



**T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



YÜKSEK LİSANS

**VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ANKİLOZAN
SPONDİLİT HASTALIĞINDA RADYOGRAFİK
PROGRESYONA ETKİLİ FAKTÖRLERİN ANALİZİ**

Yeşim ATASOY

Enformatik Anabilim Dalı

Enformatik Programı

Danışman

Yrd. Doç. Dr. Çiğdem EROL

II. Danışman

Yrd. Doç. Dr. Yalçın ÖZKAN

Haziran, 2015

İSTANBUL

Bu çalışma 11/06/2015 tarihinde aşağıdaki jüri tarafından ENFORMATİK Anabilim Dalı ENFORMATİK programında Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Jürisi:

İmza
Yrd. Doç. Dr. Çiğdem EROL (Danışman)
İstanbul Üniversitesi
Enformatik Bölümü

İmza
Prof. Dr. Sevinç GÜLSEÇEN
İstanbul Üniversitesi
Enformatik Bölümü

İmza
Prof. Dr. Ş. Alp BARAY
İstanbul Üniversitesi
Mühendislik Fakültesi

İmza
Prof. Dr. Gonca MUMCU
Marmara Üniversitesi
Sağlık Bilimleri Fakültesi

İmza
Yrd. Doç. Dr. İbrahim SÖĞÜT
İstanbul Bilim Üniversitesi
Sağlık Hizmetleri Meslek
Yüksekokulu

Bu alıřma İstanbul Üniversitesi Bilimsel Arařtırma Projeleri Yürütücü Sekreterliđinin 49824 numaralı projesi ile desteklenmiřtir.

ÖNSÖZ

Tez çalışmam süresince tüm bilgilerini benimle paylaşan, her türlü konuda desteklerini esirgemeyen, beni motive eden, hiç çekinmeden her türlü konuyu danışabildiğim danışman hocam Sayın Yard. Doç. Dr. Çiğdem Erol'a,

Veri madenciliği alanına ilgi duymamı sağlayan, deneyimlerimden ve önerilerinden faydalandığım, yardıma ihtiyacım olduğunda beni hiç geri çevirmeyen, R programını öğrenmemde ve veriyi analiz etmemde beni her türlü yönlendiren ikinci tez danışmanım Sayın Yard. Doç. Dr. Yalçın Özkan'a,

Tez çalışmasında kullanılan veri setini mesleki çabası ve bilime olan bağlılığı ile toplayıp biriktiren ve samimiyetle benimle paylaşmaktan çekinmeyen, veri setini anlamam için değerli bilgilerini ve deneyimlerini paylaşan Marmara Üniversitesi Tıp Fakültesi Romatoloji Anabilim Dalı Öğretim Üyesi Sayın Prof. Dr. M. Pamir Atagündüz ve ekibine,

Tez çalışmam boyunca ve meslek hayatımda deneyimlerine ve bilgisine her zaman ihtiyaç duyduğum, bana sürekli bir şeyler katan ve verdiği önerilerle beni yönlendiren Marmara Üniversitesi Sağlık Bilimleri Fakültesi Sağlık Yönetimi Bölümü Sağlık Bilişimi ve Teknolojileri Anabilim Dalı Başkanı Sayın Prof. Dr. Gonca Mumcu'ya,

Tez çalışmamı destekleyen İstanbul Üniversitesi Bilimsel Araştırma Projeleri (BAP) birimine ve lisans ve lisansüstü eğitimim boyunca tüm imkanlarından faydalandığım İstanbul Üniversitesi'nde eğitime katkıda bulunan tüm öğretim üyelerine ve çalışanlarına,

Tüm hayatım boyunca beni ben olduğum için seven, her zaman yanımda olan ve beni yaptığım her şeyde destekleyen sevgili dostlarıma,

Son olarak, “kız çocuklarının okuması ve meslek sahibi olması her şeyden önemlidir” diyerek beni bugünlere getiren ama artık beni gökyüzünden izleyen canım babam merhum Halis Atasoy'a, gücünü ve iyilik dolu kalbini her zaman örnek aldığım canım annem Sevgi Dede'ye ve kardeşlerime sonsuz minnet ve teşekkürlerimi sunarım.

Haziran, 2015

Yeşim ATASOY

İÇİNDEKİLER

Sayfa No

ÖNSÖZ.....	i
İÇİNDEKİLER	ii
ŞEKİL LİSTESİ.....	iv
TABLO LİSTESİ	vi
SİMGE VE KISALTMA LİSTESİ	vii
ÖZET.....	viii
SUMMARY	x
1. GİRİŞ.....	1
2. GENEL KISIMLAR	4
2.1. VERİ MADENCİLİĞİ	4
2.1.1. Veri Madenciliği Kullanım Alanları	9
2.1.1.1. Sağlık ve Tıp Alanında Veri Madenciliği	10
2.1.2. Veri Madenciliği Süreci	11
2.1.3. Sınıflandırma.....	15
2.1.3.1. Karar Ağaçları:.....	17
2.1.3.2. Karar Ağacı Oluşturan Algoritmalar.....	19
2.1.3.3. Yapay Sinir Ağları.....	30
2.1.3.4. Genetik Algoritmalar.....	31
2.1.3.5. Lojistik Regresyon	35
2.1.3.6. Bayes Ağları	35
2.1.4. Kümeleme	37
2.1.5. Birliktelik Kuralları.....	38
2.1.6. Yorumlama ve Değerlendirme.....	40
2.2. ANKİLOZAN SPONDİLİT	41
2.2.1. Epidemiyolojisi	42
2.2.2. Klinik Özellikler	43
2.2.3. Tedavi	45
2.2.4. Prognoz	45

3. MALZEME VE YÖNTEM.....	47
3.1. VERİ SETİ.....	49
3.2. ANALİZ İÇİN HAZIRLANAN VERİ SETİNDEKİ NİTELİKLERİN ÖZETİ...	52
3.3. SINIF NİTELİĞİNİN BELİRLENMESİ	53
3.4. VERİ ÖN İŞLEME.....	56
3.5. KÜMELEME ANALİZİ:	57
3.6. NİTELİK SEÇİMİ	58
3.7. VERİ MADENCİLİĞİNİN UYGULANMASI.....	59
3.7.1. Uygulamada Kullanılan Algoritmalar.....	59
3.7.1.1. C4.5 Algoritması	60
3.7.1.2. Regresyon Ağaçları Algoritması	63
3.7.1.3. Gini Algoritması.....	67
3.7.1.4. Rasgele Orman Algoritması.....	69
4. BULGULAR.....	72
4.1. C4.5 ALGORİTMASI SONUÇLARI	72
4.2. REGRESYON AĞAÇLARI ALGORİTMASI SONUÇLARI	74
4.3. GİNİ ALGORİTMASI SONUÇLARI	78
4.4. RASGELE ORMAN ALGORİTMASI SONUÇLARI	84
4.5. ALGORİTMALARIN KARŞILAŞTIRILMASI	86
5. TARTIŞMA VE SONUÇ.....	87
KAYNAKLAR	92
EKLER.....	98
EK-1: TEZ ÇALIŞMASINDA KULLANILAN VERİ SETİNİN ÖN İŞLEMEDEN SONRA .CSV UZANTILI GÖSTERİMİ.....	98
ÖZGEÇMİŞ.....	111

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1: Veri madenciliğini oluşturan disiplinler.....	5
Şekil 2.2: Veri madenciliğinin geçmişten günümüze gelişim süreci.....	6
Şekil 2.3: 2014 yılında sektörlere göre veri madenciliği kullanımının sıklığı.....	10
Şekil 2.4: Veri madenciliği uygulama süreci.....	12
Şekil 2.5: Veri madenciliği sürecinin sınıflandırma yöntemine uygun akış diyagramı.....	16
Şekil 2.6: Bir karar ağacının temel yapısı.....	18
Şekil 2.7: Karar ağacını oluşturan algoritmalar.....	18
Şekil 2.8: Gini algoritması akış diyagramı	21
Şekil 2.9: Rasgele Orman algoritması akış diyagramı.....	23
Şekil 2.10: Regresyon ağaçları algoritması akış diyagramı.....	25
Şekil 2.11: C4.5 algoritması akış diyagramı.....	27
Şekil 2.12: KNN algoritması akış diyagramı.....	29
Şekil 2.13: Biyolojik bir sinir hücresinin temel yapısı	31
Şekil 2.14: YSA algılayıcı yapısı	31
Şekil 2.15: Genetik algoritmalar akış diyagramı	34
Şekil 2.16: En çok kullanılan kümeleme yöntemleri.....	38
Şekil 2.17: Apriori algoritması akış diyagramı.....	40
Şekil 2.18: Normal omurganın ve Ankilozan spondilit hastalığında omurganın yapısı	42
Şekil 3.1: Tez çalışmasında uygulama kısmına kadar gerçekleştirilen adımlar	47
Şekil 3.2: Sınıflandırma performanslarını ölçmek için kullanılan deneme bileşenleri diyagramı	49
Şekil 3.3: Ön işlem sonrası veri setinin bir kısmının genel görünümü.....	50
Şekil 3.4: Omurganın kemikleşme dereceleri.....	54

Şekil 3.5: Omurganın kemikleşme derecelerinin radyografik görüntülenmesi	54
Şekil 3.6: Omurgadaki bölgelere göre mSASSS skorunun hesaplama tablosu.....	55
Şekil 3.7: Kategorik kümeleme analizi sonuçlarının grafiksel görüntüsü	58
Şekil 4.1: C4.5 algoritması sonucunda çıkan karar ağacı.....	73
Şekil 4.2: Regresyon ağaçları algoritması sonucunda çıkan karar ağacı.....	75
Şekil 4.3: Gini algoritması sonucunda çıkan karar ağacı	78
Şekil 4.4: Rasgele orman sınıflandırma tahminleri R program çıktısı	84
Şekil 4.5: Rasgele orman algoritması sonucunda üretilen ağaçların hata oranının grafiksel gösterimi	85
Şekil 4.6: Kullanılan algoritmalar sonucunda çıkan, AS hastalığının radyografik progresyonuna etki eden faktörlere ilişkin bulgular	86

TABLO LİSTESİ

	Sayfa No
Tablo 2.1: Ankilozan spondilit hastalık kriterleri	44
Tablo 3.1: Veri setinde analiz için seçilen nitelikler ve açıklamaları.	50
Tablo 3.2: Veri setinde yer alan niteliklerin özeti.....	52
Tablo 4.1: Veri setinde analiz için seçilen nitelikler ve açıklamaları.	72
Tablo 4.2: Regresyon ağaçları algoritması ile oluşturulan sınıflandırma ağacının hata oranlarının değişimi.	75
Tablo 4.3: Gini algoritması ile oluşturulan sınıflandırma ağacının hata oranı.....	78
Tablo 4.4: Rasgele orman algoritması hata matrisi.....	84
Tablo 4.5: Sınıflandırma algoritmasının başarı yüzdeleri.....	86

SİMGE VE KISALTMA LİSTESİ

Simgeler Açıklama

D	: Veri Tabanı
C	: Sınıf Kümesi
T	: Eğitim Veri Seti
P	: Olasılık

Kısaltmalar Açıklama

AS	: Ankilozan spondilit
İBA	: İnflamatuar Bel Ağrısı
SpA	: Spondiloartritler
KNN	: K-en yakın komşu
MIT	: Massachusetts Institute of Technology
VT	: Veri Tabanı
VM	: Veri Madenciliği
OLAP	: Online Analytical Precessing
NA	: Null Attribute
HO	: Hata Oranı
OOB	: OutOfBag
YSA	: Yapay Sinir Ağları
GA	: Genetik Algoritmalar
RO	: Rasgele Orman
RA	: Regresyon Ağaçları
DAG	: Directed Acyclic Graph

ÖZET

YÜKSEK LİSANS TEZİ

VERİ MADENCİLİĞİ YÖNTEMLERİ ile ANKİLOZAN SPONDİLİT HASTALIĞINDA RADYOGRAFİK PROGRESYONA ETKİLİ FAKTÖRLERİN ANALİZİ

Yeşim ATASOY

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Enformatik Anabilim Dalı

Danışman : Yrd. Doç. Dr. Çiğdem EROL

II. Danışman : Yrd. Doç. Dr. Yalçın ÖZKAN

Bilgi ve iletişim teknolojilerinde yaşanan gelişmeler, birçok alanda olduğu gibi sağlık alanında da mevcut verilerden yeni bilgi keşfinin kapılarını aralamaktadır. Özellikle veri madenciliği teknikleri, tıbbi verilerin analizinde sağlık çalışanlarına yeni bir ışık tutmaktadır. Bu nedenle son yıllarda sağlık alanında, özellikle hastalıklarla ilgili verilerin analizinde bu teknikler sıklıkla kullanılmaya başlanmıştır. Romatizmal bir hastalık olan ve bireye göre hastalığın seyrinin çeşitlilik gösterdiği Ankilozan spondilit hastalığına ait veri setleri de bu tekniklerin aydınlatmasına ihtiyaç duymaktadır.

Bu tez çalışmasında veri madenciliği sınıflandırma algoritmaları kullanılarak, Ankilozan spondilit hastalığına sahip bireylere ait veri seti ile hastalığın radyografik progresyonuna etki eden faktörlerin analiz edilmesi amaçlanmıştır. Çalışmada sınıflandırma karar ağaçları algoritmalarından C4.5, Gini, Regresyon ağaçları ve Random Forest algoritmaları kullanılmıştır. C4.5, Gini, Regresyon ağaçları algoritmaları ile karar ağacı modelleri oluşturulup grafiksel olarak görüntülenmiş, Random Forest algoritması ile tahmin doğruluğu incelemesi yapılmıştır.

Yapılan analizler sonucunda en yüksek sınıflandırma ve model doğruluk oranı Gini algoritması sonucu oluşturulan modelde %82 olarak bulunurken, C4.5 algoritmasında 78.7 ve regresyon ağaçlarında %74 olarak bulunmuştur. Ayrıca rasgele orman algoritmasının sınıflandırma başarı tahmini %83.11 olarak bulunmuştur. Cinsiyet, hastalık süresi, servikal tutulum, hastalığın başlangıç yaşı ve IL23rs11209032 (AA veya AG) progresyona etkili faktörler arasında bulunmuştur. Gerçek tıbbi veriler kullanılarak

retilen sonular karar vermede yardımcı model olarak kullanılabilir. alıřmanın daha ileriye tařınabilmesi iin hastalara ait veri giriřlerin eksiksiz olarak girilmesi, kayıt sayısının arttırılması gerektięi dřnlmektedir.

Haziran 2015, 121Sayfa.

Anahtar kelimeler: Veri Madencilięi, Sınıflandırma, Tıp, Ankilozan Spondilit, Veri Analizi.

SUMMARY

M.Sc. THESIS

ANALYSIS OF FACTORS AFFECTING RADIOGRAPHIC PROGRESSION IN ANKYLOSING SPONDYLITIS DISEASE BY DATA MINING METHODS

Yeřim ATASOY

İstanbul University

Institute of Graduate Studies in Science and Engineering

Department of Informatics

Supervisor : Asst. Prof. Çiğdem EROL

Co-Supervisor : Asst. Prof. Yalçın ÖZKAN

Developments in information and communication technologies pave the way for discovery of new knowledge from existing data in the health fields as in many areas. Especially data mining techniques lead the way analysis of medical data to health professionals. Therefore, in the health field in recent years, particularly in the analysis of data related to the disease, these techniques have been used frequently. Data sets that has got Ankylosing spondylitis which is a rheumatic disease and progress of the disease vary according to the individual also requires illumination of these.

In this thesis using data mining classification algorithms, with individuals who have the disease of Ankylosing spondylitis of the data set is intended to analyze the factors affecting the radiographic progression of the disease. In the study, classification decision tree algorithms which are C4.5 Gini, regression trees and Random Forest algorithm is used. C4.5 Gini, regression trees algorithms imaged decision tree models created graphically and Random Forest algorithm analysis accuracy of the estimate is made.

The analysis and model results in the highest classification accuracy rate of 82% in the model were found as a result created the Gini algorithm, C4.5 algorithm was found to be 74% in 78.7 and regression trees. In addition, random forest classification algorithm success has been estimated as 83.11%. Gender, duration of disease, cervical involvement, the age of onset of the disease and IL23rs11209032 (AA or AG) was found among the factors to progression. The results produced by using real medical data can be used as a model helps in decision making. All of the input data to be entered as

the patient to be moved further to the study, it is considered necessary to increase the number of records.

June 2015, 121 Pages.

Keywords: Data mining, Classification, Medicine, Ankylosing spondylitis, Data Analysis.

1. GİRİŞ

Günümüz toplumunda "bilgi" kavramı giderek artan bir öneme sahiptir. Bilgi en basit tanımıyla ham verilerin işlenerek anlam kazanmasıdır. Anlamli olan bilgi değerlidir ve karşımızdaki problemi çözmemizi sağlar. Bilgi aynı zamanda dönüştürülebilir bir olgudur. Yani bilgi başka bilgilere ulaşmak için bir araç olarak kullanılır ve yeni bilgileri doğurur. Bilgi günümüzü yöneten en temel unsurdur ve günümüz toplumu "bilgi toplumu" olarak adlandırılmaktadır.

Sanayi toplumlarında malzeme üretimi toplumların gücünü yansıtırken, bilgi toplumlarında bilginin üretilmesi güç unsuru olmuştur. Sanayi toplumları ile bilgi toplumlarını kıyaslayacak olursak; sanayi toplumlarındaki buhar makinelerinin yerini bilgisayarlar, maddi üretimin yerini bilgi üretimi, fiziksel emeğin yerini zihinsel emek, fiziksel sermayenin yerini beşeri sermayeler ve fabrikalardaki üretimin yerini veri bankaları ve bilgi ağlarına bağılı üretim almıştır (Kocacık, 2003).

Bilginin önemi kadar, doğru ve işe yarar bilgiye ulaşmakta bir o kadar önem arz etmektedir. Veri madenciliği verilerin analiz edilerek, veri seti içindeki gizli örüntülerin keşfedilmesini ve bilgiye ulaşılmasını sağlayan bir bilimdir (Fayyad ve Stalorz, 1997). Veri madenciliğinin klasik veri analizi tekniklerinden farkı, istatistik, makine öğrenmesi, yapay zeka gibi disiplinleri bütünleştirerek veri yığını içerisindeki gizli ve bilinmeyen örüntüleri açığa çıkarmasıdır.

Veri madenciliği 2001 yılında "MIT (*Massachusetts Institute of Technology*) *Technology Review*" dergisinde yayınlanan bir editör görüşüne göre dünyayı değiştirecek ilk on yeni teknolojiden biri olarak gösterilmiştir (Farboudi, 2009). Son on yılda veri madenciliği kullanılmış çalışmalara bakıldığında 2001 yılında yapılmış bu tahminin ne kadar doğru olduğu açıkça görülmektedir.

Klasik istatistiksel yöntemler ile farkına varılması zor ilişkilerin ve detayların belirlenmesi, teşhis, tedavi ve erken müdahaleler açısından yeni bilgi keşfi, veri madenciliği yöntemleriyle olası hale gelmiştir (Çarklı, 2010).

Veri madenciliği günümüzde telekomünikasyon, sağlık, eğitim gibi pek çok kullanım alanı bulmuştur. Tıp alanında veri madenciliği uygulamaları gün geçtikçe yaygınlaşmaktadır (Aksoy, 2009). Günümüzde teknolojinin ve depolama aygıtlarının gelişmesi sonucunda veri tabanlarında çok büyük miktarda çeşitli formatlarda ve farklı tiplerde depolanmış medikal veriler bulunmaktadır. Bu sebepten dolayı hekimlerin klasik yöntemlerle verileri işlemesi ve analiz etmesi imkansız hale gelmiştir. Veri madenciliği çok boyutlu tıp verisinin analizini mümkün kılmaktadır (Esfandiari ve diğ., 2014). Veri madenciliği tıp alanında çalışan sağlık personeline ve araştırmacılara, anomali keşifleri, yeni bilgi keşfetme, daha bilinçli karar vermek, tanı koymak, herhangi bir konuda yapılacak tahmini modelleme, kişileştirilmiş tıp ve geliştirilmiş sağlık gibi alanlarda fayda sağlamaktadır (Hays, 2012).

Ankilozan spondilit (AS) iltihaplı seyreden (inflamatuvar), aksiyal omurgayı tutan, inflamatuvar bel ağrısı ile karakterize edilen, yapısal ve fonksiyonel bozukluklara sebebiyet veren ve yaşam kalitesini azaltan kronik romatizmal bir hastalıktır. Hastalığın ilerleyişi radyografik görüntüleme ile izlenmektedir. Radyografik ilerleme ve yapısal hasarın mevcut ilaçlar ile önlenabilirliği henüz bilinmemektedir. Radyografik ilerlemeyi önleme adına, steroid ve steroid olmayan ilaçlar, anti TNF, anti inflamatuvar ilaçlar ve fizik tedavi, hastaların uzun süreli tedavileri için kullanılmaktadır (Braun ve Sieper, 2007). Radyografik ilerlemeye bağlı faktörlerin ne olduğu tam olarak bilinmemekle birlikte, literatürde HLA-B27 gen pozitifliği olan, genç erkek bireylerde ve 40 yaşın altında daha sıklıkla görüldüğü bilinmektedir (Güllüoğlu, 2003).

Bu çalışmanın amacı, veri madenciliği sınıflandırma yöntemleri karar ağacı algoritmalarından Gini, C4.5, regresyon ağaçları ve rasgele orman algoritmalarını R programında kullanılarak AS hastalığı taşıyan bireylerin verilerinin analizini yaparak, hastalığın radyografik ilerleyişine etki eden faktörlerden oluşan bir karar yapısı oluşturmaktır. Günümüzde, AS literatüründe veri madenciliği ile yapılmış bilimsel bir çalışma rastlanmamıştır.

Bu tez çalışmasında yapılan çalışmalar beş bölümden oluşmaktadır. Birinci bölümde (giriş kısmında) tez çalışmasının amacından ve kapsamından bahsedilmiştir.

İkinci bölümde (Genel Kısımlar), tezin konusunu oluşturan veri madenciliği ile ilgili temel kavramlar ve tanımlar, veri madenciliği süreci ve veri madenciliği uygulamasında kullanılan teknikler ve algoritmalar anlatılmış, daha sonra analiz için kullanılan veri setini oluşturan ankilozan spondilit hastalığı hakkında tanımlayıcı bilgilere ve tıp alanında yapılan benzer literatür çalışmalarına yer verilmiştir.

Üçüncü bölümde (Materyal ve Metot), kullanılan veri seti tanıtılmış ve veri madenciliği uygulamasında kullanılan sınıflandırma algoritmaları olan "Gini, C4.5, Rasgele Orman, Regresyon Ağaçları" algoritmalarının uygulamaları detaylı olarak anlatılmıştır.

Dördüncü bölüm olan Bulgular kısmında, algoritmaların uygulanması ile ortaya çıkan sonuçlar ve oluşturulan karar modelleri anlatılmıştır ve sonuçların karşılaştırmaları verilmiştir.

Beşinci ve son bölüm olan tartışma ve sonuç kısmında ise, yapılan tez çalışması sonuçlarının değerlendirilmesi ve geleceğe dair önerilere yer verilmiştir.

2. GENEL KISIMLAR

2.1. VERİ MADENCİLİĞİ

Bilgi ve iletişim teknolojilerinde yaşanan hızlı gelişmeler işletme ve kuruluşlarda bilgi birikimini teknolojiye bağımlı hale getirmiştir. Kurumların veri tabanlarında depolanan bilgilerin işlenerek yararlı hale getirilmesi ihtiyacı doğmuştur. Verilerin analiz edilmesi ve değerinin ortaya çıkarılması veri madenciliği teknikleri ile gerçekleştirilmektedir.

Veri madenciliği (*Data mining*), genellikle büyük ve gözlemlenebilir veri setlerinden, beklenmedik ilişkiler bularak ve veriyi daha anlaşılabilir şekilde özetleyerek, daha yararlı bir şekle getirmek için yapılan analizlerdir (Hand ve diğ., 2001).

Veri madenciliği kavramı ilk olarak 1980'lerin sonlarında kullanılmaya başlanmıştır.

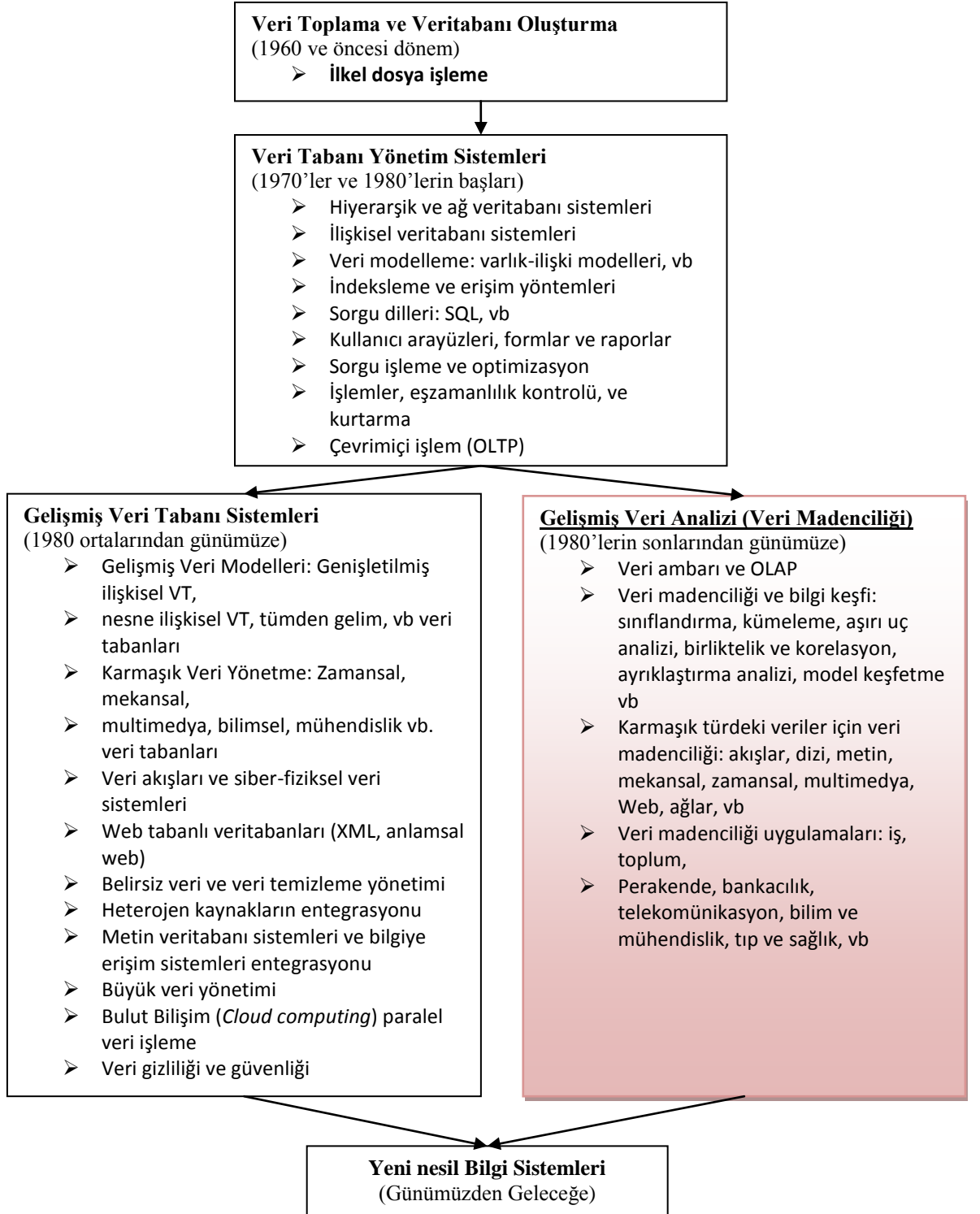
Büyük hacimdeki verilerden değerli bilginin ayıklanma süreci olarak tarif edilen veri madenciliği, büyük veri tabanlarındaki verilerden, beklenmedik modeller, değerli örüntüler ve ilginç korelasyonları keşfetme bilimidir (Ganesh, 2002).

Matematiksel ve istatistiksel modeller ile veriler üzerinde çeşitli analizler yapılabilmekte ve veriler çözümlenebilmektedir. Ancak büyük veri setlerinde özellikle ilişkisel veri tabanlarında mevcut veri setlerinin klasik istatistiksel modeller ile analiz edilmesi zorlaşmaktadır. Bu gibi büyük veri setlerinin analizi için yeni yöntemlere gereksinim duyulmaktadır. Veri madenciliği bu gibi veri setlerini analiz ederek kıymetli olan bilgiye ulaşma sürecidir (Özkan, 2013). Veri madenciliği, bilgisayar programları aracılığıyla büyük miktardaki veri kümesi içerisinde, amaçlarımıza uygun bir biçimde, veriler arasındaki ilişkileri inceleyen ve anlamlı ve gizli örüntüleri bulmayı sağlayan bir veri analizi tekniğidir. Şekil 2. 1'de görüldüğü gibi, özet olarak veri madenciliği, farklı disiplinlerin bir araya gelmesi ile oluşturulmuş disiplinler arası bir uygulama alanıdır (Savaş ve diğ., 2012). İstatistik, yapay sinir ağları, makine öğrenmesi, matematik, veri görselleştirme gibi pek çok disiplin veri madenciliği alanını etkilemektedir.



Şekil 2.1: Veri madenciliğini oluşturan disiplinler (Savaş ve diğ., 2012’den değiştirilmiştir).

Veri madenciliğinde kümeleme, sınıflandırma ve birliktelik kuralları gibi teknikler ile analizler yapılmaktadır. Şekil 2. 2’de veri madenciliğinin geçmişten günümüze ilerleyişi ve gelişim süreçleri görülmektedir.



Şekil 2.2: Veri madenciliğinin geçmişten günümüze gelişim süreci (Han ve diğ., 2012'den değiştirilmiştir).

Veri madenciliği algoritmaları kullanarak, tıbbi veri setinden bilgi elde etme konusunda çeşitli araştırmalar yapılmaktadır.

Aksoy (2009) yaptığı "*Cluster Analysis of Decompression Illness* (Dekompresyon Hastalığının Cluster Analizi)" isimli tez çalışmasında Dekompresyon (vurgun) hastalığını kümeleme analizi uygulayarak bir sınıflandırma yapmıştır. Sınıflandırma sonuçlarını klasik sınıflandırma yöntemleri, yeni yapılan istatistiksel yöntemlerle ve tedavi sonuçları ile karşılaştırma yapmıştır. Sonuçların birbiri ile uyumlu olduğunu bulmuştur.

Poyraz (2012) yaptığı "Tıp'da veri madenciliği uygulamaları: Meme kanseri veri seti analizi" isimli tez çalışmasında Winconsin veri seti'den faydalanarak meme kanserine ait veriler üzerinde karar ağacı algoritmalarından C4.5 ve Naive Bayes, lojistik regresyon ve Kstar algoritmalarını kullanarak modeller oluşturmuş ve modelleri karşılaştırmıştır.

Kumdereli (2012) yaptığı "Tıp bilişimi ve veri madenciliği uygulamaları: EEG sinyallerindeki epileptiform aktiviteye veri madenciliği yöntemlerinin uygulanması" tez çalışmasında, oluşturdukları bir yazılım ile Elektroensefalografi (EEG) cihazlarından aldığı sinyalleri analiz etmiştir. Ayrıca veriler üzerinde çeşitli sınıflandırma algoritmaları uygulayarak epileptik aktivite teşhisine yardımcı olması için bir model geliştirmiştir.

Farboudi (2009) yaptığı "Tıp bilişiminde istatistiksel veri madenciliği" tez çalışmasında üroloji polikliniğine ait gerçek verileri karar ağacı algoritmalarını kullanarak sınıflandırma yapmış ve çıkan sonuçları karşılaştırmıştır. En iyi sınıflandırma tahmin doğruluğunu % 64.66 olarak bulmuştur.

Mahajani ve Aslandoğan (2004) yaptıkları bir çalışmada deri lezyonlarını ve meme kanserlerini sınıflandırmak için k-en yakın komşu (*k-Nearest Neighbor*: KNN), Naive Bayes ve karar ağacı algoritmalarını kullanmıştır. Meme kanserini sınıflandırmak için kullanılan veri setindeki sınıflandırmadaki başarı oranı KNN için %92, Bayes sınıflayıcısı için %93, karar ağaçları için %91 olarak bulunmuştur. Deri lezyonlarını sınıflandırmak için kullanılan veri setindeki sınıflandırmadaki başarı oranı ise KNN için %42.5, Bayes sınıflayıcısı için %95, karar ağaçları %94.9 olarak bulunmuştur.

Dinçer ve Duru (2006) çalışmalarında veri madenciliği kümeleme yöntemlerini karar vermede yardımcı bir yazılım oluşturmak için kullanmışlardır. Kümeleme tekniklerinden k-ortalamlar (k-means) algoritması ile veri kümesi içindeki yoğunlaşmaları belirlemiştir.

Başka bir çalışmada Fakih ve Das (2006) hastalıklara teşhis koymada kullanılan testlerin en etkin şekilde belirlenmesi için bir program geliştirmişlerdir. Bunun için veri madenciliği yöntemlerinden kaba kümeler yöntemi ile markov karar süreçleri (*decision process*) ve destekli öğrenme (*reinforcement learning*) yöntemlerini birlikte kullanmıştır.

Kusiak (2001) yaptığı bir çalışmada akciğer tümörlerinin türünün tespiti için karar vermeye yardımcı bir sistem tasarlanmıştır. Değişik yerlerdeki kliniklerden elde edilen test verilerini kural tabanlı öğrenme algoritmaları ile analiz etmiş ve analiz sonucunda karar tablosunda %100 doğruluk saptamıştır.

Chin ve diğ., (2015) yaptıkları çalışmada ilişkisel kural madenciliği, destek vektör makineleri, C4.5 karar ağaçları ve sinir ağları da dahil olmak üzere çeşitli veri madenciliği tekniklerini kullanmışlardır. Romatoid artrit (RA) hastalığının erken teşhisinde risk faktörlerinin belirlenmesinde Nationwide Klinik veri tabanları kullanılmıştır. Çalışmada kullanılan veri tabanı 1997-2008 yılları arasında (12 yıllık izlem süresi) biriken 965.279 RA tanısı olmayan ve 1.314 RA tanısı olan toplamda 966.593 hastadan oluşmaktadır. Sınıflandırma temelli çoklu birliktelik kuralları metodunu kullanmıştır ve çıkan sonuçlarda güvenilirliği test etmek için güven eşik değeri %50 olarak belirlemiştir. Doğrulama için ise k-fold çapraz doğrulama yöntemini kullanmıştır. RA erken teşhisi için risk modelleri, duyarlılık ve özgüllük ile ölçülen on kat çapraz doğrulama kullanılarak değerlendirilmiştir. Bu çapraz doğrulamalar sonucunda ortalama duyarlılık %80.3 ve ortalama özgüllük % 72.6 olarak bulunmuştur.

Yapılan başka bir çalışmada (Yıldırım ve diğ., 2010) romatizmal hastalıklar arasındaki benzerlik tahmini için kümeleme analizi kullanılmıştır. Biyomedikal veri tabanı araçlarından Facta sistem kullanılarak web madenciliği ile MEDLINE veri tabanına erişilmiş ve bu veri tabanındaki romatizmal hastalıklara ait makaleler taranmıştır. Bu veri tabanında 17 farklı romatizmal hastalık için semptomların sıklık frekansları

belirlenmiş ve hiyerarşik kümeleme yöntemleri ile analiz edilmiştir. Hastalıklar semptomlardaki benzerliklere göre 4 kümeye ayrılmıştır. Bu çalışmaya göre yapılan kümeleme analizi sonucunda bizim çalışmamızda kullandığımız ankilozan spondilit hastalığı, romatoid artrit, fibromiyalji ve gut hastalığı aynı kümede görülmektedir. Bu dört hastalığın hastalık belirtileri birbirine benzer olarak bulunmuştur.

Bu tez çalışmasında mevcut araştırmalardan farklı olarak yeni bir veri seti kullanılmış ve daha önce veri madenciliği ile analiz edilmeyen bir hastalık olan AS hastalığı ele alınmıştır. Veri madenciliği yöntemleri açısından yeni bir metot geliştirilmesine ihtiyaç duyulmamıştır.

2.1.1. Veri Madenciliği Kullanım Alanları

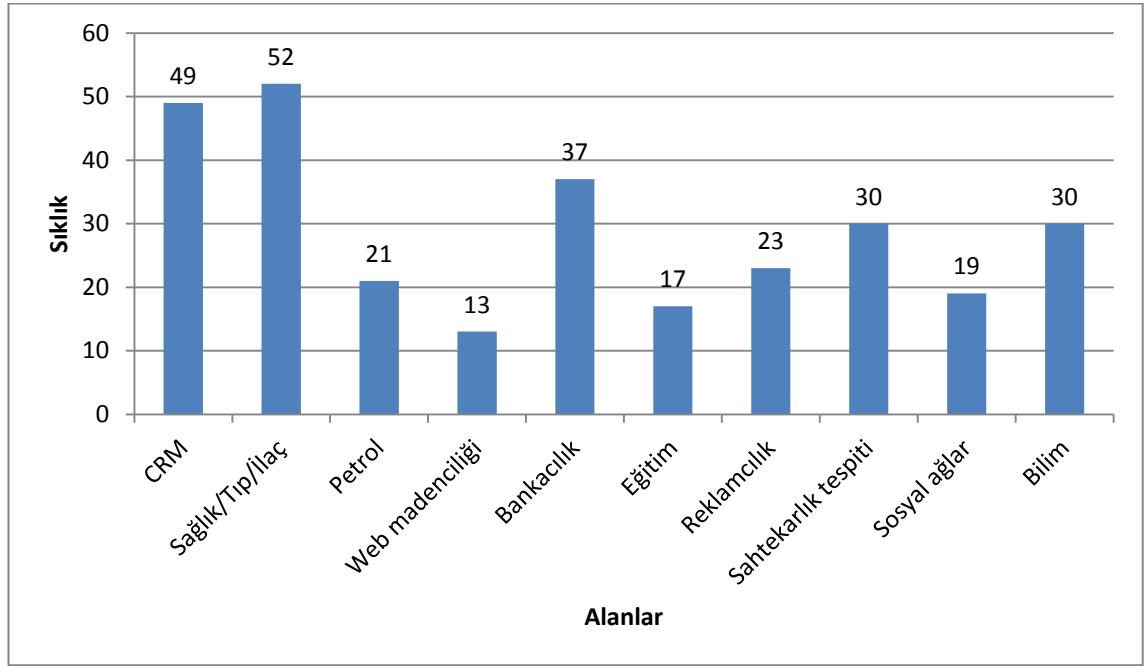
Veri madenciliği yöntemleri büyük veri tabanlarında kayıtlı veriler üzerinde uygulanabildiği için pek çok sektörde kullanılabilmektedir. Veri madenciliğinden bilimsel araştırmalar için de yararlanılmaktadır.

Veri madenciliğinin çeşitli kullanım alanları şu şekildedir (Baykal, 2006; Gülce, 2010):

1. **Pazarlama:** Müşteri analizi, sepet analizleri, satış tahminleri, çapraz satış analizleri, müşteri ilişkileri analizleri
2. **Bankacılık:** Farklı finansal göstergeler arasındaki ilişkilerin bulunması, kredi kartı dolandırıcılık tespiti, müşteri analizleri, kredi talep değerlendirme, usulsüzlük ve yolsuzluk tespiti, risk analizleri
3. **Sigortacılık:** yeni poliçe talep edecek müşteri tahmini, dolandırıcılık tespiti, riskli müşteri tespiti
4. **Perakendecilik:** Alışveriş sepeti analizleri, tedarik zinciri ve mağaza yerleşim planı analizleri, satış noktası veri analizleri
5. **Borsa:** Piyasa analizleri, hisse senedi fiyat tahmini, alım-satım stratejilerinin belirlenmesi ve en uygun strateji seçimi
6. **Telekomünikasyon:** Hisse tespitleri, hatların yoğunluk tahminleri, kalite ve iyileştirme analizleri
7. **Endüstri:** Kalite kontrol analizleri, lojistik, üretim süreçleri optimizasyonu.
8. **Bilim ve Mühendislik:** Deneysel veriler üzerinde modeller geliştirilerek bilimsel ve teknik problemlerin çözümü ve tahmini

9. Sağlık, İlaç ve Tıp: Test sonuçlarının tahmini, yeni ürün geliştirme, tıbbi teşhis ve tedavi süreçlerinin belirlenmesi

Şekil 2.3’de veri madenciliği kullanımı ile ilgili yapılan bir çalışmada 2014 yılındaki sıklıkları görülmektedir.



Şekil 2.3: 2014 yılında sektörlere göre veri madenciliği kullanımının sıklığı¹.

Şekil 2.3’de görüldüğü gibi sağlık, tıp, ilaç alanında veri madenciliği kullanımı ilk sırada yer almaktadır. Bu alanda duyulan ihtiyaç ve veri madenciliğinin katkısının sağladığı yararlar göz önüne alındığı kullanım sıklığının daha da artacağı düşünülmektedir.

2.1.1.1. Sağlık ve Tıp Alanında Veri Madenciliği

Teknolojinin hızlı ilerlemesi, pek çok alanda veri artışına sebep olmuştur. Veri artışının en hızlı olduğu alanlardan birisi de tıp ve sağlık alanlarıdır. Tıp ve sağlık alanındaki verilerin yeni teknolojiler ile analiz edilmesi kaliteli bir sağlık hizmeti sunumu için oldukça önemlidir. Tıp ve sağlık alanında veri madenciliği kullanımı giderek artan bir önem kazanmaktadır.

¹ <http://www.kdnuggets.com/polls/2014/industries-applied-analytics-data-mining-data-science.html>

Veri madenciliği uygulamaları büyük miktarda veri setleri üzerine uygulanmaktadır. Tıp alanında veri biriktirilmesi ve saklanması için imkan tanıyan uygulamalar şu şekildedir (Çarklı, 2010):

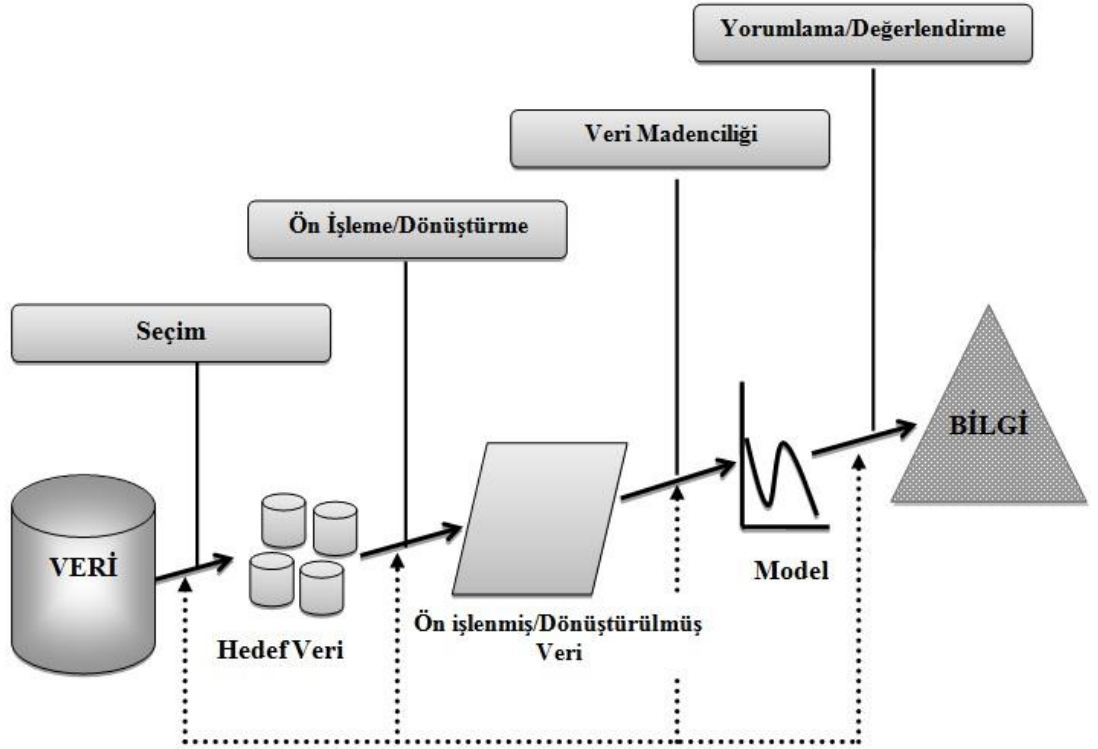
- Görüntüleme
- Teşhis koyma
- Prognoz
- Terapi
- Hastalık evrelerinin kontrolü
- Biyomedikal analizler
- Biyolojik analizler
- Epidemiyolojik çalışmalar
- Hastane Yönetimi
- Tıbbi yönerge ve eğitim

Tıbbi veriler tam, kesin ve doğrusal olmayan bulanık verilerdir. Bilgi sistemlerinin yaygınlaşması ile sağlık alanında çok fazla veri birikmiş ve işlenmesi zor ve karmaşık bir hale gelmiştir. Yukarıda belirtilen uygulamalar sonucunda toplanan veriler üzerine veri madenciliği yöntemleri uygulandığında sınıflandırılabilir, tahmin edilebilir, ilişkilendirilebilir sonuçlar üretilir. Ayrıca bir hastalığa teşhis koyma aşamasında çok fazla hastalık belirtisi ile karşılaşmaktadır ve buna bağlı olarak teşhis koymada zorlanılmaktadır (Çarklı, 2010).

2.1.2. Veri Madenciliği Süreci

Veri madenciliği veri tabanı ya da veri ambarları üzerinden bilgi keşfetme süreci olarak da tanımlanmaktadır. Veri madenciliği Şekil 2.4’de görüldüğü gibi 4 temel süreçten oluşmaktadır (Ren ve diğ., 2015; Fayyad ve diğ., 1996):

1. Seçim
2. Ön işleme – Dönüştürme
3. Veri Madenciliği Algoritmalarının ve Yöntemlerinin Uygulanması
4. Yorumlama ve değerlendirme



Şekil 2.4: Veri madenciliği uygulama süreci (Fayyad ve diğ., 1996’den değiştirilmiştir).

1-Seçim:

Veri madenciliği yaparken ne çeşit bir veri kullanılacağını belirlediği süreçtir. Veri tabanı ya da veri ambarından kullanılacak veri seçilir ve diğer süreçlere geçer.

2- Ön işleme – Dönüştürme:

Seçilen verinin yani hedef verinin veri madenciliği için hazırlandığı süreçtir. Ön işleme süreci veri madenciliğinin en uzun ve zahmetli sürecidir. Verilerin kalitesi araştırma sonucunun kalitesini de etkilemektedir.

Uygulamada analiz edilecek veriyi anlama ve veriyi hazırlama süreci, tüm veri madenciliği sürecinin % 60 ila % 90’nını kapsamaktadır (Esfandiari ve diğ., 2014).

Veri tabanlarında yer alan ham veriler işlenmemiş, eksik ya da gürültülü bir şekilde yer almaktadır. Bir veri tabanında değerler aşağıda belirtildiği gibi olabilmektedir (Larose, 2005):

- Gereksiz veri alanları

- Kayıp veriler
- Aykırı değerler
- Veri madenciliği modelleri için uygun formda olmayan değerler
- Tutarsız veriler

Uygunsuz şekildeki bu veriler ön işleme sürecine tabi tutularak, veri madenciliği uygulamaları için hazır hale getirilmektedir. Ön işleme sürecinde tanımlanan yaklaşımlar şu şekildedir:

Veri tanımlama ve özetleme (*Data description and summarization*):

Verinin mod, medyan, ortalama gibi tanımlayıcı ve betimleyici istatistiksel metotlar ile tanımlandığı yaklaşımdır (Frize ve diğ., 2001). Böylece veri setini daha iyi tanıyarak analize katılacak veya çıkartılacak alanlar daha iyi belirlenebilmektedir. Ya da bazı durumlarda veri setine mümkünse eksik verilerin ilavesi de gerçekleştirilebilmektedir.

Veri temizleme (*Data cleaning*):

Gürültülü verilerin elimine edildiği, diğer verilerden aykırı değerlerin saptandığı ve tutarsızlığın kaldırıldığı, kayıp verilerin (*missing value*) yani boş hücrelerin çıkartıldığı, yaklaşımdır (Esfandiari ve diğ., 2014). Gürültülü veri, hatalı, anlamsız veri demektir. Bu tip veriler analizin yanlış sonuçlar çıkarmasına neden olabilir. Bu sebeple gürültülü veriler belirlenerek analizden çıkartılmalıdır.

Veri setinde bilinmeyen değerler farklı şekilde kodlanmış ya da boş bırakılmış olabilir. Örneğin; “Bilinmeyen” ya da “NA” şeklinde kodlanmış olabilmektedir. Bu tip veriler eğer çok fazla ise analizden çıkarılmalı ya da sayısal değerler için diğer değerlerin ortalaması alınarak bulunan ortalama bilinmeyen hücreye yazılmalıdır (Dondurmacı ve diğ., 2014).

Veri birleştirme (*Data integration*):

Birden fazla veri setinin birleştirilerek tek bir veri setinin oluşturulmasını baz alan yaklaşımdır (Delen ve diğ., 2005). Veri bazı durumlarda farklı ortamlarda tutulabilir. Analiz öncesinde farklı ortamlardaki verileri tek ve en uygun ortamda birleştirmek gerekmektedir. Bu birleştirmeyi yaparken hatalar oluşabilmektedir. Nitelik isimleri

farklı adlandırılmış olabilir ya da bazı nitelikler diğer veri ortamında olmayabilir. Bu da birleştirme esnasında sorun olabilmektedir.

Nitelik dönüştürme (*Attribute transformation*):

Niteliklerin kategorik ya da sürekli olacak şekilde dönüştürüldüğü yaklaşımdır (Chang ve diğ., 2010). Veriler kullanılacak veri madenciliği yaklaşımına uygun bir şekilde dönüştürülmelidir.

Ayrıklaştırma ve binarizasyon (*Discretization and binarization*):

Verinin sınıflandırma ya da birliktelik kuralları gibi veri madenciliği algoritmaları için hazırlandığı yaklaşımdır (Bertolazzi ve diğ., 2008).

Veri indirgeme (*Data reduction*):

Konu dışı niteliklerin elenmesi, daha basit ve yorumlanabilir bir model oluşturabilmesine, zamandan ve bellekten tasarruf elde edilmesine olanak sağlar. Veri indirgeme ilgisiz verileri elimine eden yaklaşımdır. Literatürde pek çok veri indirgeme metotları kullanılmaktadır. En sık kullanılan veri indirgeme yöntemleri şunlardır (Esfandiari ve diğ., 2014):

1. Kümeleme
2. Nitelik Seçimi
3. Genetik Algoritmalar
4. Yapay Sinir Ağları

3- Veri madenciliği algoritmalarının / yöntemlerinin uygulanması

Veri madenciliği alanında pek çok geliştirilmiş yöntem bulunmaktadır. Bu yöntemler genel olarak şu şekilde gruplandırılabilir:

1. Sınıflandırma (*Classification*)
2. Kümeleme (*Clustering*)
3. Birliktelik Kuralları (*Association Rules*)

Bu yöntemlerden sınıflandırma yöntemi tahmin edici, gizli örüntüleri ortaya çıkarmaya yönelik geliştirilmiş bir modeldir. Kümeleme ve birliktelik kuralları ise tanımlayıcı modellerdir (Akpınar, 2000).

2.1.3. Sınıflandırma

Bir topluluk içerisinde belli ortak özellikleri olan alt gruplar oluşturmak için sınıflandırma yapmamız gerekir. Elimizde olan karmaşık bir veri kümesi içerisinde benzer verileri aynı grup altında sınıflandırmak, yeni gelen bir verinin hangi sınıfa dahil olması gerektiğini daha iyi görmemizi sağlayacaktır.

Sınıflandırma, belirli bir veri setini analiz etmek için kullanılan veri madenciliği tekniklerinden biridir. Veri setindeki her bir örneği kendisine en yakın sınıf içerisine atarak sınıflandırma hatalarını en aza indirmeye çalışır (Ragab ve diğ., 2014). Farklı bir ifade ile sınıflandırma, sınıfı bilinmeyen bir nesne için sınıf belirleme sürecidir (Özkan, 2013).

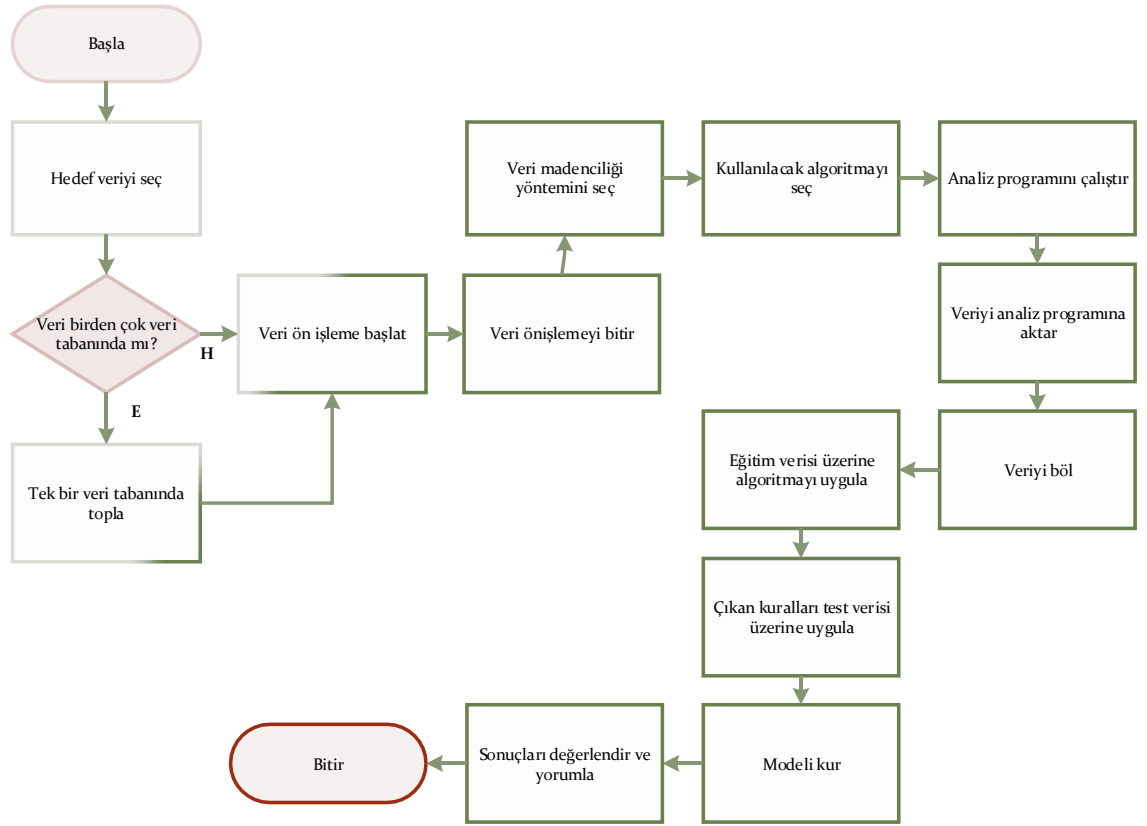
Sınıflandırma yaparken verinin tümü kullanılmaktadır. Veri “eğitim verisi” ve “test verisi” olarak ikiye ayrılmaktadır. Eğitim verisi üzerine seçilen sınıflandırma algoritması uygulanması sonucunda çıkan kurallar test verisi üzerine uygulanarak modelin doğruluğu test edilmektedir. Veri madenciliği sürecinin sınıflandırma yöntemine uygun akış diyagramı Şekil 2.5’de görüldüğü gibidir.

Sınıflandırmanın matematiksel ifadesi aşağıdaki gibidir:

Sınıflandırma yaparken kullanılan veri tabanına D ve her bir kayda t_i denilirse; $D=\{t_1, t_2, t_3, \dots, t_n\}$ tüm kayıtları içerecektir. D veri tabanındaki t_i kayıtlarının her biri özelliklerine göre bir sınıfa dahil edilmesi gerekmektedir $C= \{C_1, C_2, \dots, C_m\}$ m tane sınıftan oluşan sınıflar kümesidir. D veri tabanındaki her bir t_i kaydı mutlaka bir sınıfa dahil olmalıdır: $f: D \rightarrow C$

$$C_j = \{t_i | f(t_i) = C_j, 1 \leq i \leq n, \text{ve } t_i \in D\}$$

Yani her bir C_j ayrı bir sınıfı ifade eder ve her sınıf kendine ait kayıtları içermektedir.



Şekil 2.5: Veri madenciliği sürecinin sınıflandırma yöntemine uygun akış diyagramı.

Sınıflandırma veri madenciliği teknikleri içerisinde en çok kullanılan tekniktir. Başlıca kullanılan sınıflandırma modelleri aşağıda belirtildiği gibidir (Akpınar, 2000):

- Karar Ağaçları (*Decision Trees*)
- Yapay Sinir Ağları (*Artificial Neural Networks*)
- Genetik Algoritmalar (*Genetic Algorithms*)
- Naive-Bayes
- Lojistik Regresyondur (*Logistic Regression*)

Sınıflandırma algoritmaları pek çok alanda kullanılmaktadırlar(Larose, 2005; Prabhu, 2007):

- **Pazarlama:** Müşteri bilgilerinin yer aldığı veri tabanlarından müşterilerin özellikleri ve eski satın alma kayıtlarından faydalanarak benzer davranışlara sahip müşteri gruplarını bulmak.
- **Biyoloji:** Bitkileri ve hayvanları özelliklerine göre gruplamak
- **Kütüphaneler:** Kitap sipariş etmek
- **Sigorta:** Dolandırıcılığı belirlemek

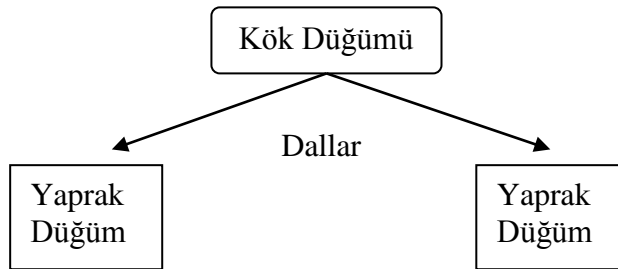
- **Şehir planlama:** Coğrafik konumları, ev tiplerine göre ev grupları belirlemek
- **Deprem çalışmaları:** Gözlemlenmiş deprem merkez üssü bölgelerine göre tehlikeli bölgeleri belirlemek
- **WWW:** Metin sınıflandırma, web bloglarına benzer erişim modellerini sınıflandırmak
- **Tıp:** Hastalık teşhis ve tedaviye yardım etmek
- **Bankacılık:** İpotek uygulamalarında yüksek ya da düşük riskli kredi durumunu belirlemek, kredi kartı dolandırıcılığı tespit etmek
- **Eğitim:** Öğrencileri özel ihtiyaçlarına göre belirli bir sınıfa yerleştirmek
- **Hukuk:** Ölen kişinin vasiyetini kendisi tarafından mı yoksa sahtekar bir kişi tarafından mı yazıldığını belirlemek

2.1.3.1. Karar Ağaçları:

Karar ağaçları bir problem ile ilgili sınıflandırma modelini görmek için yapılan bir veri madenciliği yöntemidir.

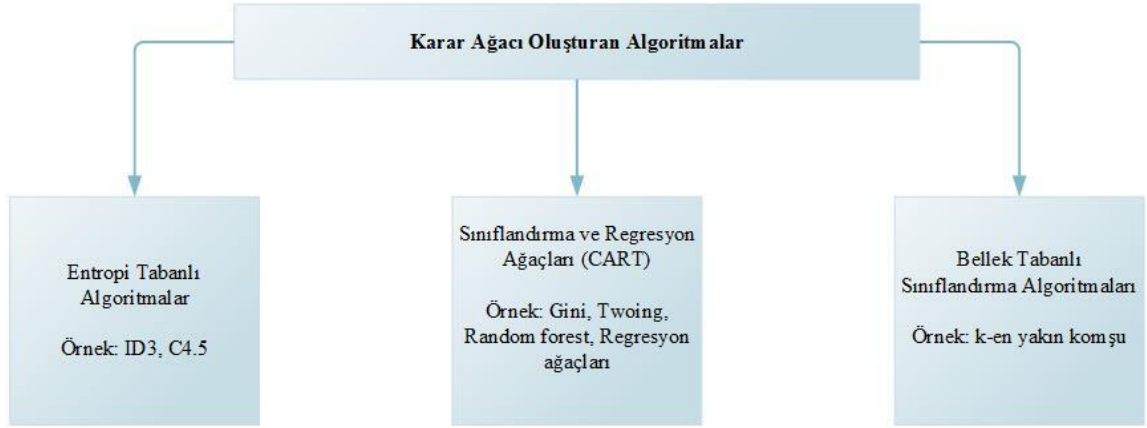
Bir karar ağacında veri setinde yer alan değişkenleri, hedef değişkeni (sınıf değişkeni) ve kestirici (prediktör) değişkenler olarak iki kısma ayırırız. Hedef değişkenimiz sınıf isimlerinin olduğu değişkenlerdir. Kestirici değişkenler ise bir veri kaydında hedef değişkenin değerini tahmin etmede kullanılan değişkenlerdir (Li ve diğ., 2001).

Bir karar ağacı nodlardan yani düğümlerden ve bu düğümleri birbirine bağlayan dallardan oluşur. Şekil 2.6'da görüldüğü gibi bir karar ağacının başlangıç noktası kök düğümdür. Bu düğümün dallanması ile oluşan farklı düğümler dalları temsil eder. En son da yaprak adı verilen yapılar ile karar ağaçları sonlandırılır.



Şekil 2.6: Bir karar ağacının temel yapısı (Pekelis, 2013'den değiştirilmiştir).

Karar ağaçlarında kökün ne olacağı, dallanmaların neye göre yapılacağı kullanılan algoritmaya göre belirlenmektedir. Karar yapısı oluşturmak için kullanılan algoritmalar temel olarak Şekil 2.7’de görüldüğü gibi gruplandırılmaktadır:



Şekil 2.7: Karar ağacını oluşturan algoritmalar.

Karar ağaçlarında kategorik veya sürekli değişkenler kullanılabilir. Eğer hedef değişkeni sürekli ise karar ağaçları genellikle regresyon ağaçları olarak, hedef değişken kategorik ise sınıflandırma ağaçları olarak adlandırılmaktadırlar. Ancak karar ağaçlarında yer alan tüm kavramlar ağaçların tümü için geçerlidir. Karar ağaçları sıklıkla, bilgisayar biliminde veri yapıları için, tıbbi alanda tanı – teşhis yöntemleri için, botanikte sınıflandırma için, psikolojide karar teorisi için ve ekonomide yatırım değerlendirme alternatifleri için kullanılmaktadır (Berry ve diğ., 2006).

Karar ağacı algoritmaları kaba tabirle şu şekilde çalışırlar (Dunham, 2002):

Algoritma. Karar Ağacı

D: Veritabanı, **T:**Ağaç

```

| T=0; Başlangıç Durumu Ağaç boş küme
| | Dallanma Kriterlerini Belirle;
| | | Dallanma Kriterlerine Göre T kök düğüm belirle;
| | | | Dallanma Kriterlerine göre Kök düğümü dallara ayır;
| | | | Tüm dallar için
| | | | Do
| | | | | Düğüm oluşturmak için değişken belirle
| | | | | If (Durdurma kriterine ulaşıldı)
| | | | | Yaprak ekle, Dur
| | | | | Else
| | | | | Return

```


Karar Ağaçlarının Test Edilmesi:

Karar ağaçlarında karar niteliği farklı olaylar (her olay karar niteliği dışındaki bir niteliği temsil eder) ile sınıflandırılarak oluşturulacak kurallar ile tahminde bulunmaktadır. Tahminin doğruluğu oluşturulacak ağacın başarısını verecektir. Ağacın başarısını anlamak için hata oranının tespit edilmesi gerekmektedir. Hata oranı yapılan sınıflandırmanın ne kadar başarılı olduğunu vermektedir. Hata oranının hesaplanması şu şekilde olmaktadır:

$$\text{Hata Oranı (HO)} = \sum (\text{Hatalı sınıflandırma}) / \sum (\text{Olaylar})$$

Karar ağaçlarının test edilmesindeki en önemli nokta veri setine yeni eklenecek kayıtların dahil olacağı sınıf için sınıflayıcının ne ölçüde başarılı olacaktır. Yeni olaylar için hata oranı ne kadar düşük olursa karar ağacının başarı oranı da bir o kadar yüksek olacaktır. Sınıflandırmadaki hata oranı tespiti yeni eklenecek veriler üzerinden yapılmalıdır. Bu sebeple analiz edilecek veri setindeki tüm verileri sınıflandırmada kullanmayıp bir kısmını test amaçlı ayırmak gerekmektedir. Karar ağacını oluşturan veri kümesi “Eğitim Veri (*Training Data*)” kümesidir. Eğitim veri kümesinin oluşturduğu ağacın doğruluğunu ya da güvenilirliğini test eden ve sınıflandırmadaki hata oranını veren veri kümesi ise “Test Veri (*Testing Data*)” kümesidir. Eğitim ve test veri kümelerinin ayrımı sonuçların doğruluğu açısından çok önemlidir ve kümeler rasgele oluşturulmalıdır.

Sınıflandırmanın doğruluğunun test edilmesi için pek çok yöntem vardır. Bunlardan en sık kullanılan yöntem “Çapraz-karşılaştırma (*Cross-validation*)” yöntemidir. Bu yöntemde eğitim kümesi ve test kümesi sırasıyla birbirinin yerine geçerek karşılaştırılır ve çapraz karşılaştırma sonucunda ağaçlardaki hata oranı tespit edilir.

2.1.3.2. Karar Ağacı Oluşturan Algoritmalar

2.1.3.2.1. Gini Algoritması:

Gini algoritması ikili bölünme yöntemi ile karar ağacı oluşturulması için kullanılan algoritmalarından biridir. Veri setinde tüm niteliklerin sağ ve sol olmak üzere iki kısma ayrılması ile oluşmaktadır. Her nitelik için ayrılan sağ ve sol grupların $Gini_{sağ}$ ve $Gini_{sol}$

değerleri hesaplanır. $Gini_{sağ}$ ve $Gini_{sol}$ değerleri denklem 2.1 ve denklem 2.2’de görüldüğü şekilde hesaplanmaktadır (McCullagh ve Nelder, 1989):

$$Gini_{sol} = 1 - \sum_1^k \left[\frac{Li}{|T_{sol}|} \right]^2 \quad (2.1)$$

$$Gini_{sağ} = 1 - \sum_1^k \left[\frac{Ri}{|T_{sağ}|} \right]^2 \quad (2.2)$$

$Gini_{sağ}$ ve $Gini_{sol}$ hesaplandıktan sonra her bir nitelik için $Gini_j$ değerleri denklem 2.3’deki formül kullanılarak hesaplanır.

$$Gini_j = \frac{1}{n} |T_{sol}| Gini_{sol} + \frac{1}{n} |T_{sağ}| Gini_{sağ} \quad (2.3)$$

Denklemlerde (1, 2, 3) yer alan sembollerin açıklamaları aşağıda belirtildiği gibidir:

k : Sınıf sayısı

T : Bir düğümdeki örnek sayısı

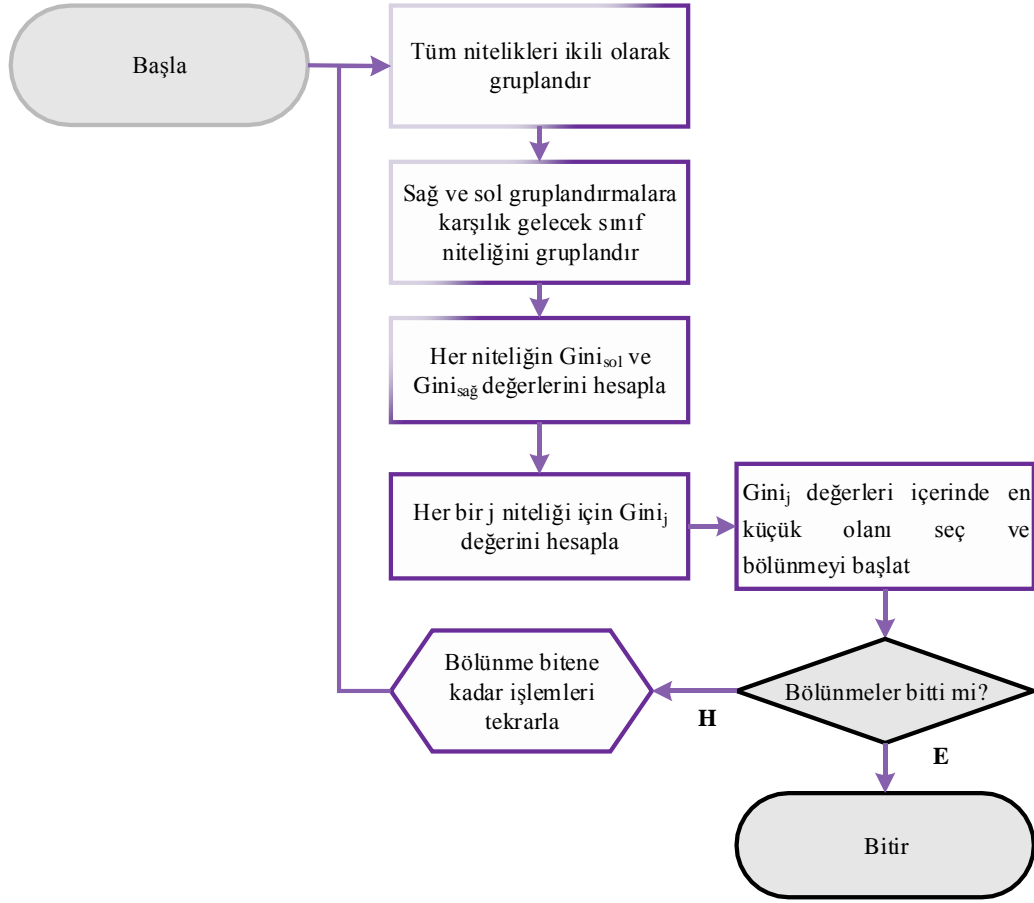
T_{sol} : Sol koldaki örneklerin sayısı

$T_{sağ}$: Sağ koldaki örneklerin sayısı

Li : Sol kolda i kategorisindeki örneklerin sayısı

Ri : Sağ koldaki i kategorisindeki örneklerin sayısı

Hesaplanan Gini değerleri arasından en küçük olanı seçilir ve bölünme bu nitelik baz alınarak gerçekleştirilir. Kalan kısım üzerinden aynı işlemler tekrar yapılarak her bir bölünme hesaplanır ve ağaç oluşturulur. Gini algoritmasının çalışması Şekil 2.8’deki akış diyagramında görülmektedir.



Şekil 2.8: Gini algoritması akış diyagramı.

2.1.3.2.2. Rasgele Orman:

Rasgele orman (*Random Forest*) binlerce karar ağacı oluşturan bir algoritmadır. Rasgele orman (RO) algoritmasını kullanabilmek için iki tane parametre tanımlanır (Schlimmer, 1985):

1. M: En iyi bölünmeyi belirlemek için her bir düğümde kullanılan değişkenlerin sayısıdır ve rasgele seçilir.
2. N: Geliştirilecek ağaçların sayısıdır.

M ve N değerleri belirlendikten sonra veri setinden bir örneklem seçilir. Tüm ağaçlar üretildiğinde (N adet) tahmin sonuçlarına göre yeni gelen kaydın sınıfı belirlenir (Liaw ve Wiener, 2002; Akar ve Güngör, 2012). RO algoritmasında belirtilmediği sürece eğitim ve test verilerinin oranlaması sabittir. Buna göre veri setinin 2/3'ü eğitim verisi (inBag) için, 1/3'ü de test verisi (OutOfBag, OOB) için ayrılmaktadır. Veri seti üzerinde bootstrap tekniği ile N tane örnek oluşturulur ve her örneğin eğitim ve test verisi ayrılır.

Eğitim verisinden yapılan tüm ağaçlar test verisi ile test edilerek hata oranı (OOB hata oranı) hesaplanır. Tüm ağaçların hata oranlarının ortalaması karar ormanının OOB hata oranını verir. Hata oranı en yüksek olan karar ağacı en düşük ağırlığı, hata oranı en düşük olan karar ağacı da en yüksek ağırlığı alır. Ağırlıklarına göre tüm ağaçlar bir oy alır ve en fazla oy alan ağaç sınıf tahmini verir (Karakoyun ve Hacıbeyoğlu, 2014).

Rasgele orman ağaçlar arasındaki korelasyon arttıkça hata oranı artar. RO ağaç yapısı oluşturmak için CART algoritmalarını kullanır (Breiman ve Cutler, 2005).

Düğümleer CART algoritması kriterlerine göre oluşturulur. Kullanılacak kriter homojenlik ölçüsünü verecektir (örneğin Gini indeksi gibi). Homojenlik ölçüsü denklem 2.4’de belirtilen formül ile bulunur (Akar ve Güngör, 2012):

$$\sum_{j \neq i} \sum (f(C_i, T)/|T|)(f(C_j, T)/|T|) \quad (2.4)$$

Denklemde (1) yer alan yer alan sembollerin açıklamaları aşağıda belirtildiği gibidir:

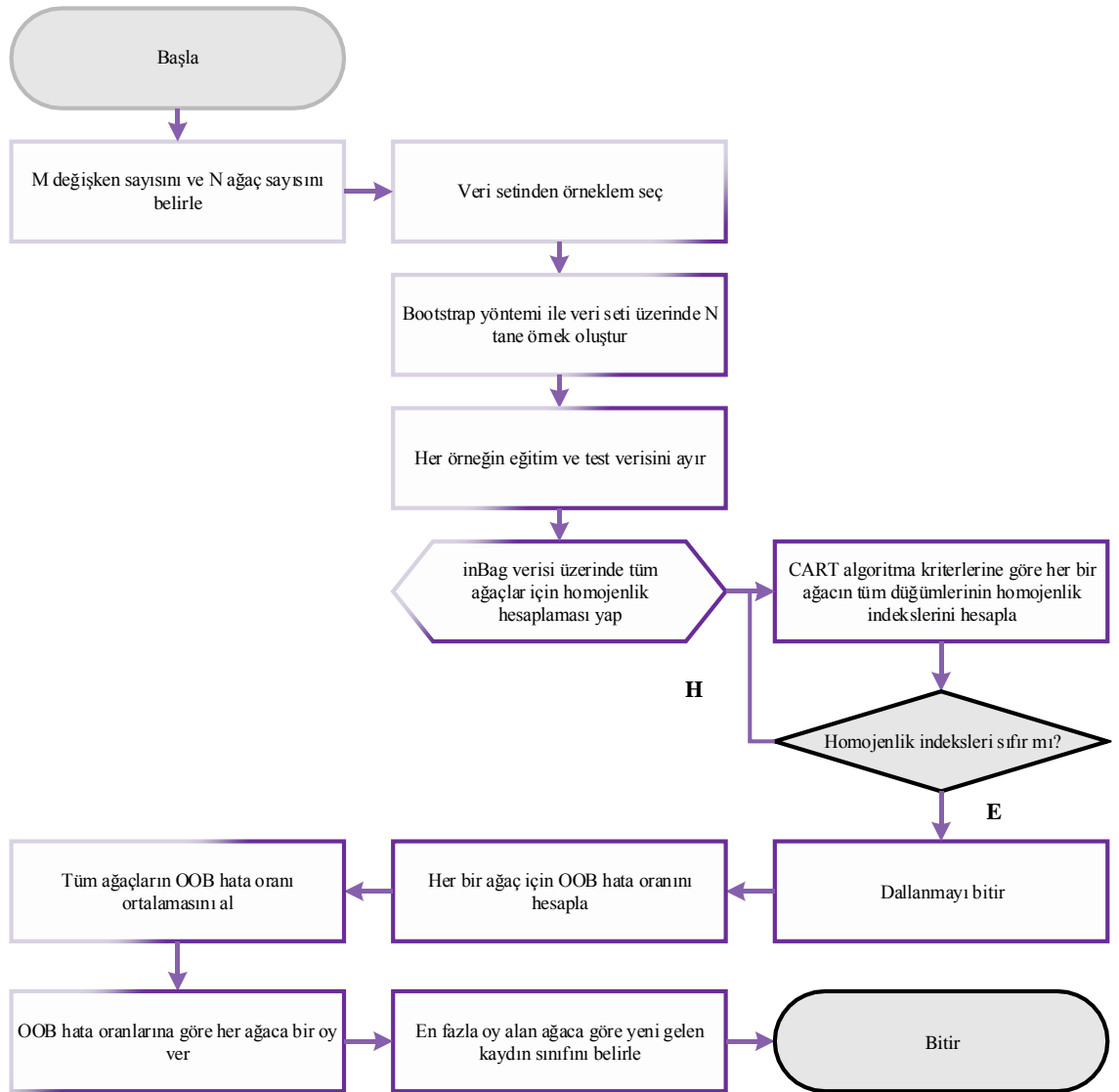
T : Eğitim veri seti

C_i : Rasgele seçilen bir niteliğin ait olduğu sınıf

f(C_i,T)/|T| : Seçilen örneğin C_i sınıfına ait olma olasılığı

Gini değeri ne kadar büyük olursa sınıf homojenliği azalacaktır. Bir ağaçtaki bir alt düğümdeki Gini indeksi bir üst düğümdeki Gini indeksinden daha az olursa o dal başarılıdır. Gini indeksi sıfıra ulaştığında ağaç sonlanacaktır (Wats ve diğ., 2011).

Rasgele orman algoritmasının çalışması Şekil 2.9’daki akış diyagramında görülmektedir.



Şekil 2.9: Rasgele Orman algoritması akış diyagramı.

2.1.3.2.3. Regresyon Ağaçları

Regresyon Ağaçları (RA) ikili bölünme yöntemine dayanan bir sınıflandırma modelidir. Regresyon ağaçları sınıflar olmadığından dolayı sınıflandırma ayırım kuralları Gini indeksi ile bulunamaz. RA’da ayrımlar artıkların karelerini azaltma algoritmasına göre gerçekleştirilir. Artıkların karelerini azaltma, sonuçlanan iki düğüm tahmin edilen toplam varyansın minimize olmasıdır (Özkan, 2012).

RA’da her düğümde minimize etme problemi denklem 2.5’deki gibi çözümlenir (Breiman ve diğ., 1998).

$$x_{j \leq x_j^R, j=1, \dots, M} \arg \min [P_1 \text{Var}(Y_l) + P_1(Y_r)] \quad (2.5)$$

P_r ve P_l : Sağ ve sol düğümlerdeki olasılıklar

M : Eğitim setindeki değişken sayısı

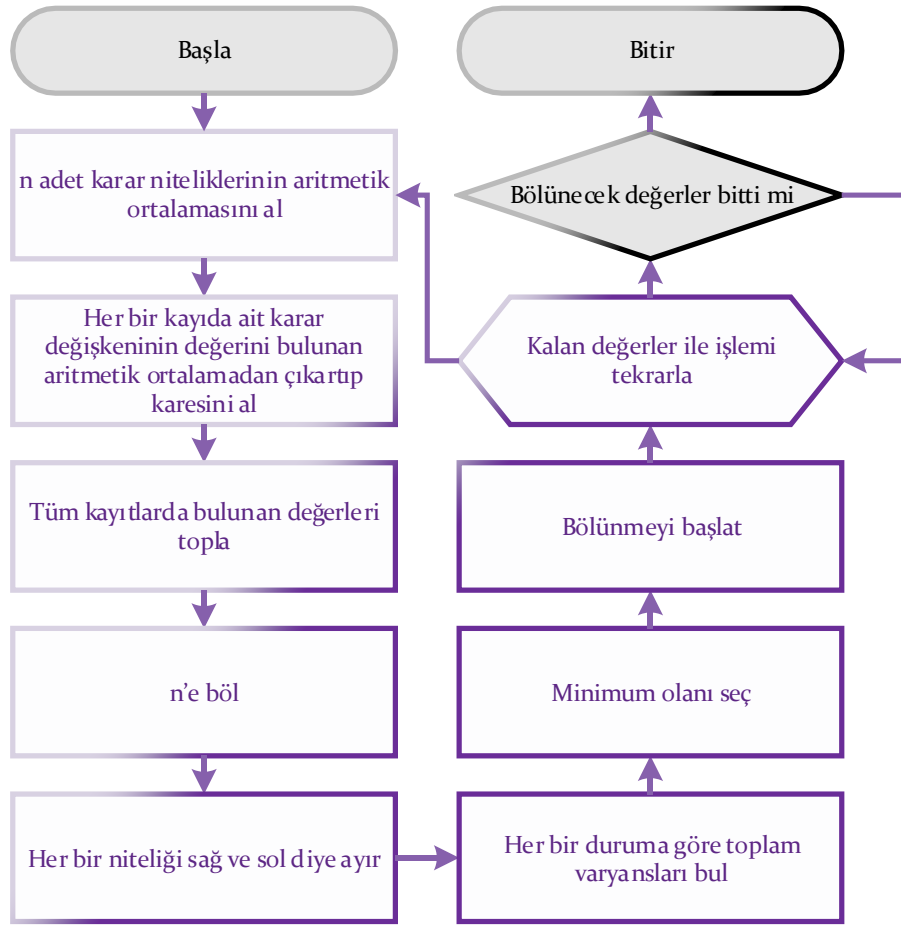
X_j : j değişkeni

X_j^R : X_j değişkeninin en iyi ayırım değeri

$\text{Var}(Y_r)$: Sağ alt düğüm için sorumlu vektör

$\text{Var}(Y_l)$: Sol alt düğüm için sorumlu vektör

RA'da ağaç, tek bir açıklayıcı değişkene dayanan basit bir kural tarafından veriyi ard arda bölerek inşa edilmiştir. Bölme prosedürü her bir gruba ayrı ayrı uygulanır. Amaç kuralı tüm gruplara homojen bir şekilde uygulamaktır. Bir ağacın boyutu son oluşan grubun sayısına eşittir. Oluşan büyük ağacın üzerinde budama işlemi yapılarak bölme işlemi devam ettirilir. Kategorik, nümerik veya hem kategorik hem nümerik veriler içeren veri setlerinde regresyon ağaçları kullanılabilir (Ath ve Fabricius, 2000). Şekil 2.10'da Regresyon ağaçları akış diyagramı görülmektedir.



Şekil 2.10: Regresyon ağaçları akış diyagramı.

2.1.3.2.4. C4.5 Algoritması:

C4.5 algoritması, Quinland (1993) tarafından ID3 algoritmasını geliştirerek ve eksiklerini gidermesi sonucunda oluşturulmuştur. C4.5 algoritması karar ağacı oluşturan algoritmalarındandır.

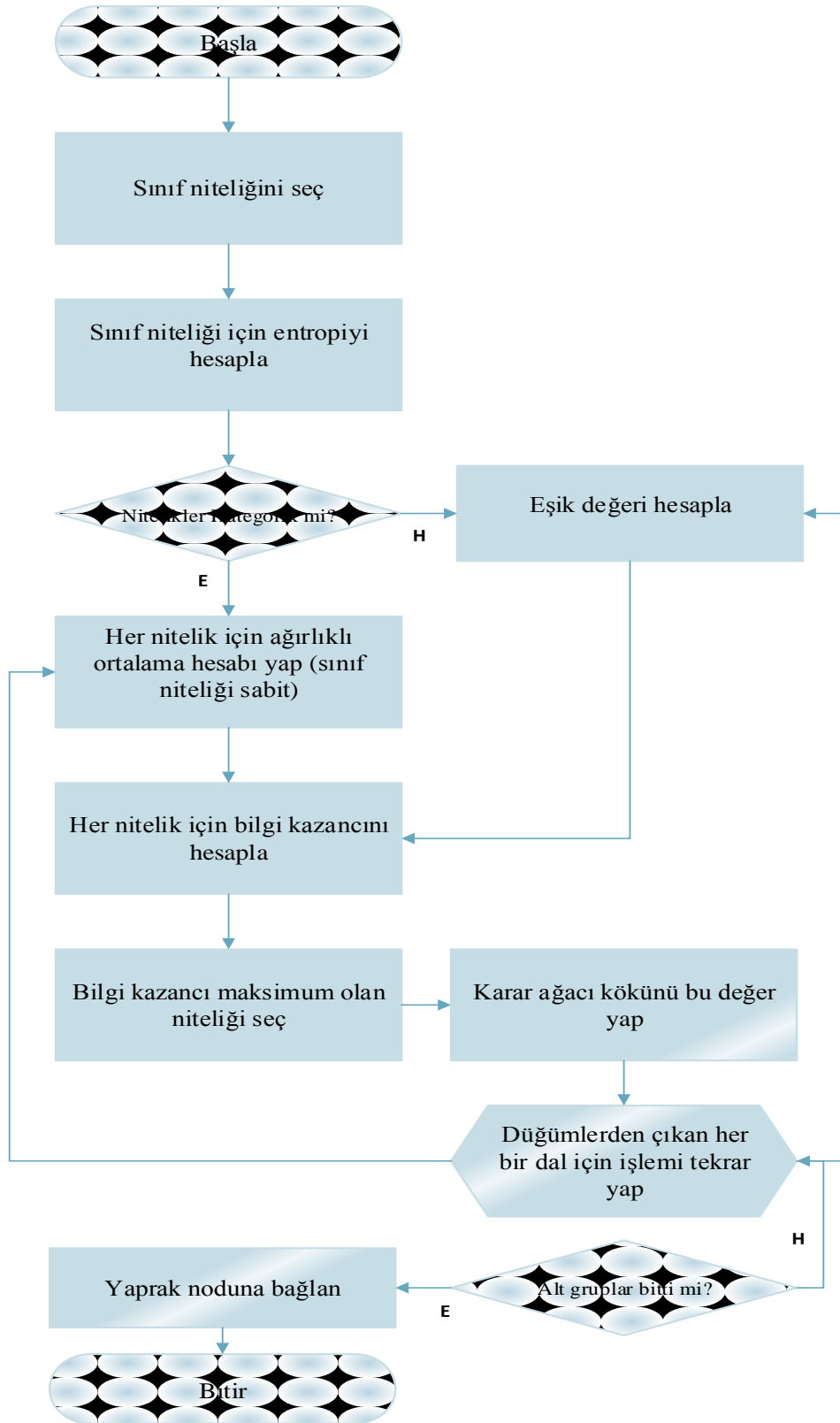
ID3 algoritması kategorik değerler için çalışan bir algoritma idi. ID3'den farklı olarak C4.5 algoritması sayısal değerleri de analize dahil ederek, ağaç yapısına katmaktadır. Entropiye dayalı algoritma çeşitlerindendir. Entropi, bir veri seti içindeki belirsizliği ve rastgeleliği ifade eder.

Kategorik veriler entropiye dayalı bilgi kazancı hesaplanması ile ağacın düğümlerini oluşturmaya başlarlar. Sürekli niteliklerde ise bir eşik değeri t_i hesaplanır ve eşik değerine göre nitelikler iki parçaya ayrılarak her bir grup kategorik değer gibi kullanılır. Eşik değeri bulmak için nitelikler sıralanır $\{V_1, V_2, \dots, V_n\}$. Niteliklerden herhangi biri seçilerek nitelik kümesi ikiye ayrılır ve her bir bölünme için bilgi kazancı

hesaplanır. Tüm bölünme olasılıkları için aynı işlem tekrarlanır ve tekrarlanan işlemler sonucunda en büyük bilgi kazancına sahip değer eşik değeri " t_i " olarak kabul edilir. Sonuç itibarıyla bulunan eşik değeri tüm değerler ile karşılaştırılır ve kategorik değerler atanır (Özkan ve Erol, 2015).

C4.5 algoritması ile CART yapılı algoritmalar arasındaki fark aşağıdaki gibidir (Larose, 2005):

- C4.5 algoritması ikili bölünme ile sınırlı değildir. CART her zaman ikili bir ağaç üretir, C4.5 ise daha fazla değişkenli bir ağaç oluşturur.
- Kategorik nitelikler için, C4.5 kategorik niteliğin her değeri için ayrı bir dal üretir. Bazı değerlerin daha düşük frekansa sahip olabileceğinden ya da diğer değerler ile ilişkili olabileceğinden dolayı bu durum istediğimizden daha çok dallanmaya sebep olabilir.
- Düğüm homojenliğini ölçmek için C4.5 algoritması CART yapılı algoritmalarından daha farklıdır.



Şekil 2.11: C4.5 algoritması akış diyagramı.

Şekil 2.11’de C4.5 algoritmasının akış diyagramında belirtildiği gibi sınıf niteliğimizi belirledikten sonra, sınıf niteliği için denklem 2.6’da görüldüğü gibi entropi hesabı yapılır (Quinland, 1993):

$$H(T) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (2.6)$$

Kategorik veriler için denklem 2.7’de görüldüğü gibi, her bir niteliğin ağırlıklı ortalama hesabı yapılır:

$$H(X, T) = \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i) \quad (2.7)$$

Sürekli veriler için denklem 2.8’de görüldüğü gibi eşik değeri hesaplaması yapılır:

$$t_i = \frac{V_i + V_{i+1}}{2} \quad (2.8)$$

Denklem 2.9’da görüldüğü gibi bilgi kazancı hesaplanır:

$$\text{Bilgi kazancı}(X, T) = H(T) - H(x, T) \quad (2.9)$$

Denklemlerde yer alan sembollerin açıklamaları aşağıda belirtildiği gibidir:

T : Sınıf değerlerini içeren küme

p_i : i tane değer olasılık dağılımı

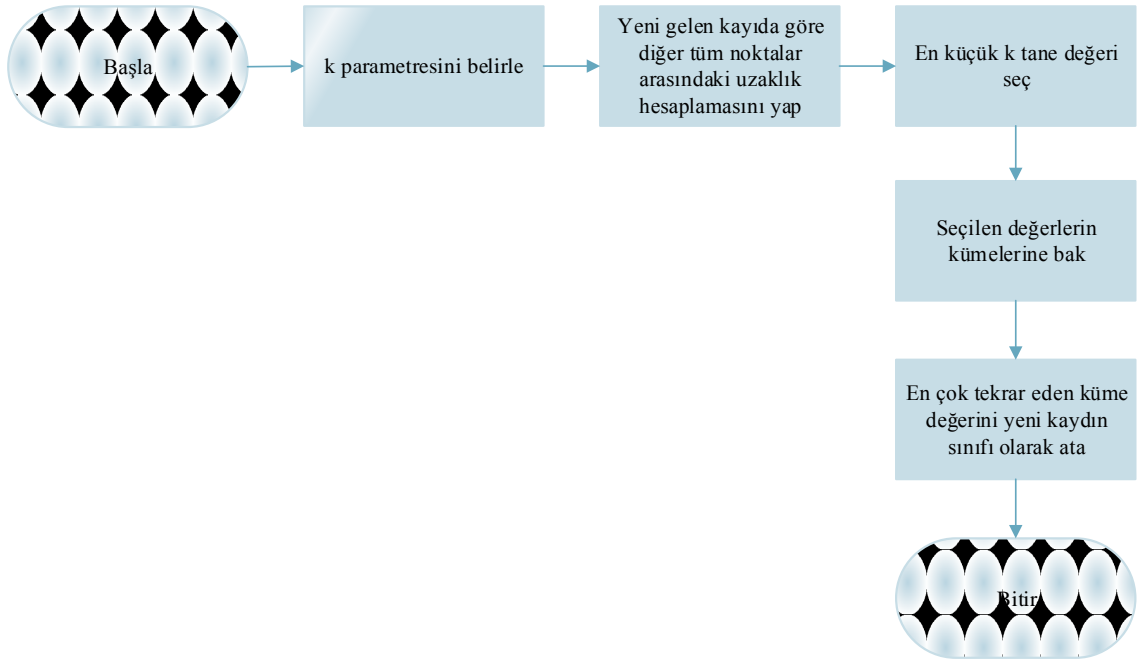
X : Hedef niteliği dışındaki herhangi bir nitelik

2.2.3.2.5. K-en Yakın Komşu Algoritması:

K-en yakın komşu (KNN) algoritması bellek tabanlı sınıflandırma algoritmalarındandır. Bu algoritmalara bir diğer örnek ağırlıklı oylama algoritmasıdır.

KNN algoritması yaygın olarak sınıflandırma alanında kullanılmaktadır. Ancak tahmin ve kestirim yapmak için de kullanılabilir (Larose, 2005). KNN algoritması sıklıkla nitelikler sürekli olduğunda tercih edilmesine rağmen kategorik nitelikler için de kullanılabilir (Bramer, 2007).

KNN algoritması nesneleri yakınlıklarına göre gruplayan bir istatistiksel sınıflandırma çeşididir. KNN algoritması hangi sınıfa ait olduğu bilinen niteliklerden oluşan veri setine eklenecek yeni kayıtların sınıfını belirlemek için kullanılır. KNN algoritması Eğitim safhası ve sınıflandırma safhası olmak üzere iki safhadan oluşur. Eğitim safhasında çalışılan veri setindeki her bir kaydın hangi sınıfa ait olduğu bilgisi saklanır. Sınıflandırma safhasında ise yeni gelen kayıtların, kullanıcı tarafından belirlenen k sayıda örneklerin her birine olan uzaklıkları hesaplanarak, en çok tekrarlanan değere göre bir sınıf etiketi atanır. Bu yöntem yeni kayıtları eğitim veri setine uzaklıklarına göre kategorize eden basit, ancak yeni kayıtları sınıflandırması açısından da bir o kadar etkili bir yöntemdir (Saini ve diğ., 2012). Şekil 2.12’de KNN algoritmasının çalışma şekli akış diyagramı görülmektedir:



Şekil 2.12: KNN algoritması akış diyagramı.

KNN algoritmasında noktalar arasındaki uzaklık hesaplaması için genellikle öklid veya Manhattan uzaklık denklemleri kullanılır (Tan ve diğ., 2006). Denklem 2.10’da Öklid uzaklık denklemi verilmektedir:

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ij} - x_{jk})^2} \quad (2.10)$$

k : Verilen bir noktaya en yakın komşuların sayısı

d : Uzaklık

KNN algoritması nispeten yavaş bir yöntemdir. Çünkü bilinmeyen bir kayıtın eğitim veri setindeki tüm kayıtlar ile arasındaki uzaklık belirlenmeli ve bunlar içerisinde en küçüğü seçilmelidir. Büyük veri tabanları düşünüldüğünde bu süreç oldukça yavaş olacaktır (Chakrabarti, 2009).

2.1.3.3. Yapay Sinir Ağları

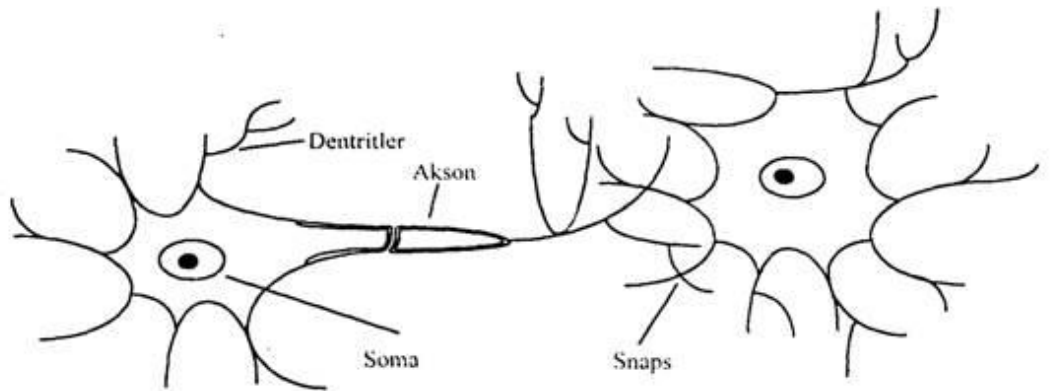
Yapay sinir ağları kendi kendine öğrenme gerçekleştiren bir veri madenciliği tekniğidir.

İnsan beyninin çalışma prensibi örnek alınarak var olan bilgilerden yeni bilgiler üretmeyi, kendi kendine öğrenmeyi amaçlayan bilgisayar sistemleridir (Prabhu, 2007).

Yapay sinir ağları, biyolojik sinir sisteminin çalışma prensibi örnek alarak oluşturulmuştur. Basitçe insan sinir sisteminin çalışma prensibi şu şekildedir:

Sinir sistemi nöron denilen sinir hücrelerinden oluşur. Şekil 2.13’de görülen biyolojik bir sinir hücresi akson ve dentrit denen iki temel kısımdan oluşur. Veri iletisi aksondan dentrite doğrudur. Dentritten gelen sinyal işlenir, ölçülür. Oluşan sinyal sayısı yeterli ise diğer nöronun aksonuna sinyal iletilir. Her akson bir dentrite bağlanarak vücudumuzda bir ağ oluştururlar.

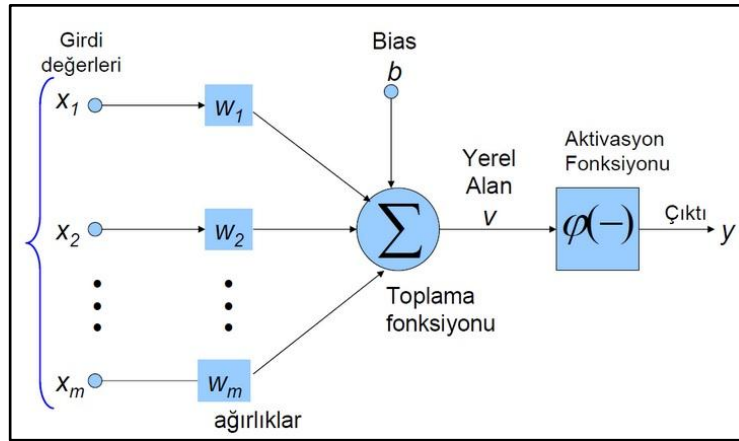
Nöronlar arasında sinaps denilen sinir boşlukları bulunmaktadır. Sinapslar öğrenmenin gerçekleştiği yerlerdir.



Şekil 2.13: Biyolojik bir sinir hücresinin temel yapısı.

Yapay sinir ağıları, bu sinir sistemini örnek alarak oluşturulmuş matematiksel bir modelledir. Sinir sistemindeki öğrenme sürecinin sonucunda mevcut verilerden öğrenme gerçekleştikten sonra yeni gözlemler için tahmin üretebilmek için modellenmiş analitik tekniklerdir (Prabhu, 2007).

İnsan sinir hücresini modelleyen bir algılayıcı (*perceptron*) Şekil 2.14’de görüldüğü gibi girdilerden, girdilerin işlenmesini sağlayan transfer fonksiyonlarından ve üretilen çıktılardan oluşur. $\{X_1, X_2, \dots, X_m\}$ sisteme herhangi bir kaynak dahil edilen girdileri temsil eder, $\{W_1, W_2, \dots, W_m\}$ ise girdilere ait ağırlıkları temsil eder. Ağırlık girdinin değeridir yani her bir girdi değeri sahip olduğu ağırlık ile çarpılır ve tüm ağırlıklar bir toplama fonksiyonu yardımıyla toplanır. Ağırlıklar toplamı eşik değerin üzerinde ise sinyal oluşturulur ve çıktı birimine gönderilir.



Şekil 2.14: YSA algılayıcı yapısı².

Yapay sinir ağıları öncelikle bir eğitim verisi ile eğitilir. Eğitim sırasında girdiler ağa defalarca gösterilir ve ağ en uygun ağırlığı bulduğunda eğitilmiş olur. Eğitilmiş bir ağ, gelecek yeni değerler için genelleme yapabilmektedir.

2.1.3.4. Genetik Algoritmalar

Bir problemin çözümü için evrim sürecinin modellenmesini örnek alarak en iyi probleme ait en iyi sonucun üretilmesini sağlayan bir veri madenciliği uygulamasıdır.

Evrime genel olarak zamana bağlı değişim demektir. Biyolojik anlamda ise canlıların zaman sürecinde çeşitli değişime uğrayarak yeni özellikler kazanmasıdır.

² <http://teknogezegen.com/yapay-sinir-aglari/>

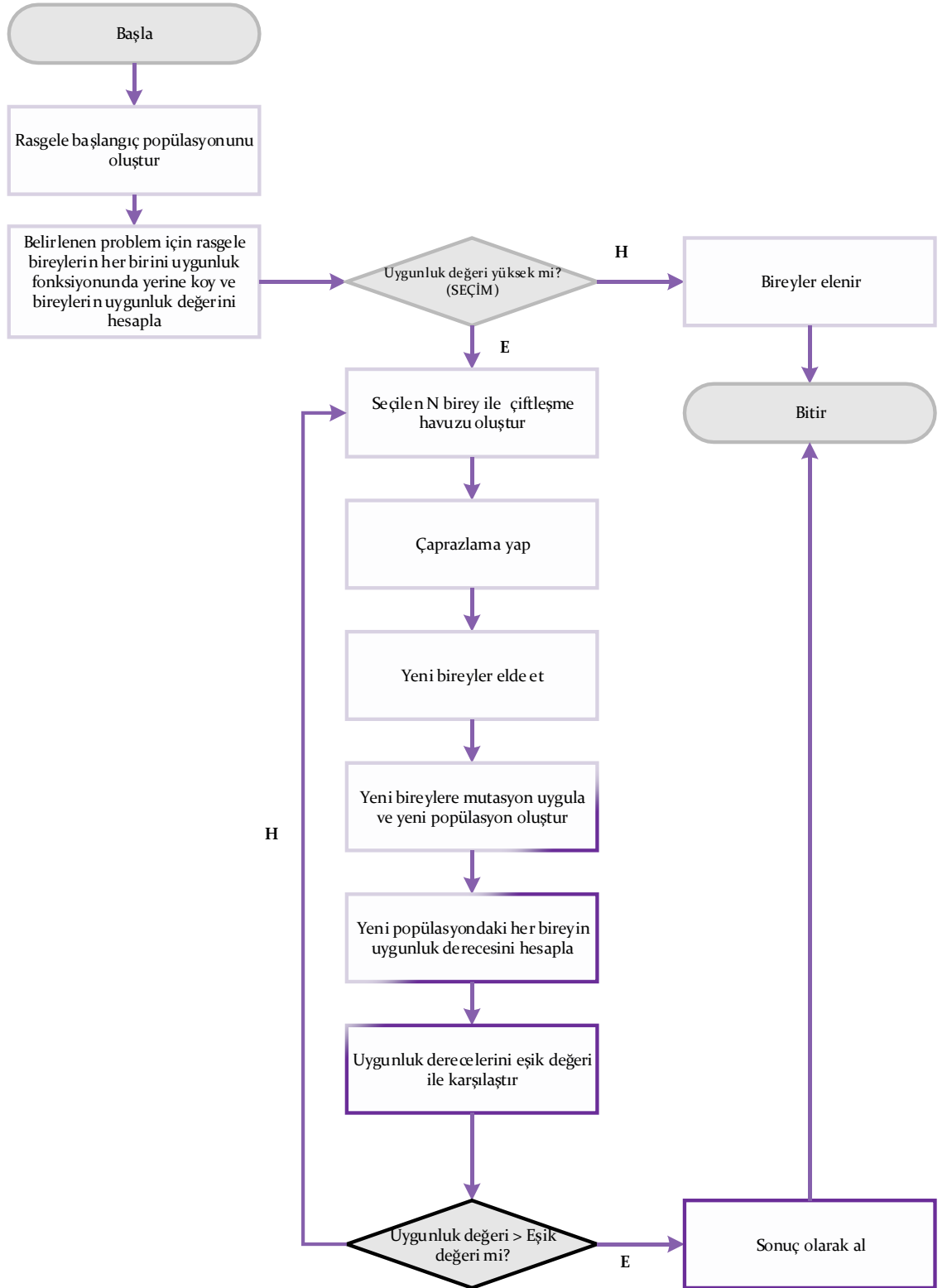
Belli bir ortamda genetik bilgileri farklı olan canlılar arasındaki rekabet sonucunda güçlü olan canlılar hayatta kalır ve çoğalarak genetik bilgilerini bir sonraki kuşağa aktarırlar. Zayıf olan canlılar ise elenerek yok olurlar. Bu olaya doğal seleksiyon yani seçim denmektedir. Güçlü olan canlıların üremeleri sonucunda çeşitlilik olacaktır. Çeşitlilik evrimleşmeyi gerçekleştirecektir. Çeşitliliğe üreme sırasında genlerde meydana gelen mutasyonlar da neden olmaktadır.

Genetik Algoritmalar (GA) karmaşık matematiksel problemlere ihtiyaç duymazlar. Bunun yerine giriş ve çıkış bilgilerine ihtiyaç duyarlar. Genellikle kompleks problemlerin optimizasyonu için tercih edilen bir veri madenciliği yöntemidir. GA’larda tek bir çözüm üretilmez. Birden fazla parametrenin en uygun çözümleri bulunacaktır. Yani GA, birden çok arama noktası üzerinden arama yapacaktır. Üretilen çözümler önceden tahmin edilemez çözümlerdir. Algoritma çalışırken arama uzayında yer alan bireylerin uygunluk indeksleri hesaplaması için olasılıksal hesaplamalar yapar. Matematiksel modelle ilgilenmez (Tuğ, 2005).

GA’lar doğal seleksiyon ilkesinden benzetim yapan bir bilgisayar uygulamasıdır. Evrimin bilgisayar üzerindeki şeklidir. Algoritmanın amacı problemleri çözerek evrimsel sistemleri modellemektir. Genetik Algoritmaların işleyişi şu şekildedir (Gülten ve Doğan, 2008):

- 1. Populasyon oluşturma:** Başlangıç için rasgele bir populasyon oluşturulur. Populasyondaki bireylerin her birinin iyi olup olmadıklarının tespit edilmesi için uygunluk fonksiyonundan geçirilir. Her problem için farklı bir uygunluk fonksiyonu kullanılır ve bu hesap olasılıksal olarak yapılır.
- 2. Seçim:** Uygunluk fonksiyonundan geçirilen bireyler seçilir diğerleri elenir.
- 3. Çaprazlama:** Seçilen bireylerin arasında gen alışverişi yani çaprazlama işlemi uygulanarak yeni nesil bireyler oluşturulur. Kaç kromozomun çaprazlanacağı çaprazlama oranına göre yapılır.
- 4. Mutasyon:** Farklı çözümlere ulaşmak ve aramanın yönünü değiştirmek amacı ile kromozomun bir elemanı değiştirilir. Kaç değişim olacağı mutasyon oranına göre karar verilir. Mutasyon sonucunda yeni bir populasyon oluşturulur.

Yeni popülasyondaki bireyler için uygunluk hesaplaması yapılır, uygunluk değeri eşik değerin altında çıkan bireyler en iyi sonuçları verecektir. Şekil 2.15’de genetik algoritmalar akış diyagramı gösterilmektedir.



Şekil 2.15: Genetik algoritmalar akış diyagramı.

2.1.3.5. Lojistik Regresyon

Analiz edilecek veri setindeki niteliklerden, bağımlı değişken ve bağımsız değişkenler sürekli (sayısal) ise bağımlı değişkenin bağımsız değişkenler ile aralarındaki ilişkinin belirlenmesi için doğrusal regresyon analizi kullanılır. Doğrusal regresyon analizinde nitelikler kategorik olamaz. Eğer nitelikler kategorik ise tahmin değişkeni ile bağımsız değişkenler arasındaki ilişkinin belirlenmesi için lojistik regresyon analizi kullanılır.

Lojistik regresyon analizinde amaç değişkenler arasındaki ilişkiyi en iyi tanımlayan bir model oluşturmaktır (Çokluk, 2010). İlişkinin matematiksel ifadesi Denklem 2.11 ve Denklem 2.12’de görüldüğü gibidir.

$$\log \left[\frac{P(y = 1)}{1 - P(y = 1)} \right] = \log \left[\frac{P(y = 1)}{P(y = 0)} \right] = \sum_{k=1}^p \beta_k x_k \quad (2.11)$$

$$P(y = 1) = \frac{\exp(\sum_{j=1}^k \beta_j x_j)}{1 + \exp(\sum_{j=1}^k \beta_j x_j)} \quad (2.12)$$

y : Bağımlı değişken

x_j : j. Bağımlı değişken

β_j : j. Bağımsız değişkene ait model parametresi

P(y=1)/P(y=0) : Odds Ratio değeri (görülme olasılığı ile görülmemeye olasılığının oranı)

Lojistik regresyonun katsayı kestirimi şu yöntemlerle bulunabilir (Dolgun ve Ersel, 2014):

- En çok olabilirlik (EÇO)
- Ağırlıklı İteratif En Küçük Kareler (AİEKK)
- Minimum Lojit Ki-Kare (ML-X²)

2.1.3.6. Bayes Ağları

İstatistiksel sınıflama modellerinden biridir. Veri setindeki her bir kaydın hangi sınıfa ait olduğunu olasılıksal olarak hesaplar ve grafiksel modeller.

Bayes grafiklerinde düğümler raslantı değişkenlerini gösterirken düğümler arası bağlar raslantı değişkenleri arasındaki olasılıksal bağımlılık durumlarını göstermektedirler.

Bayes ağları “Yönlü Dönüşsüz Grafik” (Directed Acyclic Graph [DAG]) adı verilen sezgisel olarak anlaşılabilir bir grafik yapısına sahiptirler. Bu grafik yapısı ayrıca istatistik, makine öğrenmesi ve yapay zeka alanlarında da sıklıkla kullanılmaktadır (Dolgun ve Ersel, 2014).

Biçimsel olarak bayes ağları iki bölümden oluşmaktadır $B=(G,\Theta)$. İlk bölüm olan G düğümleri ve yayları içeren bir DAG’dır. Düğümler veri setindeki değişkenler ise yaylar değişkenler arasındaki doğrudan bağımlılıkları göstermektedir. İkinci bölüm olan Θ ise koşullu olasılıkların dağılımını temsil eder. $\theta_{X_i|\Pi_{X_i}}=P_B(X_i|\Pi_{X_i})$ fonksiyonunda görüldüğü üzere olası her X_i niteliği için X_i ’nin doğrudan bağımlı olduğu üst değişkenleri vermektedir. Π_{X_i} , G içinde X_i ’nin bağlı olduğu değişken kümesidir. B ağı sonrasında aşağıdaki denklemde belirtilen birleşik olasılık dağılımını denklem 2.13’te görüldüğü gibi temsil eder (Ouali ve diğ., 2006):

$$P_B(X_1, \dots, X_n) = \prod_{i=1}^n P_B(X_i | \Pi_{X_i}) = \prod_{i=1}^n \theta_{X_i | \Pi_{X_i}} \quad (2.13)$$

Bayes ağları sınıflayıcı kümesi $V=\{X_1, \dots, X_n, C\}$ ’dir (C : Sınıf değişkeni, X_1, \dots, X_n : Diğer değişkenler). Bayes sınıflayıcısı, B ağı $P(X_1, \dots, X_n, C)$ sonsal olasılıklarını modelleyen bir bayes ağı olmak üzere B modeli üzerinden $P_B(C|X_1, \dots, X_n)$ sonsal olasılığını maksimum yapan C sınıflayıcısını belirler. Denklem 2.14’de hesaplama işlemi görülmektedir (Neapolitan, 2003).

$$\arg \max_c P_B(C|X_1, \dots, X_n) \quad (2.14)$$

Bayes ağı sınıflandırması için kullanılan birçok yaklaşım mevcuttur. Bunlardan bazıları şu şekildedir:

- *Naive Bayes*
- *Tree Augmented Naive Bayes* (TAN)
- *BN Augmented Naive Bayes* (BAN)
- *Bayesian Multinet*
- Genel Bayesci Ağ (GBN)

2.1.4. Kümeleme

Kümeleme (*Clustering*) benzer verileri aynı küme içerisine dahil etmek için kullanılmaktadır. Kümeleme analizi yapılırken veriler arasındaki uzaklıklar kullanılmaktadır. En sık kullanılan uzaklık bağıntıları aşağıdadır (MacQueen, 1967):

1. Öklid Uzaklığı: İki boyutlu Pisagor teoreminin bir uygulamasıdır ve iki nokta arasındaki hesaplamanın yapılması için denklem 2.15'deki formül kullanılır.

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.15)$$

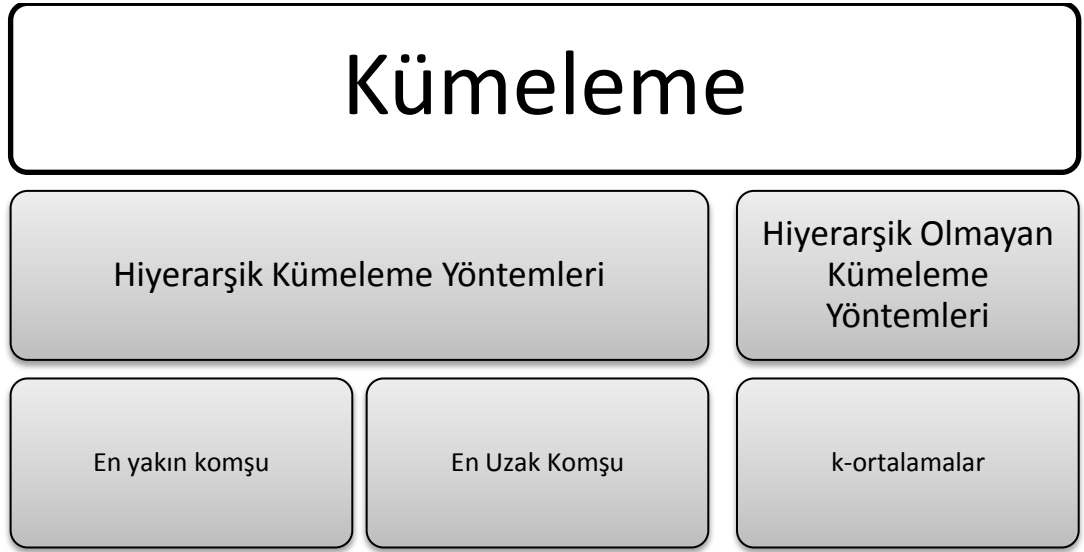
2. Manhattan Uzaklığı: Gözlemler arasındaki mutlak uzaklıkların toplamı alınarak bulunur. Bu hesaplama için denklem 2.16'daki formül kullanılır.

$$d(i, j) = \sum_{k=1}^p (|x_{ik} - x_{jk}|) \quad (2.16)$$

3. Minkowski Uzaklığı: p sayıda değişken baz alınarak gözlem değerleri arasındaki uzaklık hesaplaması yapılır. Bu hesaplama için denklem 2.17'deki formül kullanılır.

$$d(i, j) = \left[\sum_{k=1}^p (|x_{ik} - x_{jk}|^m) \right]^{\frac{1}{m}} \quad (2.17)$$

Birçok farklı kümeleme yöntemi mevcuttur. Şekil 2.16'da en çok kullanılan küme yöntemleri görülmektedir. Hiyerarşik kümeleme yöntemi, tek bir ana kümeden aşamalı olarak alt kümelere ayrılma ya da tek tek ele alınan kümelerin aşamalı olarak birleştirilmesi esasına dayanır. En sık kullanılan hiyerarşik kümeleme yöntemi algoritmaları en yakın komşu algoritması ve en uzak komşu algoritmasıdır. Hiyerarşik olmayan kümeleme yönteminde başlangıçta belirli sayıdaki kümeler için toplam ortalama hata minimize edilmektedir. Hiyerarşik olmayan kümeleme yöntemine örnek olarak k-ortalamlar yöntemi verilebilir (Han ve Kamber, 2012).



Şekil 2.16: En çok kullanılan kümeleme yöntemleri.

Kümeleme yönteminde aynı küme içerisindeki verilerin benzerlik oranlarının yüksek olması, farklı kümelerde yer alan verilerin ise benzerlik oranlarının düşük olması beklenmektedir. Aslında kümeleme analizinde de bir çeşit sınıflama yapılmaktadır. Ancak sınıflandırma yöntemlerinden farklı olarak kümeleme yapılacak veri setindeki niteliklerin dahil olacağı sınıflar önceden bilinmemektedir ve bir birleri arasındaki uzaklıklar hesaplanarak bir kümeye dahil edilirler.

2.1.5. Birliktelik Kuralları

Birliktelik kuralları (*association rules*) verilerin birlikte hareket etmesini analiz eden veri madenciliği yöntemidir. Analiz sonucunda oluşturulan kurallar neticesinde verilerin birlikte olup olmama durumlarını tespit eder.

Birliktelik kurallarına verilecek en klasik örnek sepet analizi uygulamasıdır. Bu analizde bir markette alışveriş yapan müşterilerinin alışveriş listelerinde ne gibi ürünleri birlikte almayı tercih ettiğine bakılır. Burada önemli olan nokta tercih etme oranının belli bir güven ölçütüne ve destek noktasına göre hesaplanmasıdır. Destek noktası, birlikte hareket eden ürünlerin bir alışveriş içerisinde ne kadar tekrarlandığıdır. Güven ölçütü ise x ürününü alan müşterilerin y ürününü de alma olasılığıdır. Destek ve güven ölçütleri denklem 2.18 ve denklem 2.19’da görüldüğü gibi hesaplanmaktadır (Inokuchi ve diğ., 1998):

$$destek(x \rightarrow y) = \frac{sayı(x, y)}{N} \quad (2.18)$$

$$Güven(x \rightarrow y) = \frac{sayı(x, y)}{sayı(x)} \quad (2.19)$$

$x \rightarrow y$: x ürünün alanların b ürününü de alma durumu

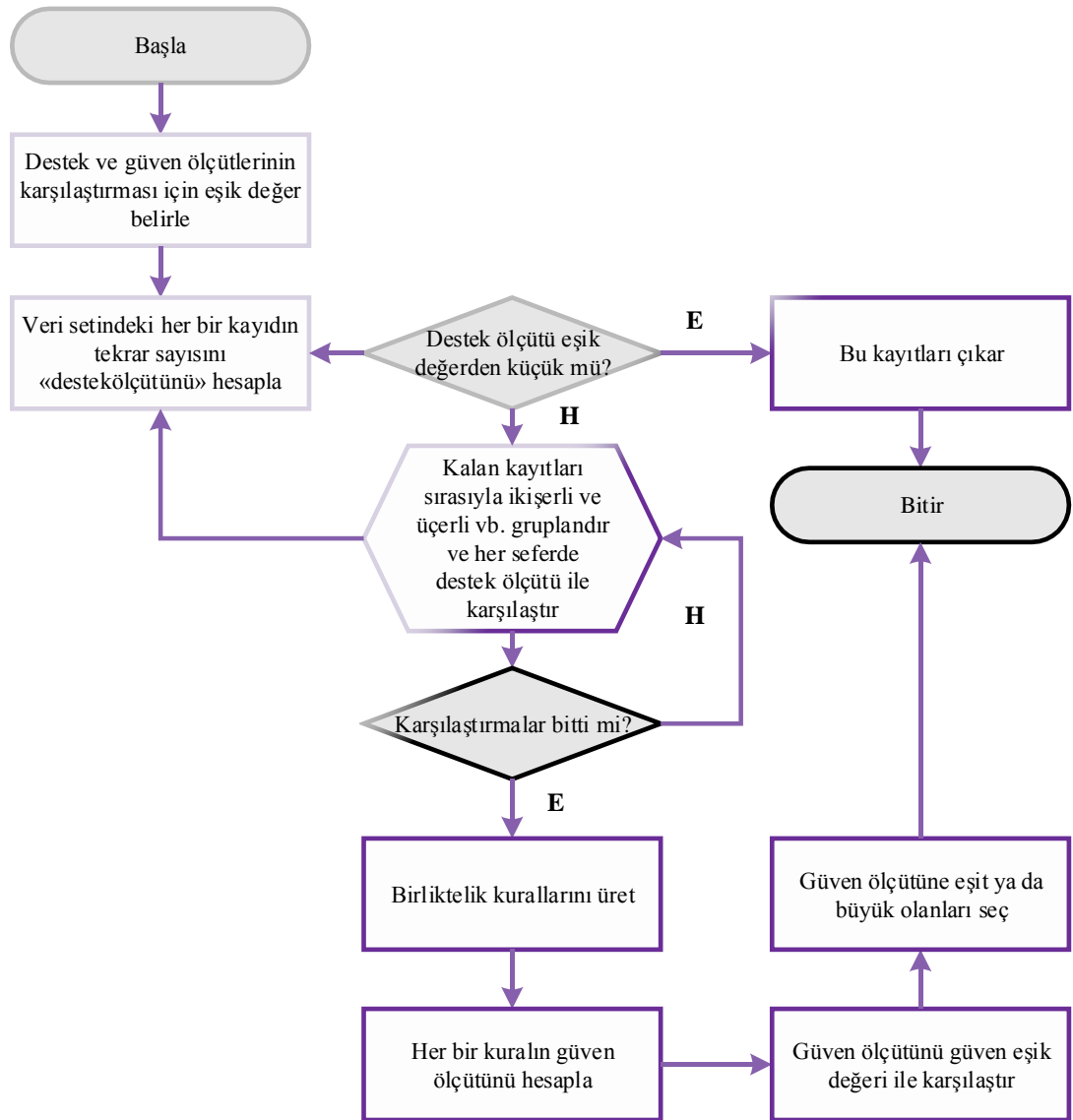
$sayı(x, y)$: x ve y ürünlerini birlikte içeren alışveriş sayısı

N : Tüm alışverişlerin sayısı

$sayı(x)$: x ürünü alanların sayısı

Destek ve güven ölçütlerinin hesaplamaları yapıldıktan sonra, analiz öncesinde belirlenen bir destek ve güven ölçütü eşik değeri ile karşılaştırılır. Sonuçlar eşik değerine ne kadar yakın ya da eşik değerinden ne kadar büyük ise analizin doğruluğu o oranda artmaktadır.

Birliktelik kuralları için en sık kullanılan algoritma apriori algoritmasıdır. Şekil 2.17’de apriori algoritmasının akış diyagramı görülmektedir.



Şekil 2.17: Apriori algoritması akış diyagramı.

2.1.6. Yorumlama ve Değerlendirme

Veri madenciliği sürecinin son aşaması yorumlama ve değerlendirme sürecidir. Bu süreçte araştırma problemine uygun veri madenciliği yönteminin uygulama sonuçları ve kurulan model ya da kurallar değerlendirilerek doğruluğu ve araştırma alanına katkısı yorumlanır.

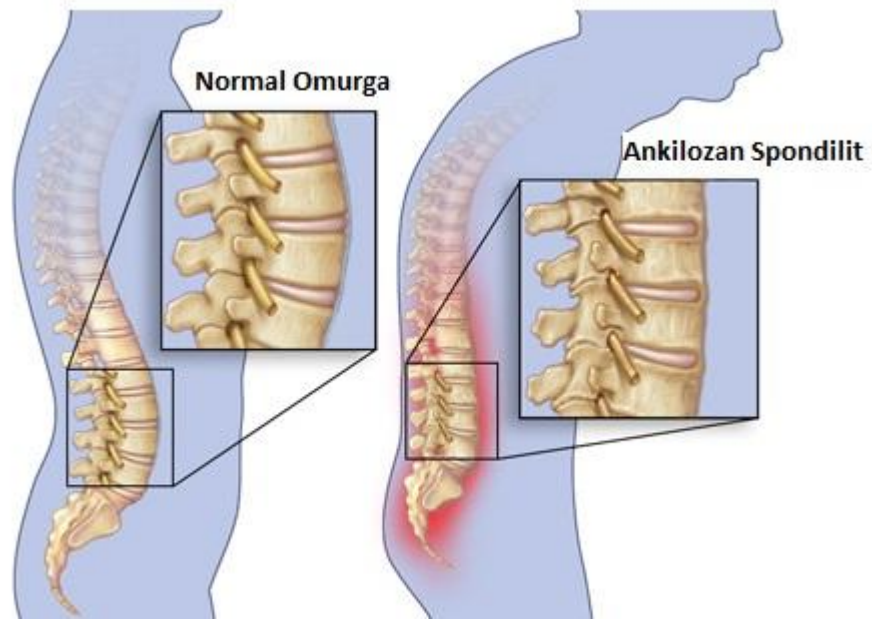
2.2. ANKİLOZAN SPONDİLİT

Spondiloartrit (SpA), sakroiliak eklemin, aksiyel omurganın, tendon ligaman yapışma bölgelerinin (Entezit) inflamasyonu (iltihaplanması) ve bazı hastalarda oligoartrit, döküntü veya inflamatuvar göz hastalığı (Üveit) ile karakterize olan bir grup romatizmal hastalıkları tanımlamak için kullanılan bir terimdir (Imboden ve diğ., 2014).

Başlıca SpA hastalıkları şunlardır:

1. Psöriatik artrit
2. Enteropatik artrit (İnflamatuvar barsak hastalığının artrit)
3. Reaktif artrit
4. Juvenil spondiloartrit
5. Ankilozan spondilit
6. Ayırt edilmemiş spondiloartritler

Ankilozan spondilit (AS), sakroiliak eklemler, omurga, periferik eklemler ve entezis bölgeleri ile karakterize edilen iltihaplı (inflamatuvar) seyreden romatizmal bir hastalıktır (Neuenschwander ve Ciurea, 2014). AS, aksiyel iskeleti etkileyen, eklemleri tutan, inflamatuvar bel ağrısına yol açan, yapısal ve fonksiyonel bozulmaya sebep olan, yaşam kalitesini azaltabilen bir hastalıktır (Orhan, 2011). Eklem ve omurganın hareketliliğini etkileyen ankilozan spondilit, fiziksel engelliliğe de yol açabilir (Lim ve diğ., 2005). Ankilozan spondilit hastalığına sahip kişilerde hastalığın doğal seyrine yönelik özellikler şu şekilde sıralanabilmektedir; başlangıcı sinsi bir hastalıktır, hastalığın ilerleyişi ya şiddetli ya da hafif olmaktadır, omurgadaki kısıtlama ve deformasyon zaman içerisinde gerçekleşmektedir, gözlerde iltihaplanma oluşup tekrarlayabilir, AS’de semptomlar 16 yaşından önce başlar ise hastalık şiddetli seyretmektedir. Cinsiyete göre ise hastalık kadınlarda erkeklere oranla daha hafiftir (Caretta ve diğ., 1983). Şekil 2.18’de AS hastalığı taşıyan omurga ile normal bir omurga arasındaki fark görülmektedir.



Şekil 2.18: Normal omurganın ve ankilozan spondilit hastalığında omurganın yapısı³.

Ankilozan spondilit hastalığına sahip bireylerde hastalığın seyri farklı olmaktadır. AS tanısı konan hastaların, omurga hareketliliğinin gelecekteki kısıtlanması konusunda endişe duymalarına karşın Ankilozan spondilit’li hastaların yaklaşık %40’ında hastalığın başlangıcından onlarca yıl sonra ciddi fonksiyon kaybına maruz kaldığı görülmektedir. Radyografik ilerleme hızı, benzer sürelerde bu hastalığa sahip hastalar arasında büyük ölçüde farklılık göstermektedir. Bu ilerlemeye hangi faktörlerin bağlı olduğuna dair bilgi bugün için azdır (Atagündüz, ve diğ., 2010). Hastalık şiddeti ilişkili HLA-B27 sıklığı ülkemizde Kuzey Amerika ve Avrupa’ya oranla daha düşük olarak bildirilmiştir (Günel ve diğ., 2008).

2.2.1 Epidemiyolojisi

AS hastalığının epidemiyolojik olarak literatür taraması sonucunda ulaşılan yaş, cinsiyet, prevalans bilgileri aşağıda sunulmaktadır.

Yaş:

Genç nüfusu etkileyen bir hastalıktır. Sıklıkla yaşamın 2. veya 3. yarısında başlar. Başlangıcı çocukluk çağına dayanan AS’li bireylerde kalça tutulumu ve periferik tutulum daha sıklıkla görülmektedir (Tüylü, 2013).

³<http://www.mayoclinic.org/diseases-conditions/ankylosing-spondylitis/multimedia/ankylosing-spondylitis/img-20008220>

Cinsiyet:

Erkek bireylerin AS hastalığına yakalanma olasılıklarının kadınlara oranla 3 kat fazla olduğu görülmektedir. Hasta popülasyonu arasında erkek hastaların yüzdeleri % 65-80 aralığında, coğrafi konuma göre değişmektedir (Alman romatolojik veri tabanında % 68.9, Fransa da yapılan bir çalışmada % 78.3). Ayrıca erkek hastalarda hastalığının klinik tablosunun daha ağır seyrettiği bilinmektedir (Sieper ve diğ., 2002).

Prevalans:

AS prevalansı beyaz Amerikalı ırkta % 0.2, Kuzey Avrupalı ırkta % 1.4 olarak bildirilmiştir. AS prevalansı coğrafi bölgeye göre farklılıklar göstermektedir (Özdemir, 2014).

Ülkemizde AS prevalansı için yapılan çalışmalarda şu gibi sonuçlar bulunmuştur:

İzmir ve çevresinde yapılan bir çalışma da (Onen ve diğ., 2008) AS prevalansı % 0.49 olarak bulunmuştur. Karadeniz bölgesinde yapılan bir çalışmaya (Karkucak ve diğ., 2011) göre AS prevalansı % 0.25, erkeklerin kadınlara oranla daha fazla AS hastalığına sahip olduğu bulunmuştur.

HLA-B27 gen pozitifliği AS hastalığının prevalansını etkileyen önemli kriterlerden biridir. Bölgesel ve coğrafi konum farklılıklarına göre değişmekle birlikte, HLA-B27 geni pozitif olanlar bireylerde AS hastalığının görülme olasılığı % 1-2 civarındadır (Güllüoğlu, 2013).

2.2.2. Klinik Özellikler

AS hastalarının %75'inde inflamatuvar bel ağrısı ilk belirtidir. Bu ağrının özellikleri ilk defa Calin ve diğ. tarafından (1977) tanımlanmıştır. Calin kriterlerine göre inflamatuvar bel ağrısı için en tipik klinik belirtiler arasında; genç bireyleri özellikle 40 yaşın altındaki bireyleri etkilemesi, sinsi bir başlangıcının olması, üç aydan daha uzun süren tanımlanmış bir bel ağrısının olması, sabah tutukluğu olması ve ağrının dinlenme ile artması ve egzersiz ile azalması sayılabilir. Bu belirtilerin en az dördünü taşıyan hasta da inflamatuvar bel ağrısı olduğu düşünülmektedir (Kınıklı, 2014).

AS hastalığının tanı kriterleri için, Roma Kriterleri (Kellgren ve diğ., 1961), *Modifiye Newyork* kriterleri (Linden ve diğ., 1984) ve son olarak ASAS (*Assessment of SpondyloArthritis International Society*) kriterleri (Rudwaleit ve diğ., 2009) Tablo 2.1’de gösterildiği gibidir (Dönmez, 2014):

Tablo 2.1: AS hastalık kriterleri.

ROMA KRİTERLERİ
<p>Klinik kriterler:</p> <ol style="list-style-type: none"> 1- 3 aydan uzun süre var olan ve istirahat ile düzelmeyen bel ağrısı ve tutukluğu 2- Torakal bölgede ağrı ve tutukluk 3- Lomber omurgada hareket kısıtlılığı 4- Göğüs ekspansiyonunda kısıtlılık 5- İritis veya sekelinin öyküsü/bulgusu <p>Radyolojik kriterler:</p> <ol style="list-style-type: none"> 6- Bilateral ankilozan spondilite özgü sakroiliak eklem değişiklikleri
MODİFİYE NEWYORK KRİTERLERİ
<ol style="list-style-type: none"> 1- En az 3 aydır var olan, egzersiz ile düzelip istirahat ile düzelmeyen bel ağrısı 2- Lomber omurganın sagittal ve frontal düzlemde hareket kısıtlılığı 3- Göğüs ekspansiyonunun yaş ve cinse göre normal değerlerin altında olması 4- Evre 2-4 bilateral sakroiliit 5- Evre 3-4 unilateral sakroiliit <p>Kesin ankilozan spondilit: Klinik kriterlerden herhangi birisi ile birlikte unilateral evre 3-4 veya bilateral evre 2-4 sakroiliit</p>
ASAS KRİTERLERİ
<p>Bel ağrısı süresi >3 ay olanlar ve başlangıç yaşı <45 yaş olan hastalarda;</p> <p>Görüntülemelerde sakroiliit + ≥ 1 SpA bulgusu veya HLA-B27 pozitifliği + ≥ 2 SpA bulgusu</p> <p>(SpA bulguları: inflamatuvar bel ağrısı/ artrit/ entezit/ üveit/ daktilit/ psöriazis/ chron-kolit/SOAIİ’lere iyi yanıt/ SpA için aile öyküsü/ HLA-B27 pozitifliği/ CRP yüksekliği)</p> <p>(Görüntülemelerde sakroiliit: MRG’da aktif (akut) inflamasyon SpA ile ilişkili sakroiliit için oldukça anlamlı / Modifiye New York kriterlerine göre kesin radyografik sakroiliit)</p>

2.2.3. Tedavi

AS hastalığında tedavinin amaçları şu şekilde sıralanabilir (Dönmez, 2014):

1. Hastanın şikayet ettiği yaşamını zorlaştıran ağrı, tutukluk gibi semptomları azaltmak
2. Eklem hareketliliğini korumak
3. İnflamasyona bağlı spinal deformasyonu engellemek
4. Hastanın günlük aktivitelerini sürdürmesini sağlamak
5. Eklemlerde meydana gelebilecek sistemik komplikasyonları önceden belirleyebilmek

AS'de tedavi hastanın eğitimi, fizik tedavisi, farmakolojik tedavi ve gerekli ise cerrahi tedavi olacak şekilde planlanmalıdır (Yazıcı ve diğ., 2004):

1. Fizik Tedavi: Hastaların postürünün korunması için, hasta fizik tedavi almalıdır. Yüzmek, yürümek gibi aktiviteler de postürün korunması için faydalıdır.
2. Farmakoloji: Hastalığın seyri boyunca kullanılan farmakolojik ilaçlar fizik tedavinin yanında hastalıktan doğan ağrının giderilmesi ve inflamasyonun azaltılması için kullanılırlar. Bu ilaçlar ilerleme evrelerine göre, steroid olmayan antiinflatuar ilaçlar, antiromatizmal ilaçlar, analjezik ilaçlar, kortikosteroidler, anti-Tümör nekrosis faktör (TNF) olarak gruplandırılırlar.
3. Cerrahi: Cerrahi müdahaleler genellikle ilerlemiş vakalarda deformasyonun veya fonksiyon kayıplarının düzeltilmesi için uygulanır. En sık yapılan cerrahi müdahale kalça eklemine yenilenmesidir. Bunun yanı sıra spinal kamburluğu düzeltmek, diz eklemlerinin yenilenmesi, servikal ve lomber osteotomi gibi ameliyatlar da uygulanmaktadır.

2.2.4. Prognoz

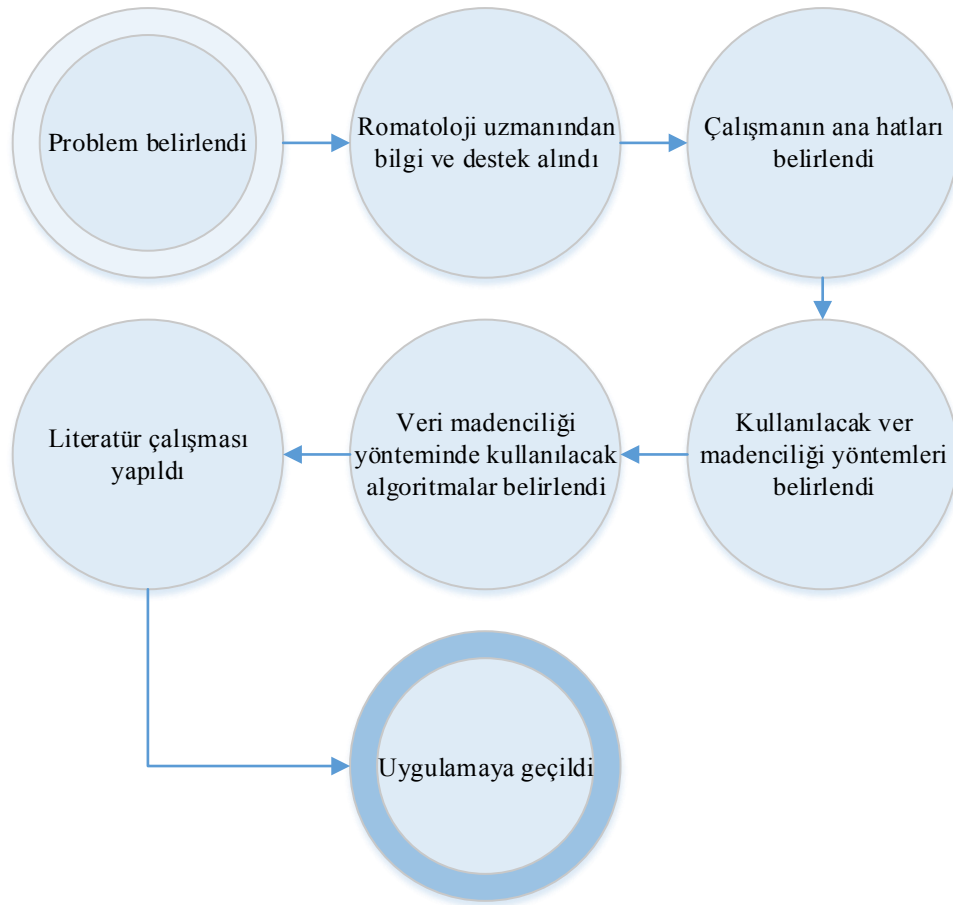
AS hastalığının seyri kişiden kişiye değişiklik göstermektedir. Bazı hastalarda ilerleme hızlı olurken bazı hastalarda çok daha yavaş bir seyir gösterebilmektedir. Hastalar fizik tedaviye devam ettiği ve ilaçlarını düzgün kullandığı sürece bir takım değişiklikler ile hayatlarını idare ettirebilirler.

Kanda romatoid faktör pozitifliği, eklemlerde tutulum, sistemik hastalık varlığı, ESH ve CRP yüksekliği, HLA-B27 gen pozitifliği, genç ve kadın hastalar, diz, dirsek ve omuz

gibi büyük eklemlerde tutulum, AS’de kötü prognoza etki eden bazı temel göstergelerdir (Öztürk, 2014).

3. MALZEME VE YÖNTEM

Tezin uygulama adımları anlatılmadan önce tez çalışması içerisinde uygulama kısmına kadar gerçekleştirilen çalışma adımları Şekil 3.1’deki gibi gerçekleşmiştir.



Şekil 3.1: Tez çalışmasında uygulama kısmına kadar gerçekleştirilen adımlar.

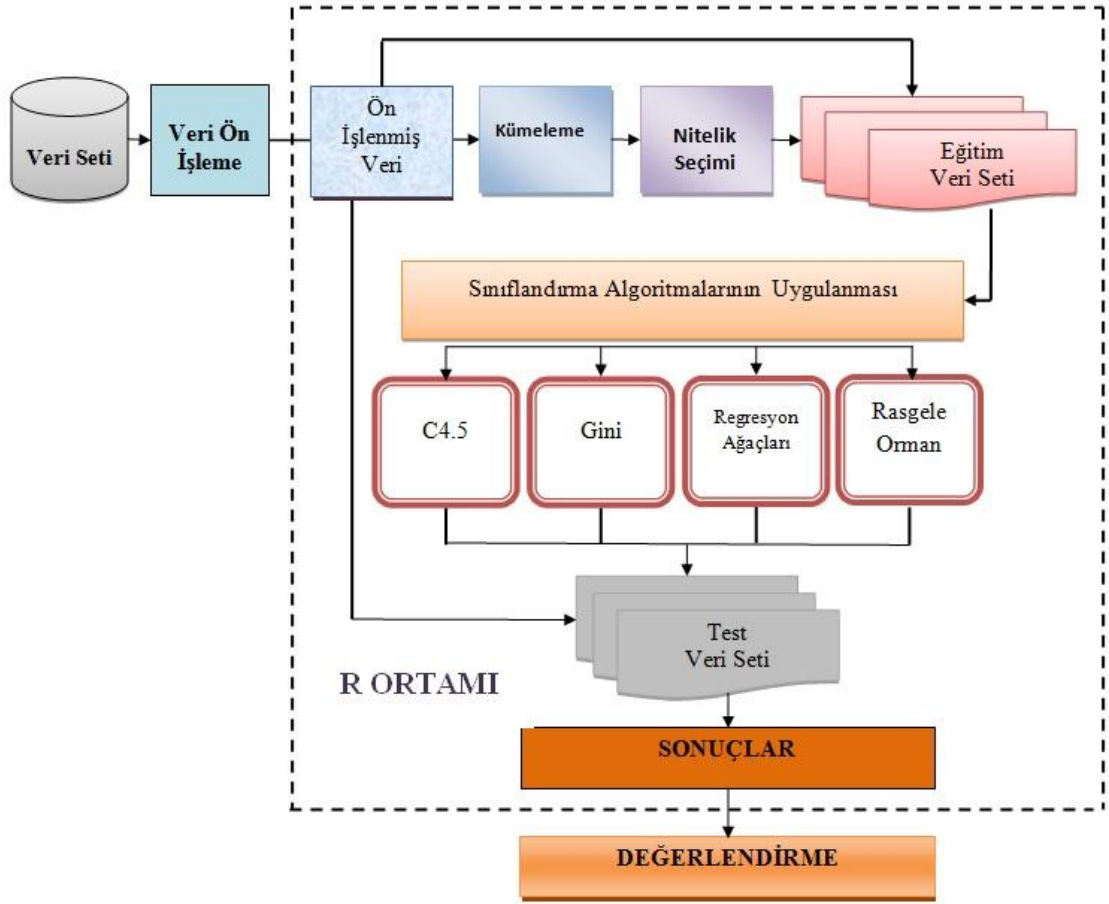
Tez çalışmasında, Marmara Üniversitesi Tıp Fakültesi İç Hastalıkları Anabilim Dalı, Romatoloji Bilim Dalı ile ortak çalışarak, bu klinikte AS tanısı almış hastalarının bilgilerinden oluşan veri seti, veri madenciliği yöntemleri ile analiz edilmiştir. Hasta grubu, Romatoloji Polikliniği’nde takipleri devam eden hastalardan oluşmaktadır. Bu çalışmanın amacı, veri madenciliği sınıflandırma yöntemlerinden karar ağacı algoritmalarını inceleyerek R programında Gini, C4.5, regresyon ağaçları ve rasgele orman algoritmaları kullanılarak AS hastalığı taşıyan bireylerin verilerinin analizini

yaparak, hastalığın radyografik ilerleyişinin bağlı olduğu faktörlerden oluşan bir karar yapısı oluşturmaktır.

Oluşturulacak model için seçilen veri madenciliği yöntemi, sınıflandırma yöntemidir. Sınıflandırma yönteminin seçilmesinin başlıca sebepleri şunlardır:

- Gelecekte veri setine eklenecek hastaların, hastalık seyrini tahmin edebilmek için kurallar oluşturmak
- Belirli alt gruplara özgü ilişkiler tanımlamak
- Belirli bir sınıfın üyesi olacak elemanları belirlenmek

Veri seti öncelikle ön işleme işlemlerine tabi tutulmuştur. Ön işleme sonucunda oluşturulan veri seti R programına aktarılması için “.csv” formatında olması gerekmektedir. Bu yüzden Excel ortamındaki veri seti “.csv” (*comma-separeted variables*: virgülle ayrılmış değerler) formatına çevrilmiştir. Veriler Ek 1’de sunulmuştur. R ortamına aktarılan veri seti üzerinde sırasıyla; kümeleme analizi ve nitelik seçimi işlemleri yapılmıştır. Bu işlemler sonrasında veri setinin üçte ikilik kısmı eğitim verisi olarak üçte birlik kısmı ise test verisi olarak ikiye ayrılmıştır. Eğitim verisi üzerine kullanılacak algoritmalar uygulanmıştır. Test verisi üzerinde sonuçlar doğrulanmıştır ve son olarak değerlendirmesi yapılmıştır. Şekil 3.2’de uygulama adımları genel olarak belirtilmektedir.



Şekil 3.2: Sınıflandırma performanslarını ölçmek için kullanılan deneme bileşenleri diyagramı (Ragab ve diğ., 2014'ten değiştirilmiştir).

3.1. VERİ SETİ

Çalışmaya dahil edilen hastaların, poliklinik takip dosyalarındaki bilgiler SPSS programından Microsoft Excel programına aktarılmıştır. Veri seti temel olarak 4 gruptan oluşmaktadır:

1. Demografik bilgiler: Yaş, cinsiyet, iletişim bilgileri, doğum yeri, doğum tarihi
2. Genetik bilgiler: HLA-B27 geninin pozitif-negatif bilgisi, polimorfizm türü ile ilgili bilgiler
3. Semptomlar: Hastanın AS dışında sahip olduğu hastalıklar ya da şikayetler ile ilgili bilgiler
4. İlaç bilgileri: Hastalık ile ilgili bazı ilaçları kullanması ya da hala kullanıyor olması ile ilgili bilgiler

Veri seti ön işleme gerçekleştirilmeden önce 256 kayıttan oluşmaktayken, ön işleme sonrasında analize dahil edilen kayıt sayısı 225'tir. Aynı şekilde veri seti ön işleme gerçekleştirilmeden önce 59 nitelikten oluşmakta iken ön işleme işlemleri sonrasında 23 kategorik, 1 lineer (tam sayı) olmak üzere toplamda 24 nitelikten oluşmaktadır. Veri setindeki her bir satır bir hastaya ait klinik, genetik ve demografik bilgileri içermektedir. Veri setinin boyutu nedeni ile bir kısmının görseli örnek olarak Şekil 3.3'de sunulmaktadır.

	A	B	C	D	E	F	G	H	I	J	K	L
1	ERAPrs30187	rs30187Talel	PTGER4rs10440635	rs10440635Aalel	IL23rs11209032	rs11209032Aalel	ERAPrs27044	rs27044Galel	Sex	İlksemptonkategorik	BaslangicYasiKategorik	HLAB27
2	TC	Var	AA	Var	AG	Var	CC	Yok	Erkek	Aksiyal	Orta	Negatif
3	TC	Var	AA	Var	AG	Var	CC	Yok	Erkek	Aksiyal	Orta	Pozitif
4	CC	Yok	GG	Yok	AA	Var	CC	Yok	Erkek	EkstraArt	Orta	Pozitif
5	CC	Yok	AG	Var	AG	Var	CC	Yok	Erkek	PeriferikAksiyal	Orta	Pozitif
6	TC	Var	GG	Yok	AG	Var	CC	Yok	Erkek	PeriferikAksiyal	Orta	Negatif
7	TC	Var	GG	Yok	GG	Yok	CC	Yok	Erkek	PeriferikAksiyal	Orta	Pozitif
8	TC	Var	AG	Var	AA	Var	CC	Yok	Erkek	Aksiyal	Orta	Pozitif
9	TC	Var	AA	Var	AA	Var	CC	Yok	Erkek	PeriferikAksiyal	Yasli	Pozitif
10	TC	Var	AA	Var	AG	Var	CC	Yok	Erkek	PeriferikAksiyal	Orta	Pozitif
11	TC	Var	AG	Var	AG	Var	CC	Yok	Erkek	Aksiyal	Orta	Negatif
12	CC	Yok	AG	Var	AG	Var	CC	Yok	Kadin	Aksiyal	Orta	Negatif
13	TC	Var	AA	Var	AA	Var	CC	Yok	Kadin	Aksiyal	Orta	Negatif
14	TC	Var	AG	Var	GG	Yok	CC	Yok	Kadin	Aksiyal	Orta	Negatif
15	CC	Yok	AG	Var	AA	Var	CC	Yok	Kadin	Aksiyal	Orta	Negatif

Şekil 3.3: Ön işlem sonrası veri setinin bir kısmının genel görünümü.

Veri setinde yer alan toplam nitelik sayısı 59'dur. Ön işlemeden sonra bu sayı 24'e indirilmiştir. Ön işleme sonrasında analiz için seçilen nitelikler ve açıklamaları Tablo 3.1'de gösterildiği gibidir:

Tablo 3.1: Veri setinde analiz için seçilen nitelikler ve açıklamaları.

NİTELİK ADI	AÇIKLAMASI
1 ERAPrs30188	Endoplazmik retikulum aminopeptidaz 1 genindeki rs30187 tek nükleotid polimorfizminin genotipi (CC, TC, TT)
2 rs30187Talel	ERAP1'in rs30187 polimorfizmi T aleli içeriyorsa riskli kabul edilir (genotip CT veya TT ise var, CC ise yok)
3 PTGER4rs10440636	Prostaglandin E2 (PGE2) reseptör tip4 genindeki rs10440635 tek nükleotid polimorfizminin genotipi (AA, AG, GG)
4 rs10440635Aalel	PTGER4'ün rs10440635 polimorfizmi Aaleli içeriyorsa riskli kabul edilir (genotip AA veya AG ise var, GG ise yok)
5 IL23rs11209033	İnterlökin 23reseptör geninin rs11209032 tek nükleotid polimorfizminin genotipi (AA, AG, GG)

Tablo 3.1 (devam):

6	rs11209032Aalel	IL23'ün rs 11209035 polimorfizmi A aleli içeriyorsa riskli kabul edilir (genotip AA veya AG ise var, GG ise yok)
7	ERAPrs27045	Endoplazmik retikulum aminopeptidaz 1 genindeki rs27044 tek nükleotid polimorfizminin genotipi (CC, CG, GG)
8	rs27044Galel	ERAP1'in rs27044 polimorfizmi G aleli içeriyorsa riskli kabul edilir (genotip CG veya GG ise var, CC ise yok)
9	AgirVShafif	mSASSS radyografik skorlama (Veri setinde mSASSS değeri 20 ve üstü hastalar ağır hasta olarak, 20'den küçük olan hastalar hafif hasta grubu olarak kabul edilmiştir)
10	Sex	Cinsiyet (Kadın-Erkek)
11	Ilksemptomkategorik	İlk Semptomun ne olduğu kategorik belirtilmiştir: 1: Aksiyal 2: Periferik Artrit 3: Periferik + Aksiyal 4: Ekstra Artiküler
12	BaslangicYasiKategorik	2015 yılına göre hastaların yaşları hesaplanarak; 20 yaşından küçük olan hastalar "Genc", 20-40 yaş arasında olan hastalar "Orta", 40 yaş ve üzeri olan hastalar ise "Yasli" olarak belirtilmiştir.
13	HastalikSuresi	Tanıdan itibaren geçen süre
14	HLAB27	HLA-B27 geninin pozitif - negatif olma durumu
15	KalcaTutulumu	Kalça tutulumu hastalık esnasında hiç geçirdi mi? (Var - Yok)
16	PerArt	Periferik Artirit hastalık esnasında hiç geçirdi mi? (Var - Yok)
17	ServTutulum	Servikal tutulum hastalık esnasında hiç geçirdi mi? (Var- Yok)
18	ExtraArtEver	Ekstraartiküler tutulum hastalık esnasında hiç geçirdi m? (Var - Yok)
19	Uveit	Uveit hastalık esnasında hiç geçirdi mi? (Var- Yok)
20	Entezit	Entezit hastalık esnasında hiç geçirdi mi? (Var- Yok)
21	NSAID	Non steroid antienflamatuar ilaçlar
22	CurrentTNF	TNF ilaç grubunu hala kullanıyor mu (Evet - Hayır)
23	TNFi	Tümör nekrozis faktör alfa inhibitör tedavi alanlar
24	LEF	Leflunomid tedavisi alanlar

3.2. ANALİZ İÇİN HAZIRLANAN VERİ SETİNDEKİ NİTELİKLERİN ÖZETİ

Veri madenciliği adımlarına geçmeden önce R programında verilerin özet istatistikleri belirlendi. Verilerin özet istatistiklerinin belirlenmesi için R’de yazılan kod aşağıdaki gibidir:

```
>veri<-read.table("C:/VS/AS.csv", sep=";", header=TRUE)

>summary(veri)
```

Yukarıdaki kodda belirtildiği gibi C diskinde VS isimli klasörün altında .csv uzantılı AS isimli dosya içerisindeki veri seti veri isimli bir değişkene atanmıştır. Veri setinin özetini vermek için summary() fonksiyonu kullanılmıştır. Tablo 3.2.’de summary(veri) sonuçları görülmektedir.

Tablo 3.2: Veri setinde yer alan niteliklerin özeti.

Değişken adı	Değişkenlerde yer alan değerler	Özeti
ERAPrs30187	CC	72
	TC	118
	TT	35
rs30187Talel	Var	153
	Yok	72
PTER4rs10440635	AA	51
	AG	114
	GG	60
rs10440635Aalel	Var	165
	Yok	60
IL23rs11209032	AA	54
	AG	103
	GG	68
rs11209035Aalel	Var	157
	Yok	68
ERAPrs27044	CC	109
	CG	102
	GG	14
rs27044Galel	Var	116
	Yok	109
Sex	Erkek	132
	Kadin	93

Tablo 3.2 (devam):

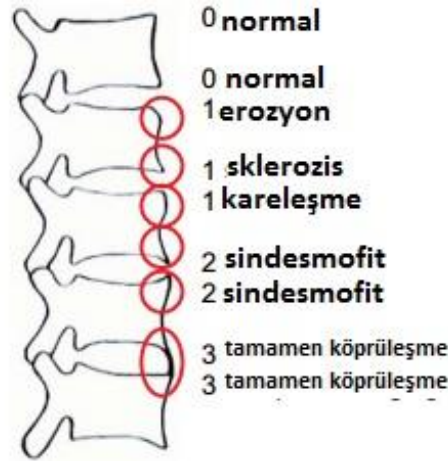
İlkseptomkategorik	Aksiyal	134
	EkstraArt	1
	PeriferikAksiyal	76
	PeriferikArtrit	14
BaslangicYasiKategorik	Genc	39
	Orta	177
	Yasli	9
HLAB-27	Negatif	90
	Pozitif	135
Kalcitutumumu	Var	61
	Yok	164
Periferikartrit	Var	104
	Yok	121
Servikaltutulum	Var	105
	Yok	120
Ekstraartever	Var	90
	Yok	135
Uveit	Var	35
	Yok	190
Entezit	Var	61
	Yok	164
NSAID	Evet	214
	Hayir	11
LEF	Evet	3
	Hayir	222
CurrentTNFi	Evet	99
	Hayir	126
TNFi	Evet	104
	Hayir	121
Hastaliksuresi	Min	1000
	1st Qu.	4000
	Median	7000
	Mean	8182
	3rd Qu.	10000
	Max	49000
AgirVShafif	Agir	62
	Hafif	163

3.3. SINIF NİTELİĞİNİN BELİRLENMESİ

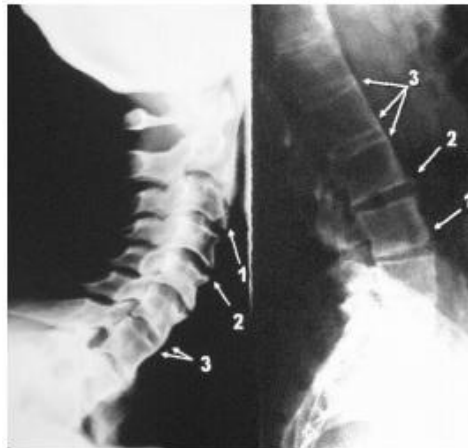
Veri setinde “AgirvsHafif” olarak belirtilen nitelik sınıf niteliğidir. Sınıf niteliği hastaların mSASSS (*Modified Stoke Ankylosing Spondylitis Spinal Score*) denilen

radyografik skor değerinin kategorik karşılığıdır. mSASSS değeri omurlar arası kaynaşmanın derecesini ifade eder. Bu değer (0-72 aralığında) uzmanlar tarafından verilmektedir. Kaynaşma derecesi hastanın radyografik çekimi sonucunda belli olmaktadır.

Sağlıklı bir insanda omurlar arası mesafe sabittir. AS'li bireylerde omurgaların birbirine yaklaşmasıyla en düşük derecede aşınma, daha sonra sertleşme, kareleşme, ince kemik köprülerin oluşması ve son safhada tamamen köprüleşme ya da kemikleşme oluşur. Kaynaşmanın başlaması ve ilerleyişi Şekil 3.4'de görüldüğü gibi derecelendirilir. Şekil 3.5'de radyografik olarak derecelendirme görülmektedir.



Şekil 3.4: Omurganın kemikleşme dereceleri⁴.



Şekil 3.5: Omurganın kemikleşme derecelerinin radyografik görüntülenmesi⁴.

⁴ <http://www.asif.info/en/index.cfm/assessment-forms/modified-stoke-ankylosing-spondylitis-spinal-score-m-sasss/msasss-english/>

	0	1	2	3
C2				
C3				
C4				
C5				
C6				
C7				
T1				
T12				
L1				
L2				
L3				
L4				
L5				
S				

x 0 =	0
x 1 =	
x 2 =	
x 3 =	

Check: sum must be 24

total = m-SASSS =

Şekil 3.6: Omurgadaki bölgelere göre mSASSS skorunun hesaplama tablosu⁵.

Şekil 3.6’da yer alan omurların ait olduğu bölgeler aşağıda belirtildiği gibidir:

1. C2, C3, C4, C5, C6, C7, T1: Servikal bölge omurları
2. T12, L1, L2, L3, L4, L5: Lomber bölge omurları
3. S: Sakral bölge omurları

Omurgada yukarıda karşılık gelen alanlara skor yazılarak her bir skorun toplam değeri belirlenerek toplamdaki mSASSS değeri hesaplanır.

Çalışmada kullanılan veri setinde konunun uzmanları tarafından yapılan değerlendirme sonucunda mSASSS değeri 20 ve üzeri olan hastalar ağır hasta grubunda, 20’nin altında olan hastalar ise hafif hasta grubunda kabul edilmiştir. Sayısal olan nitelikler kategorik değere dönüştürülmüştür. Veri setinde 225 hastanın 62 tanesi ağır grup hasta 163 tanesi hafif grup hasta’dır.

⁵ <http://www.asif.info/en/index.cfm/assessment-forms/modified-stoke-ankylosing-spondylitis-spinal-score-m-sasss/msasss-english/>

3.4. VERİ ÖN İŞLEME

Veri seti analize dahil edilmeden önce ön işleme sürecine tabi tutulmuştur. Veri setinde, tek bir alan içerisinde birden fazla değer girişi, giriş yapılmamış hücreler, kısaltılmış tıbbi terimler, alan adlarının gereğinden uzun veya yanlış yazılması, aynı kaydın farklı hücrelerde farklı şekilde adlandırılması gibi analize uygun olmayan girişlere rastlanmıştır. Bu sebeplerden dolayı veri seti içerisindeki eksiklikler ve hatalar giderilmiş ve şu şekilde düzeltilmiştir:

1. **Veri Temizleme:** veri temizleme aşamasında aşağıda belirtilen niteliklere sahip kısımlar silinmiştir:
 - Hasta gizliliği ve mahremiyeti ilkesi göz önünde bulundurularak hastalara ait özel bilgileri içeren kısımlar (ad, soyad, iletişim bilgileri)
 - Boş bırakılan ya da tekrarlı satırlar
 - Yeterince bilgi içermeyen sütunlar
2. **Veri Birleştirme:** Bazı sütunlar (nitelikler) birleştirilerek yeni ve veri madenciliği analiz algoritmaları için daha değerli hale getirilmiştir.
3. **Veri Dönüşümü:** Veri dönüşümü gerçekleştirirken yapılan bazı işlemler aşağıda belirtildiği gibidir:
 - Hastanın çeşitli semptomlar gösterip göstermemesi durumuna göre; gösteren hastalar var, göstermeyen hastalar yok olarak belirtilmiştir.
 - Hastalığın ağır veya hafif şiddette olmasını ağır hastalar “Ağır”, hafif hastalar “Hafif” olarak belirtilmiştir.
 - HLA-B27 geni için; gen aktif olan hastalar pozitif, olmayan hastalar negatif olarak belirtilmiştir.
 - Hastalığın başlangıç yaşı genç, orta ve yaşlı olarak kategorik hale dönüştürülmüştür. Bu dönüşüm yapılırken 20 yaşından küçük olan hastalar “Genc”, 20-40 yaş arasında olan hastalar “Orta”, 40 yaş ve üzeri olan hastalar ise “Yasli” olarak belirtilmiştir.
4. **Veri İndirgeme:** Veri indirgeme işlemi yaparken veri madenciliği yöntemlerinden kümeleme ve nitelik seçimi işlemleri yapılmıştır.

Ön işleme sonucunda veri analiz için en uygun hale getirilmiştir.

3.5. KÜMELEME ANALİZİ:

Sınıflandırma algoritmalarını uygulamadan önce, hedef niteliğimizle ne derece doğru kümelendiğine bakmak için kümeleme yöntemi kullanılmıştır

Kümele analizinde R paketlerinden Cluster paketi (Maecher ve diğ., 2015) ve paket içindeki Pam fonksiyonu kullanılmıştır. Kümeleme analizi için R’de yazılan kod aşağıda belirtildiği gibidir.

```
> setwd("C:/")

> require(cluster)

> library(cluster)

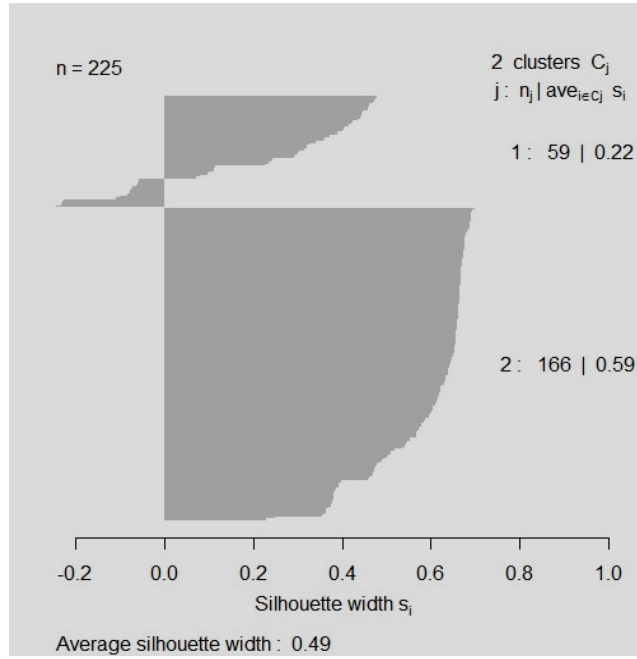
> veri<-read.table("C:/VS/AS.csv", sep=";", header=TRUE)

> y<-veri[, -c(1, 2, 11, 27)]

> Kume<-pam(veri=y, k=2, keep.diss=TRUE, keep.data=TRUE)

> plot(Kum, which.plots=2, main="")
```

Buna göre iki küme oluşturuldu ve hafif grup 2. küme, ağır grup 1. küme olarak belirlendi. 225 örnek üzerinde yapılan kategorik kümeleme yöntemi sonucunda 52 değer kendi kümelerine ait olarak bulunmadı. Yani 225 veride 52 hatalı kümeleme vardır ve 52/225 olduğundan, %23.1 olarak tutarsızlık saptanmıştır. %76.9 kümeleme tutarlıdır. plot fonksiyonu ile yapılan çizim Şekil 3.7’de görülmektedir.



Şekil 3.7: Kategorik kümeleme analizi sonuçlarının grafiksel görüntüsü.

3.6. NİTELİK SEÇİMİ

Nitelik seçimi işlemi, veri集中的 nitelikler içerisinde analize uygun olanların seçilmesi, analiz için daha değerli niteliklerin belirlenmesi için yapılmaktadır. Nitelik seçimi işleminde R paketlerinden FSelector (Kononenko, 1994) paketi ve cfs fonksiyonu kullanıldı. Nitelik seçimi için R’de yazılan kod aşağıda belirtildiği gibidir.

```
> setwd("C:/")
> veri<-read.table("C:/VS/ AS.csv", sep=";", header=TRUE)
> library(FSelector)
> nitelikler<-cfs(AgirVShafif~.,veri)
> show(nitelikler)
```

Nitelik seçimi işlemi sonucunda 4 nitelik çıkmıştır::

- "Sex": Cinsiyet niteliği
- "Servikaltutulum": Semptom olarak hastanın servikal tutulumu gösterip göstermediği bilgisi

- "CurrentTNFi": Şu anda TNF kullanıp kullanmama durumu
- "Hastaliksuresi": Hastalığın süresi

3.7. VERİ MADENCİLİĞİNİN UYGULANMASI

3.7.1. Uygulamada Kullanılan Algoritmalar

Tez kapsamında sınıflandırma algoritmaları kullanılarak en uygun modelin bulunması amaçlanmıştır. R programı üzerinde uygulama gerçekleştirilmiştir. Tez çalışmasında R programının seçilmesinin sebepleri şunlardır:

1. Open source-açık kaynak kodludur
2. Ücretsizdir
3. Kullanımı kolaydır
4. Analize uygun paketleri kendi verinize adapte etmek kolaydır
5. Akademik çalışmalar için uygundur
6. Dokümantasyonu çok iyi

Analize dahil edilen niteliklerin veri ön işleme aşaması gerçekleştirildikten sonra verinin üçte birlik kısmı test işlemleri için ayrılarak, geri kalan üçte ikilik kısma deneme verisinin özelliklerine uygun olarak belirlenen sınıflandırma algoritmalarından aşağıdakiler uygulanmıştır:

1. C4.5 Algoritması
2. Gini Algoritması
3. Regresyon Ağaçları
4. Random Forest

R programında kullanılan temel fonksiyonlar ve açıklamaları şu şekildedir:

install.pacages(): R programında bir paket yüklemek için kullanılmaktadır. Paket kurulumu yaparken parantez içerisinde kullanmak istenilen paket çift tırnak işareti (" ") içerisinde yazılarak kurulum gerçekleştirilir. Bir paketin kurulumunu bir kez yapmak o paketin R kütüphanesi içerisinde entegre olmasını sağlayacaktır. Paketi kullanmak istediğimiz her seferde paketi tekrar kurmaya gerek yoktur.

library(): R programında içerisinde komutların saklandığı bir kütüphane bulunmaktadır. Yüklenen paketlerdeki komutları kullanabilmek için R kütüphanesine

bağlanmak gerekmektedir. Programın her başlatılışında kurulu paketleri kullanmak için kütüphaneye bağlanılmalıdır. `library()` fonksiyonu R programında kütüphaneye bağlanmak için kullanılan fonksiyondur.

setwd(): Çalışılacak dizini belirlemek için `setwd()` komutu kullanılır. Örneğin C diskinden bir dosya çalışacaksak C diskini R ortamına tanıtmak gerekmektedir. Bunun için parantez içindeki kısma çift tırnak işareti (" ") içerisinde dizin ismi yazılmalıdır.

read.table(): Çalışılacak dosyayı tablo formatında okur ve veri tablolarını R’de kullanmak için *data frame*’e (veri çerçevesi) kaydeder.

plot(): Sonuçların çizimle gösterilmesi için kullanılan genel bir fonksiyondur.

cbind(): Matris, vektör ya da *data frame* olan değişkenlerin sırasıyla satır ve sütunlarını birleştiren genel bir fonksiyondur.

c(): Analiz edilecek veri seti üzerinde birleştirmek, bölmek, çıkarmak istediğimiz sütun veya satırları liste ya da vektör olarak göstermeye yarar.

set.seed(): Rasgele sayı üretmek için kullanılır. Fakat fonksiyonda tohum denilen bir başlangıç belirtilir. Bu başlangıç ile her seferinde aynı sonuçları almak için önemlidir.

3.7.1.1. C4.5 Algoritması

C4.5 algoritması genel kısımlarda belirtildiği gibi entropiye dayalı hesaplama yapan bir sınıflandırma algoritmasıdır. C4.5 algoritması hem kategorik hem de sürekli verileri sınıflandırma ağacına dahil ettiğinden güçlü bir algoritmadır. Karmaşık ve yoğun verilerde iyi sonuçlar vermektedir.

R programında C4.5 algoritması kodları yazılırken “RWeka” ve “partykit” paketleri kullanılmıştır.

"RWeka" veri ön işleme, sınıflandırma, regresyon, kümeleme, birliktelik kuralları, görselleştirme gibi veri madenciliği algoritmaları için araçlar (*tools*) içeren Java’da yazılmış bir pakettir (Ian ve Frank, 2005; Hornik ve diğ., 2009). “partykit” regresyon ve sınıflandırma modellerini özetlemek, sunmak ve görselleştirmek için altyapı sunan bir araçtır (Hothorn ve Zeileis, 2015).

C4.5 algoritmasının adımlarının R programında kullanımı aşağıda belirtildiği gibidir:

İlk olarak RWeka ve partykit paketlerinin R programı üzerinde kurulumu gerçekleştirilir. Paketleri kullanabilmek için R kütüphanesine bağlantı library() fonksiyonu ile yapılır.

```
> install.packages("RWeka")

> instal.packages("partykit")

> library(RWeka)

> library(partykit)
```

Kullanılacak paketler belirlendikten ve kütüphaneye tanıtıldıktan sonra veri seti .csv uzantılı olarak, read.table() fonksiyonu kullanılarak R programına aktarıldı.

```
> setwd("C:/")

> veri<-read.table("C:/VS/AS.csv",header=TRUE,sep=";")
```

Kullanılan veri seti kategorik ve sürekli verilerden oluştuğu için bu verilerin türünü R programında belirtmemiz gerekmektedir. Kategorik veriler “as.factor” olarak, sürekli veriler “as.numeric” olarak belirtilmektedir. “veri” nesnesini 1:22 sütunları kategorik olarak “veri1” isimli nesneye, 23. sütun nümerik olarak “veri2” isimli nesneye aktarıldı. veri1 ve veri2 nesneleri data.frame içerisine kaydedildi ve veri3 nesnesinde kategorik ve nümerik veriler birleştirildi.

```
> veri1<-data.frame(sapply(veri[,c(1:22)],as.factor))

> veri2<-data.frame(sapply(veri[,c(23)],as.numeric))

> veri3<-c(veri1,veri2)
```

Karar niteliğimiz olan 24. sütun “sinif” isimli nesneye as.factor olarak atandıktan sonra tüm veriler “veriT” isimli data.frame’de birleştirilmiştir.

```
> sinif<-sapply(veri[,c(24)],as.factor)

> veriT=data.frame(veri3,sinif)
```

Veri hazırladıktan sonra verinin üçte ikilik kısmı eğitim verisi olarak ayrılıp bu veri üzerinde sınıflandırma algoritmaları uygulandı. “oran” isimli bir nesne oluşturulup 2/3 olarak verilerin ayrılacağı oran belirtildi. `set.seed()` komutu ile rasgele bir başlangıç noktası seçilerek bölünme başlatıldı

```
> oran<-2/3
```

```
> set.seed(111)
```

Öğrenme verisi “Overi” isimli nesnenin içine aktarıldı. Bunun için `sample()` fonksiyonu kullanılarak belirttiğimiz oranda bir örnek alındı.

```
> Overi<-sample(length(sinif),
+ size=floor(oran*length(sinif)))
```

Test verisi öğrenme verisinin dışında kalan kısım olarak belirlendi ve veri seti üzerinden ayrılan kısımları “training_data”, “testing_data” ve “testing_sinif” isimli değişkenlere aktarıldı.

```
> test=-Overi
```

```
> training_data=veriT[Overi,]
```

```
> testing_data=veriT[test,]
```

```
> testing_sinif=sinif[test]
```

RWeka içinde J48 fonksiyonu budanmış ya da budanmamış C4.5 karar ağaçlarını üretir (Quinlan,1993). Bu fonksiyon kullanılarak oluşturulan eğitim verisinde sınıflandırma yapıldı. J48 fonksiyonu içinde Weka_control objesi ile veri seti içerisindeki boş (NULL) hücrelerin kontrolü yapılmaktadır ve bir hata ile karşılaşıldığında uyarı mesajı olarak ekranda gösterilmektedir. J48 fonksiyonu içerisine yazılan “M” değeri yaprak başına düşen minimum örnek sayısını ifade eder. M değerinin artması yaprak başına düşen örnek sayısını arttıracığından ağacın boyutunu da küçültecektir. U (unpruned tree) değeri oluşturulacak sınıflandırma ağacının budanıp budanmamasının belirtildiği

değerdir. U=TRUE ise budanmamış (*unpruned tree*), U=FALSE ise budanmış ağaç (*pruned tree*). "plot()" fonksiyonu ağacın çizimini yapmak için kullanılan fonksiyondur.

```
> m1<-J48(sinif~.,training_data,control=Weka_control
+ (M=8,U=TRUE))

> plot(m1)
```

Eğitim verisi üzerinde sınıflandırma yapıldıktan sonra ortaya çıkan ağacın tahminlemesi "predict()" fonksiyonu ile yapıldı. Verinin test için ayrılan üçte birlik kısmında modelin nasıl davrandığına yani ne ölçüde doğrulandığına bakıldı. "mean()" fonksiyonu ile de test verisinin kestirim sonucunun oluşturulan ağaçtan farkının karşılaştırma oranı bulundu. Bu oran bize "yanlış sınıflandırma oranını" (misclassification error) vermektedir.

```
> tree_pred=predict(m1,testing_data,type="class")

> mean(tree_pred!=testing_sinif)
```

3.7.1.2. Regresyon Ağaçları Algoritması

Regresyon ağaçları ikili bölünme temeline dayanan CART algoritmalarındandır. Çalışmada tercih edilmesinin sebebi, her bir bölünmede homojenliğin bir önceki bölünmeye oranla daha fazla olması ve değişken sayısının fazla olduğu veri kümeleri üzerine uygulanabilirliğidir. Ayrıca regresyon ağaçları yorumlaması kolay bir algoritmadır.

R programında regresyon ağaçları algoritması kodları yazılırken "tree" paketi kullanılmıştır.

"tree" paketi sınıflandırma regresyon modelleri için kullanılan bir araçtır (Ripley, 2014).

Regresyon ağaçları algoritmasının adımlarının R programında kullanımı aşağıda belirtildiği gibidir:

İlk olarak "tree" paketini kullanabilmek için R programı üzerinde kurulumu gerçekleştirilir. Yüklenen paketler library() fonksiyonu kullanılarak R programına tanıtılma işlemini gerçekleştirmektedir.

```
> install.packages("tree")
```

```
> library(tree)
```

Verilerden türlerine göre, kategorik, nümerik ve sınıf değişkeni (veri1, veri2, sınıf) oluşturuldu ve data.frame içerisine kaydedildi ve veri3 nesnesinde kategorik ve nümerik veriler birleştirildi.

```
> veri<-read.table("C:/VS/AS.csv",
```

```
+ sep=";", header=TRUE)
```

```
> veri1<-data.frame(sapply(veri[,c(1:22)], as.factor))
```

```
> veri2<-data.frame(sapply(veri[,c(23)], as.numeric))
```

```
> veri3<-cbind(veri1, veri2)
```

Karar niteliğimiz olan 24. sütun "sınıf" isimli nesneye as.factor olarak atandıktan sonra tüm veriler "veriT" isimli data.frame'de birleştirilmiştir.

```
> sınıf<-sapply(veri[,c(24)], as.factor)
```

```
> veriT<-data.frame(cbind(veri3, sınıf))
```

Veri hazırladıktan sonra verinin üçte ikilik kısmı eğitim verisi olarak ayrılıp bu veri üzerinde sınıflandırma algoritmaları uygulandı. "oran" isimli bir nesne oluşturulup 2/3 olarak verilerin ayrılacağı oran belirlendi. set.seed() komutu ile rasgele bir başlangıç noktası seçerek bölünme başlatıldı.

```
> set.seed(2)
```

```
> train<-sample(2:nrow(veriT), nrow(veriT)/3)
```

Test verisi, öğrenme verisinin dışında kalan kısım olarak belirlendi ve veri seti üzerinden ayrılan kısımları “training_data”, “testing_data” ve “testing_sinif” isimli değişkenlere aktarıldı.

```
> test<--train

> training_data<-veriT[train,]

> testing_data<-veriT[test,]

> testing_sinif<-sinif[test]
```

tree paketinde tree fonksiyonu belirtilen bağımlı değişkene göre bağımsız değişkenler ile arasında ağaç oluşturmaya yarar. Veri setimizdeki sinif değişkenine göre öğrenme verisi üzerinde ağaç oluşturuldu ve oluşturulan ağaç tree_model isimli nesneye aktarıldı. Bu ağacın çizimi plot() fonksiyonu ile gösterildi. Çizilen grafik üzerinde metinleri göstermek için ayrıca text() fonksiyonu kullanıldı.

```
> tree_model<-tree(sinif~.,training_data)

> plot(tree_model)

> text(tree_model,pretty=0)
```

Eğitim verisi üzerinde sınıflandırma yapıldıktan sonra ortaya çıkan ağacın test verisi üzerindeki tahminlemesi "predict()" fonksiyonu ile yapıldı. Verinin test için ayrılan üçte birlik kısmında modelin nasıl davrandığına yani ne ölçüde doğrulandığına bakıldı. "mean()" fonksiyonu ile de test verisinin kestirim sonucunun oluşturulan ağaçtan farkının karşılaştırma oranı bulundu. Bu oran bize "yanlış sınıflandırma oranını" (*misclassification error*) vermektedir.

```
> tree_pred<-predict(tree_model,testing_data,type="class")

> mean(tree_pred!=testing_sinif)
```

Oluşturulan ağaç üzerinde budama işlemi gerçekleştirildi. Budama işleminde amaç ağacı daha sade bir şekle dönüştürmektir. Bu sebeple tekrar bir rasgele başlangıç noktası seçildi ve cv.tree() fonksiyonu ile ağaç üzerinde çapraz doğrulama (*cross validation*)

işlemi gerçekleştirildi. FUN budama yapmak için kullanılan bir fonksiyondur. Kategorik sınıf niteliğine sahip olduğumuz için FUN=prune.misclass fonksiyonu kullanıldı.

```
> set.seed(3)
```

```
> cv_tree<-cv.tree(tree_model,FUN=prune.misclass)
```

Üretilen “cv_tree” nesnesinin özgün değerlerini görmek için names() fonksiyonu kullanıldı. Nesnenin özgün değerleri "size" "dev" "k" "method" değerleri olarak çıkmaktadır. Bu özgün değerlerin herhangi birini kullanmak istediğimizde “\$” işareti ile kullanmak istediğimiz değerin simgesini yazmak yeterlidir. Çapraz doğrulama yaptığımız ağaçta farklı boyutlardaki hata oranları grafiksel olarak gösterildi. “dev”, ağaç oluşturulurken oluşan hata oranını ifade etmektedir. “type=”b” ile çizdirdiğimiz grafikteki noktaları birleştiren çizgi eklendi.

```
> names(cv_tree)
```

```
> plot(cv_tree$size,cv_tree$dev,type="b")
```

Çıkan grafikte minimum hata oranı görülmektedir. Bu hata oranı ağacı budamak için kullanılmaktadır.

prune.misclass fonksiyonu ile budama gerçekleştirilmiş ağaç oluşturuldu. prune.misclass fonksiyonu misclass metodunu kullanarak en düşük sınıflandırma hatasını alarak en az hatalı ağacı üretir. Ürettiğimiz hata oranı az budanmış ağacın, plot() fonksiyonu ile çizimi gerçekleştirildi ve ağaç üzerinde text() fonksiyonu kullanılarak metinler ağaç üzerinde gösterildi.

```
> pruned_model<-prune.misclass(tree_model,best=6)
```

```
> plot(pruned_model)
```

```
> text(pruned_model,pretty=0)
```

Eğitim verisi üzerinde sınıflandırma yapıldıktan sonra ortaya çıkan ağacın tahminlemesi "predict()" fonksiyonu ile yapıldı. Verinin test için ayrılan üçte birlik kısmında modelin

nasıl davrandığına yani ne ölçüde doğrulandığına bakıldı. "mean()" fonksiyonu ile de test verisinin kestirim sonucunun, oluşturulan ağaçtan farkının karşılaştırma oranı bulundu.

```
> tree_pred<-predict(pruned_model,
+ testing_data,type="class")
> mean(tree_pred!=testing_sinif)
```

3.7.1.3. Gini Algoritması

Gini Algoritması da regresyon ağaçları gibi ikili bölünme temeline dayanan CART algoritmalarındandır. R programında gini ağaçları algoritması kodları yazılırken "rpart" paketi kullanılmıştır.

"rpart" paketi sınıflandırma regresyon modelleri için kullanılan bir araçtır (Therneau ve diğ., 2015).

Gini algoritmasının adımlarının R programında kullanımı aşağıda belirtildiği gibidir:

İlk olarak "rpart" paketini kullanabilmek için R programı üzerinde kurulumu gerçekleştirilir. Yüklenen paketler library() fonksiyonu kullanılarak R programına tanıtılma işlemini gerçekleştirmektedir.

```
> install.packages("rpart")
> library(rpart)
```

Verilerden türlerine göre, kategorik, nümerik ve sınıf değişkeni (veri1, veri2, sinif) oluşturuldu ve data.frame içerisine kaydedildi ve veri3 nesnesinde kategorik ve nümerik veriler birleştirildi.

```
> veri<-read.table("C:/VS/AS.csv",
+ sep=";",header=TRUE)
> matris<-as.data.frame(veri)
> veri1<-data.frame(sapply(veri[,c(1:22)],as.factor))
```

```
> veri2<-data.frame(sapply(veri[,c(23)],as.numeric))

> veri3<-c(veri1,veri2)
```

Karar niteliğimiz olan 24. sütun “sinif” isimli nesneye as.factor olarak atanmıştır.

```
> sınıf<-sapply(veri[,c(24)],as.factor)
```

Veriyi hazırladıktan sonra verinin üçte ikilik kısmı eğitim verisi olarak ayrılıp bu veri üzerinde sınıflandırma algoritmaları uygulandı. “oran” isimli bir nesne oluşturulup 2/3 olarak verilerin ayrılacağı oran belirlendi. set.seed() komutu ile rasgele bir başlangıç noktası seçilerek bölünme başlatıldı. Veri belirlediğimiz oran doğrultusunda bölünerek öğrenme verisi oluşturuldu. sample() fonksiyonu bir örnek oluşturmak için kullanılır. size () fonksiyonu verinin boyutunu belirler. floor bölünmeleri yuvarlar.

```
> oran<-2/3

> set.seed(111)

> Overi<-sample(length(sınıf),

+ size=floor(oran*length(sınıf)))
```

Test verisi, öğrenme verisinin dışında kalan kısım olarak belirlendi ve veri seti üzerinden ayrılan kısımları “training_data”, “testing_data” ve “testing_sınıf” isimli değişkenlere aktarıldı.

```
> test==Overi

> veriT=data.frame(veri3,sınıf)

> training_data=veriT[Overi,]

> testing_data=veriT[test,]

> testing_sınıf=sınıf[test]
```

Sınıf değişkeninin öğrenme verisi üzerinde modeli oluşturulur. Veri kategorik olduğundan metot olarak "class" kullanılır. minimum bölünme sayısı 4 olarak belirlendi.

parms bölünme fonksiyonu için isteğe bağlı bir parametre seçmek istenildiğinde kullanılan bir fonksiyondur. Bölünmenin Gini indeksine göre olmasını istediğimiz için "split" değeri Gini olarak seçilir ve kategorik veri seti olduğundan "list" parametre olarak belirlenir.

```
> tree_model<-rpart(sınıf~.,training_data,method="class",
+ minsplit=4,parms=list(split="gini"))
```

Oluşturduğumuz ağaç modelini kuralları ile görüntülemek istediğimizde show() fonksiyonu kullanılır.

```
> show(tree_model)
```

prp() fonksiyonu rpart içerisinde kullanılan bir çizim fonksiyonudur. rpart paketinin plot() çizim fonksiyonu büyük ağaçları çok karmaşık gösterdiğinden buradaki çizimde prp() fonksiyonu kullanılmıştır.

```
> prp(tree_model)
```

Eğitim verisi üzerinde sınıflandırma yapıldıktan sonra ortaya çıkan ağacın tahminlemesi "predict()" fonksiyonu ile yapıldı. Verinin test için ayrılan üçte birlik kısmında modelin nasıl davrandığına yani ne ölçüde doğrulandığına bakıldı. "mean()" fonksiyonu ile de test verisinin kestirim sonucunun oluşturulan ağaçtan farkının karşılaştırma oranı bulundu.

```
> tree_pred=predict(tree_model,testing_data,type="class")
> mean(tree_pred!=testing_sınıf)
```

3.7.1.4. Rasgele Orman Algoritması

Rasgele orman algoritması binlerce ağaç oluşturan bir algoritmadır. Her ağaçta bir karar yapısı oluşur ve sınıflandırmada bu karar yapıları kullanılır. Rasgele orman algoritması yeni gelecek örnekler için tahmin yapan güçlü bir algoritmadır.

R programında RO algoritması kodları yazılırken "randomForest" paketi kullanılmıştır.

"randomForest" paketi sınıflandırma regresyon modelleri için kullanılan bir araçtır (Liaw ve Wiener, 2002).

Rasgele Orman algoritmasının adımlarının R programında kullanımı aşağıda belirtildiği gibidir:

İlk olarak "randomForest" paketini kullanabilmek için R programı üzerinde kurulumu gerçekleştirilir. Yüklenen paketler library() fonksiyonu kullanılarak R programına tanıtılma işlemi gerçekleştirilmektedir.

```
> install.packages("randomForest")
```

```
> library(randomForest)
```

Veri .csv uzantılı olarak depolandığı yerden R ortamına aktarıldı. Karar niteliği olan AS hastalığının ağır ya da hafif olmasını simgeleyen nitelik "sinif" isimli nesneye aktarıldı ve diğer nitelikler "veri1" isimli nesneye aktarıldı.

```
> veri<-read.table("C:/VS/AS.csv",
```

```
+ header=TRUE, sep=";")
```

```
> veri1<-veri[, -24]
```

```
> sinif<-veri[, 24]
```

randomForest() fonksiyonu sınıflandırma yapmak ve veri noktaları arasındaki yakınlıkları değerlendirmek için kullanılmaktadır. veri1'in sınıf değişkenine göre sınıflandırılması "rf" isimli nesneye aktarıldı ve show() fonksiyonu ile sınıflandırma sonuçları görüntülendi.

```
> rf<-randomForest(veri1, sinif)
```

```
> show(rf)
```

Algoritmanın yaptığı tahmini görüntülemek için predict() fonksiyonu kullanıldı.

```
> predict(rf)
```

Tahmin sonuçlarının grafiksel gösterimi için `plot()` fonksiyonu kullanıldı.

```
> plot(rf,main="Raslantısal Orman")
```

4. BULGULAR

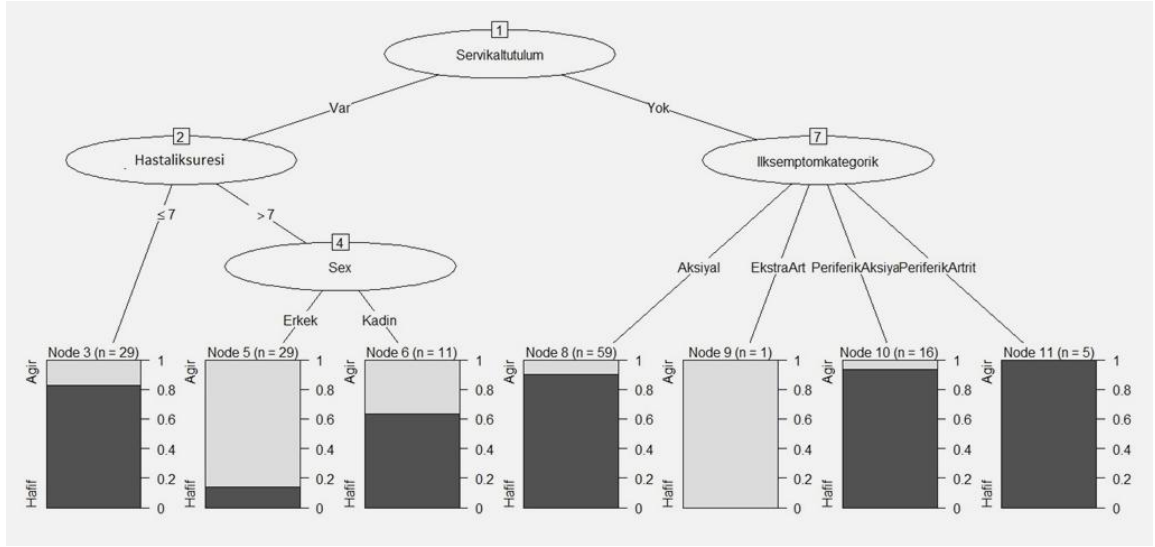
4.1. C4.5 ALGORİTMASI SONUÇLARI

C4.5 algoritmasının uygulanması sonucunda çıkan en iyi yanlış sınıflandırma oranı 0,213 yani %21,3 olarak bulunmuştur. Bu hata oranı C4.5 sınıflandırmasında J48 fonksiyonu içinde kullanılan M değeri 5 ve U=FALSE olduğunda yani ağaç üzerinde budama işlemi gerçekleştirildiğinde bulunmuştur. M değerinin değiştirilmesinde ve ağaç budama işlemi olup olmaması durumuna göre sınıflandırma hata oranındaki değişim Tablo 4.1’de görülmektedir.

Tablo 4.1: C4.5 algoritması ile oluşturulan sınıflandırma ağacının hata oranlarının değişimi.

M değeri	Ağaç budandı mı?	Yanlış sınıflandırma oranı
3	Evet (U=FALSE)	0,25 (%25)
5	Hayır (U=TRUE)	0,31 (%31)
5	Evet (U=FALSE)	0,213 (%21,3)
8	Hayır (U=TRUE)	0,213 (%21,3)
8	Evet (U=FALSE)	0,213 (%21,3)

Tablo 4.1’deki sonuçlara bakıldığında budama işlemi yapılan ve yapılmayan ağaçlarda M değerindeki değişiminin sınıflandırma hata oranını değiştirdiğini görmekteyiz. Bu durumda C4.5 algoritmasında elde edilen sınıflandırma hata oranı %21.3 olarak belirlenmiştir.



Şekil 4.1: C4.5 algoritması sonucunda çıkan karar ağacı.

Şekil 4.1’de görülen C4.5 algoritması sonucunda çıkan karar ağacından oluşturulan karar kuralları aşağıda belirtildiği gibidir.

KURAL 1 (29.0/5.0):

Eğer Servikaltutulum = Var ise ve

Eğer Hastaliksuresi ≤ 7 : Hafif ise AS=Hafif

KURAL 2 (29.0/4.0):

Eğer Servikaltutulum = Var ise ve

Eğer Hastaliksuresi > 7 : Hafif ve

Eğer Sex=Erkek ise AS=Agir

KURAL 3 (11.0/4.0):

Eğer Servikaltutulum = Var ise ve

Eğer Hastaliksuresi > 7 : Hafif ve

Eğer Sex=Kadin ise AS=Hafif

KURAL 4 (59.0/6.0):

Eğer Servikaltutulum = Yok ise ve

Eğer İlkseptomkategorik=Aksiyal ise AS=Hafif

KURAL 5 (1.0):

Eğer Servikaltutulum = Yok ise ve

Eğer İlkseptomkategorik=EkstraArt ise AS=Agir

KURAL 6 (16.0/1.0):

Eğer Servikaltutulum = Yok ise ve

Eğer İlkseptomkategorik=PeriferikAksiyal ise AS=Hafif

KURAL 7 (5.0):

Eğer Servikaltutulum = Yok ise ve

Eğer İlkseptomkategorik=PeriferikArtrit ise AS=Hafif

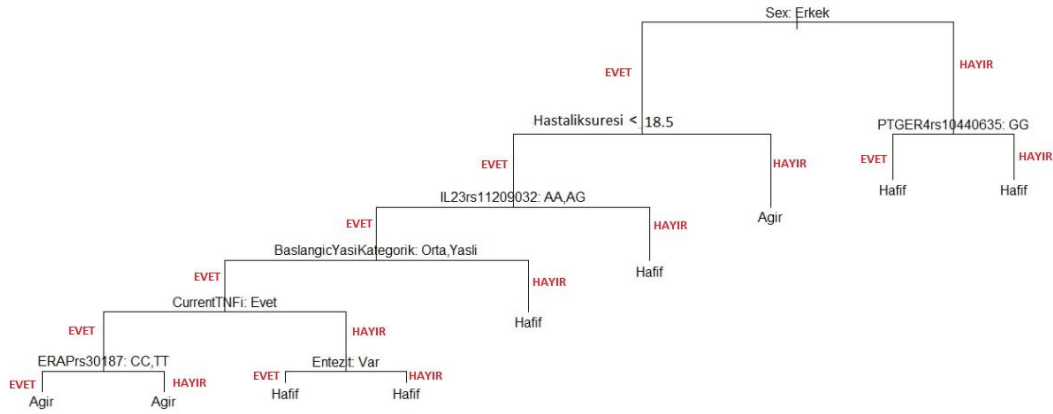
4.2 REGRESYON AĞAÇLARI ALGORİTMASI SONUÇLARI

Regresyon ağaçları algoritmasının uygulanması sonucunda çıkan en iyi yanlış sınıflandırma oranı 0,26 yani %26 olarak bulunmuştur. Oluşturulan ağaç üzerinde budama işlemi gerçekleştirildiğinde de hata oranı aynı çıkmıştır. Budama yapılmadan önce ve yapıldıktan sonra bulunan sınıflandırma hata oranındaki değişim Tablo 4.2’de görülmektedir.

Tablo 4.2: Regresyon ağaçları algoritması ile oluşturulan sınıflandırma ağacının hata oranlarının değişimi.

Ağaç budandı mı?	Yanlış sınıflandırma oranı
Evet	0,26 (%26)
Hayır	0,26 (%26)

Tablo 4.2’deki sonuçlara bakıldığında budama işleminin sınıflandırma hata oranını değiştirmedikini görmekteyiz. Bu durumda regresyon ağaçları algoritması sonucu yapılan sınıflandırmasında elde edilen sınıflandırma hata oranı %26 olarak belirlenmiştir.



Şekil 4.2: Regresyon ağaçları algoritması sonucunda çıkan karar ağacı.

Şekil 4.2’de görülen regresyon ağaçları algoritması sonucunda çıkan karar ağacından oluşturulan karar kuralları aşağıda belirtildiği gibidir:

KURAL 1:

Eğer Sex = Erkek ise ve

Eğer Hastaliksuresi < 18,5 ise ve

Eğer IL23rs11209032 = AA veya AG ise

Eğer Baslangicyasikategorik = Orta veya Yasli ise

Eğer CurentTNFi = Evet ise ve

Eğer ERAPrs30187 = CC veya TT ise AS=Agir

KURAL 2:

Eğer Sex = Erkek ise ve

Eğer Hastaliksuresi < 18,5 ise ve

Eğer IL23rs11209032 = AA veya AG ise

Eğer Baslangicyasikategorik = Orta veya Yasli ise

Eğer CurentTNFi = Evet ise ve

Eğer ERAPrs30187 = TC ise AS=Agir

KURAL 3:

Eğer Sex = Erkek ise ve

Eğer Hastaliksuresi < 18,5 ise ve

Eğer IL23rs11209032 = AA veya AG ise

Eğer Baslangicyasikategorik = Orta veya Yasli ise

Eğer CurentTNFi = Hayir ise ve

Eğer Entezit = Var ise AS=Hafif

KURAL 4:

Eğer Sex = Erkek ise ve

Eğer Hastaliksuresi < 18,5 ise ve

Eğer IL23rs11209032 = AA veya AG ise

Eğer Baslangicyasikategorik = Orta veya Yasli ise

Eğer CurentTNFi = Hayir ise ve

Eğer Entezit = Yok **ise** AS=Hafif

KURAL 5:

Eğer Sex = Erkek **ise ve**

Eğer Hastaliksuresi < 18,5 **ise ve**

Eğer IL23rs11209032 = AA **veya** AG **ise**

Eğer Baslangicyasikategorik = Genc **ise** AS=Hafif

KURAL 6:

Eğer Sex = Erkek **ise ve**

Eğer Hastaliksuresi < 18,5 **ise ve**

Eğer IL23rs11209032 = GG **ise** AS=Hafif

KURAL 7:

Eğer Sex = Erkek **ise ve**

Eğer Hastaliksuresi > 18,5 **ise** AS=Agir

KURAL 8:

Eğer Sex = Kadın **ise ve**

Eğer PTGER4rs10440635 = GG **ise** AS=Hafif

KURAL 9:

Eğer Sex = Kadın **ise ve**

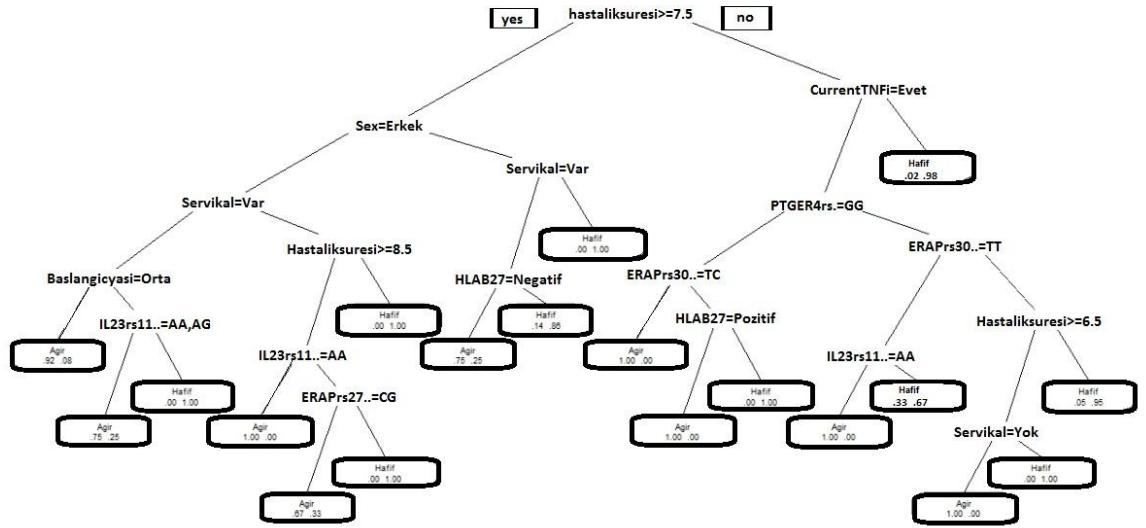
Eğer PTGER4rs10440635 = AA **veya** AG **ise** AS=Hafif

4.3 GİNİ ALGORİTMASI SONUÇLARI

Gini algoritmasının uygulanması sonucunda çıkan en iyi yanlış sınıflandırma oranı Tablo 4.3’de görüldüğü gibi 0,18 yani %18 olarak bulunmuştur. Oluşturulan ağaç üzerinde budama işlemi gerçekleştirilmemiştir.

Tablo 4.3: Gini algoritması ile oluşturulan sınıflandırma ağacının hata oranı.

Ağaç budandı mı?	Yanlış sınıflandırma oranı
Hayır	0,18 (% 18)



Şekil 4.3: Gini algoritması sonucunda çıkan karar ağaçları.

Şekil 4.3’de görülen regresyon ağaçları algoritması sonucunda çıkan karar ağacından oluşturulan karar kuralları aşağıda belirtildiği gibidir:

KURAL 1:

Eğer Hastaliksuresi $\geq 7,5$ **ise ve**

Eğer Sex = Erkek **ise ve**

Eğer Servikaltutulum = Var **ise ve**

Eğer Baslangicyasikategorik = Orta **ise** AS=Agir

KURAL 2:

Eğer Hastaliksuresi $\geq 7,5$ ise ve

Eğer Sex = Erkek ise ve

Eğer Servikaltutulum = Var ise ve

Eğer Baslangicyasikategorik = Orta ise ve

Eğer IL23rs11209032 = AA veya AG ise AS=Agir

KURAL 3:

Eğer Hastaliksuresi $\geq 7,5$ ise ve

Eğer Sex = Erkek ise ve

Eğer Servikaltutulum = Var ise ve

Eğer Baslangicyasikategorik = Orta ise ve

Eğer IL23rs11209032 = GG ise AS=Hafif

KURAL 4:

Eğer Hastaliksuresi $\geq 7,5$ ise ve

Eğer Sex = Erkek ise ve

Eğer Servikaltutulum = Yok ise ve

Eğer Hastaliksuresi $\geq 8,5$ ise ve

Eğer IL23rs11209032 = AA ise AS=Agir

KURAL 5:

Eğer Hastaliksuresi $\geq 7,5$ ise ve

Eğer Sex = Erkek ise ve

Eğer Servikaltutulum = Yok ise ve

Eğer Hastaliksuresi $\geq 8,5$ ise ve

Eğer IL23rs11209032 = AG veya GG ise ve

Eğer ERAPrs27044 = CG ise AS=Hafif

KURAL 6:

Eğer Hastaliksuresi $\geq 7,5$ ise ve

Eğer Sex = Erkek ise ve

Eğer Servikaltutulum = Yok ise ve

Eğer Hastaliksuresi $\geq 8,5$ ise ve

Eğer IL23rs11209032 = AG veya GG ise ve

Eğer ERAPrs27044 = CC veya GG ise AS=Hafif

KURAL 7:

Eğer Hastaliksuresi $\geq 7,5$ ise ve

Eğer Sex = Erkek ise ve

Eğer Servikaltutulum = Yok ise ve

Eğer Hastaliksuresi $< 8,5$ ise AS=Hafif

KURAL 8:

Eğer Hastaliksuresi $\geq 7,5$ ise ve

Eğer Sex = Kadın ise ve

Eğer Servikaltutulum = Var ise ve

Eğer HLAB27 = Negatif ise AS=Agir

KURAL 9:

Eğer Hastaliksuresi $\geq 7,5$ ise ve

Eğer Sex = Kadın ise ve

Eğer Servikaltutulum = Var ise ve

Eğer HLAB27 = Pozitif ise AS=Hafif

KURAL 10:

Eğer Hastaliksuresi $\geq 7,5$ ise ve

Eğer Sex = Kadın ise ve

Eğer Servikaltutulum = Yok ise AS=Hafif

KURAL 11:

Eğer Hastaliksuresi $< 7,5$ ise ve

Eğer CurrentTNFi = Hayir ise AS=Hafif

KURAL 12:

Eğer Hastaliksuresi $< 7,5$ ise ve

Eğer CurrentTNFi = Evet ise ve

Eğer PTGER4rs10440635 = GG ise ve

Eğer ERAPrs30187 = TC ise AS=Agir

KURAL 13:

Eğer Hastaliksuresi $< 7,5$ ise ve

Eğer CurrentTNFi = Evet ise ve

Eğer PTGER4rs10440635 = GG ise ve

Eğer ERAPrs30187 = CC veya TT ise ve

Eğer HLAB27 = Pozitif ise AS=Agir

KURAL 14:

Eğer Hastaliksuresi < 7,5 ise ve

Eğer CurrentTNFi = Evet ise ve

Eğer PTGER4rs10440635 = GG ise ve

Eğer ERAPrs30187 = CC veya TT ise ve

Eğer HLAB27 = Negatif ise AS=Hafif

KURAL 15:

Eğer Hastaliksuresi < 7,5 ise ve

Eğer CurrentTNFi = Evet ise ve

Eğer PTGER4rs10440635 = AA veya AG ise ve

Eğer ERAPrs30187 = TT ise ve

Eğer IL23rs11209032 = AA ise AS=Agir

KURAL 16:

Eğer Hastaliksuresi < 7,5 ise ve

Eğer CurrentTNFi = Evet ise ve

Eğer PTGER4rs10440635 = AA veya AG ise ve

Eğer ERAPrs30187 = TT ise ve

Eğer IL23rs11209032 = AA veya AG ise AS=Hafif

KURAL 17:

Eğer Hastaliksuresi < 7,5 ise ve

Eğer CurrentTNFi = Evet ise ve

Eğer PTGER4rs10440635 = AA veya AG ise ve

Eğer ERAPrs30187 = TC veya CC ise ve

Eğer Hastaliksuresi >= 6,5 ise ve

Eğer Servikaltutulum = Yok ise AS=Agir

KURAL 18:

Eğer Hastaliksuresi < 7,5 ise ve

Eğer CurrentTNFi = Evet ise ve

Eğer PTGER4rs10440635 = AA veya AG ise ve

Eğer ERAPrs30187 = TC veya CC ise ve

Eğer Hastaliksuresi >= 6,5 ise ve

Eğer Servikaltutulum = Var ise AS=Hafif

KURAL 19:

Eğer Hastaliksuresi < 7,5 ise ve

Eğer CurrentTNFi = Evet ise ve

Eğer PTGER4rs10440635 = AA veya AG ise ve

Eğer ERAPrs30187 = TC veya CC ise ve

Eğer Hastaliksuresi < 6,5 ise AS=Hafif

4.4 RASGELE ORMAN ALGORİTMASI SONUÇLARI

Rasgele orman algoritması ile 500 raslantısal ağaç oluşturulmuştur. Değişken sayısı her bölünme için 4 olarak belirlenmiştir. Test verisi (OutOfBag OOB) hata oranı tahmini %16,89 olarak çıkmıştır. Sınıflandırma tahmini sonucunda Tablo 4.4'deki hata matrisinde, sınıf grubu kategorik değerlerinin yanlış sınıflandırılması görülmektedir. Sınıflandırma hatası ağır hasta grubu için 0,46 hafif grup için 0,05 olarak bulunmuştur.

Tablo 4.4: Rasgele orman algoritması hata matrisi.

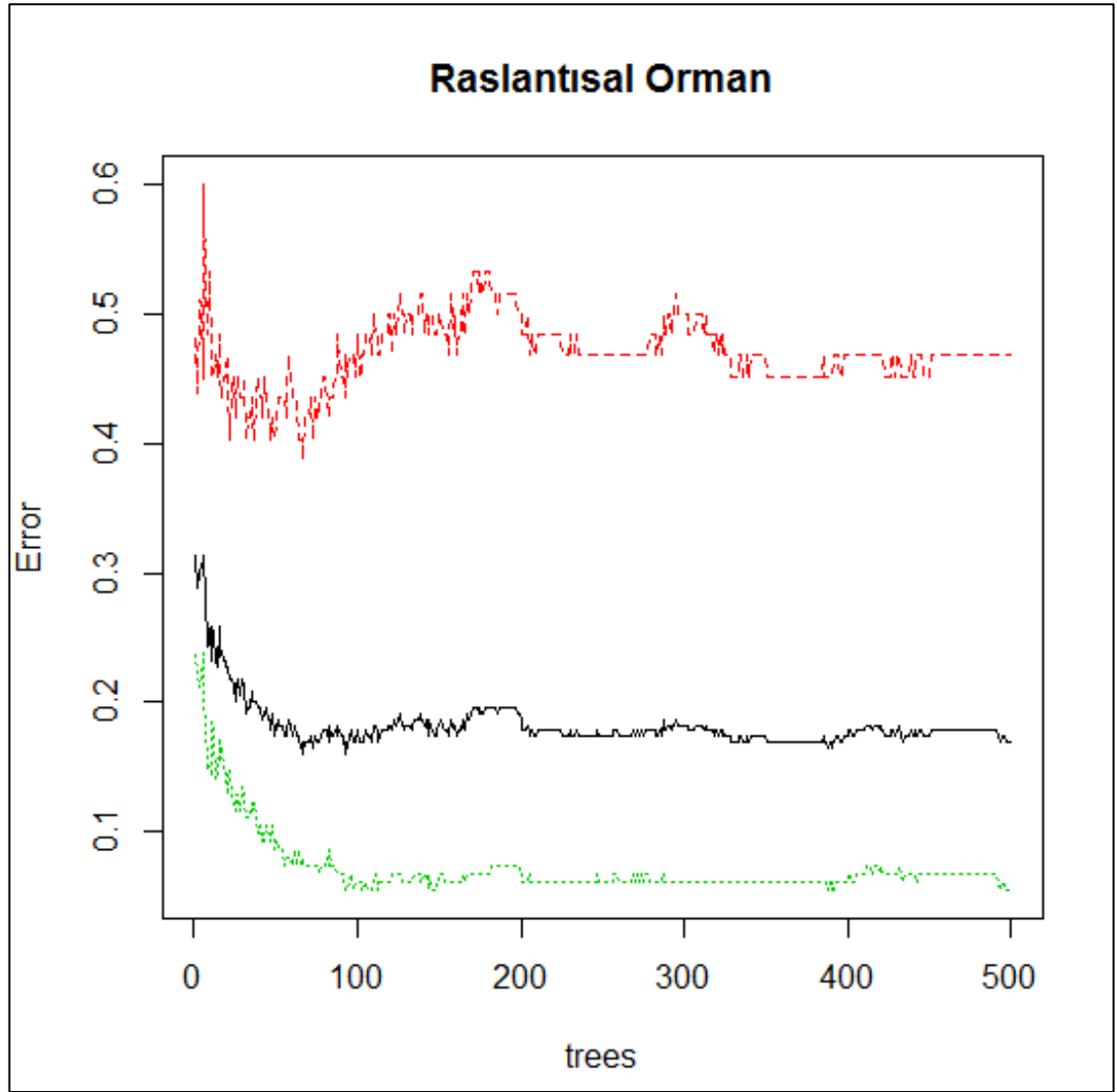
	Agir	Hafif	Sınıf hatası (class.error)
Agir	33	29	0.46774194
Hafif	9	154	0.05521472

Sınıflandırma tahminlerinin R programındaki görüntüsü Şekil 4.4'de görüldüğü gibidir.

```
> predict(rf)
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Agir Hafif Hafif Agir Hafif Agir Hafif Agir Agir Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif
28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
Agir Hafif Agir Agir Agir Hafif Hafif Agir Hafif Hafif Hafif Hafif Hafif Agir Hafif Agir Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Agir Hafif
55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81
Agir Hafif Agir Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Agir Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Agir
82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108
Agir Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Agir Agir Hafif Agir Hafif Hafif Hafif Agir Hafif Hafif Hafif Agir Agir Hafif Hafif Agir Hafif
109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135
Agir Agir Hafif Hafif Agir Hafif Hafif Hafif Hafif Hafif Agir Hafif Hafif Hafif Agir Hafif Hafif Hafif Hafif Agir Hafif Hafif Hafif Hafif Hafif Agir
136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162
Hafif Hafif Hafif Hafif Hafif Hafif Agir Hafif Hafif Hafif Hafif Hafif Hafif Agir Hafif Agir Hafif Hafif Hafif Hafif Agir Hafif Hafif Hafif Agir Hafif
163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189
Hafif Agir Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif
190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216
Hafif Hafif Hafif Hafif Hafif Hafif Agir Hafif Hafif Agir Hafif Hafif Agir Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif Agir Hafif Hafif Hafif Hafif
217 218 219 220 221 222 223 224 225
Hafif Hafif Hafif Hafif Hafif Hafif Hafif Hafif
```

Şekil 4.4: Rasgele orman algoritması sınıflandırma tahminleri R program çıktısı.

Rasgele orman algoritması ile oluşturulan ağaçların sınıflandırma hatalarının grafiksel gösterimi Şekil 4.5'de görülmektedir.



Şekil 4.5: Rasgele orman algoritması sonucunda üretilen ağaçların hata oranının grafiksel gösterimi.

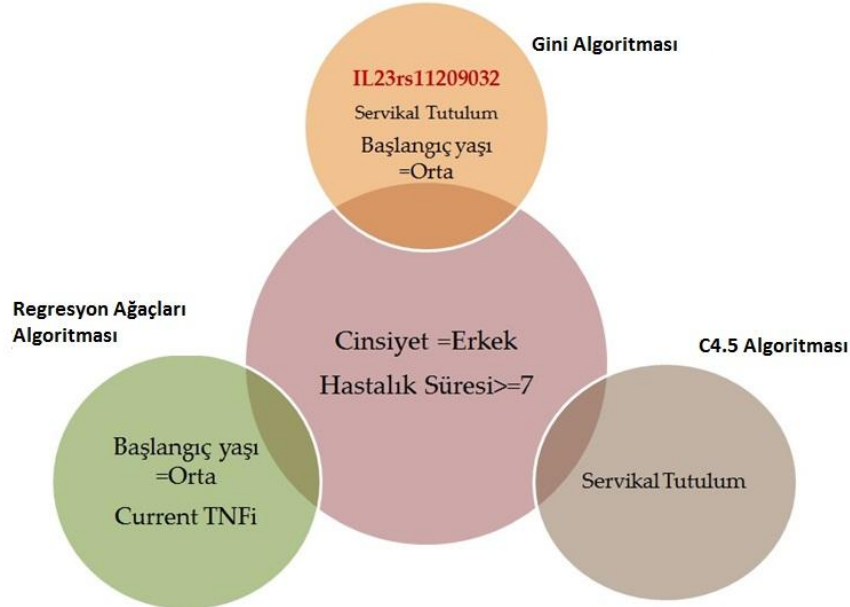
4.5. ALGORİTMALARIN KARŞILAŞTIRILMASI

Tablo 4.5’de algoritmaların başarı yüzdelere göre sınıflandırma tahmini veren RO’nun başarı yüzdesi 83.11%, Gini algoritmasının 82%, C4.5 algoritmasının 78.7% ve Regresyon ağaçlarının 74% olarak görülmektedir.

Tablo 4.5: Sınıflandırma algoritmalarının başarı yüzdeleri.

Algoritma	Başarı Yüzdesi
Rasgele Orman	83.11%
Gini Algoritması	82%
C4.5	78.7%
Regresyon Ağaçları	74%

Algoritmalar sonucunda sıklıklar da göz önüne alındığında AS hastalığının radyografik progresyonuna etki eden faktörlere ilişkin bulgular Şekil 4.6’da görüldüğü gibidir.



Şekil 4.6: Kullanılan algoritmalar sonucunda çıkan, AS hastalığının radyografik progresyonuna etki eden faktörlere ilişkin bulgular.

5. TARTIŞMA VE SONUÇ

Bu tez çalışmasında, veri madenciliği sınıflandırma yöntemlerinden karar ağaçları incelenmiş ve sınıflandırma algoritmalarından Gini, regresyon ağaçları, C4.5 ve rasgele orman algoritmalarından detaylı bir şekilde bahsedilmiştir. Sınıflandırma yöntemlerinden karar ağacı algoritmalarının veri madenciliği ile nasıl kullanıldığı ile ilgili tıp alanında bir uygulama yapılmıştır. Bu uygulamada Marmara Üniversitesi Tıp Fakültesi İç Hastalıkları Anabilim Dalı, Romatoloji Bilim Dalı'nda ankilozan spondilit hastalığına sahip hastalardan oluşan veri seti kullanılmıştır.

Çalışmada kullanılan veri seti hastalık süresi minimum 1 yıl, maksimum 49 yıl olan, hastalara ait verileri içermektedir. Ele alınan veri setinde hastalığın ilerleyişinin izlendiği radyografik skorlama değeri olan mSASSS değeri hastalığın ilerleyişi ve omurların kaynaşma derecelerine göre değişmektedir. Çalışmada hastalar mSASSS skorlarına göre “ağır” ve “hafif” olarak ayrılmıştır ve sınıflandırma yapılmıştır. Gini, regresyon ağaçları, C4.5 ve rasgele orman algoritmaları kullanılarak hastaların “ağır” ya da “hafif” olma durumlarının sınıflandırma tahmini yapılmıştır. Ayrıca Gini, regresyon ağaçları, C4.5 algoritmaları ile üç adet karar yapısı modeli oluşturulmuştur. Oluşturulan modellerde hastalığın “ağır” ya da “hafif” olmasının bağlı olduğu faktörler belirlenmiştir.

Yapılan analizler sonucunda rasgele orman algoritmasının sınıflandırma başarı tahmini %83.11 olarak bulunurken, sınıflandırma ve model doğruluk oranı Gini algoritmasında %82, C4.5 algoritmasında 78.7 ve regresyon ağaçlarında %74 olarak bulunmuştur. Gini algoritması en iyi sınıflandırma ve model tahmin oranını vermektedir. Gini, regresyon ağaçları, C4.5 algoritmaları sonucunda oluşturulan karar ağaçları incelendiğinde ağır olan hasta grubunun bağlı olduğu faktörler şunlardır:

C4.5 algoritması için: Hastalık süresi (7 yıl'dan fazla olması durumu), cinsiyet (erkek hastalar), servikal tutulum (tutulunun varlığı)

Regresyon ağaçları algoritması için: Hastalık süresi (18 yıl'dan fazla olanlar), hastalığın başlangıç yaşı (orta yaş grubu), Current TNFi (Tümör nekrosis faktör alfa inhibitör tedavisi devam edenler), cinsiyet (erkek hastalar)

Gini algoritması için: Hastalık süresi (7,5 yıl'dan fazla olması durumu), cinsiyet (erkek hastalar), servikal tutulum (tutulumun varlığı), hastalığın başlangıç yaşı (orta yaş grubu), IL23rs11209032 (interlökin 23 geninin rs11209032 tek nükleotid polimorfizminin genotipinin AA veya AG olması durumu)

Karar yapıları incelendiğinde üç algoritmanın da ortak olarak karar yapılarında çıkan faktörler erkek cinsiyetteki hastalar ve hastalık süresinin 7 yıl ve üzerinde olması durumudur.

Ankilozan spondilit hastalığının radyografik ilerleyişine etki eden faktörlere bakıldığında literatürde erkek bireylerin kadın bireylere oranla hastalığı daha hafif atlattığı görülmektedir ve hastalığın sinsi olarak ilerleyiş gösterdiği bilinmektedir (Caretta ve diğ., 1983). Servikal tutulumun da bazı hastalarda görüldüğü bilinmektedir (Us ve diğ., 2014). Başlangıç yaşı için ASAS kriterlerinde 45 yaş ve altı olarak belirtilmiştir. Yapılan çalışma sonucunda veri setinde 20 ila 40 yaş arası bireylerden oluşan grup orta yaş olarak kabul edilmiştir. Karar ağaçlarında çıkan faktörlerden hastalığın başlangıç yaşı ortaya yaş olarak çıkmıştır. Literatürde yer alan kötü prognostik faktörlerin (Kalça tutulumu, üveit, cinsiyet vb.) veri madenciliği sonrasında da saptanması AS literatürüne anlamlı katkıda bulunacaktır.

İlerleme evresi fazla olan hastalarda TNF (tümör nekrosis faktör) kullanımı yapılan farmakolojik tedavi yöntemlerinden birisidir (Yazıcı ve diğ., 2004). IL23rs11209032 - interlökin 23 geninin rs11209032 tek nükleotid polimorfizminin genotipinin AA veya AG olması durumunun progresyona etkisi ile ilgili bilgiye literatürde rastlanmamıştır.

Çalışma sonuçları AS hastalığının ilerleyişine etkili olan faktörleri genel anlamda desteklemektedir.

Doğruluk oranı, yani modellerin performansının %90 nın altında çıkmasının nedeni analizde kullanılan veri kümesi nitelik sayısının az olmasındandır. Yani hasta sayısının az olmasından ziyade nitelik sayısının az olması bu oranı düşürmektedir.

Tıp alanında veri madenciliği uygulamalarının giderek yaygınlaştığı bilinmektedir. Veri madenciliğinin tıp alanına katkısının artırılabilmesi için; alan uzmanlarının mümkünse veri toplama aşamasından önce bir araya gelip beraber çalışması, veri girişlerinin dikkatli bir şekilde yapılması, sağlık çalışanlarının bu araştırmalara olumlu yaklaşması ve çalışmanın veriminin artırılması açısından çeşitli kurum ve kuruluşlar tarafından maddi olarak desteklenmesi önem arz etmektedir.

Sonuç olarak, veri madenciliği C4.5, Gini, Regresyon ağaçları ve Random Forest sınıflandırma algoritmaları kullanılarak, Ankilozan spondilit hastalığına sahip bireylere ait veri seti analiz edildiğinde; hastalığın radyografik progresyonuna etki eden faktörler olarak; hastalık süresi (7.5’den fazla olması durumu), cinsiyet (erkek hastalar), servikal tutulum (tutulumun varlığı), hastalığın başlangıç yaşı (orta yaş grubu), IL23rs11209032 (interlökin 23 geninin rs11209032 tek nükleotid polimorfizminin genotipinin AA veya AG olması durumu), CurrentTNFi faktörleri bulunmuştur. Bu faktörlerden cinsiyet (erkek hastalar) ve hastalık süresi (7.5’den fazla olması durumu) tüm algoritmalarda ortak yer alırken, servikal tutulum (tutulumun varlığı), hastalığın başlangıç yaşı (orta yaş grubu), IL23rs11209032 (interlökin 23 geninin rs11209032 tek nükleotid polimorfizminin genotipinin AA veya AG olması durumu) Gini algoritma analizinde, hastalığın başlangıç yaşı (orta yaş grubu), Current TNFi (Tümör nekrosis faktör alfa inhibitör tedavisi devam edenler) regresyon ağaçları algoritma analizinde, servikal tutulum (tutulumun varlığı) C4.5 algoritma analizinde çıkmıştır.

Bu çalışmanın istatistiksel analizleri mevcut literatür ile karşılaştırıldığında:

Cinsiyetle ilgili yapılan çalışmalarda erkek bireylerin hastalığa yakalanma ve hastalığın daha ağır seyretme oranının, kadınlara göre yaklaşık 3 kat daha fazla olduğu bilinmektedir (Atagündüz ve diğ., 2010). Veri madenciliği uygulaması sonucunda erkek bireylerin ağır hasta grubuna bağlanan faktörler arasında çıkmış olması literatürü desteklemektedir.

AS hastalık kriterlerinin belirlenmesi için kullanılan ASAS (Rudwaleit ve diğ., 2009) kriterlerine göre hastalık yaşı 45 yaşında altında olması hastalık seyrinin kötü ilerleyişine sebep olan faktörler arasında bilinmektedir. Çalışmamızda kullanılan veri setinde bireylerin yaşları 20 – 40 aralığında olanlar “Orta” yaş grubu olarak

sınıflandırılmıştır. Veri madenciliği uygulaması sonucunda “Orta” yaş, radyografik ilerlemeye etki eden faktörler arasında çıkmış olup literatürde var olan bilgiyi desteklemektedir.

AS hastalığının sinsi ilerleyen bir hastalık olduğu bilinmektedir (Arasıl ve diğ., 2000) hastalık süresinin fazla olması hastalığın kötü prognostik ilerlemesine etki etmektedir. Veri madenciliği uygulaması sonucunda “hastalık süresi 7 yıl ve üzeri” olması ağır hasta grubuna doğru ilerletmektedir.

AS hastalığında omurga tutulumu genel olarak lomber (bel) bölgesinden başlar, hastalığın seyri sırasında torasik (göğüs) ve servikal (boyun) bölgeler de tutulum gerçekleşir. Servikal bölgede tutulum oksiputtan (başın arka kısmı) başlayarak aşağı doğru ankiloza yani köprüleşmeye sebep olur (Us ve diğ., 2014). Radyografik ilerleme de köprüleşme oranı arttıkça hastanın hareket yeteneği kısıtlanmakta ve böylece hastalığı ağır seyretmektedir. Veri madenciliği uygulaması sonucunda servikal tutulumun ağır hasta grubuna bağlanan faktörler arasında çıkmış olması literatürü destekler niteliktedir.

Hastalığın tedavi yöntemleri arasında yer alan farmakolojik tedavide kullanılan TNFi (Tümör nekrosis faktör alfa inhibitör) ilaç grubu ağır hasta grubunda daha çok tercih edilmektedir. TNFi tedavisi devam eden hastalar doğal olarak ağır hasta grubunda çıkması beklenmektedir. Nitekim veri madenciliği uygulaması sonucunda da “Current TNFi” (Tümör nekrosis faktör alfa inhibitör tedavisi devam edenler) ağır grubu içerisinde yer alan faktörlerden çıkmış olup literatür ile uyumlu olduğu görülmektedir.

IL23rs11209032 (interlökin 23 geninin rs11209032 tek nükleotid polimorfizminin genotipinin AA veya AG olması durumu) faktörünün ağır radyografik progresyon ile ilişki durumu ile ilgili bir bilgiye AS literatüründe rastlanmamıştır. Veri madenciliği uygulaması sonucunda “IL23rs11209032” faktörü ağır grup içerisinde bulunmuş ve çalışmamız bu sonuç ile mevcut literatürden ayrılmıştır. Farklı olan faktörlere yönelik daha geniş veri setleri ile araştırma yapılmasının uygun olacağı düşünülmektedir.

Literatürde var olan HLA-B27 gen pozitifliği, kalça tutulumu, üveit gibi AS hastalığının kötü prognostik ilerlemesine etkiettiği bilinen faktörler veri madenciliği uygulaması sonucunda karar yapısında çıkmamıştır.

Ankilozan spondilit hastalığına ait verilerin veri madenciliği yöntemleri kullanılarak analiz edilmesine ilişkin bir araştırmaya literatürde rastlanmamıştır. Bu nedenle yapılan ilk çalışma niteliğinde olan bu tez çalışmasından elde edilen bulguların özgünlüğü, konuya ilgi duyan araştırmacılara bir yol haritası oluşturabilir.

KAYNAKLAR

- AbdelAal, R.E., 2005, Gmdh-Based Feature Ranking And Selection For Improved Classification Of Medical Data, *Journal of Biomedical Informatics*, 38 (2005), 456-468.
- Adams, E., 2014, *Ankylosing Spondylitis & Related Diseases*, <http://www.spondylitis.org/about/overview.aspx>, [Ziyaret tarihi: 2 Şubat 2015].
- Akar, Ö., Güngör, O., 2012, Rasgele Orman Algoritması Kullanılarak Çok Bantlı Görüntülerin Sınıflandırılması, *Journal of Geodesy and Geoinformation*, 1 (2), 139-146.
- Akpınar, H., 2000, Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29 (1), 1-22.
- Aksoy, B., 2009, *Cluster Analysis of Decompression Illness*, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü.
- Atagunduz, P., Aydın, S.Z., Bahadır, C., Erer, B. ve Direskeneli, H., 2010, Determinants of Early Radiographic Progression in Ankylosing Spondylitis, *The Journal of Rheumatology*, 1-19.
- Ath, G. D., Fabricius, K.E., 2000, Classification and Regression Trees: A Powerful yet Simple Technique For Ecological Data Analysis, *Ecological Society of America*, 81 (11), 3178-3192.
- Baykal, A., Veri Madenciliği Uygulama Alanları, *Dumlupınar Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 7, 95-107.
- Berry, M. W., Browne, M., 2006, *Lecture Notes in Data Mining*, World Scientific, Danvers, ISBN: 981-256-802-6.
- Bertolazzi, P., Felici, G., Festa, P., Lancia, G., 2008, Logic classification and feature selection for biomedical data, *Computers and Mathematics with Applications*, 55, 889-899.
- Bramer, M., 2007, *Principles of Data Mining*, Springer, London, ISBN: 1-84628-765-0.
- Braun, J., Sieper, J., 2007, Ankylosing spondilitis, *Lancet*, 369, 1379-1390.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1998, *Classification and Regression Trees*, Chapman&Hall.
- Breiman, L., Cutler, A., 2005, *Random Forest*, https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#workings, [Ziyaret Tarihi: 5 Mayıs 2015].
- Calin, A., Porta, J., Fries J. F., Schurman D. J., 1977, Clinical History as a Screening Test for Ankylosing Spondylitis, *Journal of the American Medical Association*, 237 (24), 2613-2614.

- Caretta, S., Graham, D., Little, H., Rubenstein, J. ve Rosen, P., 1983, The Natural Disease Course of Ankylosing Spondylitis, *Arthritis Rheum*, pp. 186-190.
- Chakrabarti, S., Cox, E., Frank, E., Güting, R.H., Han, J., Jiang, X., Kamber, M., Lightstone, S. S., Nadeau, T. P., Neapolitan, R. E., Pyle, D., Refaat, M., Schneider, M., Teorey, T. J., Witten, I. H., 2009, *Data Mining Know It All*, Elsevier, United States,.
- Chang, R.I., Chu, C.C., Wu, Y.Y., Chen, Y.L., 2010, Gene Clustering By Using Query-Based Self-Organizing Maps, *Expert System with Application*, 37, 6689-6694.
- Chin, C. Y., Weng, M. Y., Lin, C. T., Cheng, S., Y., Yang, H. K. Y., Tseng V. S., 2015, Mining Disease Risk Patterns from Nationwide Clinical Databases for the Assessment of Early Rheumatoid Arthritis, *Plos One*, 10 (4), 1-20.
- Çarklı, B., 2010, *Sağlık Sektöründe Apriori Algoritması ile Bir Veri Madenciliği Uygulaması*, Yüksek Lisans, Fen Bilimleri Enstitüsü.
- Çataloluk, H., 2012, *Gerçek Tıbbi Veriler Üzerinde Veri Madenciliği Yöntemlerini Kullanarak Hastalık Tespiti*, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü.
- Çokluk, Ö., 2010, Lojistik Regresyon: Analizi Kavram ve Uygulama, *Kuram ve Uygulamada Eğitim Bilimleri / Educational Sciences: Theory & Practice*, 10 (3), 1357-1407.
- Delen, D., Oztekin, A., Kong, Z., 2010, A Machine Learning-Based Approach To Prognostic Analysis of Thoracic Transplantation, *Artificial Intelligence in Medicine*, 49, 33-42.
- Dincer, E., Duru N., 2006, Gırtlak Kanseri Ameliyat Verilerinin K-Means Yöntemiyle Analizi, *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 2 (2), 1-10.
- Dolgun M. Ö., Ersel D., 2014, Doğrudan Pazarlama Stratejilerinin Belirlenmesinde Veri Madenciliği Yöntemlerinin Kullanımı, *İstatistikçiler Dergisi: İstatistik&Aktüerya*, 7, 1-13.
- Dondurmacı, G. A., 2014, Finans Sektöründe Veri Madenciliği Uygulaması, *The Journal of Academic Social Science*, 2 (1), 258-271.
- Dönmez, Ü., 2014, *Ankilozan Spondilit Hastalarında Global Postural Reeducasyon Programı Ve Radar Tedavisinin Hastalık Aktivitesi, Fonksiyonellik, Spinal Mobilite, Yaşam Kalitesi Ve Depresyon Üzerine Etkisi*, Uzmanlık Tezi, Tıp Fakültesi.
- Dunham, M.H., 2002, Data Mining, Introductory and Advanced Topics, <http://www.general.nsysu.edu.tw/gena/gena02/dm/part2.pdf>, [Ziyaret tarihi: 10 Mayıs 2015].
- Esfandiari, N., Babavalian, M.R., Moghadam, A.M.E., Tabar, V.K., 2014, Knowledge Discovery in Medicine: Current Issue And Future Trend, *Expert System with Applications*, 41, 4434-4463.
- Fakih S., Das T., 2006, A Methodology For Learning Efficient Approaches To Medical Diagnosis, *Institute of Electrical and Electronics Engineers*, 10 (2), 220-228.

- Farboidi, S., 2009, *Tıp Bilişiminin İstatistiksel Veri Madenciliği*, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü.
- Fayyad, U., Shapiro, G.P., Smyth, P., 1996, From Data Mining To Knowledge Discovery in Databases, *AI Magazine*, 17 (3), 37-54.
- Fayyad, U., Stolorz, P., 1997, Data Mining and KDD: Promise and Challenges, *Future Generation Computer System*, 13, 99-115.
- Frize, M., Ennett, C.M., Stevenson, M., Trigg, H.C.E., 2001, Clinical Decision Support System For Intensive Care Units: Using Artificial Neural Networks, *Medical Engineering & Physics*, 23, 217-225.
- Ganesh, S., 2002, Data Mining: Should It Be Included in The Statistics Curriculum?, *The Sixth International Conference on Teaching Statistics*, New Zealand.
- Gölce, A.C., 2010, *Veri Madenciliğinde Apriori Uygulaması ve Apriori Uygulamasının Farklı Veri Kümelerinde Uygulanması*, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü.
- Güllüoğlu, Ç. H., 2013, Türkbio Veri tabanında Kayıtlı Biyolojik İlaç Tedavisi Kullanan Ankilozan Spondilit Hastalarının Değerlendirilmesi, Dokuz Eylül Üniversitesi Tıp Fakültesi, syf 4.
- Gülten, A., Doğan, Ş., 2008, Genetik Algoritmalar Yönteminin Biyomedikal Veriler Üzerindeki Uygulamaları, *DAUM*, 7 (1), 12-16.
- Gunal, E. K., Sarvan, F. O., Kamali, S., Gul, A., Inanc, M., Carin, M., Konice, M., Aral, O., Ocal, L., 2008, Low Frequency of HLA-B27 in Ankylosing Spondylitis Patients From Turkey, *Joint Bone Spine*, 75 (3), 299-302.
- Han, J., Kamber, M., Pei, J., 2012, *Data Mining Concepts and Techniques*, Morgan Kaufmann, USA, ISBN: 978-0-12-381479-1.
- Hand, D., Mannila, H., Smyth, P., 2001, *Principles of Data Mining*, Massachusetts Institute of Technology, America, ISBN: 0-262-08290
- Hays, T., 2012, Medical Data Mining, <http://www.nist.gov/healthcare/upload/Hays-Medical-Data-Mining-slides-for-web.pdf>, [Ziyaret tarihi: 20 Mayıs 2012].
- Hornik, K., Buchta, C., Zeileis, A., 2009, Open-Source Machine Learning: R Meets Weka. *Computational Statistics*, 24(2), 225-232. doi:10.1007/s00180-008-0119-7
- Hothorn, T., Zeileis, A., 2015, *partykit: A Modular Toolkit for Recursive Partytioning in R*. *Journal of Machine Learning Research*, <http://EconPapers.RePEc.org/RePEc:inn:wpaper:2014-10>, [Ziyaret tarihi: 11 Mayıs 2015].
- Ian, H., Frank, E., 2005, *Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition*, Morgan Kaufmann, San Francisco.
- Imboden, J. B., Hellmann, D. B., Stone, J. H., (2014). *Romatoloji Tanı ve Tedavi* (T. Arasil, Çev.). Ankara: Güneş Tıp Kitapevleri. (2013).

- Inokuchi, A., Washio, T., Motoda, H., 1998, An apriori-based algorithm for mining frequent substructures from graph data, In Proc. 2000 *European Symp. Principle of Data Mining and Knowledge Discovery (PKDD'00)*, Lyon, France, 13–23.
- Karakoyun, M., Hacıbeyoğlu, 2014, M., Biyomedikal Veri Kümeleri ile Makine Öğrenmesi Sınıflandırma Algoritmalarının İstatistiksel Olarak Karşılaştırılması, *DEÜ Mühendislik Fakültesi Mühendislik Bilimleri Dergisi*, 16 (48), 30-41.
- Kellgren, J. H., Jeffrey, M. R., Ball, J., 1963, The Epidemiology of Chronic Rheumatism: Atlas of Standard Radiographs of Arthritis, Blackwell Scientific Publications, Oxford.
- Kınıklı, G., 2014, Spondiloartropatiler, <http://ichastaliklariromatoloji.medicine.ankara.edu.tr/files/2014/02/Spondiloartropatiler.pdf>, [Ziyaret tarihi: 25 Nisan 2015].
- Kocacık, F., 2003, Bilgi Toplumu ve Türkiye, *Cumhuriyet Üniversitesi Sosyal Bilimler Dergisi*, 27 (1), 1-10.
- Kononenko, I., 1994, Estimating Attributes: Analysis and Extensions of RELIEF, In: *European Conference on Machine Learning*, 171-182.
- Kumdereli, Ü. C., 2012, *Tıp Bilişimi ve Veri Madenciliği Uygulamaları: EEG Sinyallerindeki Epileptiform Aktiviteye Veri Madenciliği Yöntemlerinin Uygulanması*, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü.
- Kusiak, A., 2001, Decomposition in Data Mining: A Medical Case Study, *SPIE conference on Data Mining and Knowledge Discovery*, Orlando.
- Larose, D.T., 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley & Sons, Incorporated, New Jersey, ISBN: 0-471-66657-2.
- Li, X., Ye, N., 2001, Decision Tree Classifiers for Computer Intrusion Detection, *Parallel and Distributed Computing Practices*, 4 (2), 179-190.
- Liaw, A., Wiener, M., 2002, Classification And Regression By Random Forest, *The Newsletter of The R Project*, 2 (3), 18-22.
- Lim, H. J., Young, I. M., Lee, M. S., 2005, Effects Of Home-Based Daily Exercise Therapy On Joint Mobility, *Rheumatol*, 225–229.
- Linden, V.D.S., Valkenburg H.A., Cats, A., 1984, Evaluation of Diagnostic Criteria for Ankylosing Spondylitis. A Proposal for Modification of the New York Criteria, *Arthritis Rheum*, 27 (4), 361-368.
- MacQueen, J.B., 1967, Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, California, University of California Press, 281– 297.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, 2015, Cluster: Cluster Analysis Basics and Extensions. R package version 2.0.1.
- Mahajani, G., Aslandoğan, Y.A., Stan, T., 2003, Evidence Combination in Medical Data Mining, *International Conference on Information Technology: Coding and Computing (ITCC'04)*, 2, 465 – 469.

- McCullagh, P., Nelder, J.A., 1989, *Generalized Linear Models*, London Chapman and Hall.
- Neapolitan, R.E., 2003, *Learning Bayesian Networks*, Prentice Hall, Chicago.
- Neuenschwander, R., Ciurea, A., 2014, Gender Differences in Axial Spondyloarthritis, *World Journal of Rheumatology*, 4 (3), 35-43.
- Orhan, M., 2011, *Erken Aksiyal Spondiloartrit Hastalarda Hastalık Aktivitesini Değerlendirmede Bath Ankilozan Spondilit Hastalık Aktivite İndeksinin Geçerliliği*, Uzmanlık Tezi, Tıp Fakültesi
- Onen, F., Akar, S., Birlik, M., Sari, I., Khan, M.A., Gurler, O., Ergor, A., Manisali, M., Akkoc, N., 2008, Prevalence of Ankylosing Spondylitis and Related Spondyloarthritides in An Urban Area of Izmir, Turkey. *The Journal of Rheumatology*, 35 (2), 305-309.
- Ouali, A., Cherif, A. R., Krebs M. O., 2006, Data Mining Based Bayesian Networks for Best Classification, *Computational Statistics & Data Analysis*, 51 (2006), 1278-1299.
- Özdemir, S., 2014, *Ankilozan Spondilit ve Behçet Hastalığında Gözün Ön ve Arka Segment Yapılarının Değerlendirilmesi*, Uzmanlık Tezi, Tıp Fakültesi.
- Özkan, K., 2012, Sınıflandırma Ve Regresyon Ağacı Tekniği (SRAT) İle Ekolojik Verinin Modellenmesi, *SDÜ Orman Fakültesi Dergisi*, 13, 1-4.
- Özkan, Y., Erol, Ç., 2015, *Biyoenformatik DNA Mikrodizi Veri Madenciliği*, İstanbul, Papatya Yayıncılık.
- Özkan, Y., 2013, *Veri Madenciliği Yöntemleri*. İstanbul, Papatya Yayıncılık, ISBN: 978-975-6797-82-2.
- Öztürk, O., 2014, *Sigara Bağımlılığının Romatoid Artrit ve Ankilozan Spondilit Hastalarının Yaşam Kalitesi Üzerine Etkisi*, Uzmanlık Tezi, Tıp Fakültesi.
- Paget, S. A., Gibofsky, A., Beary, J.F., 2004, *Romatoloji ve Klinik Ortopedi El Kitabı*, (Y. Yazıcı, D. Erkan ve A. İnce, Çev.), İstanbul: Nobel Tıp Kitapevi, (2004).
- Pekelis, L., 2013, *Classification and Regression Trees: A Practical Guide for Describing a Dataset*, http://statweb.stanford.edu/~lpekelis/talks/13_datafest_cart_talk.pdf, [Ziyaret tarihi: 12 Nisan 2015].
- Poyraz, O., 2012, *Tıp'da Veri Madenciliği Uygulamaları: Meme kanseri veri seti analizi*, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü.
- Praphu, S., Venkatesan, N., 2007, *Data Mining and Warehousing*, New Age International, New Delhi, ISBN: 978-81-224-2432-4.
- Quinlan, J. R., 1993, *C4.5 programs for machine learning*, Morgan Kaufmann, London, ISBN: 1-55860-238-0
- Ragab, A.H.M., Noaman, A.Y., Ghamdi, A.S., Madbouly, A.I., 2014, A Comparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining, *IDEE'14 Proceedings of the 2014 Workshop on Interaction Design in Educational Environments*, 09 June 2014, Spain, New York, ACM, ISBN: 978-1-4503-3034-3, 106.

- Ren, X., Yan, D., Hong, T., 2015, Data Mining of Space Heating System Performance in Affordable Housing, *Building and Environment*, 89, 1-13.
- Ripley, B., 2014, *Tree: Classification and Regression Trees, R package version 1.0-35*, <http://CRAN.R-project.org/package=tree>, [Ziyaret tarihi: 1 Mayıs 2015].
- Rudwaleit M., van der Heijde D., Landewé R., Listing J., Akkoc N., Brandt J., Braun J., Chou C.T., Collantes E.E., Dougados M., Huang F., Gu J., Khan M.A., Kirazli Y., Maksymowych W.P., Mielants H., Sorensen I.J., Ozgocmen S., Roussou E., Valle O.R., Weber U., Wei J., Sieper J., 2009, The Development of Assessment of Spondyloarthritis International Society Classification Criteria for Axial Spondyloarthritis (Part II): Validation and Final Selection, *Annals of The Rheumatic Diseases*, 68 (6), 777-783.
- Saini, I., Singh D., Khosla, A., 2013, QRS Detection Using K-Nearest Neighbor Algorithm (Knn) And Evaluation On Standart ECG Databases, *Journal of Advanced Research*, 4, 331-344.
- Savaş, S., Topaloğlu, N., Yılmaz, M., 2012, Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, *İstanbul Ticeret Üniversitesi Fen Bilimleri Dergisi*, no. 21, 1-23.
- Schlimmer, J.C., 1985, *Model Import Car and Truck Specifications*, 1985 Ward's Automotive Yearbook.
- Sieper, J., Braun, J., Rudwaleit, M., Boonen, A., Zink, A. 2002, Ankylosing Spondylitis: An Overview, *Annals of the Rheumatic Diseases*, no 61 (3:iii), 8-18.
- Tan, P.N., Steinbach, M., Kumar, V., 2006 , *Introduction to Data Mining*, Addison-Wesley.
- Therneau, T., Atkinson, B., Ripley, B., 2015, *rpart: Recursive Partitioning and Regression Trees, R package version 4.1-9.*, <http://CRAN.R-project.org/package=rpart>, [Ziyaret tarihi: 10.05.2015].
- Tuğ, E., 2005, *Genetik Algoritmalar ile Tıbbi Veri Madenciliği*, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü.
- Tüylü, T. B., 2013, *Ankilozan Spondilit Hastalarında Sindesmofit Gelişimi ile İlişkili Çeşitli Biyobelirteçlerin Değerlendirilmesi*, Uzmanlık Tezi, Tıp Fakültesi.
- Us, A.K., Armangil, M., Kalem, M., 2014, *Romatolojik Hastalıklarda Omurga Tutulumu ve Cerrahi Yaklaşım*, <http://ichastaliklariromatoloji.medicine.ankara.edu.tr/files/2014/02/Romatolojik-Hastal%C4%B1klarda-Omurga-Tutulumu.pdf>, [Ziyaret tarihi: 10 Mayıs 2015].
- Watts, J., Powell S., Lawrence R. L., Hilker, T., 2011, Improved Classification of Conservation Tillage Adoption Using High Temporal And Synthetic Satellite Imagery, *Remote Sensing of Environment*, 115, 66-75.
- Yıldırım, P., Çeken, Ç., Hassanpour, R., Tolun, M. R., 2010, Prediction of Similarities Among Rheumatic Diseases, *Journal of Medicine Systems*, 36, 1485-1490.

EKLER**EK-1: TEZ ÇALIŞMASINDA KULLANILAN VERİ SETİNİN ÖN İŞLEMEDEN SONRA .CSV UZANTILI GÖSTERİMİ:**

ERAPrs30187;rs30187Talel;PTGER4rs10440635;rs10440635Aalel;IL23rs11209032;rs11209032Aalel;ERAPrs27044;rs27044Galel;Sex;Ilksemptomkategorik;BaslangicYasiKategorik;HLAB27;Kalcitutulumu;Periferikartrit;Servikaltutulum;Extraarterver;Uveit;Entezit;NSAID;LEF;CurrentTNFi;TNFi;Hastaliksuresi;AgirVShafif

TC;Var;AA;Var;AG;Var;CC;Yok;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;18;Agir

TC;Var;AA;Var;AG;Var;CG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;10;Hafif

CC;Yok;GG;Yok;AA;Var;CC;Yok;Erkek;EkstraArt;Orta;Pozitif;Var;Var;Yok;Var;Var;Yok;Evet;Evet;Hayir;Hayir;10;Agir

CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiyal;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Hayir;Hayir;Evet;Evet;8;Agir

TC;Var;GG;Yok;AG;Var;CC;Yok;Erkek;PeriferikAksiyal;Orta;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;6;Hafif

TC;Var;GG;Yok;GG;Yok;CC;Yok;Erkek;PeriferikAksiyal;Orta;Pozitif;Var;Var;Var;Var;Var;Yok;Evet;Hayir;Hayir;Hayir;16;Agir

TC;Var;AG;Var;AA;Var;CG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;11;Agir

TC;Var;AA;Var;AA;Var;CC;Yok;Erkek;PeriferikAksiyal;Yasli;Pozitif;Var;Var;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;23;Agir

TC;Var;AA;Var;AG;Var;CG;Var;Erkek;PeriferikAksiyal;Orta;Pozitif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;20;Hafif

TC;Var;AG;Var;AG;Var;CG;Var;Erkek;Aksiyal;Orta;Negatif;Var;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;8;Hafif

CC;Yok;AG;Var;AG;Var;CC;Yok;Kadin;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;8;Hafif

TC;Var;AA;Var;AA;Var;CG;Var;Kadin;Aksiyal;Orta;Negatif;Yok;Yok;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;5;Hafif

TC;Var;AG;Var;GG;Yok;CG;Var;Kadin;Aksiyal;Orta;Negatif;Yok;Var;Var;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif

- CC;Yok;AG;Var;AA;Var;CC;Yok;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;5;Hafif
- TT;Var;GG;Yok;AG;Var;CG;Var;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Var;Var;Yok;Evet;Hayir;Hayir;Hayir;6;Hafif
- CC;Yok;AA;Var;AG;Var;CG;Var;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;6;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Kadin;PeriferikArtrit;Orta;Pozitif;Var;Var;Yok;Var;Var;Yok;Evet;Hayir;Hayir;Hayir;15;Hafif
- TC;Var;AG;Var;AA;Var;CC;Yok;Kadin;Aksiya;Orta;Negatif;Var;Var;Yok;Var;Yok;Var;Evet;Hayir;Evet;Evet;3;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Kadin;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Var;Yok;Yok;Evet;Hayir;Evet;Evet;14;Agir
- TC;Var;AG;Var;GG;Yok;CG;Var;Kadin;PeriferikAksiya;Orta;Pozitif;Yok;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;5;Hafif
- TT;Var;GG;Yok;AG;Var;GG;Var;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;4;Hafif
- TT;Var;AG;Var;AA;Var;CG;Var;Kadin;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;8;Hafif
- CC;Yok;AA;Var;AG;Var;CC;Yok;Kadin;PeriferikAksiya;Orta;Negatif;Var;Var;Var;Var;Yok;Var;Hayir;Hayir;Evet;Evet;12;Agir
- TT;Var;GG;Yok;AG;Var;CG;Var;Kadin;PeriferikAksiya;Orta;Negatif;Var;Var;Var;Var;Yok;Var;Hayir;Hayir;Evet;Evet;11;Hafif
- TC;Var;AG;Var;AG;Var;CG;Var;Kadin;Aksiya;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;9;Hafif
- TT;Var;AG;Var;AA;Var;CG;Var;Kadin;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;5;Hafif
- TC;Var;AG;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiya;Genc;Pozitif;Var;Yok;Var;Var;Var;Yok;Evet;Hayir;Hayir;Hayir;12;Agir
- TC;Var;AG;Var;AA;Var;CC;Yok;Erkek;Aksiya;Genc;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;14;Agir
- TC;Var;AG;Var;AG;Var;CC;Yok;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;4;Hafif
- CC;Yok;AG;Var;AA;Var;CC;Yok;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;25;Agir
- TC;Var;AG;Var;AG;Var;CG;Var;Erkek;Aksiya;Orta;Pozitif;Yok;Var;Var;Var;Var;Var;Evet;Hayir;Hayir;Hayir;15;Hafif

- TC;Var;AG;Var;AG;Var;CG;Var;Erkek;Aksiyal;Genc;Pozitif;Yok;Yok;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;14;Hafif
- CC;Yok;GG;Yok;AA;Var;CC;Yok;Kadin;Aksiyal;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;4;Agir
- TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;PeriferikAksiyal;Genc;Pozitif;Var;Var;Yok;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;3;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiyal;Orta;Pozitif;Var;Var;Var;Var;Var;Var;Evet;Hayir;Hayir;Evet;8;Agir
- TC;Var;GG;Yok;AG;Var;CG;Var;Erkek;PeriferikAksiyal;Orta;Negatif;Yok;Var;Yok;Var;Yok;Var;Evet;Evet;Hayir;Hayir;8;Hafif
- TC;Var;AG;Var;GG;Yok;CG;Var;Kadin;Aksiyal;Genc;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;11;Hafif
- CC;Yok;GG;Yok;GG;Yok;CC;Yok;Kadin;Aksiyal;Genc;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- TC;Var;AG;Var;AA;Var;CG;Var;Kadin;PeriferikAksiyal;Orta;Negatif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;5;Hafif
- TT;Var;GG;Yok;AG;Var;GG;Var;Kadin;Aksiyal;Genc;Negatif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;3;Hafif
- TT;Var;GG;Yok;AA;Var;GG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;27;Agir
- TT;Var;AG;Var;AA;Var;GG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;6;Agir
- TC;Var;AG;Var;AA;Var;CC;Yok;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;11;Agir
- TC;Var;AA;Var;GG;Yok;CG;Var;Kadin;PeriferikAksiyal;Orta;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;6;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Kadin;Aksiyal;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;9;Hafif
- CC;Yok;AA;Var;AG;Var;CC;Yok;Kadin;Aksiyal;Yasli;Pozitif;Yok;Var;Var;Var;Var;Var;Evet;Hayir;Evet;Evet;10;Hafif
- TT;Var;AG;Var;AG;Var;CC;Yok;Kadin;Aksiyal;Genc;Pozitif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- TC;Var;AA;Var;AG;Var;CC;Yok;Kadin;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;5;Hafif
- CC;Yok;GG;Yok;GG;Yok;CC;Yok;Kadin;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;5;Hafif

- TC;Var;GG;Yok;AA;Var;CC;Yok;Kadin;PeriferikAksiyal;Orta;Pozitif;Yok;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;12;Hafif
- TC;Var;AG;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiyal;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- TC;Var;GG;Yok;AG;Var;CG;Var;Erkek;PeriferikArtrit;Genc;Pozitif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;11;Hafif
- TC;Var;AA;Var;GG;Yok;CG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Var;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;8;Agir
- TC;Var;AG;Var;AA;Var;CG;Var;Erkek;PeriferikAksiyal;Orta;Negatif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;1;Hafif
- TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Var;Var;Yok;Var;Hayir;Hayir;Evet;Evet;11;Hafif
- TC;Var;AG;Var;AG;Var;CG;Var;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;6;Agir
- TC;Var;GG;Yok;AG;Var;CG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;7;Agir
- CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;PeriferikArtrit;Genc;Negatif;Yok;Yok;Var;Var;Var;Var;Hayir;Hayir;Evet;Evet;14;Agir
- CC;Yok;AG;Var;AA;Var;CC;Yok;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;6;Hafif
- CC;Yok;GG;Yok;AG;Var;CC;Yok;Erkek;PeriferikAksiyal;Genc;Pozitif;Var;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif
- CC;Yok;AG;Var;AG;Var;CG;Var;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;9;Agir
- TC;Var;AA;Var;AG;Var;CC;Yok;Kadin;PeriferikAksiyal;Orta;Pozitif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- TC;Var;GG;Yok;AG;Var;CG;Var;Kadin;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;4;Hafif
- TC;Var;AA;Var;AG;Var;CG;Var;Kadin;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Evet;5;Hafif
- TC;Var;AG;Var;GG;Yok;CG;Var;Kadin;PeriferikAksiyal;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;10;Hafif
- TT;Var;AG;Var;AG;Var;GG;Var;Kadin;PeriferikAksiyal;Orta;Negatif;Yok;Var;Var;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;4;Hafif
- TC;Var;AG;Var;AG;Var;CG;Var;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;8;Agir

- TC;Var;AG;Var;GG;Yok;CG;Var;Kadin;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Var;Var;Var;Evet;Hayir;Evet;Evet;2;Hafif
- TC;Var;AG;Var;AG;Var;CG;Var;Kadin;PeriferikAksiya;Orta;Pozitif;Yok;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;12;Hafif
- CC;Yok;GG;Yok;GG;Yok;CC;Yok;Erkek;PeriferikAksiya;Genc;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;4;Hafif
- TC;Var;AG;Var;AG;Var;CC;Yok;Kadin;PeriferikAksiya;Orta;Negatif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;13;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Kadin;PeriferikAksiya;Orta;Pozitif;Yok;Var;Var;Var;Var;Var;Evet;Hayir;Evet;Evet;7;Hafif
- TT;Var;AG;Var;GG;Yok;CC;Yok;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Var;Var;Var;Evet;Hayir;Hayir;Hayir;33;Hafif
- TC;Var;AG;Var;AG;Var;CG;Var;Kadin;Aksiya;Orta;Pozitif;Var;Yok;Yok;Var;Yok;Var;Evet;Hayir;Evet;Evet;3;Hafif
- TT;Var;AG;Var;AG;Var;CG;Var;Kadin;PeriferikAksiya;Orta;Pozitif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;11;Agir
- TC;Var;GG;Yok;AG;Var;CC;Yok;Kadin;PeriferikAksiya;Orta;Negatif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;13;Hafif
- CC;Yok;AA;Var;AG;Var;CC;Yok;Kadin;PeriferikAksiya;Orta;Negatif;Var;Var;Var;Yok;Yok;Yok;Evet;Evet;Evet;Evet;6;Hafif
- TC;Var;AA;Var;GG;Yok;CG;Var;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;7;Agir
- CC;Yok;GG;Yok;AA;Var;CC;Yok;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;9;Hafif
- TC;Var;AG;Var;AA;Var;CC;Yok;Erkek;Aksiya;Yasli;Pozitif;Yok;Yok;Var;Var;Var;Yok;Evet;Hayir;Hayir;Hayir;4;Agir
- CC;Yok;AG;Var;AG;Var;CG;Var;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;7;Hafif
- CC;Yok;GG;Yok;AA;Var;CC;Yok;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;25;Agir
- CC;Yok;GG;Yok;AG;Var;CC;Yok;Kadin;Aksiya;Orta;Pozitif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;12;Hafif
- TC;Var;AG;Var;AG;Var;CG;Var;Kadin;PeriferikAksiya;Orta;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;5;Hafif
- TC;Var;AG;Var;GG;Yok;CC;Yok;Kadin;PeriferikAksiya;Orta;Pozitif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;16;Hafif

TC;Var;GG;Yok;AG;Var;CG;Var;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Ev
et;Hayir;Hayir;Hayir;4;Hafif

TC;Var;AA;Var;AG;Var;CG;Var;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Ev
et;Hayir;Evet;Evet;11;Agir

TT;Var;AA;Var;GG;Yok;CG;Var;Erkek;Aksiya;Genc;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Ev
et;Hayir;Hayir;Hayir;9;Hafif

CC;Yok;GG;Yok;AA;Var;CC;Yok;Kadin;Aksiya;Genc;Negatif;Yok;Yok;Yok;Var;Var;Yok;E
vet;Hayir;Evet;Evet;2;Hafif

TT;Var;GG;Yok;GG;Yok;CG;Var;Kadin;PeriferikAksiya;Orta;Negatif;Var;Var;Var;Var;Var;
Var;Evet;Hayir;Hayir;Hayir;14;Agir

CC;Yok;GG;Yok;AA;Var;CC;Yok;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;E
vet;Hayir;Evet;Evet;8;Hafif

TC;Var;AA;Var;AG;Var;CG;Var;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet
;Hayir;Evet;Evet;8;Agir

TC;Var;GG;Yok;AA;Var;CG;Var;Erkek;Aksiya;Orta;Pozitif;Yok;Var;Var;Var;Yok;Var;Evet;
Hayir;Evet;Evet;11;Agir

TC;Var;GG;Yok;AA;Var;CC;Yok;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Yok;Var;Yok;Yok;
Yok;Evet;Hayir;Evet;Evet;7;Agir

TT;Var;AA;Var;GG;Yok;GG;Var;Erkek;PeriferikAksiya;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;
Yok;Evet;Hayir;Evet;Evet;6;Hafif

TC;Var;AG;Var;GG;Yok;CC;Yok;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Var;Yok;V
ar;Evet;Hayir;Hayir;Hayir;6;Hafif

CC;Yok;AA;Var;GG;Yok;CC;Yok;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Yok;Yok;Yok;
Yok;Evet;Hayir;Evet;Evet;3;Agir

CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiya;Yasli;Pozitif;Yok;Yok;Var;Yok;Yok
;Yok;Evet;Hayir;Hayir;Hayir;7;Hafif

CC;Yok;AG;Var;GG;Yok;CC;Yok;Erkek;PeriferikAksiya;Genc;Pozitif;Var;Var;Var;Yok;Yok
;Yok;Evet;Hayir;Evet;Evet;9;Hafif

TT;Var;AG;Var;AG;Var;CC;Yok;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Ev
et;Hayir;Evet;Evet;3;Hafif

TT;Var;AA;Var;AA;Var;CG;Var;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Ev
et;Hayir;Evet;Evet;4;Hafif

CC;Yok;AG;Var;AA;Var;CC;Yok;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Eve
t;Hayir;Evet;Evet;8;Agir

TC;Var;AG;Var;AG;Var;CG;Var;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Var;Var;Yok;Var;Evet;
Hayir;Evet;Evet;9;Agir

TT;Var;AA;Var;AA;Var;GG;Var;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;29;Agir

TC;Var;AA;Var;AA;Var;CG;Var;Kadin;Aksiya;Genc;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif

CC;Yok;GG;Yok;AA;Var;CC;Yok;Kadin;PeriferikAksiya;Yasli;Negatif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;5;Hafif

CC;Yok;GG;Yok;GG;Yok;CC;Yok;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;9;Agir

CC;Yok;GG;Yok;AG;Var;CC;Yok;Kadin;PeriferikArtrit;Orta;Negatif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;7;Hafif

TC;Var;AG;Var;AG;Var;CC;Yok;Erkek;Aksiya;Yasli;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;17;Agir

TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;15;Agir

CC;Yok;AA;Var;GG;Yok;CG;Var;Erkek;PeriferikArtrit;Orta;Negatif;Var;Yok;Yok;Var;Var;Yok;Evet;Hayir;Hayir;Hayir;5;Hafif

TC;Var;AG;Var;AA;Var;CC;Yok;Erkek;Aksiya;Genc;Pozitif;Yok;Yok;Yok;Var;Var;Yok;Evet;Hayir;Evet;Evet;4;Hafif

TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;Aksiya;Orta;Pozitif;Var;Var;Var;Var;Var;Yok;Evet;Hayir;Evet;Hayir;13;Agir

CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;3;Hafif

TT;Var;AG;Var;AG;Var;CG;Var;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;16;Hafif

TC;Var;GG;Yok;AA;Var;CG;Var;Kadin;PeriferikAksiya;Orta;Negatif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;3;Hafif

TC;Var;GG;Yok;GG;Yok;CG;Var;Kadin;PeriferikAksiya;Orta;Negatif;Yok;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Hayir;4;Agir

CC;Yok;AA;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiya;Orta;Negatif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;5;Hafif

TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;22;Agir

TC;Var;AA;Var;AG;Var;CC;Yok;Kadin;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;11;Hafif

CC;Yok;AG;Var;GG;Yok;CC;Yok;Erkek;PeriferikAksiya;Orta;Pozitif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;5;Hafif

- TC;Var;AA;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiyal;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;3;Hafif
- TC;Var;AG;Var;AG;Var;CG;Var;Erkek;Aksiyal;Genc;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;21;Agir
- TC;Var;AA;Var;GG;Yok;CC;Yok;Erkek;PeriferikArtrit;Orta;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;7;Hafif
- TC;Var;AG;Var;GG;Yok;CC;Yok;Erkek;PeriferikAksiyal;Genc;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;4;Hafif
- TT;Var;AG;Var;AG;Var;CG;Var;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;7;Hafif
- TT;Var;AG;Var;GG;Yok;GG;Var;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;1;Hafif
- TC;Var;AG;Var;AA;Var;CG;Var;Erkek;Aksiyal;Orta;Pozitif;Var;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;8;Hafif
- TC;Var;GG;Yok;AG;Var;CC;Yok;Erkek;Aksiyal;Genc;Pozitif;Var;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;9;Hafif
- CC;Yok;AG;Var;AA;Var;CC;Yok;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;6;Hafif
- CC;Yok;AA;Var;GG;Yok;CC;Yok;Erkek;PeriferikAksiyal;Orta;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;6;Hafif
- CC;Yok;AG;Var;AG;Var;CG;Var;Kadin;PeriferikAksiyal;Orta;Pozitif;Var;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;9;Hafif
- TC;Var;GG;Yok;AG;Var;CG;Var;Kadin;PeriferikArtrit;Orta;Negatif;Yok;Var;Yok;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;7;Hafif
- TC;Var;AG;Var;AG;Var;CG;Var;Kadin;PeriferikAksiyal;Orta;Pozitif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;7;Hafif
- CC;Yok;AA;Var;GG;Yok;CC;Yok;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;14;Agir
- TC;Var;AA;Var;AA;Var;CG;Var;Kadin;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;9;Hafif
- CC;Yok;GG;Yok;AA;Var;CC;Yok;Kadin;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;7;Hafif
- TT;Var;AG;Var;GG;Yok;CG;Var;Erkek;PeriferikArtrit;Orta;Pozitif;Var;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;8;Hafif
- TC;Var;GG;Yok;AA;Var;CG;Var;Kadin;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif

- CC;Yok;AG;Var;AG;Var;CC;Yok;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;9;Hafif
- TC;Var;AA;Var;GG;Yok;CG;Var;Kadin;Aksiya;Genc;Pozitif;Var;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;4;Hafif
- TC;Var;GG;Yok;GG;Yok;CG;Var;Kadin;Aksiya;Genc;Pozitif;Yok;Yok;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;23;Hafif
- TT;Var;AG;Var;AA;Var;GG;Var;Erkek;PeriferikAksiya;Orta;Pozitif;Yok;Var;Var;Var;Yok;Yok;Evet;Hayir;Evet;Evet;5;Agir
- TC;Var;AG;Var;AG;Var;CG;Var;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;10;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;Aksiya;Genc;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;Aksiya;Genc;Pozitif;Yok;Yok;Yok;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;2;Hafif
- TC;Var;GG;Yok;AG;Var;CC;Yok;Erkek;PeriferikAksiya;Genc;Pozitif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;7;Hafif
- TC;Var;AA;Var;GG;Yok;CC;Yok;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;7;Hafif
- CC;Yok;AG;Var;AA;Var;CC;Yok;Kadin;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;16;Hafif
- TT;Var;AG;Var;GG;Yok;CC;Yok;Erkek;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;17;Agir
- TC;Var;AG;Var;AA;Var;CG;Var;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- TC;Var;GG;Yok;AA;Var;CG;Var;Erkek;Aksiya;Genc;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Hayir;Hayir;Evet;Evet;35;Agir
- TT;Var;AA;Var;GG;Yok;GG;Var;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif
- TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif
- TC;Var;AG;Var;GG;Yok;CC;Yok;Erkek;PeriferikAksiya;Genc;Pozitif;Var;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- CC;Yok;GG;Yok;GG;Yok;CC;Yok;Erkek;Aksiya;Orta;Negatif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;3;Hafif
- TT;Var;GG;Yok;AG;Var;CG;Var;Erkek;Aksiya;Orta;Pozitif;Var;Var;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;26;Agir

- TC;Var;AG;Var;AA;Var;CC;Yok;Kadin;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Hayir;Hayir;Hayir;Hayir;12;Hafif
- TC;Var;GG;Yok;AG;Var;CG;Var;Kadin;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Kadin;PeriferikAksiya;Orta;Pozitif;Var;Var;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;8;Hafif
- TC;Var;GG;Yok;GG;Yok;CG;Var;Erkek;Aksiya;Orta;Pozitif;Var;Var;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;15;Agir
- TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;Aksiya;Genc;Pozitif;Yok;Yok;Yok;Var;Var;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- CC;Yok;AA;Var;AG;Var;CC;Yok;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;9;Hafif
- CC;Yok;AG;Var;GG;Yok;CC;Yok;Erkek;Aksiya;Orta;Pozitif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Evet;49;Agir
- TC;Var;AA;Var;AA;Var;CG;Var;Erkek;PeriferikAksiya;Orta;Negatif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;8;Hafif
- CC;Yok;GG;Yok;AG;Var;CC;Yok;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Var;Var;Yok;Evet;Hayir;Hayir;Hayir;4;Hafif
- TC;Var;AG;Var;AA;Var;CG;Var;Erkek;PeriferikArtrit;Genc;Pozitif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;6;Hafif
- TT;Var;AA;Var;GG;Yok;GG;Var;Kadin;PeriferikArtrit;Genc;Pozitif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Evet;3;Hafif
- TC;Var;AG;Var;GG;Yok;CC;Yok;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Var;Var;Yok;Yok;Evet;Hayir;Hayir;Hayir;9;Agir
- TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;Aksiya;Genc;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;5;Hafif
- TC;Var;AA;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiya;Genc;Pozitif;Var;Var;Yok;Var;Var;Yok;Evet;Hayir;Evet;Evet;6;Hafif
- CC;Yok;AA;Var;GG;Yok;CC;Yok;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;6;Hafif
- TC;Var;AA;Var;AA;Var;CG;Var;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Hayir;Hayir;Evet;Evet;7;Agir
- TC;Var;AG;Var;AG;Var;CG;Var;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Var;Yok;Var;Evet;Hayir;Evet;Evet;7;Agir
- TC;Var;GG;Yok;AG;Var;CG;Var;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;6;Hafif

TC;Var;AG;Var;AA;Var;CC;Yok;Erkek;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Evet;4;Hafif

CC;Yok;GG;Yok;GG;Yok;CC;Yok;Kadin;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif

TT;Var;GG;Yok;AG;Var;GG;Var;Erkek;PeriferikAksiya;Orta;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;8;Hafif

TC;Var;AG;Var;AG;Var;CG;Var;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;5;Agir

TC;Var;AG;Var;AA;Var;CG;Var;Erkek;PeriferikAksiya;Genc;Negatif;Yok;Var;Yok;Var;Var;Yok;Hayir;Hayir;Evet;Evet;3;Hafif

TC;Var;AG;Var;AG;Var;CG;Var;Erkek;PeriferikArtrit;Genc;Pozitif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;14;Hafif

TC;Var;AG;Var;AG;Var;CG;Var;Kadin;PeriferikAksiya;Orta;Pozitif;Yok;Yok;Var;Var;Var;Var;Evet;Hayir;Evet;Evet;4;Hafif

CC;Yok;GG;Yok;GG;Yok;CC;Yok;Kadin;Aksiya;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif

TC;Var;AG;Var;GG;Yok;CG;Var;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;6;Hafif

CC;Yok;AA;Var;GG;Yok;CC;Yok;Kadin;Aksiya;Genc;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif

CC;Yok;AA;Var;GG;Yok;CC;Yok;Kadin;PeriferikAksiya;Orta;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;9;Hafif

CC;Yok;AA;Var;AG;Var;CC;Yok;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;5;Hafif

TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;PeriferikArtrit;Orta;Negatif;Yok;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;6;Hafif

CC;Yok;GG;Yok;GG;Yok;CC;Yok;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Var;Var;Yok;Evet;Hayir;Hayir;Hayir;4;Hafif

CC;Yok;AA;Var;GG;Yok;CG;Var;Erkek;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif

CC;Yok;AG;Var;GG;Yok;CC;Yok;Kadin;Aksiya;Orta;Pozitif;Var;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;6;Hafif

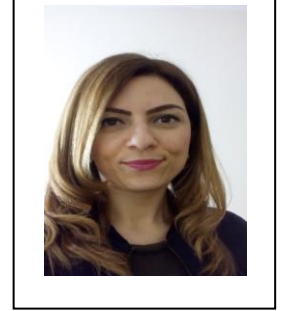
TC;Var;AA;Var;AA;Var;CG;Var;Erkek;PeriferikArtrit;Genc;Pozitif;Var;Var;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;14;Hafif

CC;Yok;AG;Var;AA;Var;CC;Yok;Kadin;Aksiya;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;5;Hafif

- TC;Var;AA;Var;GG;Yok;CC;Yok;Kadin;PeriferikAksiyal;Orta;Pozitif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;6;Hafif
- TC;Var;AG;Var;AG;Var;CG;Var;Erkek;PeriferikAksiyal;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;12;Agir
- CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;20;Agir
- TT;Var;GG;Yok;GG;Yok;GG;Var;Kadin;PeriferikAksiyal;Orta;Negatif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;4;Hafif
- TC;Var;AG;Var;AG;Var;CC;Yok;Kadin;Aksiyal;Orta;Pozitif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiyal;Orta;Pozitif;Var;Var;Var;Var;Var;Yok;Evet;Hayir;Evet;Evet;9;Agir
- CC;Yok;AG;Var;AG;Var;CC;Yok;Kadin;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif
- TC;Var;AG;Var;GG;Yok;CG;Var;Kadin;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- CC;Yok;AG;Var;AG;Var;CC;Yok;Erkek;PeriferikAksiyal;Orta;Pozitif;Var;Var;Var;Var;Var;Var;Evet;Hayir;Hayir;Evet;9;Agir
- TC;Var;GG;Yok;AG;Var;CC;Yok;Kadin;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;4;Hafif
- TT;Var;AG;Var;AG;Var;CG;Var;Kadin;PeriferikArtrit;Genc;Negatif;Var;Var;Var;Var;Yok;Var;Hayir;Hayir;Hayir;Evet;5;Hafif
- TC;Var;AA;Var;AG;Var;CG;Var;Erkek;PeriferikAksiyal;Orta;Pozitif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;2;Hafif
- TT;Var;GG;Yok;GG;Yok;CG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif
- CC;Yok;AG;Var;GG;Yok;CC;Yok;Erkek;Aksiyal;Genc;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;21;Hafif
- CC;Yok;GG;Yok;AG;Var;CC;Yok;Erkek;PeriferikAksiyal;Orta;Negatif;Yok;Yok;Yok;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;2;Hafif
- TT;Var;GG;Yok;AG;Var;GG;Var;Erkek;Aksiyal;Yasli;Pozitif;Yok;Yok;Yok;Var;Yok;Var;Evet;Hayir;Evet;Evet;3;Agir
- TC;Var;AG;Var;AA;Var;CG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Var;Yok;Yok;Yok;Hayir;Hayir;Hayir;2;Hafif
- CC;Yok;AA;Var;GG;Yok;CC;Yok;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;1;Hafif

- TC;Var;AG;Var;AA;Var;CC;Yok;Erkek;Aksiyal;Orta;Pozitif;Var;Var;Var;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;10;Agir
- CC;Yok;GG;Yok;AG;Var;CC;Yok;Erkek;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;5;Hafif
- TC;Var;AG;Var;AA;Var;CG;Var;Erkek;PeriferikAksiyal;Orta;Negatif;Yok;Var;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;5;Hafif
- TC;Var;GG;Yok;AG;Var;CG;Var;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Var;Yok;Var;Evet;Hayir;Hayir;Hayir;3;Hafif
- TT;Var;GG;Yok;AG;Var;CG;Var;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;9;Agir
- TT;Var;AG;Var;AG;Var;CG;Var;Erkek;PeriferikAksiyal;Orta;Pozitif;Var;Var;Var;Yok;Yok;Yok;Evet;Hayir;Evet;Evet;6;Agir
- TT;Var;AA;Var;GG;Yok;CG;Var;Erkek;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- TC;Var;AG;Var;GG;Yok;CG;Var;Erkek;Aksiyal;Orta;Pozitif;Yok;Var;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;9;Agir
- CC;Yok;AG;Var;AG;Var;CC;Yok;Kadin;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif
- TC;Var;AA;Var;AA;Var;CG;Var;Kadin;Aksiyal;Yasli;Pozitif;Yok;Yok;Var;Var;Yok;Var;Evet;Hayir;Evet;Evet;7;Hafif
- TC;Var;AA;Var;AA;Var;CG;Var;Kadin;Aksiyal;Yasli;Pozitif;Yok;Yok;Yok;Var;Yok;Yok;Evet;Hayir;Evet;Evet;11;Hafif
- TC;Var;GG;Yok;AG;Var;CC;Yok;Kadin;Aksiyal;Orta;Negatif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;3;Hafif
- TC;Var;GG;Yok;AG;Var;CC;Yok;Kadin;Aksiyal;Genc;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif
- TC;Var;AG;Var;GG;Yok;CG;Var;Kadin;Aksiyal;Orta;Pozitif;Yok;Yok;Yok;Yok;Yok;Yok;Evet;Hayir;Hayir;Hayir;2;Hafif

ÖZGEÇMİŞ



Kişisel Bilgiler

Adı Soyadı	Yeşim ATASOY
Uyruğu	T.C.
Doğum tarihi, Yeri	1984, İstanbul
Telefon	(0)555 621 43 85
E-mail	yesim.atasoy41@gmail.com

Eğitim

Derece	Kurum/Anabilim Dalı/Programı	Yılı
Yüksek Lisans	İ.Ü. Fen Bilimleri Enstitüsü / Enformatik	2015
Tezsiz Yüksek Lisans	İ.Ü. Fen Bilimleri Enstitüsü / Biyoloji Öğretmenliği	2010
Lisans	İ.Ü. Fen Fakültesi / Biyoloji	2008

Makaleler / Bildiriler

Özdemir Ş., Lanpir E., Atasoy Y., Gülseçen S., 2014, "New Technology Experience In Turkey: The Case Of Bitcoin ", 5th International Future-Learning Conference on Innovations in Learning for the Future 2014: e-Learning, İstanbul.
Atasoy Y., Erol Ç. S., Reis Z. A., Gülseçen S., 2014, Bilgi Ekonomisi Açısından Biyoçeşitliliğin Ekonomik Değeri,. Akademik Bilişim 2014, Mersin.
Çatar Ö., Kaymaz R., Özata S., Atasoy Y., Kitapçı N. Ş., Köksal L., Mumcu G., 2013, Elektronik Tıbbi Kayıtlarda Kalite ve Kullanıcı Tatmini, 4. Uluslararası Sağlık ve Hastane Yönetimi Kongresi.
Atasoy Y., Çatar Ö., Şişman N., Köksal L., Mumcu G, Tarım M., 2012, Sağlık Yönetimi Eğitiminde Sağlık Bilişimi, 3. Uluslararası Sağlık ve Hastane Yönetimi Kongresi, Antalya.