaktadır.



Şekil 7.2. İki nokta arasındaki d uzaklığı

A ve B noktaları arasındaki Öklid uzaklığı şu şekilde olacaktır:

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (7.1)$$

Bu bağıntı genelleştirilecek olursak, i ve j noktaları için şu şekilde bir bağıntıya ulaşılır:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (7.2)$$

Manhattan uzaklığı: Diğer bir uzaklık ölçüsü Manhattan uzaklığıdır. Bu uzaklık, gözlemler arasındaki mutlak uzaklıkların toplamı alınarak hesaplanır. Söz konusu uzaklık şu şekilde ifade edilir:

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad i, j = 1, 2 \dots n; k = 1, 2 \dots p \quad (7.3)$$

Minkowski uzaklığı: p sayıda değişken göz önüne alınarak gözlem değerleri arasındaki uzaklığın hesaplanması söz konusu ise Minkowski uzaklık bağıntısı kullanılabilir. Söz konusu uzaklık şu şekilde hesaplanır:

$$d_{ij} = \left[\sum_{k=1}^p (|x_{ik} - x_{jk}|^m) \right]^{\frac{1}{m}} \quad i, j = 1, 2 \dots n; k = 1, 2 \dots p \quad (7.4)$$

7.2. Giriş Değerler

Veriler girilmeden önce min-max normalleştirme yöntemi uygulanır. Bölüm 2’de min-max normalleştirilmesi ayrıntılı bir şekilde anlatılmıştır.

Uygulamada giriş kısmı dokuz parametreden oluşmaktadır, bu parametrelere girildiği an var olan veri setinde yeni bir veri eklenir ve bu yeni girilen veri teşhis edilir. Teşhis edildikten sonra veri setinden silinir ve gerçek veri seti olduğu gibi kalır.

Veri setindeki bütün parametreler göz doktoru tarafından seçilmiştir, parametrelerin kategorik değerleri aşağıdaki tabloda verilmiştir.

Tablo 7.1 Giriş değerlerin kategorik değerleri

Parametre	Hasta (1)
DM (yıl) ²	>= 5
HGB	<= 12
LDL	>= 130
HbA1C	>= 6.5
ÜRE	<= 45
Kreatin	<= 1.2
Trigliserid	<= 150
HDL	<= 40
Glukoz	>= 140

Tabloda görüldüğü gibi bütün bu parametrelerin kendilerine özgü değerleri bulunmakta olup, kategorik değerler göz doktoru tarafından seçilmiştir. Bu değerler Naive Bayes sınıflandırma yöntemi için kullanılmıştır.

² Bu parametrede hastanın diyabet süresi girilir

Veri seti Sakarya Üniversitesi Eğitim ve Araştırma Hastanesi Göz Polikliniğinde tedavi olunan hastalardan alınan gerçek veriler üzerinden oluşmuştur.

7.3. Sonuç Bölümü

Bu bölüm, sınıflandırma yöntemlerin sonuçları ve Çapraz Geçerlilik (Cross Validation) sonuçlarından oluşmaktadır.

7.3.1. Sınıflandırma yöntemlerinin sonuçları

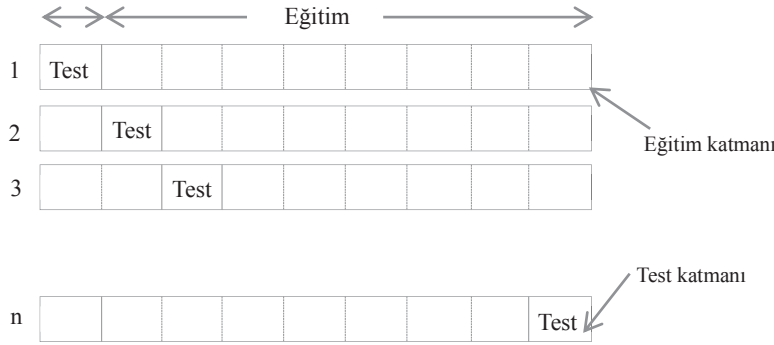
Uygulamada bildiğimiz gibi K-en Yakın Komşu, Ağırlıklı Oylama KNN ve Naive Bayes sınıflandırma yöntemleri kullanılmıştır. Uygulamanın arayüzünde kullanıcı üç ayrı sınıfa birer birer seçebilir. K-en Yakın Komşu ve Ağırlıklı Oylama KNN yöntemleri bildiğimiz gibi tüm örnekler bir örüntü uzayında saklanır, bilinmeyen bir örneğin hangi sınıfa dahil olduğunu belirlemek için örüntü uzayını araştırarak bilinmeyen örneğe en yakın olan k örneğini bulur. Bunun üzerinde uygulamada çapraz geçerlilik (cross validation) yöntemini kullanarak $k = 7$ doğruluk oranında en iyi sonuç alınmıştır. Buna göre yazılımda $k = 7$ olarak kullanılmıştır. Arayüzde görüldüğü gibi kullanıcı verileri girdiği zaman, yeni verilerinin veri setteki bütün verilerle mesafesi ölçülür, bu ölçümlerden en yakın 7 tanesi olanı sonuç bölümünde gösterilir, bunun yanı sıra en yakın 7 verilerinin teşhis sonuçları gösterilmektedir. Burada veritabanında eğer teşhis hasta ise 1 olarak kaydedilir, aksi taktirde 0 olarak kaydedilir. Ve en sonunda hastalıkla ilgili teşhis sonucu gösterilmektedir.

7.3.2. Çapraz geçerlilik ve karışım matrisi

7.3.2.1. Çapraz geçerlilik

Bu yöntemde veriler n gruba ayrılır, n katlı çapraz testi tercih edilebilir. Verilerin örneğin 10 gruba ayrıldığı bu yöntemde, ilk aşamada birinci grup test, diğer gruplar eğitim için kullanılır. Bu süreç her defasında bir grubun test, diğer grupların eğitim

amaçlı kullanılması ile sürdürülür. Aşamalardaki elde edilen doğruluk oranlarının ortalaması alınarak, uygulanan sınıflandırma yönteminin doğruluk oranı bulunur.



Şekil 7.3. n Katlamalı Çapraz Geçerlilik

7.3.2.2. Karışım matrisi

Karışım matrisi bir sınıflandırma yöntemi tarafından yapılan güncel ve tahmin edilen sınıflandırmalar hakkında bilgi içerir. Sınıflandırma yöntemlerinin performansı genellikle matristeki verilere göre değerlendirilir. Literatürde genellikle karışım matrisi iki tür sınıf olup, aşağıdaki tablodaki gibi gösterilir.

Tablo 7.2. Karışım matrisinin tablosu

		Tahmin edilen	
		Negatif	Pozitif
Gerçek	Negatif	a	b
	Pozitif	c	d

Bu tabloda Negatif ve Pozitif sınıf olarak gösterilmiştir. Tablodaki değerler şu anlama gelmektedir:

- **a** gerçek örneğin negatif olduğunda doğru tahmin sayısıdır.
- **b** gerçek örneğin negatif olduğunda yanlış tahmin sayısıdır.
- **c** gerçek örneğin pozitif olduğunda yanlış tahmin sayısıdır.
- **d** gerçek örneğin pozitif olduğunda doğru tahmin sayısıdır.

Doğruluk oranı doğru tahmin edilen sayıların toplam oranıdır, aşağıdaki denklemi kullanarak belirlenmektedir:

$$d = \frac{a+d}{a+b+c+d} \quad (7.5)$$

7.3.2.3. Çapraz geçerlilikten bir örnek

Bu çalışmada 10 - katlamalı çapraz geçerlilik yöntemi (10 – fold cross validation) kullanılmıştır. Bu yöntemin sonuçlarını göstermek için her aşamada matris karışımı yöntemi kullanılmıştır. Gerçek veri seti bildiğimiz gibi 385 verilerden oluşmaktadır, buna da yeni girilen hasta verisi eklendiği zaman 385 + 1 olur. Bunlar rasgele 10 bölünür ve her bir katlamada (fold) biri test diğer dokuzu eğitim için kullanılır. Aşağıda verilen örnek sonucunda görüldüğü gibi her katlamada diyagonal toplamı, satır sayısı ve doğruluk oranı gösterilmektedir. Ve en sonunda genel doğruluk oranı gösterilmiştir. Burada Satır Sayısı sütunu her aşamada test setini göstermektedir. Diyagonal Toplamı sütunu ise uygulamada doğru tahmin sayısını ifade etmektedir.

Tablo 7.3. Çapraz Geçerlilik ve Karışım Matrisinin uygulamada bir örneği

Katlama	Diyagonal Toplamı	Satır Sayısı	Doğruluk Oranı (%)
1	34	39	87.18
2	35	39	89.74
3	36	39	92.31
4	36	39	92.31
5	37	39	94.87
6	36	39	92.31
7	34	39	87.18
8	34	39	87.18
9	34	39	87.18
10	33	35	94.29
Doğruluk Oranı: 90.45%			

7.3.2.4. Sonuçların değerlendirilmesi

Veri madenciliği sınıflandırma yöntemlerinin test sonuçlarını elde etmek ve değerlendirmek için, 10-katlamalı çapraz doğrulama yöntemi kullanılmıştır. Bu nedenle veri seti 10 kez eğitim seti ve test setine ayrılmıştır.

7-NN için ve Bayes sınıflandırmalar için, test sonuçları aşağıdaki tabloda sunulmuştur.

Tablo 7.3. Sınıflandırıcıların teşhis sonuçları

Doğruluk oranı (%)			
	K-en yakın komşu	Ağırlıklı oylama KNN	Naive Bayes
Öklid mesafesi	92.28	92.48	88.89
Manhattan mesafesi	92.19	92.79	
Minkowski mesafesi	91.65	91.71	

Yukarıdaki tabloda görüldüğü gibi K-en yakın komşu ve ağırlıklı oylama KNN yöntemleri %90 değerini aşarak en iyi doğruluk oranı vermiş olup, bunun yanında Naive Bayes sınıflandırıcısı %88 doğruluk oranında K-en yakın komşu yöntemini takip ediyor. Mesafe ölçümlerinde ise Öklid ve Manhattan ölçümleri test sonuçlarında en başarılı olarak gözlenmişlerdir.

BÖLÜM 8. TARTIŞMA VE SONUÇ

Bu çalışmada göz hastalığı diyabetik retinopati için bir karar destek sistemi tasarlanmıştır. Bu yazılım hastane bilgi sistemlerinin bir alt modülü olması amacıyla yapılmıştır, bu sayede doktorlara yardımcı olabilecek amacıyla tasarlanmıştır. Veri madenciliği adımlarının tek tek uygulanmasıyla geçmiş verilere dayanarak teşhiste bulunmaktadır.

Doktor, uzman bilgileri ile oluşturulan yazılımın kullanıcı arayüzüne hasta ile ilgili dokuz parametreyi girerek programı çalıştırmakta ve yazılımın önerdiği teşhisi görebilmektedir. Bu sonuçla, uzman doktorun göz hastasına uygulayabileceği tedavide yardımcı ve fikir vermektedir. Bu çalışmada veri madenciliğinin tıp alanında tedavi belirlenmesinde kullanılabileceğini sonucunu ortaya koymuştur.

K-en yakın komşu, ağırlıklı oylama KNN ve bayes veri madenciliği sınıflandırma yöntemleri uygulanmıştır. Mesafe ölçümleri Öklid, Manhattan ve Minkowski ölçümleri kullanılmıştır. Bu yöntemler Çapraz Geçerlilik (Cross Validation) yöntemi ile test edilerek değerlendirilmiştir. Bu uygulamada veri madenciliğinin üç farklı yöntemini kullandığımız için aralarında karşılaştırma yapılmasını gerekli kılmıştır. Değerlendirmede K-en yakın komşu yöntemi teşhiste daha iyi doğruluk oranı verdiği gözlenmiştir. Mesafe ölçümlerinde ise Öklid ve Manhattan ölçümleri doğruluk oranında daha iyi sonuç vermişlerdir. Yapılan testlerde programın her yöntemle yüksek doğrulukta çalıştığı saptanmıştır.

Geliştirilen bu yazılım esnek bir yapıya sahip olduğundan, yazılıma yeni hastalıklar eklenebilir ve yazılım geliştirilebilir. Ayrıca uygulama bir doktor kontrolünde web tabanlı olarak geliştirildiği için, teknik donanım eksikliği ile olumsuz koşullar altında bir karar destek sistemi olarak çalışılabilir. Sistemin ticari açılardan kullanılabilmesi için istisnai durumların ele alınması, güvenlik açısından geliştirilmesi gerekmektedir.

KAYNAKLAR

- [1] <http://www.florence.com.tr/saglik-rehberi/diyabetik-repinotapi.html>, Eriřim Tarihi: 10.05.2013.
- [2] NABİYEV, V., BAHÇEKAPILI, S., Diabetik Retinopatinin Otomatik Algılanması Amacıyla Göz Görüntüsünden Kan Damarlarının Eşiklenmesi, Elektrik-Elektronik-Bilgisayar ve Biyomedikal Mühendisliği Ulusal Kongresi, 2007.
- [3] DEMİREL, B., Meme Kanseri Tedavi Yöntemlerinin Veri Madenciliği ile Belirlenmesi, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü, Isparta, 2008.
- [4] DOĞAN, Ş., Veri Madenciliği Kullanarak Biyokimya Verilerinden Hastalık Teşhisi, Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ, 2007.
- [5] SONİ, J., ANSARI, U., SHARMA, D., Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction, International Journal of Computer Applications, 2011.
- [6] ÇATALOLUK, H, KESLER M, A Diagnostic Software Tool for Skin Diseases with Basic and Weighted K-NN, Innovations in Intelligent Systems and Applications (INISTA), 2012 International Symposium on, 2012.
- [7] RİCHARDS, G., RAYWARD-SMITH V.J, SÖNKSEN, P.H., CAREY, S., WENG C., Data Mining for Indicators of Early Mortality in a Database of Clinical Records, Artificial Intelligence in Medicine, 2001.
- [8] BALAKRISHNAN, V., SHAKOURI M., HOODEH, H., Predictions Using Data Mining and Cased-based Reasoning: A Case Study for Retinopathy, World Academy of Science, Engineering and Technology 63, 2012.
- [9] TAŞKIN, G., ÜSKÜPLÜ S., Bilişimin Türkçe 'si, İstanbul Teknik Üniversitesi, Bilişim Enstitüsü, 2004.
- [10] Biyomedikal Cihaz Teknolojileri Tıbbi Bilişim, Milli Eğitim Bakanlığı, Ankara, 2008.
- [11] BAYKAL, N., Tıp Bilişimi Yüksek Lisans ve Doktora Programı, ODTÜ Enformatik Enstitüsü Sağlık Bilişimi Ana Bilim Dalı.

- [12] ÖZATA, M., ASLAN, Ş., Klinik Karar Destek Sistemleri ve Örnek Uygulamalar, Kocatepe Tıp Dergisi, 2004.
- [13] MESUT, A., Veri Tabanı Yönetimi, Trakya Üniversitesi Bilgisayar Mühendisliği.
- [14] ÖZKAN, Y., Veri Madenciliği Yöntemleri, Papatya Yayıncılık Eğitim, 2008.
- [15] ALPAYDIN, E., Zeki veri madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri, Bilişim 2000 Eğitim Semineri, 2000.
- [16] SAVAŞ, S., TOPALOĞLU, N., YILMAZ, M., Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 2012.
- [17] YILDIRIM, P., ULUDAĞ, M., GÖRÜR, A., Hastane Bilgi Sistemlerinde Veri Madenciliği, Akademik Bilişim 2008 Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, 2008.
- [18] <http://burakisikli.wordpress.com/tag/veri-madenciligi/>, Erişim Tarihi: 20.04.2013.
- [19] TEKEREK, A., Veri Madenciliği Süreçleri ve Açık Kaynak Kodlu Veri Madenciliği Araçları.
- [20] ÇALIŞKAN, B., Veri Madenciliği ve Müşteri İlişkileri Yönetimi (CRM), 2006.
- [21] ARSLAN, H., Sakarya Üniversitesi Web Sitesi Erişim Kayıtlarının Web Madenciliği ile Analizi, Yüksek Lisans Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, 2008.
- [22] ÖZDEMİR, A., YAVUZ, U., AYIK, Z., Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte ile İlişkisinin Veri Madenciliği Tekniğiyle Analizi.
- [23] ÖZTEMEL, E., Yapay Sinir Ağları, Papatya İstanbul, 2006.
- [24] <http://www.slideshare.net/nesibeyalcin/apriori-algoritmas-15209923>, Erişim Tarihi: 13.04.2013.
- [25] <http://www.orbitgozmerkezi.com/tr/goz-hastaliklari-5102/i3972/diabet-ve-goz-hastaliklari>, Erişim Tarihi: 28.04.2013.
- [26] <http://www.anatoliagoz.com/17-diyabet.html>, Erişim Tarihi: 15.05.2013.
- [27] <http://www.hattathospital.com/goz-jinekoloji-genel-cerrahi.asp>, Erişim Tarihi: 12.05.2013.

- [28] KÜÇÜKŞAHİN, H., Diabetik Maküla Ödeminde İntravitreal Triamsinolon Asetonid Tedavisi, Uzmanlık Tezi, 2005.
- [29] http://www.kaskaloglu.com/de/index.php?page=hastalik_bilgi_icerik&ID=4&description=diabetik-retinopati, Erişim Tarihi: 14.05.2013.
- [30] UYANIK, F.Y., Diyabetik Retinopatili Hastalarda Oküler Yüzey Sağlığı ve kuru göz, Uzmanlık Tezi, Pamukkale Üniversitesi Tıp Fakültesi Göz Hastalıkları Anabilim Dalı, 2010.
- [31] AKYOKUŞ, S., Veri Madenciliği Yöntemlerine Genel Bakış, 2006

ÖZGEÇMİŞ

Menduh Çerkezi, 30.01.1988 de Vıçıtırın' da doğdu. İlk, orta ve lise eğitimini kendi şehrinde tamamladı. 2006 yılında FEN Lisesinden mezun oldu. 2006 yılında başladığı Priştine Üniversitesi Mühendislik Fakültesi, Bilgisayar Mühendisliği bölümünü 2011 yılında bitirdi. 2012 yılında Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar ve Bilişim Mühendisliği Bölümüne girdi.