

T.C.
KAFKAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI

TÜRKİYE SAĞLIK ARAŞTIRMASININ VERİ
MADENCİLİĞİ TEKNİKLERİ İLE İNCELENMESİ

YÜKSEK LİSANS TEZİ

HAZIRLAYAN
ONUR CAN

Danışman
DOÇ. DR. ÖTÜKEN SENER

Kars – 2017



T.C.

**KAFKAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI**

**TÜRKİYE SAĞLIK ARAŞTIRMASININ VERİ
MADENCİLİĞİ TEKNİKLERİ İLE İNCELENMESİ**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN

ONUR CAN

Danışman ve Jüri Üyeleri

Doç. Dr. Ötüken SENER

Prof. Dr. Erkan OKTAY

Yrd. Doç. Dr. Mehmet Emin KARABAYIR

Kars – 2017

KAFKAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Onur CAN tarafından hazırlanan *Türkiye Sağlık Araştırmasının Veri Madenciliği Teknikleri ile İncelenmesi* başlıklı bu çalışma, 29/06/2017 tarihinde yapılan tez savunma sınavı sonucunda [başarılı/başarısız] bulunarak jürimiz tarafından İşletme Anabilim Dalı'nda Yüksek Lisans tezi olarak (oy çokluğuyla/birliğiyle) kabul/red edilmiştir.

TEZ JÜRİSİ ÜYELERİ (Unvanı, Adı ve Soyadı)

Başkan	: Prof. Dr. Erkan OKTAY	İmza:
Üye	: Doç. Dr. Ötügen ŞENER	İmza:
Üye	: Yrd. Doç. Dr. Mehmet Emin KARABAYIR	İmza:

ONAY

Bu tezin kabulü Sosyal Bilimler Enstitüsü Yönetim Kurulunun/...../..... tarih ve/..... sayılı kararı ile onaylanmıştır.

Unvanı Adı Soyadı
Enstitü Müdürü

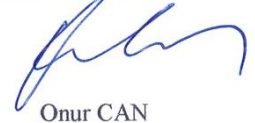
BİLİMSEL ETİK BİLDİRİMİ

Yüksek Lisans tezi olarak hazırladığım *Türkiye Sağlık Araştırmasının Veri Madenciliği Teknikleri ile İncelenmesi* adlı çalışmanın öneri aşamasından sonuçlanmasına kadar geçen süreçte bilimsel etiğe ve akademik kurallara özenle uyduğumu, tez içindeki tüm bilgileri bilimsel ahlak ve gelenek çerçevesinde elde ettiğimi, tez yazım kurallarına uygun olarak hazırladığım bu çalışmamda doğrudan veya dolaylı olarak yaptığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu beyan ederim.

Scientific Ethic Statement

I declare that I complied with the rules of academic and scientific ethics from the proposal stage to the process of completion of the study titled *Investigation of Turkey Healty Survey with Data Mining Techniques* as a Master Thesis I prepared, that I obtained all information in term Project with the framework of scientific ethics and traditions, that I showed sources to the each quotation I made directly or indirectly in this study I prepared as a term Project in accordance with the writing rules and works which I used have been shown in the bibliography.

29/06/2017



Onur CAN

İÇİNDEKİLER

ÖZET.....	v
ABSTRACT	vi
ÖNSÖZ.....	vii
KISALTMALAR	viii
TABLO LİSTESİ.....	ix
ŞEKİL LİSTESİ.....	x
GİRİŞ	1
1. GENEL KAVRAMLAR VE BİLGİ KEŞFİ	2
1.1. Veriden Bilgiye Giden Yol	2
1.2. Veri Tabanı ve Veri Ambarı Kavramları	4
1.3. Veri Tabanlarından Bilgi Keşfi	6
1.3.1. Problemin Tanımlanması	7
1.3.2. Verilerin Elde Edilmesi.....	7
1.3.3. Veri Seçme ve Dönüştürme	8
1.3.4. Veri Madenciliği Adımı	8
1.3.5. Bilgi Sunumu.....	8
2. VERİ MADENCİLİĞİ	9
2.1. Veri Madenciliğinin Tarihsel Gelişim Süreci	11
2.2. Veri Madenciliğinin İlişkili Olduğu Disiplinler	12
2.3. Veri Madenciliği ve İstatistik	13
2.4. Veri Madenciliği Süreci.....	14
2.4.1. Problemi Anlama	16
2.4.2. Veriyi Anlama	16
2.4.3. Veriyi Hazırlama	16
2.4.3.1 Verinin Toplanması	17
2.4.3.2. Verinin Temizlenmesi.....	17
2.4.3.3. Verinin Bütünleştirilmesi	17
2.4.3.4. Verinin Dönüştürülmesi	18
2.4.3.5. Verinin İndirgenmesi	18
2.4.4. Modelleme.....	18

2.4.5. Modelin Değerlendirilmesi	19
2.4.6. Modelin Kullanılması	19
2.5. Veri Madenciliği Tekniklerinin Kullanıldığı Alanlar.....	19
2.6. Veri Madenciliği Çalışmalarında Karşılaşılan Sorunlar.....	22
2.6.1. Veri Tabanı Boyutu	23
2.6.2. Kayıp Veri.....	23
2.6.3. Gürültülü Veri	23
2.6.4. Dinamik Veri.....	24
2.6.5. Artık Veri	24
2.7. Veri Madenciliğinde Kullanılan Yazılımlar	24
2.7.1. IBM SPSS Modeller	24
2.7.2. Enterprise Miner	26
2.7.3. Intelligent Miner	26
2.7.4. WEKA.....	26
2.7.5. Analysis Manager	27
2.7.6. IBM SPSS Statistics	27
2.8. Veri Madenciliği Uygulama Örnekleri.....	27
2.8.1. Tıp Alanında Uygulanan Veri Madenciliği Çalışmaları.....	28
2.8.2. Mühendislik Alanında Uygulanan Veri Madenciliği Çalışmaları.....	28
2.8.3. Eğitim Alanında Uygulanan Veri Madenciliği Çalışmaları.....	29
2.8.4. Finans, Borsa ve Bankacılık Alanında Uygulanan Veri Madenciliği Çalışmaları	30
2.8.5. Ticari Alanda Uygulanan Veri Madenciliği Çalışmaları	30
3. VERİ MADENCİLİĞİ YÖNTEMLERİ.....	31
3.1. Sınıflama ve Regresyon Modelleri.....	32
3.1.1. Karar Ağaçları	33
3.1.1.1. Karar Ağaçlarında Kullanılan Algoritmalar.....	37
3.1.1.1.1. CLS Algoritması	37
3.1.1.1.2. ID3 Algoritması	37
3.1.1.1.3. C4.5 Algoritması.....	38
3.1.1.1.4. C5.0 Algoritması.....	39

3.1.1.1.5. C&R Tree Algoritması.....	39
3.1.1.1.6. CHAID Algoritması.....	40
3.1.1.1.7. QUEST Algoritması.....	41
3.1.1.1.8. Random Forest Algoritması	42
3.1.2. Yapay Sinir Ağları.....	42
3.1.3. Genetik Algoritmalar	44
3.1.4. K-En Yakın Komşu	45
3.1.5. K-Ortalamlar Kümeleme Algoritması.....	47
3.1.6. Naive Bayes Sınıflandırıcısı.....	47
3.2. Kümeleme	49
3.3. Birliktelik Kuralları	50
3.4. Ardışık Zaman Örüntüler	51
3.5. Veri Madenciliği Çalışmalarında Kullanılan İstatistiksel Teknikler	52
3.5.1 Regresyon Analizi	52
3.5.2. Lojistik Regresyon.....	53
3.5.3. Diskriminat Analizi	54
3.5.4. Faktör Analizi.....	55
3.5.5. Ki-Kare Testi.....	56
3.5.6. Zaman Serileri Analizi.....	57
4. GEREÇ VE YÖNTEM	58
4.1. Türkiye Sağlık Araştırması	58
4.2. Problemin Tanımlanması	59
4.3. Analiz Öncesi Veri Seti	59
4.4. Veri Ön İşleme	60
4.4.1. Veri Temizleme	61
4.4.2. Veri Dönüştürme	61
4.4.3. Veri İndirgeme	62
4.5. Veri Madenciliği Sürecinde Kullanılacak Değişkenlerin Özellikleri ve Tanımlayıcı İstatistikler	64
4.6. Veri Setinin Bölünmesi.....	72
4.7. Veri Seti Üzerinde Karar Ağacı Algoritmalarının Uygulanması	73

4.7.1. CHAID Algoritması.....	74
4.7.2. QUEST Algoritması	96
4.7.3. C&R Tree Algoritması.....	103
5. SONUÇ	111
KAYNAKLAR.....	114
EKLER.....	127
ÖZGEÇMİŞ	137



KAFKAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI
TÜRKİYE SAĞLIK ARAŞTIRMASININ VERİ MADENCİLİĞİ TEKLİKLERİ İLE
İNCELENMESİ
YÜKSEK LİSANS TEZİ
Onur CAN
Doç. Dr. Ötüken SENER
2017 –XI + 137

ÖZET

Bu çalışmanın temel amacı, bireylerin genel sağlık durumunu etkileyen faktörleri veri madenciliği teknikleri ile belirlemektir. Bu amaç doğrultusunda, 2014 yılında Türkiye İstatistik Kurumu tarafından yapılan Türkiye Sağlık Araştırmasına ait B grubu mikro veri seti temin edilmiştir. Daha sonra çeşitli manipülasyon yöntemleriyle veri madenciliği analizlerine uygun hale getirilen veri seti üzerinde CHAID, QUEST ve C&R Tree algoritmaları uygulanmıştır. Analiz sonuçlarına göre kullanılan üç algoritma da genel sağlık durumu üzerinde en çok etkiye sahip faktörlerin, 6 ay ya da daha uzun süren veya sürmesi beklenen sağlık problemi ve herhangi bir sağlık problemi nedeniyle günlük faaliyetlerin 6 ay ve daha uzun süre kısıtlanma durumu olduğunu tespit etmiştir. Karar ağacı algoritmaları birbirleri ile kıyaslandığında ise CHAID algoritmasının hem QUEST hem de C&R Tree algoritmasına göre daha iyi sonuçlar verdiği gözlenmiştir.

Anahtar Kelimeler: Veri Madenciliği, Genel Sağlık Durumu, Karar Ağaçları

Kafkas University
Graduate School of Social Sciences
Department of Business Administration
Investigation of Turkey Healthy Survey with Data Mining Techniques
Master's Thesis
Onur CAN
Assoc. Prof. Dr. Ötüken SENER
2017 – XI + 137

ABSTRACT

The main purpose of this study is to determine the factors affecting general state of health of individuals by data mining techniques. In accordance with this purpose, B group micro data set of the Turkish Health Research conducted by the Turkish Statistical Institute in 2014 was provided. Afterwards, CHAID, QUEST and C&R Tree algorithms were applied on the data set that was made suitable for data mining analysis with data pre-processing methods. The three algorithms used according to the results of the analysis found that the factors with the greatest impact on the general state of health, the health problem which is expected to last for 6 months or longer, and the daily activities are restricted for 6 months or longer due to any health problem. When the decision tree algorithms are compared with each other, it is observed that the CHAID algorithm gives better results than both the QUEST and C&R Tree algorithms.

Keywords: Data Mining, General State of Health, Decision Trees

ÖNSÖZ

Çalışmamın her aşamasında bilgisi ve tecrübesiyle desteklerini esirgemeyen değerli hocam Doç. Dr. Ötüken Senger'e teşekkürlerimi ve şükranlarımı sunarım. Yüksek lisans eğitimim boyunca desteği ve uzman görüşleri ile yardımcı olan TÜİK Kars Bölge Müdürü Adnan BEDLEK'e, tez aşamasında yardımlarını esirgemeyen Dr. Ali Kemal Çelik'e ve Mustafa ERDEN'e teşekkürlerimi sunarım.

Beni bu günlere getiren ve her zaman yanımda hissettiğim annem Ayşe CAN, babam Şerafettin CAN, kardeşim Oğuz CAN'a, tanıştığımız ilk günden beri sonsuz sevgisi, sabrı ve anlayışıyla hayatıma güzellik katan kıymetli eşim Duygu CAN'a ve üzerimde emeği olan, isimlerini buraya sığdıramadığım herkese teşekkürü bir borç bilirim.

...../...../2017

Onur CAN

KISALTMALAR

CHAID	: Chi-Square Automatic Interaction Detector
CLS	: Concept Learning System
CRISP-DM	: The Cross Industry Standart Process for Data Mining
C&R Tree	: Classification and Regression Trees
IBM	: International Business Machines
ID3	: Iterative Dichotomiser 3
QUEST	: Quick, Unbiased, Efficient, Statistical Tree
SAS	: Statistical Analysis Software
SPSS	: Statistical Package for the Social Sciences
TDIDT	: Top Down Induction of Decision Trees
TSA	: Türkiye Sağlık Araştırması
TÜİK	: Türkiye İstatistik Kurumu

TABLO LİSTESİ

<u>Tablo No</u>	<u>Sayfa</u>
Tablo 1.1.1: Veri, Enformasyon ve Bilgi Arasındaki Farklar	3
Tablo 2.1.1: Veri İşleme Tekniklerinin Gelişimi.....	11
Tablo 2.3.1: İstatistik ile Veri Madenciliğinin Farkları	14
Tablo 4.4.2.1: Veri Dönüştürme İşlemi	61
Tablo 4.4.3.1: Bağımlı Değişken ile Diğer Değişkenler Arasındaki İlişkinin İncelenmesi.....	63
Tablo 4.5.1: Değişken Tablosu.....	65
Tablo 4.7.1.1: CHAID Algoritmasında Kullanılan Değişkenlerin Önem Değerleri .	75
Tablo 4.7.1.2: CHAID Algoritmasının Performans Ölçüm Sonuçları.....	96
Tablo 4.7.2.1: QUEST Algoritmasında Kullanılan Değişkenlerin Önem Değerleri .	97
Tablo 4.7.2.2: QUEST Algoritmasının Performans Ölçüm Sonuçları	102
Tablo 4.7.3.1: C&R Tree Algoritmasında Kullanılan Değişkenlerin Önem Değerleri	104
Tablo 4.7.3.2: C&R Tree Algoritmasının Performans Ölçüm Sonuçları.....	110

ŞEKİL LİSTESİ

<u>Sekil No</u>	<u>Sayfa</u>
Şekil 1.2.1: Veri ambarı kullanım süreci	5
Şekil 1.3.1: Bilgi Keşfinin Aşamaları.....	7
Şekil 2.2.1: Veri Madenciliğinin İlişkili Olduğu Disiplinler	12
Şekil 2.4.1: Veri Madenciliği Adımları	15
Şekil 2.7.1.1: IBM SPSS Modeller Çalışma Ekranı.....	25
Şekil 3.1.1.1: Basit Bir Karar Ağacı Yapısı.....	35
Şekil 3.1.1.2: Veri Sınıflandırma İşlemi.....	36
Şekil 3.1.2.1: Yapay Sinir Ağı	43
Şekil 3.1.4.1: K-En Yakın Komşu Yöntemi	46
Şekil 3.2.1: Kümeleme Analizinde Kümelerin Konumları.....	49
Şekil 4.3.1: Veri Kalitesinin İncelenmesi	60
Şekil: 4.4.3.1: Veri Ön İşlemi Adımından Sonra Veri Setinin Kalitesi	64
Şekil 4.5.1: Değişkenlerin Seçilmesi.....	65
Şekil 4.6.1: Veri Setinin Bölünmesi	72
Şekil 4.7.1: Auto Classifier İşlemcisi ile Karar Ağacı Algoritmalarının Belirlenmesi	73
Şekil 4.7.2: Kurulan Modelin Ekran Görüntüsü	74
Şekil 4.7.1.1: CHAID Algoritması İle Oluşturulan Karar Ağacının Görünümü	76
Şekil 4.7.1.2: CHAID algoritması ile Oluşan Karar Ağacının Başlangıç Düğümü ...	76
Şekil 4.7.1.3: CHAID Algoritmasının 1. Düzümü Göre Dallanması.....	78
Şekil 4.7.1.4: CHAID Algoritmasının 6. Düzümü Göre Dallanması.....	82
Şekil 4.7.1.5: CHAID Algoritmasının 7. Düzümü Göre Dallanması.....	86
Şekil 4.7.1.6: CHAID Algoritmasının 8. Düzümü Göre Dallanması.....	88
Şekil 4.7.1.7: CHAID Algoritmasının 9. Düzümü Göre Dallanması.....	91
Şekil 4.7.1.8: CHAID Algoritması İle Üretilen Kural Setinin Bir Kısmı	94
Şekil: 4.7.2.1: Quest Algoritması ile Oluşturulan Karar Ağacı.....	98
Şekil 4.7.2.2: QUEST Algoritması İle Üretilen Kural Seti	101

Şekil 4.7.3.1: C&R Tree Algoritması ile Oluşturulan Karar Ağacı	105
Şekil 4.7.3.2: C&R Tree Algoritması İle Üretilen Kural Seti.....	108



GİRİŞ

Hayatımıza hızla giren bilgisayar teknolojileri, günümüzde insanoğlunun vazgeçilmezi olmuştur. Bilgisayarlar internetle buluştuktan sonra her an her yerde veri üretilmekte ve bu veriler depolanmaktadır.

Bilim ve teknolojinin baş döndüren bir hızla ilerlemesiyle veri depolama sistemleri de gelişmiştir. Kısa bir süre önce kilobyte, megabyte ve gigabyte ile ifade edilen depolama kapasiteleri artık terabyte, petabyte hatta exabyte ile ifade edilmektedir.

Sürekli büyüyen veriler, veri yığınlarını oluşturmakta ve kullanılmayan veri yığınları veri çöplüklerine dönüşmektedir. Bu noktada ortaya çıkan veri madenciliği, büyük veri yığınları içinde gizli kalmış bilgileri gün ışığına çıkartmak için özel yöntemler sunmaktadır. Veri madenciliği çalışmalarında kullanılan yöntemlerden karar ağaçları, büyük veri setlerine kolayca uygulanabilmesi ve yüksek güvenilirlikte sonuçlar vermesi sebebiyle en çok tercih edilen tekniklerinden birisidir.

Bu çalışma beş bölümden oluşmaktadır. Birinci bölümde veri madenciliğine giriş yapmadan önce gerekli ön bilgiler verilmiş ve bilgi keşfinin aşamaları gösterilmiştir. İkinci bölümde, veri madenciliği detaylı bir şekilde anlatılmış ve literatürde yer alan başarılı veri madenciliği uygulamaları aktarılmıştır. Üçüncü bölümde ise veri madenciliğinde kullanılan teknikler açıklanarak formüllerin gösterimi yapılmıştır. Dördüncü bölümde, gerçek bir veri seti üzerinde karar ağacı algoritmaları uygulanmış ve analiz sonuçları ayrıntılı bir şekilde incelenmiştir. Beşinci bölümde, kullanılan karar ağacı algoritmalarının performansları karşılaştırılıp hangi algoritmanın daha iyi sonuç verdiği yorumlanmıştır.

Bu tez çalışmasında, Türkiye İstatistik Kurumunun (TÜİK) 2014 yılında yapmış olduğu Türkiye Sağlık Araştırmasına (TSA) ait mikro veri seti üzerinde karar ağacı algoritmalarını uygulamak ve kişinin genel sağlık durumunu etkileyen faktörleri belirlemek amaçlanmıştır.

1. GENEL KAVRAMLAR VE BİLGİ KEŞFİ

1.1. Veriden Bilgiye Giden Yol

İnsanoğlu, sahip olduğu bilgiyi kullanma yeteneği ile diğer canlılardan ayrılmaktadır. Bilgiyi üretip etkin olarak kullanabilen toplumlar yüksek refah düzeyine ulaşabilmektedir. Bu sebeple toplumların bilgi seviyesi, analiz becerilerini geliştiren eğitim sistemleri ve bilgi temelli karar verme yetkinlikleri gelişmişlik düzeyi açısından belirleyici rol oynamaktadır (Argüden ve Erşahin, 2008: 9).

Faydalı bilgiye ulaşmak için izlenen yol, veri toplama işlemi ile başlar. Veri manasında kullanılan data kelimesi çoğul yapıdadır ve kökü datum tekilinden gelmektedir (<http://en.wikipedia.org/wiki/Data>, Erişim tarihi: 04 Ağustos 2016). İstatistik terimleri sözlüğünde veri; “deneyler ya da gözlemler sonucunda elde edilen nicel ya da nitel değerler” olarak tanımlanmaktadır. (Akman, 2010: 1). Kavramsal anlamda ise veri, kayıt altına alınmış her türlü durum, fikir ve olaylardır. Veri, günlük yaşam döngüsü içinde doğal olarak oluşmakta ya da çeşitli cihazların olağan aktiviteleri sonucu meydana gelmektedir. Bu durumda, çevremizdeki her nesne bir veri olarak algılanabilir (Öğüt 2005: 1-2). Kısaca; sayım, ölçüm, deney, gözlem veya araştırma ile elde edilen değerlere veri diyebiliriz. Eğer veri sayısal bir değere sahipse nicel veriler, sayısal bir değere sahip değilse nitel veriler olarak adlandırılır. Nicel veriler, kaynağından sayım veya ölçüm yolu ile toplanır. Nitel veriler ise metin, resim, ses ya da video gibi verilerdir (Şık 2014: 2). Verinin öğelerini tanımlayan verilere ise metaveri adı verilmektedir. Kısaca metaveri, verinin verisidir.

Veriler anlamlandırılmamış ham kayıtlar oluşu için tek başlarına herhangi bir anlam ifade etmezler. Verilerin anlam kazanması için analiz aşamasından geçmeleri gerekir. Veri analizi; verileri toplama, düzenleme ve istatistiksel yöntemler uygulayarak anlamlı kararlar verebilme ve geçerli sonuçlar çıkarabilme süreci olarak tanımlanabilir (Büyüköztürk 2015).

Enformasyon ise ilişkilendirilerek anlamlandırılmış veri olarak tanımlanmaktadır. Veriden daha fazla içeriğe sahip olan enformasyon, sözlü, görsel ya da yazılı bir mesajdır (Kiremitçi, 2005: 7).

Veriden bilgiye giden yolun son noktası olan bilgi, en genel anlamıyla veriyi ve enformasyonu işe yarar faaliyete dönüştürme yeteneğidir (Özer, 2013: 73). Bilgi, kişinin enformasyonu algılayıp onu anlamlandırmasıyla ortaya çıkar. Örneğin, bir markette satılan herhangi bir ürünün fiyatı ve stok miktarı birer veridir. Markette, son bir ayda hangi ürünün ne kadar satıldığı gibi işlenmiş veriler enformasyon örneği olabilir. Firmanın sattığı bir ürün için yapılan kampanya sonucu, satılan ürün miktarındaki oransal değişim ise bilgi olarak kabul edilebilir.

Bu bilgiler üzerinden, veri, enformasyon ve bilginin farkları Tablo 1.1.1’de olduğu gibi özetlenebilir (Canbazoglu, 2000: 9).

Tablo 1.1.1: Veri, Enformasyon ve Bilgi Arasındaki Farklar

Nitelik	Veri	Enformasyon	Bilgi
Ayrıntı Düzeyi	Düşük düzeyde ayrıntı	Bir araya getirilmiş, tümleşik ayrıntı	Yüksek düzeyde soyutlanmış ayrıntı
Bağlam	Bağlam yok	Tam bağlam	Genişletilmiş bağlam
Kapsam	Dar kapsam	Belirli bağlam ile sınırlı kapsam	Enformasyon kapsamının ötesinde genişletilmiş kapsam
Zamanlılık	Zamanlılık yok	Sınırlı zamanlılık	Sonsuz

Kaynak: Canbazoglu (2000: 9)

1.2. Veri Tabanı ve Veri Ambarı Kavramları

Hızla gelişen teknoloji sayesinde büyük boyutlarda, karmaşık verileri saklamak oldukça kolaylaşmıştır. İşletmeler, kamu kuruluşları ve bilgisayar teknolojilerini kullanan fertlerin sakladıkları veriler devasa boyutlara ulaşmıştır. Çeşitli veri kaynaklarından akan bu veriler düzenli bir şekilde saklanmalı ve kullanıcı istediği an ihtiyacı olan verilere rahatlıkla ulaşabilmelidir.

Aralarında ilişki olan veri yığınlarını içeren, veriye ulaşım imkânı veren ve veriyi yönetmeye destek olan yazılım programları kümesine veri tabanı denir (Hand and Kamber 2006: 9). Veri tabanı meydana getirmek, çoğaltmak, saklamak, güncellemek ve yönetmek için kullanılan sistemlere, “veri tabanı yönetim sistemleri” adı verilir. Bir veri tabanı yönetim sisteminin yapısını tasarlayan ve özelliklerini inceleyen bilim ise bilgi bilimidir (Döşlü 2008: 4).

Banka müşterilerinin kredi kartı ile yaptıkları harcamalar, şirketlerin ödeme ve teslimat bilgileri, bir okulun eğitim planlaması, hatta telefon rehberinde kayıtlı kişilerin numaraları birer veri tabanı örneğidir. Veri tabanları doğru kullanıldığında kullanışlı bir depo görevi üstlenebilir ancak yanlış kullanımı sonucunda gereksiz veri yığınları halini alır.

Veri madenciliğinin amacı büyük boyutlardaki verileri incelemek olduğundan veri tabanları ile doğrudan ilişkilidir. Veri tabanları birbirlerinden farklı yerdedir. Veriler birleştirilip analiz edilerek raporlanma ihtiyacı hissedildiğinde, verilerin bulunduğu kaynakları birleştirmeye yardımcı olan veri ambarları devreye girer (Ukuş 2014: 4).

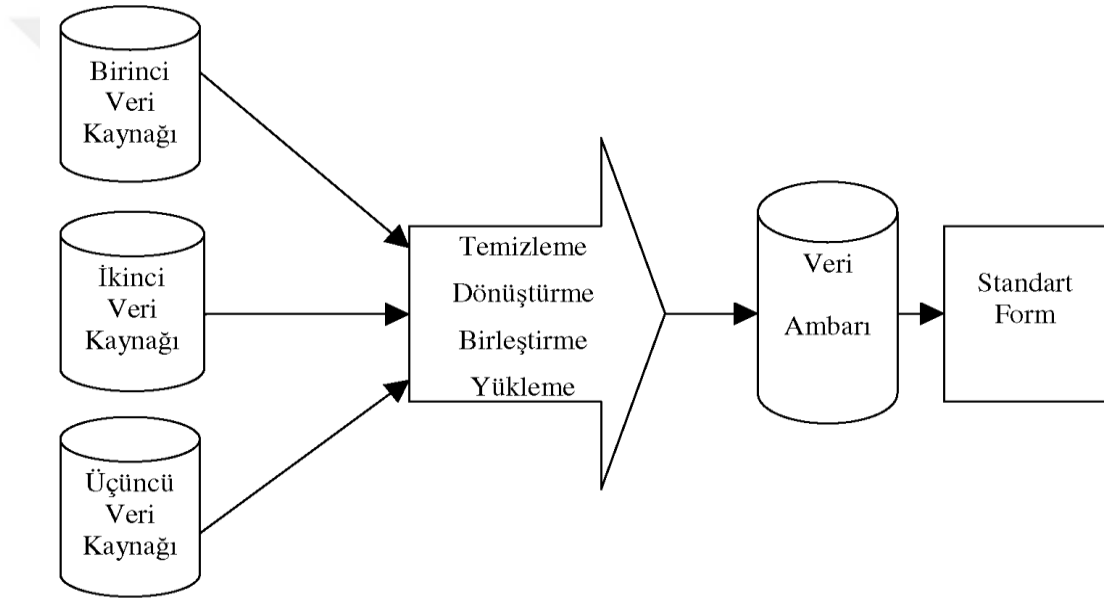
Veri ambarları, birden fazla veri tabanından alınıp birleştirilen verilerin toplandığı depolardır (Mackinnon ve Ned, 1999: 260). Veri ambarları, verileri okumak için oluşturulmuştur. Veri ambarında yer alan veriler, kullanıcının analiz yapmasını kolaylaştıracak biçimde depolanmaktadır. Burada analiz; istatistiki hesaplamalar, raporlama, sorgulama ya da karar destek sistemlerini kapsamaktadır (Akın, 2008:

40). Veri ambarındaki verilerin en önemli özelliği, kullanıcının karar almasına yardımcı olmasıdır (Ballou ve Tayi 1999: 73).

Veri tabanlarında bulunan veri genellikle gerçek zamanlıdır ve değiştirilebilir ancak veri ambarında tarihsel veriler mevcuttur, bu nedenle değişken yapıda değildir. Veri ambarındaki tarihsel veriler raporlama, analiz ve geleceği tahmin etmek amacıyla kullanılır (Düzgünoğlu 2006: 4).

Veri ambarlarının kullanım süreci Şekil 1.2.1’ de gösterilmiştir.

Şekil 1.2.1: Veri ambarı kullanım süreci



Kaynak: Akın (2008: 40)

Çeşitli veri kaynaklarında depolanmış veriler; temizlenme, dönüştürme ve birleştirme işlemlerinden geçtikten sonra veri ambarlarına yüklendiği için kullanıma hazır haldedir. Veri ambarlarındaki veriler, çeşitli işlemlerden geçip kullanıcıya sunulduğundan geçerliliğini uzun süre korumaktadır.

Veri ambarı ve veri tabanı arasındaki farklar şunlardır (Şeker, 2015: 6-7) :

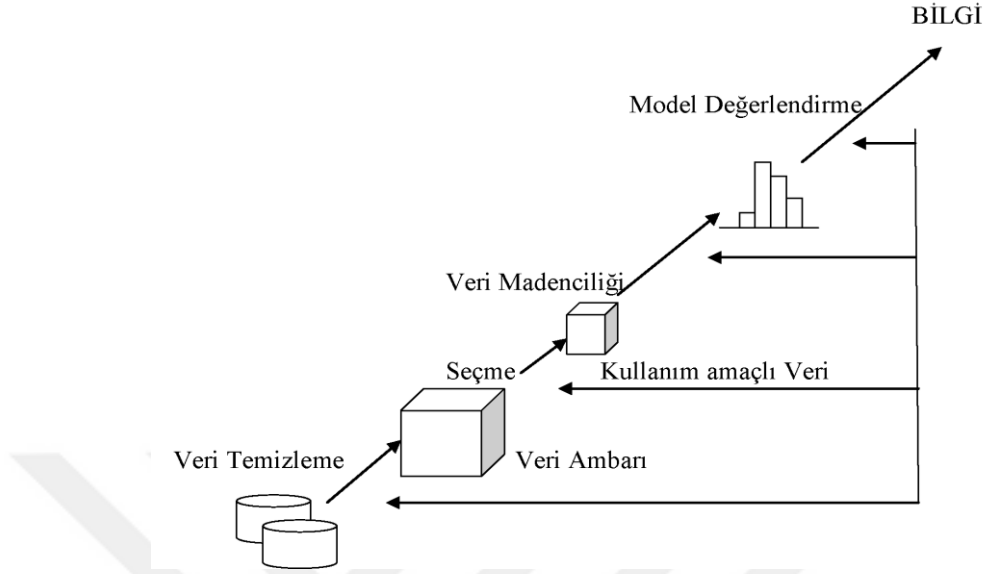
- Veri ambarları ile veri tabanları arasındaki en önemli fark, değişik kaynaklardan besleniyor olmalarıdır. Örneğin banka şubelerinden gelen veriler ve kredi kartlarından yapılan harcama verileri, veri tabanlarında tablo formatında farklı veri kaynakları halinde depolanır. Depolanan bu veri kaynakları birleştirilip bir amaç doğrultusunda raporlanması ve verinin daha faydalı kullanılması için geliştirilmiş sistemler ise veri ambarlarıdır.
- Veri ambarı, özelleşmiş ve işlenmiş verileri saklarken veri tabanları, bütün verileri saklamaktadır.
- Veri ambarları değişimleri takip edip kendini zaman içinde günceller. Veri tabanlarında güncelleme yapabilmek için çeşitli algoritmalar kullanılır.
- Veri ambarlarına, veri tabanlarından veri akışı tek yönlüdür.

1.3. Veri Tabanlarından Bilgi Keşfi

Uzun zamandır veri analizleri, istatistiksel tekniklerle yapılmaktadır. Hızla artan veri miktarı karşısında istatistiksel yöntemlerin tek başına yetersiz kaldığı görülmüştür. Ham verilerin artış hızı, bu verilerden elde edilmiş faydalı bilgilerin artış hızının çok üstündedir. Veri tabanlarından bilgi keşfi olarak bilinen süreç, sürekli artan verilerden daha fazla ve yararlı bilgiler üretebilmek için ortaya çıkmıştır (Kılınç, 2015: 11).

Veri tabanlarında gizlenmiş faydalı bilgiyi ortaya çıkartmak oldukça zor görünse de günümüzde gelişen teknoloji sayesinde bilgiye ulaşmak oldukça kolaylaşmıştır. Sumathi ve Sivanandam (2006: 9), veri tabanlarından bilgi keşfini; bilgisayar yardımıyla büyük verilerin analizi sırasında yapılan işlemler olarak tanımlamıştır. Veri tabanlarından bilgi keşfinin aşamaları Şekil 1.3.1’de gösterilmiştir.

Şekil 1.3.1: Bilgi Keşfinin Aşamaları



Kaynak: Ukuş (2014: 11)

Bilgi keşfinin aşamaları, aşağıda maddeler açıklanmıştır.

1.3.1. Problemin Tanımlanması

Problemin ne olduğunun tanımlanması ve uygulamanın hangi amaçla yapılacağının ifade edilmesi, veri madenciliği çalışmalarının en önemli adımıdır. Bu nedenle var olan problem üzerine yoğunlaşarak hedef açıkça belirtilmelidir. Daha sonra elde edilecek sonuçların başarı yüzdelerinin hangi yöntemle ve nasıl hesaplanacağı tanımlanmalıdır. Ayrıca yanlış tahminlerde artacak maliyetlere ve doğru tahminlerde elde edilen kazanımlara ait öngörülere de burada yer verilmelidir (Şengür, 2013: 14).

1.3.2. Verilerin Elde Edilmesi

Verilerin elde edilmesi aşamasında araştırmacı, işine yaramayan verileri elemektedir. Eğer başka bir kaynaktan da işine yaramayacak veri gelirse, daha önce elediği verilerle bu verileri birleştirip tek bir veri seti haline getirmektedir. Ayrıca veri seti, gürültülü ve tutarsız verilerden temizlenerek bir sonraki aşama için hazır hale getirilmektedir (Elabiad, 2013: 8).

1.3.3. Veri Seçme ve Dönüştürme

Veri seçme, çeşitli veri kaynaklarından gelen verilerin seçilip üzerinde çalışılacak bir veri dosyası oluşturulması işlemidir. Veri seçme işlemi, kayıp, hatalı ve uç değerlerin düzeltilmesi işlemlerini de içerir. Veri dönüştürme ise birleştirme, düzeltme, normalleştirme ve genelleştirme işlemlerinden bir ya da birkaçını içerebilir. En çok kullanılan veri dönüştürme işlemi, veri normalleştirmedir (Albayrak ve Yılmaz, 2009: 35). Bu aşama, veri setinin veri madenciliği algoritmalarının uygulanabileceği hale dönüştüğü aşamadır.

1.3.4. Veri Madenciliği Adımı

Sürecin bu adımında, hangi veri madenciliği tekniğinin kullanılacağına karar verilir ve faydalı bilgiye ulaşmak için veri analizi gerçekleştirilir.

Veri madenciliği adımı, kullanıcı ve bilgi tabanı ile etkileşim halindedir. İlginç örüntüler kullanıcıya gösterilir ve istenirse bu örüntüler, bilgi tabanına da kaydedilebilir. Veri madenciliği adımının bu süreçte nerede başlayıp nerede bittiği oldukça bulanıktır. Genellikle, kendinden önceki ve sonraki adımları kapsamaktadır (Pektaş, 2013: 101).

1.3.5. Bilgi Sunumu

Veri madenciliği adımında elde edilen bilgiler, bu aşamada kullanıcılara sunulmaktadır. Bilgi; tablo, şekil ve grafiklerle zenginleştirilerek de sunulabilir.

2. VERİ MADENCİLİĞİ

Bilgisayar teknolojilerinin keşfi ile dijital ortamda veri saklama süreci başlamıştır. Veri depolama teknolojilerinin hızlı gelişimi, devasa boyutta ve çok çeşitlilikte veri depolanmasına imkân tanımıştır. Kredi kartı kullanım verileri, bankacılık işlemleri, marketlerden yapılan alışveriş bilgileri ve telefon ile yapılan görüşmelerin detayları gibi günlük hayatta ortaya çıkan verilerin yanı sıra tıbbi veriler ve uydu gözlemlerinden elde edilip depolanan çeşitli veriler farklı amaçlar için kurulmuş veri tabanı yönetim sistemleri sayesinde dijital ortamlarda depolanmakta ve yönetilmektedir. Ayrıca özel işletmeler ve devlet kurumlarında geliştirilmiş veri tabanları da, ilgili kurum ve kuruluşların işleyişini sağlıklı bir şekilde devam ettirmeleri için gerekli verileri depolamaktadır (Aydın, S. 2007: 3).

Günümüzde verilerin çığ gibi büyümesi, onlara anlam kazandırmayı daha da zorlaştırmaktadır. Bu sebeple geleneksel yöntemlerle veri yığınlarının içinde gizlenmiş ve kurum stratejileri için fayda sağlayacak bilgileri ortaya çıkarmak her geçen gün daha da zorlaşmıştır. Geline nokta veri madenciliği çalışmaları oldukça önem kazanmıştır (Uzar, 2016: 25-26).

Madenciliğin kelime anlamı, yeryüzünün gizli ve değerli kaynaklarının ortaya çıkarılması sürecidir. Veri ile madencilik kelimelerinin birlikteliği, veri yığınları içinde ilk anda fark edilemeyen değerli bilgilerin bulunup çıkartılması düşüncesini çağrıştırmaktadır (Giudici, 2003: 1). Veri madenciliği, kaynaklardan veri toplayıp onların üzerinde çeşitli algoritmalar yardımıyla analiz yöntemleri uygulamanın ötesinde yer alan bir süreçtir.

Veri madenciliğinin geleneksel araştırma yöntemlerinden farkı, büyük miktarda veri içeren veri tabanları içindeki değerli bilgiye ulaşma çabasıdır. Buna rağmen veri madenciliği yöntemlerinin temelini istatistiksel teknikler oluşturmaya devam etmektedir (Olson ve Delen, 2008: 5). Veri madenciliğinin literatürde farklı tanımları bulunmaktadır. Bu tanımlardan bazılarına aşağıda yer verilmiştir.

Veri madenciliği konusunun önde gelen isimlerinden Hand (1997: 112) veri madenciliğini; istatistik, makine öğrenimi, veri tabanı teknolojisi ve örüntü tanıma ile ilişkili yeni bir disiplin olarak tanımlamıştır.

Jacobs (1999: 8), ham verilerin sunamadığı faydalı bilgiyi ortaya çıkaran analiz sürecine, veri madenciliği demiştir.

D.T. Larose ve C.D. Larose (2005: 42) ise veri madenciliğini, veriler arasındaki anlamlı ve değerli ilişkileri saptayıp veriyi yeni bir şekilde özetlemek olarak tanımlamıştır.

Statistical Package for the Social Sciences (SPSS) firması veri madenciliğini, büyük miktarlarda veri yığınları içinde gizlenen bilgilerin çıkarılması ve bu bilgilerin ilgili kuruma stratejik kararlar vermede yardımcı olarak kullanılması şeklinde ifade etmiştir (<https://www.ibm.com/us-en/marketplace/learning-lab/courses/coursera-data-mining>, Erişim Tarihi: 09 Ağustos 2016).

Statistical Analysis Software (SAS) firması ise veri madenciliğini, kurum yöneticilerinin ihtiyaç duyduğu verilerin veri yığınları içinden seçilmesi, düzenlenmesi ve modellenmesini içeren süreçler olarak tanımlamıştır (http://www.sas.com/en_us/software/enterprise-miner.html, Erişim Tarihi: 09 Ağustos 2016).

Veri madenciliği, büyük veri setlerinden oluşmuş veri tabanları içinde gizlenmiş bilgileri gün ışığına çıkarır. Bu işlem; istatistik ve matematik disiplinleri, veri tabanı teknolojisi, bilgisayar programları ve modelleme algoritmaları ile yapılır. Veri madenciliği tek başına bir çözüm üretmez. Çözüme ulaşmak için işe yarar bilgilere ulaşan ve karar sürecini destekleyen bir araçtır. Ayrıca veri madenciliği, araştırmacıya çözüme giden yolda ortaya çıkan veriler arasındaki ilişkileri ve şablonları bulması için de yardım etmektedir (Baykal, 2006: 96).

2.1. Veri Madenciliğinin Tarihsel Gelişim Süreci

İlk bilgisayarlar 1950’li yıllarda sayım yapma amacıyla kullanılmaya başlanmıştır. 1960’lı yıllarda ise veri tabanı ve verilerin depolanması kavramları teknoloji dünyasında yer bulmuştur. 1960’lı yılların sonunda bilim insanları basit öğrenmeli bilgisayarlar geliştirmeyi başarmışlardır. 1980’li yıllarda şirketler; ürünlerinin, rakiplerinin ve müşterinin verilerinden oluşan veri tabanlarını kurmuşlardır. Bu yıllarda veri tabanı yönetim sistemlerinin bilimsel platformlarda kullanımı yaygınlaşmıştır. 1990’lı yıllarda ise veri tabanlarından yararlı bilgilerin nasıl ortaya çıkarılabileceği düşünölmeye başlanmıştır. 1992 yılında veri madenciliği çalışmalarında kullanılmak üzere ilk yazılım üretilmiştir. 2000’li yıllara gelindiğinde veri madenciliği teknikleri sürekli gelişim göstermiş ve alınan iyi sonuçlar neticesinde bu alana olan ilgi artmıştır (Savaş vd., 2013: 5). Tablo 2.1.1’de veri madenciliğinin gelişim süreci özetlenmektedir.

Tablo 2.1.1: Veri İşleme Tekniklerinin Gelişimi

Gelişme Basamakları	İmkân Sağlayan Teknolojiler	Ürün Sağlayıcıları	Karakteristikler
Veri Toplama (1960’lar)	Bilgisayarlar, Kasetler	IBM, CDC	Statik Veri Ulaşımı
Veriye Ulaşım (1980’ler)	İlişkili Veri Tabanları (RDMS), SQL, ODBC	Oracle, Sybase, IBM, Informix, Microsoft	Kayıt Esnasında Veriye Dinamik Ulaşım
Veri Depolama ve Karar Destek (1990’lar)	OLAP, Çok Boyutlu Veri Tabanları, Veri Ambarları	Pilot, Cornshare, Arbor, Cognos, Microstrategy	Her Veriye Dinamik Ulaşım
Veri Madenciliği (Günümüzde)	Gelişmiş Algoritmalar, Çoklu İşlemcili Bilgisayarlar, Büyük Hacimli Veri Tabanları	Pilot, Lockheed, IBM, SGI	Bilgiye Önceden Ulaşım

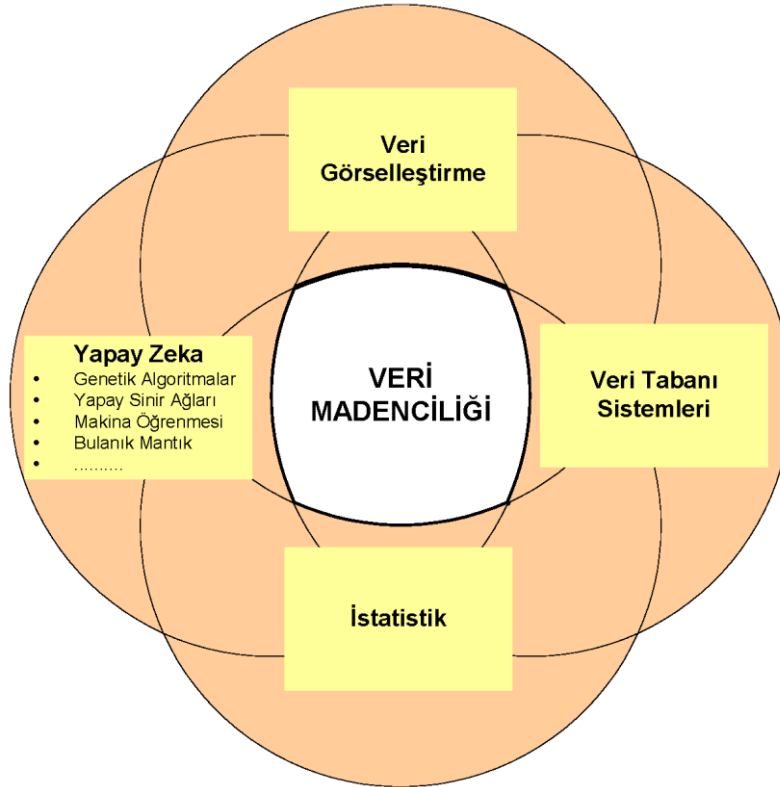
Kaynak: Çoban (2006: 35)

Günümüzde, gelişen teknoloji ile birlikte depolama kapasiteleri artmış ve büyük verilerin olduğu hemen hemen her alanda veri madenciliği teknikleri kullanılmaya başlanmıştır.

2.2. Veri Madenciliğinin İlişkili Olduğu Disiplinler

Veri madenciliği, sadece ham verinin analizinden ibaret olmadığı için çalışma sırasında farklı disiplinlerden beslenmektedir. Veri madenciliğinin birçok disiplinin kesişmesiyle ortaya çıkmış ve sürekli gelişimini sürdüren bir yapıda olduğu söylenebilir.

Şekil 2.2.1: Veri Madenciliğinin İlişkili Olduğu Disiplinler



Kaynak: Baykasoylu (2005: 2)

Veri madenciliği, Şekil 2.2.1’de gösterildiği gibi, istatistik, veri tabanı sistemleri, veri görselliği ve yapay zekâ gibi disiplinlerle yakın ilişkilidir.

2.3. Veri Madenciliği ve İstatistik

Veri madenciliği ve istatistik bilimi, belli olmayanın üstesinden gelmek ve gelecekteki olayları daha iyi tahmin etmek amacıyla ortaya çıkmıştır. Veri madenciliği ile istatistik, verileri anlamlandırmak ile ilgilenir. Veriyi bilgiye dönüştürme amacı, iki yaklaşımın da ortak özelliğidir (Tüzüntürk, 2010: 72).

Literatürde yer alan birçok tanım ve uygulamada veri madenciliği çalışmaları için kullanılan temel aracın istatistik olduğu söylenmektedir. İstatistik; verilerin toplanması, sınıflandırılması, özet haline getirilmesi, tablo ve grafiklerle kullanıcıya sunulmasıdır. İstatistik ile veriler analiz edilerek kitle hakkında anlamlı bilgiler elde edilir ve yorumlanır. Veri ve bilgi kavramları her iki disiplinin de temel ögesidir. Bu sebeple veri madenciliği ve istatistik birbirleriyle fazlaca örtüşen konulardır (Özmen, 2001: 2).

Veri madenciliği ve istatistiğin birbirlerinden ayrıldığı noktalar Tablo: 2.3.1’de özetlenmiştir.

Tablo 2.3.1: İstatistik ile Veri Madenciliğinin Farkları

İstatistiksel Analiz	Veri Madenciliği
İstatistikçiler çalışmalarına genellikle hipotez kurarak başlarlar.	Veri madenciliği hipotez kurmaz.
İstatistikçiler, hipotezlerini eşleştirmek için kendi eşitliklerini geliştirmek zorundadırlar.	Veri madenciliği algoritmaları eşitliklerini otomatik olarak geliştirir.
İstatistiksel analizler, niteliksel ve niceliksel veriler kullanır.	Veri madenciliği niteliksel ve niceliksel verilerin yanı sıra ses ve metin gibi farklı tipte veriler de kullanır.
İstatistikçiler, kirli verileri analiz esnasında bulup filtre ederler.	Veri madenciliği temiz veri ile çalışır.
İstatistikçiler, kendi sonuçlarını yorumlar ve bu yorumlarını yöneticilere iletirler.	Veri madenciliğinin sonuçlarını yorumlamak kolay değildir. Veri madenciliği sonuçlarını analiz etmede, yorumlamada ve bulguları yöneticilere iletmede mutlaka istatistikçiye ihtiyaç duyulmaktadır.

Kaynak: Koyuncugil (2006: 48)

Veri madenciliği, verileri işlerken çok sayıda veri analizinden faydalanır. Bu nedenle istatistik ile veri madenciliği iç içedir. Veri madenciliği, problemlere bir ya da daha fazla istatistiksel yöntem kullanarak çözüm ürettiğinden veri madenciliği analistleri, istatistiksel tekniklere ve istatistik konularına hakim olması gerekir. İstatistik, veri madenciliğinin vazgeçilmezidir (Tüzüntürk, 2010: 86).

2.4. Veri Madenciliği Süreci

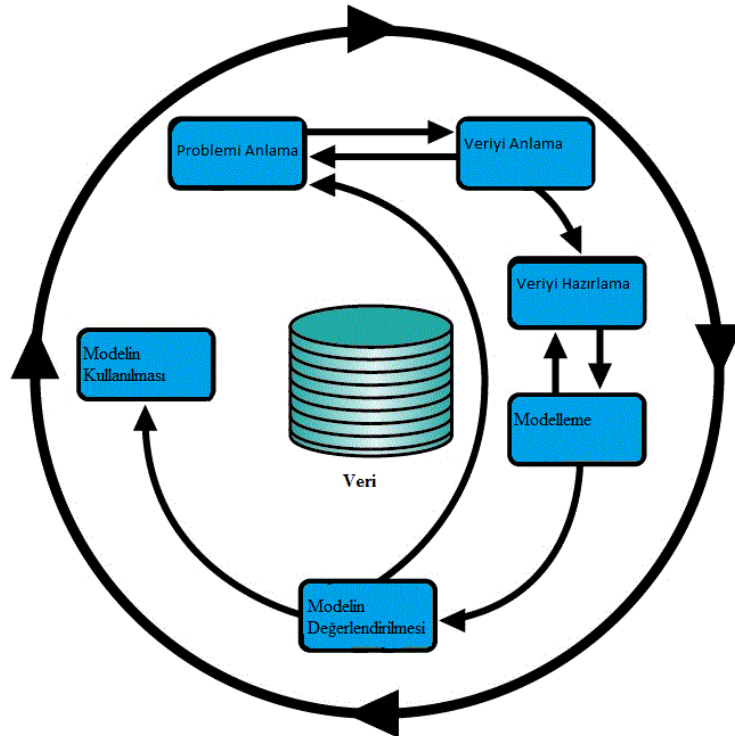
Veri madenciliğinin sadece veri analizinden oluştuğu düşünülmemelidir. Veri madenciliği; projenin amacının belirlenmesi, amaç doğrultusunda veri tabanının hazırlanması, ilgili veri tabanından verilerin değerlendirilip hazırlanması ve

hazırlanan verilerin analiz edilip bilgiye dönüştürülmesini kapsayan uzun ve meşakkatli bir süreçtir.

Veri madenciliği çalışmalarının bir süreç halinde devam edebilmesi için bütün adımlar dikkatle izlenmelidir. Çünkü veri madenciliği sürecinde her adım kendisinde bir önceki adımın sonuçlarıyla bağlantılıdır (Bilgin, 2013: 6).

Veri madenciliği süreci, belirli bir sistem içinde yönetilmelidir. Veri hacminin büyüklüğü, karmaşık yapıda olması ve süreç sınırlarının belirgin olmaması çalışma esnasında yapılan ufak hataları bile geri dönülemez kılmaktadır. Bu durum araştırmacıya ciddi kaynak ve zaman kaybı olarak dönebilir. Bu nedenle veri madenciliği süreçlerinin belli bir standartta işlemesine ihtiyaç duyulmuştur. Veri madenciliği süreci, The Cross Industry Standart Process for Data Mining (CRISP-DM) konsorsiyumu tarafından standartlaştırılmıştır (Gökmen, 2014: 40).

Şekil 2.4.1: Veri Madenciliği Adımları



Kaynak: <http://crisp-dm.eu/reference-model/>, Erişim Tarihi: 26 Aralık 2016

CRISP-DM süreci; problemi anlama, veriyi anlama, veriyi hazırlama, modelleme, modelin değerlendirilmesi ve modelin kullanılması adımlarından oluşmaktadır.

2.4.1. Problemi Anlama

Veri madenciliği süreci problemi anlama ile başlar. Araştırmacı bu aşamada çok iyi çözümleme yaptığı takdirde süreç sağlıklı bir şekilde işleyecektir. Veri madenciliği süreci tesadüfen ilerlememektedir. Sürecin sonunda ulaşmak istenilen hedefler ve yönetici beklentileri net bir şekilde ortaya konulmalıdır (Aytaç ve Bilge, 2013: 29). Araştırmanın amaçları, kısıtları ve ihtiyaçları belirlendikten sonra hedefe doğru bir şekilde ulaşabilmek için strateji oluşturulur.

2.4.2. Veriyi Anlama

Bu adım; verilerin toplanması, eldeki verilerin uygun olup olmadığının değerlendirilmesi, model kurmak için ihtiyaç duyulan farklı verilerin saptanması, veri boyutunun yeterliliğin değerlendirilmesi gibi veri kalitesinin düşünce süzgecinden geçtiği aşamadır. Problemi anlama ve veriyi anlama, iç içe geçmiş adımlardır. Araştırmacı, problemi anladıkça değişik verilere bakabilir ve verilere baktıkça iş ile ilgili değişik bakış açılarına sahip olabilir. Çalışmada kullanılacak veriler, bu döngü içinde netleşir (Yurdakul, 2015: 24).

2.4.3. Veriyi Hazırlama

Veriyi hazırlama aşamasında eldeki veriler işlenerek veri madenciliğinde kullanılabilir duruma getirilir. Veriyi hazırlama adımı, veri madenciliğinin en fazla zaman alan ve iş gücü gerektiren aşamasıdır. Araştırmacı bu aşamada; verinin toplanması, temizlenmesi, bütünleştirilmesi, dönüştürülmesi ve indirgenmesi işlemleri yapmaktadır.

2.4.3.1 Verinin Toplanması

Araştırmacı, belirlediği problemin çözümü için ihtiyaç duyduğu veri kaynaklarını bu aşamada saptar ve verileri toplar. İşletmeler bu adımda kendi veri tabanları dışında başka veri kaynaklarından da yararlanabilir.

Verilerin nasıl, ne kadar ve nereden toplandığı işletmeler için oldukça önemlidir. Çünkü ihtiyaçtan daha az veriyle çalışmak çalışmayı eksik bırakabilirken gereğinden fazla veri, maliyeti arttırıp ve süreci uzatabilir (Bilgin, 2013: 11). Veri madenciliği sürecinin faydalı bir şekilde sonlanması için veri kaynaklarının yeterli ve güvenilir olması gerekir.

2.4.3.2. Verinin Temizlenmesi

Veri kaynaklarından temin edilmiş veri setlerinin kusursuz olması beklenemez. Bu adımda veri madenciliğinin başarı oranını arttırmak için veri setinin içinde bulunan aykırı, uç, hatalı ve tutarsız değerler silinir. Eksik olan veriler ise uygun yöntemlerle tamamlanır.

2.4.3.3. Verinin Bütünleştirilmesi

Birden fazla veri kaynağından alınan farklı türden veriler ile çalışabilmek için verilerin ortak bir türe dönüştürülmesi gerekir. Örneğin veri setinde, “ferdin genel sağlık durumu” diye bir değişken “i” ve “k” harfleriyle kodlanmış olabilir. “i” kodu iyi, “k” kodu ise kötü sağlık durumunu simgelemektedir. Bir başka veri setinde söz konusu sağlık durumu, “1” ve “0” rakamları ile farklı bir veri setinde ise “iyi” ve “kötü” kelimeleri ferdin genel sağlık durumunu belirtmek amacıyla kullanılmış olabilir. Bu tip farklı türden veriler, veri seti üzerinde çalışmayı olanaksız hale getirebilir. Bu sebeple farklı tipteki veriler, üzerinde analiz yapmaya başlamadan önce bütünleştirilmeli yani ortak bir türe dönüştürülmelidir (Gölce, 2010: 16).

2.4.3.4. Verinin Dönüştürülmesi

Veriler dönüştürmeden veri madenciliği analizlerine dâhil edilirse araştırmacıyı yanlış sonuçlara götürebilir. Veri setindeki değişkenlerin varyansları ve ortalamaları birbirinden büyük ölçüde farklı olabilir. Böyle durumlarda büyük varyans ve ortalamaya sahip olan değişkenler diğer değişkenlerin üstünde fazlaca baskı oluşturur ve onların etkinliğini önemli derecede azaltır. Ayrıca veri setindeki değişkenler, farklı ölçü birimleri kullanılarak elde edilmiş ise veriler tek bir ölçü birimine dönüştürülür. Nitel değişkenleri nicel değişkenlere dönüştürmek için de veri dönüştürme işlemi kullanılır (Vatansever, 2008: 21).

2.4.3.5. Verinin İndirgenmesi

Veri madenciliği büyük veri yığınları içindeki bilgileri gün ışığına çıkartmak için yapılan işlemler dizisidir. Veri madenciliği sürecinde kullanıcının elindeki veri seti çok büyük boyutta ise çözüm süreci uzun sürebilir. Sürecin uzaması işletmeler ve organizasyonlar açısından ekstra maliyet getirebilmektedir.

Veri setinin bütünlüğünü korumak şartıyla ve elde edilecek sonucun değişmeyeceği öngörülüyorsa veri sayısı veya değişken sayısında azaltma işlemi yapılabilir. Yapılan bu işlemler verinin indirgenmesi olarak ifade edilmektedir (Kartum, 2014: 46).

2.4.4. Modelleme

Modelleme aşaması, veri madenciliği sürecinin en karışık adımıdır. Bu adımda verilerin içindeki gizli kalmış bilgileri ortaya çıkartmak için ileri çözümleme algoritmaları kullanılır. Bu adım; uygun modelleme yönteminin seçimi, test tasarımının üretimi, modelin geliştirmesi ve tahmin işlemlerini kapsar (Erol, 2013: 24-25). Problemin çözümünde uygun olan teknik ya da tekniklerin belirlenmesi için birden fazla model denenir ve içlerinden en uygunu kullanılır.

2.4.5. Modelin Değerlendirilmesi

Veri madenciliği sürecinin önceki adımları, modelin doğruluğu ve genelliği gibi özellikleriyle ilgilenmektedir. Modelin değerlendirilmesi adımıyla araştırmacının elinde kurulmuş bir model vardır. Araştırmacı, modelin istediği sonuca uygun olup olmadığını değerlendirir (Yakut, 2012: 13).

Basit geçerlilik testi, bir modelin doğruluğunun test edilmesinde kullanılan en kolay yöntemdir. Basit geçerlilik testinde verilerin % 5 ile % 33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısmın üzerinde modelin öğrenimi gerçekleştirilir. Model öğreniminden sonra veriler üzerinde test işlemi yapılır (Çoban, 2006: 42).

Bir sınıflandırma modelinde hata oranı, yanlış sınıflandırılmış olay sayısının tüm olay sayısına bölümüyle hesaplanır. Doğruluk oranını ise doğru sınıflandırılmış olay sayısının tüm olay sayısına bölümü verir (Çoban, 2006: 42). Veri madenciliğinin kullanımına ait kararlar bu adımın sonunda verilir.

2.4.6. Modelin Kullanılması

Modeli oluşturmak, veri madenciliği çalışmasının bitmesi anlamına gelmemektedir. Veri yığınlarından elde edilen bilgi, kullanıcıların faydalanacağı şekilde düzenlenip ifade edilmelidir. Modelin kullanılması adımı, bir rapor üretmek kadar basit ya da modelin başka sistemlerin içine dahil edilmesi karmaşık olabilir (Göral, 2007: 27-28).

2.5. Veri Madenciliği Tekniklerinin Kullanıldığı Alanlar

Verilerin depolandığı her alanda veri madenciliği çalışması yapmak mümkündür. Dünyada yapılmış çalışmalar incelendiğinde veri madenciliği yöntemlerinin en fazla biyoloji, tıp ve genetik alanlarında uygulandığı görülmektedir. Amerika Birleşik Devletleri son yirmi yılda vergi kaçakçılığının tespitinden gizli dinlemeye kadar birçok alanda veri madenciliği tekniklerini kullandığı bilinmektedir (Savaş vd., 2012:

5). Veri madenciliği tekniklerinin yaygın olarak kullanıldığı alanlar aşağıda gruplandırılmıştır.

Tıp ve İlaç:

- Tıbbi teşhis sürecinde,
- Test sonuçlarının tahmin edilmesinde,
- Tedavi süreçlerinin belirlenmesinde,
- Ürün geliştirme sürecinde kullanılmaktadır (Tiryaki, 2006: 12).

Kimya:

- Yeni ilaç çeşitlerinin araştırılması, keşfi ve geliştirilmesinde,
- Kimyasal moleküllerin keşfi ve sınıflandırılması çalışmalarında kullanılmaktadır (Dinçer, 2016: 11).

Genetik ve Biyoloji:

- Genetik hastalıkların tespitinde,
- Gen haritasının analiz edilmesinde,
- Bitki türlerinin ıslah edilmesinde,
- Virüs türlerinin keşfinde ve sınıflandırılmasında,
- Fizyolojik parametrelerin analiz edilmesi ve değerlendirilmesinde kullanılmaktadır (Dinçer, 2016: 11).

Pazarlama:

- Müşteri değerlendirmede ve çeşitli müşteri özelliklerinin analizinde,
- Satış tahminlerinde,
- Pazarlama kampanyalarında,
- Pazar sepeti analizinde,
- Müşteri kaybetmemek için pazarlama stratejileri geliştirilmesinde,
- Çapraz satış analizinde kullanılmaktadır (Tiryaki, 2006: 10).

Bankacılık:

- Banka müşterilerini değerlendirmede ve çeşitli müşteri özelliklerinin analizinde,
- Kredi kartı dolandırıcılığı tespitinde,
- Kredi taleplerinin değerlendirilmesinde,
- Birbirinden farklı finansal değişkenler arasındaki bilinmeyen korelasyonların tespitinde,
- Usulsüzlüklerin tespitinde,
- Risk yönetiminde,
- Risk analizlerinde kullanılmaktadır (Tiryaki, 2006: 11).

Borsa:

- Piyasanın genel durum analizlerinde,
- Alış ve satış için optimum stratejilerinin belirlenmesinde,
- Hisse senedi fiyatlarının tahmininde kullanılmaktadır (Tiryaki, 2006: 11-12).

Sosyal Bilimler ve Davranış Bilimleri:

- Seçim tahminlerinde,
- Kamuoyu araştırmalarında,
- Demografik özelliklerin belirlenmesinde,
- İnsanların genel eğilimlerinin tespitinde kullanılmaktadır (Dinçer, 2016: 12).

Perakendecilik

- Optimum mağaza yerleşim yerlerinin belirlenmesinde,
- Ürün tedarik optimizasyonunda,
- Satış noktası tespiti için yapılan veri analizlerinde,
- Alışveriş sepet analizlerinde kullanılmaktadır (Tiryaki, 2006: 11).

Sigortacılık:

- Risk teşkil eden müşteri profillerinin tespitinde,
- Yeni sigorta poliçesi talebinde bulunacak olası müşterilerinin tahmininde,
- Sigorta dolandırıcılıklarının tespitinde (Tiryaki, 2006: 11).

Robot Görüş Sistemleri ve Görüntü Tanıma Teknolojisi:

- Parmak izi ve yüz tanıma teknolojisinde,
- Yol ve engel tanıma teknolojisinde kullanılmaktadır (Dinçer, 2016: 12).

Endüstri:

- Lojistik faaliyetlerinin optimize edilmesinde,
- Kalite kontrol çalışmalarında,
- Üretim sürecinin optimizasyonunda kullanılmaktadır (Tiryaki, 2006: 12).

Uzay Bilimleri:

- Yıldızların konumlarına göre sınıflandırılmasında,
- Gezegenlerin konumları ve yüzey şekillerinin belirlenmesinde,
- Yeni galaksilerin keşfedilmesinde kullanılmaktadır (Dinçer, 2016: 12).

Metin Madenciliği:

- Büyük boyutlu ve anlamsız metin yığınlarından anlamlı ilişkiler çıkarmada kullanılmaktadır (Dinçer, 2016: 12).

İnternet Madenciliği:

- İnternette bulunan veri hacmi katlanarak artmaktadır. Artan veriler her geçen gün karmaşık bir hal almaktadır. İnternet madenciliğinin amacı, interneti kategorilere ayırarak veriye erişimin daha hızlı olmasını sağlamaktır (Dinçer, 2016: 12).

2.6. Veri Madenciliği Çalışmalarında Karşılaşılan Sorunlar

Küçük veri setlerinde kararlı ve hızlı çalışan bir sistem, büyük veri tabanlarına uygulandığı zaman tamamen farklı çalışabilir. Bir veri madenciliği sistemi tutarlı veriler ile çok iyi çalışırken aynı veri setine gürültü eklendiği zaman sistem kayda değer bir şekilde kötüleşebilir (Sever ve Hayri, 2003: 179). Veri madenciliği çalışmalarında birçok sorunla karşılaşma olasılığı vardır. Bu sorunlar aşağıda özetlenmiştir.

2.6.1. Veri Tabanı Boyutu

Teknolojik gelişmelere paralel olarak veri tabanlarında kayıt altında tutulan bilgi sayısı her geçen gün katlanarak artmaktadır. Çoğu algoritma, küçük örneklemeler üzerinde çözüm üretebilecek yapıda tasarlanmıştır. Üzerinde çalışılan veri kümesimin büyük boyutta olması elde edilen çözümlerin doğruluğu açısından bir avantajdır ancak araştırmacı, algoritmaları çok dikkatli bir şekilde uygulamalıdır. Veri tabanı boyutu yüzünden işlem süresi uzayıp çalışma karmaşık bir hal alabilir. Bu durumda kullanıcı veri tabanından örneklem çekebilir ya da veri tabanı boyutunu küçültebilir.

2.6.2. Kayıp Veri

Veri tabanında bazı kayıtlar hiç girilmemiş olabilir. Kayıp veriler, veri setinin doğasından ya da büyüklüğünden kaynaklanmaktadır. Kayıp veriler ile karşılaşıldığında (Savaş vd., 2012: 6);

- Kayıp veriye sahip girdi ya da girdiler veri setinden çıkartılabilir,
- İlgili değişkenin ortalaması, kayıp verilerin yerine atanabilir,
- Veri setindeki veriler göz önüne alınarak en uygun değer atanabilir.

Kayıp verilerin tahmininde verilerin özelliği ve gözlem sayısı çok önemlidir. Veri madenciliği çalışmalarında sıklıkla karşılaşılan kayıp veriler, araştırmacılar tarafından dikkate alınmadığında yanlış sonuçlar ortaya çıkabilir (Sezgin ve Çelik, 2013: 26).

2.6.3. Gürültülü Veri

Veri tabanındaki verilerin sınıflarında ya da özelliklerindeki hatalara gürültü adı verilir. Araştırmacının üzerinde çalıştığı veri setinin gürültüsüz olma ihtimali çok düşüktür. Hatalı veriler veri madenciliği çalışmalarında ciddi sorunlara yol açabilir. Kullanılan veri madenciliği yöntemi, gürültülü veriyi tespit edip sistem dışına atmalıdır (Akgöbek ve Çakır, 2009, 804).

2.6.4. Dinamik Veri

Dinamik veri tabanlarında kayıtlar sürekli güncellenir. Bu tip veri tabanlarının bünyesinde bulunan veriler; düzeltme, ekleme ve silme işlemlerine maruz kalmaktadır. Dinamik yapıya sahip veri tabanları, veri madenciliği çalışmalarını zorlaştırmaktadır. Araştırmacı, veri tabanındaki verilerin güncellenme ihtimaline karşı üzerinde çalıştığı verileri veri tabanından ayırabilir.

2.6.5. Artık Veri

Veri madenciliği çalışmalarında analize giren veri setinde hiç kullanılmayacak gereksiz değişkenler olabilir. Artık verilerin olduğu veri seti, çalışma süresini uzatabilir. Örneğin veri setinde hem yaş hem de doğum yılı değişkeni varsa doğum yılı değişkeni veri setinden çıkarılabilir.

2.7. Veri Madenciliğinde Kullanılan Yazılımlar

Veri madenciliği çalışmalarında doğru çözümlemeler yapabilmek için bilgisayar programlarından faydalanılmaktadır. Günümüzde birbirlerinden farklı algoritmalara sahip birçok veri madenciliği programı geliştirilmiştir.

Veri madenciliğinde araştırmacı için en uygun yazılım, en ileri özelliklere sahip yazılım değildir. Programın kullanımının kolay olması, kabul edilebilir ölçüde doğruluk sağlaması ve veri madenciliği projesindeki tüm ortak görevleri yapabilmesi program seçimini etkileyen temel özelliklerdir (Dolgun, 2006: 50).

Veri madenciliği problemlerinin çözümüne yönelik yazılımların sayısı her geçen gün artmaktadır. Bu yazılımlardan bazılarına aşağıda yer verilmiştir.

2.7.1. IBM SPSS Modeler

SPSS firması tarafından veri madenciliği çalışmaları için geliştirilen SPSS Clementine yazılımı, SPSS'nin International Business Machines (IBM) firması

IBM SPSS Modeler yazılımının çalışma alanına tuval adı verilir ve pencerenin alt kısmında paletler bulunur. Araştırmacı, ilgili işlemcileri tuval üzerine yerleştirerek veri madenciliği uygulamasını yapar (TÜİK, 2014: 23). Şekil 2.7.1.1’de IBM SPSS Modeler yazılımının çalışma ekranından bir görüntü yer almaktadır.

- Favorites paletinde sık kullanılan işlemciler,
- Sources paletinde analiz edilecek verileri okuyan işlemciler,
- Record Operators paletinde verileri birleştirme, filtreleme ve sıralama gibi düzenlemelerin yapılmasını sağlayan işlemciler,
- Field Operator paletinde veri setinde yer alan değişkenleri seçme, tipini belirleme ve yeni değişken türetme işlemlerini yapan işlemciler,
- Graphs paletinde değişkenleri görsel olarak incelenmesine olanak tanıyan işlemciler,

- Modelling paletinde veri setinin modellenmesinde kullanılan işlemciler,
- Output paletinde sonuçların görüntülenmesini sağlayan işlemciler,
- Export paletinde ise analiz sonuç tablolarının görüntülenmesini ve farklı formatta kaydedilmesini sağlayan işlemciler yer almaktadır.

Açık ve anlaşılır bir yapıda olan IBM SPSS Modeler; yapay sinir ağları, karar ağaçları, birliktelik kuralları, zaman serileri ve regresyon analizlerini yapabilen tüm algoritmaları bünyesinde barındırır. Modeler sayesinde araştırmacı, veri madenciliği problemlerine ileri bir teknoloji ile çözüm arar.

2.7.2. Enterprise Miner

Enterprise Miner, SAS firması tarafından geliştirilmiş veri madenciliği yazılımıdır. Regresyon analizi, kümeleme analizi, zaman serileri analizi, karar ağaçları ve yapay sinir ağları gibi problemlere çözümler sunabilmektedir. Programda istatistiksel analiz araçlarının fazla olması ve programın kullanım kolaylığı tercih edilmesindeki en önemli etkenlerdir.

2.7.3. Intelligent Miner

Intelligent Miner, IBM firması tarafından geliştirilmiştir. Sınıflama, ilişkilendirme, gruplama, tahmin modelleme, regresyon analizi, sıralı desen analizi gibi algoritmaları bünyesinde bulundurur. IBM firmasının veri tabanı ile birleşik yapıda olması ve ölçeklendirilebilir yapısı ile Intelligent Miner, diğer yazılımlardan ayrılır.

2.7.4. WEKA

WEKA, Java programlama dili ile geliştirilmiş ve kullanıcıların ücretsiz olarak faydalandığı veri madenciliği yazılımıdır. Ücretsiz olması sebebiyle veri madenciliği problemlerinin çözümünde sıklıkla tercih edilmektedir. Ancak büyük veri setleri üzerinde yetersiz kalması ve Java programlama dili ile yazıldığından yavaş çalışması olumsuz özellikleridir.

2.7.5. Analysis Manager

Analysis Manager, Microsoft firması tarafından geliştirilmiş veri madenciliği yazılımıdır. Karar ağacı ve kümeleme analizi algoritmalarını bünyesinde barındırır. Kullanımı kolay ve görsel yönden güçlü bir yazılımdır ancak farklı veri madenciliği algoritmalarını desteklememesi en büyük eksikliğidir.

2.7.6. IBM SPSS Statistics

SPSS'in temelleri, 1960'ların sonlarında Standford Üniversitesinde C. Hadlia, Norman H. Nie ve Dale H. Bent adlı üç bilim insanı tarafından atılmıştır. Bilimsel araştırmaları sırasında çok büyük miktarda verinin analizini yapmak zorunda olan üç kişi, işin altından ancak kod yazarak kalkabileceklerini fark edince ortaya hatırı sayılır seviyede bir yazılım çıkmıştır. Bu yazılıma mimarları tarafından SPSS ismi verilmiştir. Yazılım kısa sürede üniversitedeki diğer bölümlerde kullanılmaya başlanmıştır. 1970'lerde yazılımın mimarları kullanıcılar için el kitabı yazınca açık kaynaklı olarak dağılan yazılım bir anda diğer üniversitelere de yayılmış ve ABD devlet dairelerinde kullanılmaya başlanmıştır. 1980'lerde kişisel bilgisayarlara uyumlu versiyonları üretilmiştir. 2009 yılında SPSS, IBM tarafından satın alınınca yazılımın adı IBM SPSS Statistics haline gelmiştir (Pektaş, 2013: 2). Günümüzde en son versiyonu olan SPSS 24.0 kullanılmaktadır. Başta anket şirketleri, eğitim kurumları ve hükümetler olmak üzere pek çok kuruluşun kullandığı bir yazılımdır. Özellikle sosyal bilimlerde, sağlık araştırmalarında ve pazar araştırmalarında kullanılmaktadır (<https://tr.wikipedia.org/wiki/SPSS>, Erişim tarihi: 12 Aralık 2016).

2.8. Veri Madenciliği Uygulama Örnekleri

Veri madenciliği yöntemleri, etkili sonuçlar vermesi ve kolay uygulanabilirliği sayesinde her geçen gün daha fazla çalışmada kullanılmaktadır. Yapılan çalışmalar incelendiğinde veri madenciliği tekniklerinin ne kadar farklı alanda kullanıldığı anlaşılmaktadır. Literatürde yer alan veri madenciliği çalışmaları; tıp, mühendislik, eğitim, borsa, ticaret ve bankacılık başlıkları altında sınıflandırılarak özetlenmiştir.

2.8.1. Tıp Alanında Uygulanan Veri Madenciliği Çalışmaları

Muhammet Akman 2010 yılında, periodontoloji alanında oluşturulmuş veri seti üzerinde veri madenciliği çalışması yapmıştır. Random Forest, Bagging ve C&R Tree yöntemleri ile sınıflandırma yapan Akman, en başarılı sonucu Random Forest yöntemi ile elde etmiştir. Bu yöntem % 95,4 doğruluk ile sınıflandırma yapmıştır (Akman, 2010).

Yeşim Atasoy 2015 yılında gerçekleştirdiği çalışmasında, ankilozan spondilit hastası fertlere ait veri setini kullanmış ve hastalığın radyografik progresyonuna etki eden faktörleri; C4.5, Regresyon ağaçları, Gini ve Rndom Forest algoritmaları ile analiz etmiştir. Yaptığı analiz sonucu, Gini algoritmasının en yüksek sınıflandırma ve model doğruluk oranına sahip olduğunu tespit etmiştir (Atasoy, 2015).

Fatih Elmas 2014 yılında, kalp krizi geçiren kişilerin kan değerlerindeki ortak değişimler belirlenerek Gri ve Apriori algoritmaları ile analiz gerçekleştirilmiştir. Analiz sonucunda Gri algoritması ile % 68, Apriori algoritması ile % 72 oranında kalp krizi durumu tespit etmiştir (Elmas, 2014).

2.8.2. Mühendislik Alanında Uygulanan Veri Madenciliği Çalışmaları

Chen, Wu ve Zhu 2004 yılında interaktif bir veri madenciliği sistemi geliştirmişlerdir. Kümeleme, sınıflandırma ve ilişkilendirme modellerinin kullanıldığı sistem, veri madenciliği sürecinde araştırmacıya en uygun veri madenciliği tekniği seçmesine yardım etmektedir (Chen vd., 2004).

Eren Berk Aytaç 2014 yılında, ATM'den çekilebilecek günlük para miktarını veri madenciliği teknikleri ile tahmin etmiştir. Bu sayede ATM'lerde fazla miktarda para tutulmasının engellenmesi ve banka karının artması amaçlanmıştır. Çalışmasında özel bir bankanın zaman serileri biçimindeki ATM'den çekilen para miktarını veri seti olarak kullanmış ve veri setine, veri madenciliği teknikleri uygulamıştır (Aytaç, 2014).

Fatma Turna 2011 yılında, Türkiye’de faaliyet gösteren bir tramvay şirketinden temin ettiği 4 yıllık tramvay arıza kayıtlarını veri seti olarak kullanmıştır. Veri madenciliği teknikleri ile tramvay seferlerine engel olan arızaların meydana gelmesi ile ilgili anlamlı ve yararlı kurallar çıkartmıştır (Turna, 2010).

Erkan Karaca 2010 yılında, uçuş güvenliğini arttırmak amacıyla veri madenciliği yöntemlerinden faydalanmıştır. Amerika Birleşik Devletleri’nde 2005–2010 yılları arasında meydana gelen sivil uçak kazalarını incelemiştir. Çalışmanın sonunda ise sivil havacılık ve devlet kurumlarına uçuş güvenliğinin artırılması ve kazaların önlenmesi konularında önerilerde bulunmuştur (Karaca, 2010).

2.8.3. Eğitim Alanında Uygulanan Veri Madenciliği Çalışmaları

2015 yılında, Kırıkkale İli Anadolu Liselerinde okuyan 11. sınıf öğrencileriyle yaptığı anket sonucu veri setini oluşturan Semra Yurdakul, veri madenciliği teknikleriyle öğrenci performansına etki eden faktörleri belirleyip başarı veya başarısızlığı etkileyen faktörlerin birbirleriyle ilişkilerini araştırmıştır. Çalışmasının sonucunda; öğrencinin öğretmenlerle ve arkadaşlarıyla olan iletişiminin iyi olması, kendisine ait çalışma odasının olması arasında başarıyı arttıran bir ilişki tespit etmiştir (Yurdakul, 2015).

Y. Ziya Ayık vd., Atatürk Üniversitesini kazanacak öğrenci profilini tahmin etmek amacıyla Atatürk Üniversitesi öğrencilerinin lise başarıları ve lise türleri ile tercih ettikleri fakülteler arasındaki ilişkiyi, veri madenciliği yöntemleri ile incelemiştir. Çalışma sonucunda lise türü ve lise mezuniyet derecelerinin istenilen bir fakültenin kazanılmasında aynı derecede ve çok önemli olduğunu tespit etmişlerdir (Ayık vd., 2007).

Esra Aksoy 2014 yılında yaptığı çalışmada, öğrencilerin çoklu zekâ alanı, cinsiyet, öğrenme stili, sınıf düzeyi ve kişilik tipi değişkenlerini kullanarak matematik alanında üstün yetenekli ve zekâlı olup olmadıklarını veri madenciliği teknikleriyle saptamaya çalışmıştır (Aksoy, 2014).

2.8.4. Finans, Borsa ve Bankacılık Alanında Uygulanan Veri Madenciliği Çalışmaları

Ercan Özbay 2007 yılındaki çalışmasında, veri madenciliği tekniklerini 12773 adet internet bankacılığı müşteri işlemine uygulayarak internet bankacılığı dolandırıcılığını belli ölçülerde önleyen bir sistem tasarlamıştır (Özbay, 2007).

Mustafa Aykut Göral, kredi kartı başvurularındaki sahtekârlıkları tespit etmek için 2007 yılında bir veri madenciliği modeli kurmuştur. Kurulan model sayesinde yakalanan sahtekârlık sayısı 3,3 katına çıkmıştır (Göral, 2007).

Hüseyin Aktürk 2008 yılında, borsa ile ilgilenen kişiler üzerindeki riski aşağı yöne çekmeyi amaçlamıştır. Çalışmasında, Türkiye Cumhuriyeti Merkez Bankası döviz kurları üzerinde veri madenciliği, volatilité ve korelasyon tekniklerini kullanmıştır (Aktürk, 2008).

2.8.5. Ticari Alanda Uygulanan Veri Madenciliği Çalışmaları

Furkan Yener 2014 yılında yaptığı çalışmasında, perakende sektöründe faaliyet gösteren bir markanın, bir bölgedeki mağazalarından gelen talepleri karşıladığı ana dağıtım deposunu incelemiş ve yeni bir tasarım önermiştir. Veri madenciliği, çok boyutlu ölçekleme tekniği ve matematiksel bir model kullanıp ana dağıtım deposunun içindeki ürünlerin yerleri yeniden düzenlenmiştir. Yeni düzenleme sayesinde elde edilen kazanç istatistiki olarak kanıtlanmıştır (Yener, 2014)

Simge Gizem Ukuş 2014 yılında, unlu gıda ve buğday ürünleri üreticisi bir markanın 4 yıllık satış miktarı kayıtlarını veri seti olarak kullanmış ve veri madenciliği teknikleriyle markanın gelecek 6 aylık dönemine ilişkin satış miktarlarının tahminlerini yapmıştır. Çalışmanın sonunda, tahmin sonuçları üzerinden markaya çeşitli pazarlama ve kampanya stratejileri önermiştir (Ukuş, 2014).

3. VERİ MADENCİLİĞİ YÖNTEMLERİ

Veri madenciliği, var olan probleme uygun çözüm ürettiği için istatistik ve matematik temelli çok sayıda model geliştirilmiştir. Veri madenciliği çalışmalarında araştırmacı, elindeki veri setine ve tanımladığı probleme göre en uygun modeli seçer.

Veri madenciliği çalışmaları için geliştirilen modeller, aşağıdaki gibi tahmin edici ve tanımlayıcı modeller olmak üzere iki başlıkta toplanabilir (Asilkan, 2008: 28).

- Tahmin Edici Modeller
 - Sınıflama ve Regresyon
- Tanımlayıcı Modeller
 - Kümeleme
 - Birliklik Kuralları
 - Ardışık Zaman Örüntüler

Tahmin edici modeller ile sonuçları bilinen veri kümelerini kullanılır ve uygun bir model kurulur. Kurulan bu model üzerinden sonuç değerleri bilinmeyen veri setleri için sonuç değerleri tahmin edilir (Gürbüz, 2009: 45). Borsada işlem gören bir hisse senedinin geçmiş yıllardaki değerinden yola çıkıp gelecekte ulaşacağı muhtemel değerlerin tahmin edilmesi, tahmin edici bir model tipine örnek olabilir.

Tanımlayıcı modeller ise veriler arasında gizlenmiş özellikleri ve ilişkileri araştırır (Anane, 2001: 300). Bir süpermarketin, 30 yaş altı bekar müşterileri ile 30 yaş üzeri evli müşterilerinin satın alma performanslarının birbirleri ile benzerlik gösterip göstermediğinin belirlenmesi tanımlayıcı model tipine örnek gösterilebilir.

Veri madenciliğinde tanımlayıcı ve tahmin edici modeller kesin çizgilerle birbirlerinden ayrılmazlar. Eğer model anlaşılabilir ise tahmin edici model

tanımlayıcı model yerine, tanımlayıcı modeller ise tahmin edici model yerine kullanılabilir (Velickov ve Solomatine, 2000: 4).

3.1. Sınıflama ve Regresyon Modelleri

Sınıflama ve regresyon modelleri, gelecekteki veri eğilimlerini öngören ya da önemli veri sınıfları oluşturabilen veri analiz teknikleridir. Regresyon modelleri sürekli değerleri tahmin ederken sınıflama modelleri kategorik tipteki değerleri tahmin eder (Hand vd., 2011: 19). Bir sınıflama modeli, işletmede çalışan personelleri kadın veya erkek olarak kategorize etmek için kurulurken regresyon modeli, çalışan personelin performanslarına göre yılsonunda alacakları maaş zammını tahmin etmek için kurulabilir.

Kategorik değişkenler, veri madenciliğinin temel fonksiyonlarından biridir. Kategorik sonuçları tahmin etmek için sınıflandırma modelleri kullanılmaktadır. Sınıflandırma, birbirlerinden farklı özelliklerine göre verilerin bir sınıflayıcı tarafından sınıflara atanması sürecidir. Sınıflandırma ile eldeki verilerin bir sınıfa ait olup olmadığı tespit edilir. Eğer veriler bir sınıfa aitse hangi sınıfa atanacağı, sınıflandırma algoritmaları ile belirlenir. Sınıflandırma iki aşamada sonuçlanır. Birinci aşamada verilerin eğitimi gerçekleştirilir, ikinci aşama da ise model test edilir. Sınıflandırma girdileri, her biri bir sınıf etiketi ile etiketlenecek örnek ya da gözlemlerden meydana gelmiş bir eğitim kümesidir. Sınıflandırma modelinin her bir örneği için niteliklerine dayalı olarak atandığı sınıf etiketi ise sınıflandırma çıktısıdır (Ukuş 2014: 23).

Sınıflama yöntemleri; pazarlama, hastalık tanılarını, dolandırıcılık tespiti ve daha birçok konuda yaygın olarak kullanılmaktadır. Bir tıp doktorunun, şeker hastası olan hastasının alabileceği dört farklı tedaviden hangisini almasını gerektiğini saptamak için önceki şeker hastalığı verilerini analiz etmek istemesi ya da kampanya düzenleyip satışlarını arttırmak isteyen bir süpermarketin, kampanyaya katılma ihtimali olan müşterilerinin önceki müşteri verilerini kullanarak tespit etmesi sınıflandırma problemlerine örnek olabilir.

Veri madenciliği çalışmalarında en çok kullanılan modeller regresyon modelleridir. Veri madenciliği programlarının hemen hepsi regresyon fonksiyonlarını içermektedir. Regresyon modelleri matematiksel ve tanısal ölçümlere uyumludur, yüksek kalitede model üretir ve kolay anlaşılabilir yapıdadır. Regresyon modelleri; mühendislik, tıp, askeri, ekonomi ve tarım gibi birçok büyük çaplı problemlere uygulanabilmektedir (Refaat, 2007: 15).

Sınıflandırma ve regresyon modellerinde kullanılan başlıca yöntemler şu şekilde sıralanabilir.

- Karar Ağaçları
- Yapay Sinir Ağları
- Genetik Algoritmalar
- K-En Yakın Komşu Algoritması
- K-Ortalamalar Kümeleme Algoritması
- Naive Bayes

3.1.1. Karar Ağaçları

Karar ağaçları yöntemi; kural çıkarımının kolaylığı, yorumlanmasının basit olması ve yüksek güvenilirliği sebepleriyle son zamanlarda geniş bir uygulama alanı yakalayan veri madenciliği yöntemidir (Sun ve Li, 2008: 2). Karar ağaçlarında bazı algoritmalar geleceği tahmin etme özellikleri içerse de genel olarak bir sınıflandırma aracıdır. Bu açıdan bakıldığında karar ağaçları, istatistik disiplinde bulunan diskriminant analizi ile büyük benzerlikler göstermektedir (Akpınar, 2014: 200).

Karar ağaçları teorisi genellikle birbirinden bağımsız çalışmalarla ortaya çıkmıştır. Gelişim adımları aşağıda verilmiştir (Dolgun: 2014: 21-22).

- 1962 yılında Hunt, yapısal bilgiler kazanmak için karar ağaçları tekniğini kullanmıştır,

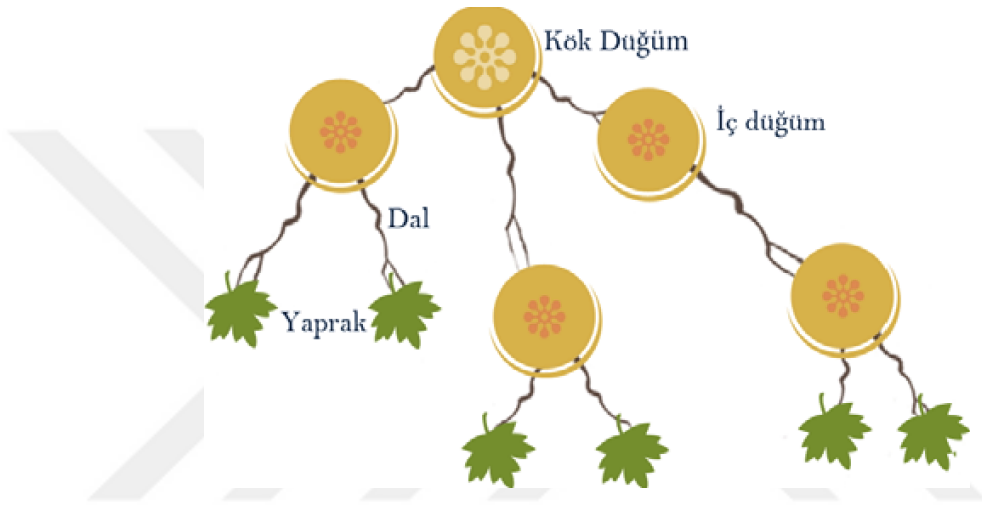
- 1963 yılında Feingbaum ve Simon; karar ağaçlarını, diskriminant ağları şeklinde kullanmıştır,
- 1967 yılında Samuel, oyunlarda kendini geliştirmek için karar ağaçlarını kullanmıştır,
- 1969 yılında Quinlan, problem çözümlerinde kullanmıştır,
- 1980 yılında Kass, Chi-Square Automatic Interaction Detector (CHAID) algoritmasını geliştirmiştir,
- 1983 yılında Carbonell, Mitchell ve Michalski en basit algoritma mantığı olan Top–Down Induction of Decision Trees (TDIDT) algoritmasını geliştirmişlerdir,
- 1984 yılında Breiman, Olshen, Stone ve Freidman, Classification and Regression Trees (C&R Tree) algoritmasını geliştirmişlerdir,
- 1986 yılında Quinlan, TDIDT algoritmasını temel alarak Iterative Dichotomiser 3 (ID3) algoritmasını geliştirmiştir,
- 1987 yılında Riverst R. Boolean, Decision List algoritmasını geliştirmiştir,
- 1993 yılında Quinlan, ID3 algoritmasının gelişmiş hali olan C4.5 algoritmasını geliştirmiştir,
- 1998 yılında Quinlan, C4.5 algoritmasını da geliştirerek C5.0 algoritmasını elde etmiştir.

Karar ağaçları algoritması ile oluşturulan model, ters dönmüş bir ağaca benzemektedir. Bu ağaç, karar verme noktaları olan düğümler ve düğümleri birbirine bağlayan dallardan meydana gelir. En üstte kök düğüm yer alır. Kök düğümde bir takım özellikler test edilir ve testin farklı sonuçlarına göre dallar oluşur. Karar ağacı yapısının en alt noktasında ise artık kendisinden dal oluşmayan yaprak düğümler yer alır (Seyrek ve Ata, 2010: 72). Karar ağacı işlemi sırasında oluşan yaprak, veri üzerinde belirlenmek istenen sınıflardan biridir. Karar ağacı işleminin başlangıç noktası kök düğümdür. Yaprğa ulaşana kadar yukarıdan aşağıya ardışık düğümler takip edilerek karar ağacı işlemi gerçekleşir (Ayık vd., 2007: 445). Yaprğa ulaşıldığında bir karar verilmiştir. Karar ağacının yapısında aynı sınıflamayı yapan

yapraklar olabilir. Bazı yapraklar aynı sınıflandırmayı yapsa da bunu farklı sebeplerle yaparlar.

Karar ağacı işlemi sırasında kullanılan algoritma değişirse ağacın yapısı da değişir. Buna bağlı olarak sınıflama da değişecektir. Şekil 3.1.1.1’de örnek bir karar ağacı yapısı yer almaktadır.

Şekil 3.1.1.1: Basit Bir Karar Ağacı Yapısı



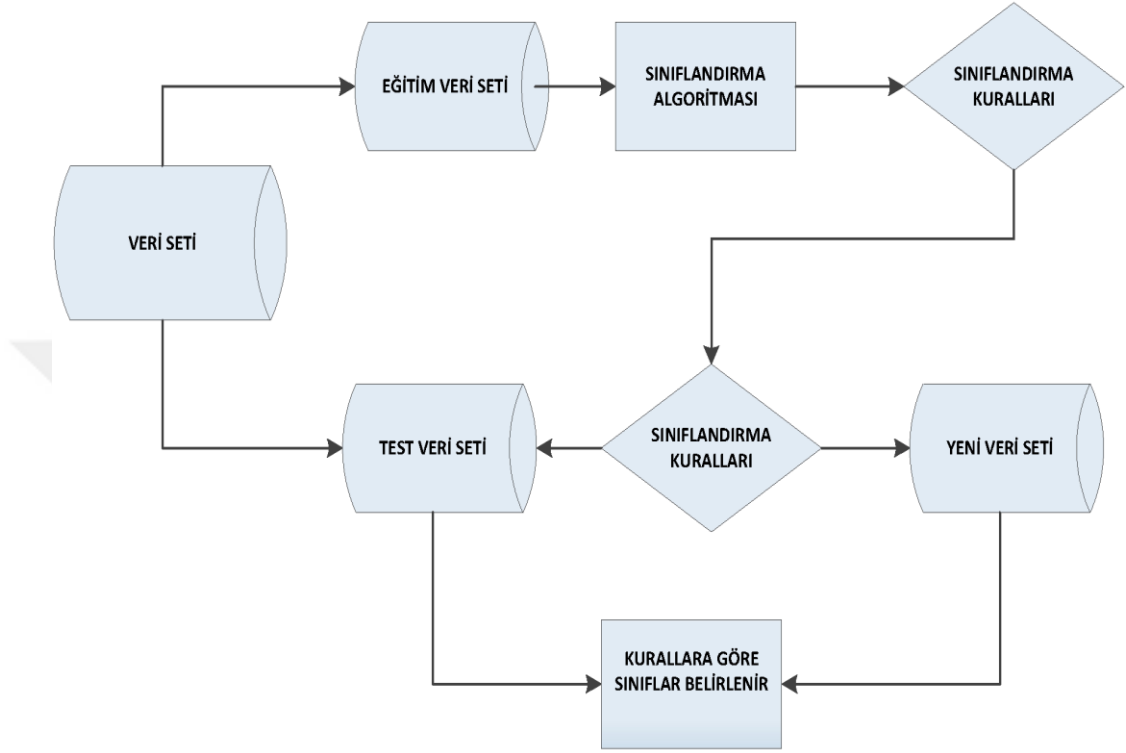
Kaynak: Kavzoğlu vd. (2012: 26).

Karar ağacı oluşturma işleminde sınıflamaya etkisi olmayan gereksiz dalların ağaçtan çıkarılmasına budama işlemi denir. Budama, ağaç kurulduktan sonra yapılabileceği gibi ağaç kurulurken de yapılabilir. Karar ağacı algoritmalarının büyük kısmında budanan değerler, varsayılan değerden daha düşük anlamlılığı olan değerlerdir (Silahtaroglu, 2008: 51).

Karar ağacı yöntemi ile verilerin sınıflandırılması iki aşamalı bir işlemdir. İlk aşama, eğitim aşamasıdır. Eğitim aşamasında daha önce bilinen bir eğitim verisi, model kurmak için sınıflama algoritması tarafından analiz edilir. Öğrenilen model, karar ağacı ya da sınıflama kuralları şeklinde gösterilir. İkinci aşama ise sınıflama aşamasıdır. Bu aşamada test verisi, karar ağacının ya da sınıflama kurallarının doğruluğunu saptamak amacıyla kullanılır. Eğer doğruluk kabul edilebilir oranda ise

kurallar yeni verilerin sınıflanması için kullanılır (Özekeş, 2003: 68). Karar ağaçları ile verilerin sınıflandırılması işlemi Şekil 3.1.1.2’de gösterilmiştir.

Şekil 3.1.1.2: Veri Sınıflandırma İşlemi



Kaynak: Pala (2013: 9)

Karar ağaçları veri setine kolay uygulanır ve normalleştirme gibi hazırlık yöntemlerine ihtiyaç duymaz. Büyük veri tabanlarında başarılı sonuçlar verir. Hem kategorik hem de sayısal veriler üzerinde güvenilir sonuçlar alır. Bu nedenlerle karar ağaçları, veri madenciliği çalışmalarında yaygın olarak kullanılmaktadır (Khoonsari ve Motie, 2012: 540). Ancak sürekli veriler üzerinden tahmin yapılırsa karar ağaçları uygun sonuçlar üretememektedir (Tantuğ, 2002: 17).

Oğuzlar (2004: 80), karar ağacı uygulamalarının yaygın olarak kullanıldığı alanları aşağıdaki gibi sıralamıştır,

- Belirli bir sınıfa atanacak değişkenlerin tespiti,

- Çeşitli olayların düşük, orta ve yüksek risk grupları şeklinde kategorize edilmesi,
- Gelecekte gerçekleşmesi muhtemel olayları tahmin edebilmek için kurallar oluşturulması,
- Parametrik model kurmak için kullanılacak çok sayıda veri ve değişken kümesinden önemli olanların seçilmesi,
- Kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikli değişkenlere dönüştürülmesi,
- Sadece belirli alt gruplara özgü ilişkilerin tanımlanmasında kullanılır.

3.1.1.1. Karar Ağaçlarında Kullanılan Algoritmalar

Karar ağaçlarını esas alan birçok algoritma geliştirilmiştir. Bu algoritmalar, karar ağacı oluşurken kök, düğüm ve dallanma işlemlerinde izledikleri yol ile birbirlerinden ayrılırlar.

Çalışmanın bu bölümünde CLS, ID3, C4.5, C5.0, C&R Tree, CHAID, QUEST ve Random Forest algoritmaları incelenecektir.

3.1.1.1.1. CLS Algoritması

Concept Learning System (CLS) algoritması, karar ağaçlarını kullanan ilk algoritmadır. CLS serisinde toplam 9 algoritma vardır. Bu algoritmaların ilk sekizi yalnızca iki sınıflı örnekler çözebiliyorken CLS9 algoritması çok sınıflı örnek setleri ile çalışabilmektedir. (Akgöbek ve Öztemel, 2006: 5).

3.1.1.1.2. ID3 Algoritması

ID3 algoritması, karar ağacı oluşturan bir sınıflama algoritmasıdır. Bu algoritma, J. Ross Quinlan tarafından 1986 yılında Sydney Üniversitesinde geliştirilmiştir. Bu algoritma, örneklerin nitelik değerlerini test ederek sınıflandırma yapar. Örnekler kümesinin niteliklerinin tanımlanmasıyla karar ağacı, yukarıdan aşağıya doğru oluşmaya başlar. Karar ağacının bütün düğümlerinde bir nitelik test edilir ve

sınıflandırma işlemi testin sonuçlarına göre yapılır. Sürecin sonunda yaprak düğüm oluşur (Collier vd., 1998: 23).

ID3 algoritması, en iyi bölen özelliği seçerken elindeki bilgileri sayısallaştırır. Diğer bir ifadeyle entropi yönteminden faydalanır. Entropi yönteminin amacı, sistemdeki rastgeleliği ve belirsizliği ölçmektir (Dunham 2006: 77). Bir özniteliğin kategorilerinin entropi değerleri $H(t)$, iki temelli logaritma kullanılarak aşağıdaki verilen eşitlik ile hesaplanır (Akpınar, 2014: 208).

$$H(t) = -\sum_{i=0}^{c-1} p\left(\frac{i}{t}\right) * \log_2 p\left(\frac{i}{t}\right) \quad (3.1)$$

Bu eşitlikte;

t: Veri dizisindeki toplam gözlem sayısını,

i: İlgili kategorideki gözlem sayısını,

c: Toplam kategori sayısı,

$p(i/t)$: İlgili kategorideki gözlem sayısının toplam gözlem sayısına olan oranını temsil etmektedir.

ID3 algoritması, sayısal örneklerin çözümünde yetersiz kaldığından C4.5 algoritması geliştirilmiştir.

3.1.1.1.3. C4.5 Algoritması

C4.5 algoritması, 1993 yılında Quinlan tarafından geliştirilmiş karar ağacı algoritmasıdır. Bu algoritma, ID3 algoritmasını esas almış ve onun kısıtlarını ortadan kaldırmıştır. C4.5 algoritması sürekli ve kategorik veriler üzerinde çalışabilmektedir (Ongun, 2015: 12). Karar ağacı oluştururken eksik verileri kullanmadığından ID3 algoritmasından daha anlamlı çözümler üretebilir. Eksik veriler, diğer örnekler kullanılarak tahmin edilir (Tapkan vd., 2011: 249).

3.1.1.1.4. C5.0 Algoritması

C5.0 karar ağacı algoritması, C4.5 algoritmasının geliştirilmiş halidir. C4.5 algoritmasına göre daha hızlı ve doğru sonuçlar vermektedir (Govindarajan, 2007: 3460). C5.0 algoritması, özellikle büyük veri setlerinde başarılı ağaçlar oluşturmaktadır.

3.1.1.1.5. C&R Tree Algoritması

C&R Tree algoritması; 1984 yılında Breidman, Freidman, Olshen ve Stone tarafından kamuoyuna duyurulmuştur. C&R Tree algoritması, veri madenciliği çalışmalarında hem sınıflama hem de regresyon problemlerini çözme kabiliyetine sahip parametrik olmayan istatistiksel bir tekniktir. Örnek kümesindeki bağımlı değişken kategorik ise sınıflama ağacı, sürekli ise regresyon ağacı adını alır (Deconinck vd., 2005: 92).

C&R Tree algoritması, karar ağacı oluştururken verileri ham hali ile işleyerek tekrarlı ya da ikili bölmeli yol izler. Değişkenler kategorik veya sürekli yapıda olabilir. Veri, kök düğümünden sonra iki çocuk düğüme (alt düğüm) ayrılır. Daha sonra her bir çocuk düğüm tekrar iki torun çocuk alt düğüme ayrılır. Bu algoritma ile karar ağacı, durdurma kuralı olmaksızın en yüksek büyüklüğe ulaşır. En yüksek büyüklüğe ulaşan ağaç, maliyet karmaşıklığı budama metodu ile kök kısmına doğru budanmaya başlar (Kuzey, 2012: 82).

C&R Tree algoritması, dallanma kriteri olarak gini endeksinden faydalanmaktadır. Gini endeksi $g(t)$, aşağıda verilen eşitlik ile hesaplanır (Akpınar, 2014: 212).

$$g(t) = 1 - \sum_{i=1}^c p_i^2 \quad (3.2)$$

p_i : Her bir kategorinin olasılığı,

c: Kategori sayısını temsil etmektedir.

C&R Tree algoritması, bağımlı değişken ile bağımsız değişken arasındaki ilişkiyi araştırmanın yanında bağımsız değişkenler arasında etkileşimleri de açığa çıkartmaktadır (Güner, 2014: 62).

3.1.1.1.6. CHAID Algoritması

CHAID algoritması, karar ağacı düğümlerinde en iyi bölünmeyi hesaplamak amacıyla 1980 yılında Kass tarafından geliştirilmiş (SPSS, 1999: 2). CHAID algoritması, en iyi bölünmeyi hesaplarken entropi yöntemi ya da gini endeksinden faydalanmaz. Değişken tipi sürekli ise F testini, kategorik ise ki-kare testini kullanır. CHAID algoritması, C&R Tree algoritması gibi hem sürekli hem de kategorik değişkenler ile çözümleme yapabilmektedir. ID3, C4.5, ve C&R Tree algoritmaları ikili karar ağaçları üretirken CHAID algoritması çoklu ağaçlar üretmektedir (Ture vd., 2008: 3).

CHAID algoritmasının üstünlükleri aşağıdaki gibi sıralanabilir (Ünveren ve Doğan, 2010: 42-43).

- Kategorik ve sürekli değişkenler üzerinde aynı anda çalışabilir,
- Bağımlı ve bağımsız değişkenler arasındaki ilişkileri daha ayrıntılı bir şekilde ele alır,
- Bağımsız değişkenleri, ağaç modeli halinde resmederek gösterebilir,
- Ağaç modeli sonuçları, diğer analiz sonuçlarına göre görsel anlamda daha kolay yorumlanabilir,
- Elde edilen sonuçlar kolay anlaşılabilir.

CHAID algoritmasının çalışma adımları aşağıda açıklanmıştır (SPSS, 2002: 240).

1. Veri setindeki bağımlı ve bağımsız değişkenlerin kategorileri arasında çapraz tablo oluşturulur. Bunların arasında en büyük p değerini alan (en az öneme sahip olan) kategori çiftleri bulunur ve bağımlı değişkenin türüne göre p

değeri hesaplanır. Bağımlı değişken sürekli yapıda ise F testi, Kategorik yapıda ise ki-kare testi ile p değeri hesaplanır.

2. P değeri en büyük olan bağımsız kategori çifti için önceden belirlenmiş alfa (α) düzeyi ile p değeri kıyaslanır;
 - a. $p > \alpha$ ise iki kategori birleştirilir ve 1. adıma geri dönülür,
 - b. $p < \alpha$ ise 3. adıma geçilir.
3. Bonferroni düzeltmesi yardımıyla bağımlı ve bağımsız kategori kümeleri için düzeltilmiş p değeri hesaplanır.
4. En yüksek öneme sahip bağımsız değişken, en küçük düzeltilmiş p değerine sahip olanlar arasından seçilir ve bu değer önceden belirlenmiş α düzeyi ile kıyaslanır;
 - a. $p \leq \alpha$ ise, belirlenmiş bağımsız kategori kümesi temel alınarak düğüm bölünür,
 - b. $p > \alpha$ ise, bölünme işlemi gerçekleşmez ve bu düğüm uç düğüm olur.
5. Bu işlem, durma kuralları sağlanıncaya kadar devam eder.

3.1.1.1.7. QUEST Algoritması

Quick, Unbiased, Efficient, Statistical Tree (QUEST) algoritması, 1997 yılında Loh ve Shih tarafından geliştirilmiştir. Karar ağacı oluştururken ikili ağaç üretir. QUEST algoritması, bölünme noktası belirlerken uygun bölünme için değişken de seçmektedir. Sınıflandırma problemlerinin çözümünde C&R Tree ve CHAID algoritmalarından daha hızlı çözüm üretmektedir. Ancak regresyon tipindeki problemleri çözme kabiliyeti yoktur (Karaibrahimoğlu, 2014: 40).

3.1.1.1.8. Random Forest Algoritması

Random Forest algoritması, çok sayıda karar ağacından oluşan bir veri madenciliği yöntemidir. Veri kümesinde yer alan sınıf değişkeni sürekli ise regresyon ağacı, kategorik ise sınıflama ağacı oluşturmaktadır. Random forest tekniğinde, budama işlemi yapılmaksızın karar ağaçları oluşturulur (Akman vd., 2011: 37). Oluşan her bir ağaç kendi kararını verir ve verilen kararlar içinde en yüksek oyu alan sınıf nihai karar olarak seçilir (Doğan ve Diri, 2010: 14).

Random forest algoritmasının çalışma adımları aşağıdaki gibidir (Akman, 210: 35-36).

1. Veri setinden n tane örnekleme yap. Her örneklemin $2/3$ 'ünü ağaç kurmak için öğrenme verisi olarak kullan.
2. Her örneklem için budama işlemi yapmaksızın karar ağacını şu şekilde oluştur;
 - a. Öğrenme veri setinden her düğüm için rastgele m tane tahmin değişkeni seç, içlerinden en çok bilgi kazancı sağlayacak olanı tespit et.
 - b. Tespit edilen tahmin değişkeni için gini indeksi ile en iyi dallanma kriterini hesapla. Hesaplanan değere göre veri setini her düğümde iki alt dala ayır.
 - c. a ve b'deki işlemleri yaprak düğüme ulaşana kadar her düğümde tekrarla.
3. n tane ağacın ayrı ayrı yapmış olduğu tahminleri bir araya getirip yeni tahminde bulun;
 - Regresyon ağaçları için yapılan oylamanın ortalamasını alıp final tahmini yap,
 - Sınıflama ağaçları için ise en çok oyu alan sınıfı tahmin olarak seç.

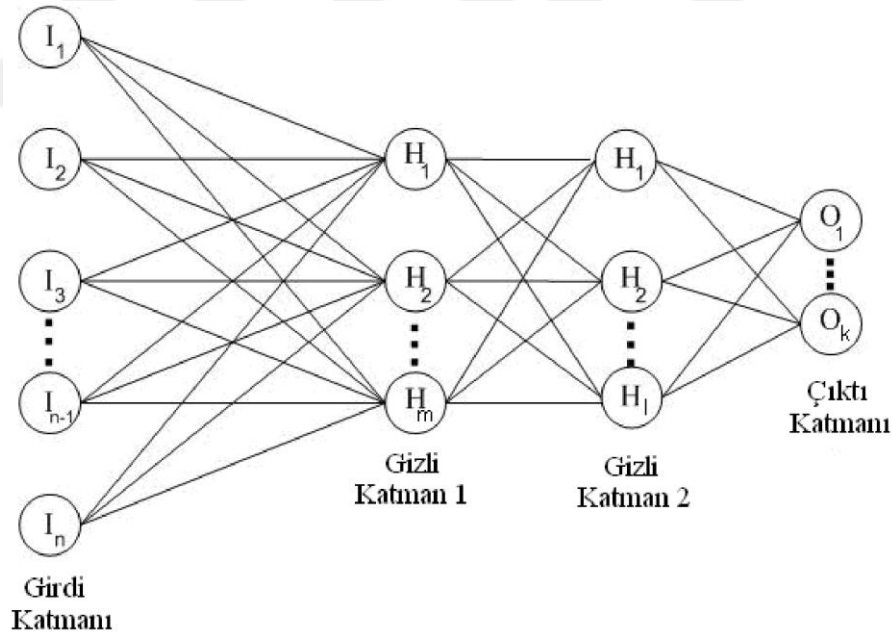
3.1.2. Yapay Sinir Ağları

Yapay sinir ağları, karmaşık yapıdaki problemlere çözüm üretmek için geliştirilmiş bir yapay zeka yöntemidir (Çakar vd., 1996: 77). Yapay sinir ağları, insan beynini temel alır. İnsan beyninin temel fonksiyonu olan öğrenme yolu ile bilgi keşfedebilme

ve yeni bilgiler oluşturabilme özelliklerini dışardan yardım almadan yapabilmek amacıyla geliştirilmiş bilgisayar sistemleridir (Akkaya vd., 2009: 192).

Yapay sinir hücreleri (nöronlar), katmanlar içinde bir araya gelip yapay sinir ağını oluşturur. Yapay sinir ağları; girdi katmanı, gizli katman ve çıktı katmanı olmak üzere üç ana katmandan meydana gelmektedir. Girdi katmanı ve çıktı katmanı bir tanedir ancak gizli katman, sinir ağının yapısına göre birden fazla olabilir. Girdi katmanından alınan girişler, girdi ve gizli katmanı arasında yer alan bağlantı ağırlıkları ile çarpılıp gizli katmana iletilir. Gizli katmandaki nöronlara gelen girişler toplanıp aynı şekilde gizli ile çıktı bağlantı ağırlıkları ile çarpılarak çıktı katmanına iletilir. Çıktı katmanında yer alan nöronlar, kedisine gelen girişleri toplar ve uygun çıkış üretir (Efe ve Kaynak, 2000: 148). Şekil 3.1.2.1’de bir girdi, bir çıktı katmanı ve 2 gizli katmandan oluşan bir yapay sinir ağı gösterilmiştir.

Şekil 3.1.2.1: Yapay Sinir Ağı



Kaynak: Kutlu ve Badur (2009: 29)

Yapay sinir ağları; sınıflandırma, tahmin ve genelleme gibi yeteneklerinden dolayı veri madenciliği çalışmalarında yaygın olarak kullanılmaktadır. Yüksek

sınıflandırma doğrulukları elde eden yapay sinir ağları, ürettikleri bilgileri kullanıcının anlayabileceği şekilde sunamadığı için kullanımı eğilim ve kural çıkarımına yöneliktir. Yapay sinir ağlarında saklı bilgilerin aktivasyonlar ve nöron bağıntılarına dağılmış olması tanımlama yeteneklerini sınırlar (Kulluk vd., 2011: 226). Veri madenciliği yöntemlerinin en çok bilinen ancak en az anlaşılan tekniği olan yapay sinir ağları günümüzde; denetim, sistem modelleme, ses tanıma, el yazısı tanıma, parmak izi tanıma, meteorolojik tahminler ve otomatik araç denetimi alanlarında kullanılmaktadır (Elmas, 2003: 25).

3.1.3. Genetik Algoritmalar

Genetik algoritmalar, karmaşık bir probleme en uygun çözümü ararken kullanılan ve doğal seleksiyon metotlarını temel alan bir yöntemdir. Genetik algoritmalar, biyolojik organizmaların genetik sürecini esas alır ve gerçek dünya problemlerinin çözümü için kullanılır (Bolat, 2006: 33).

Genetik algoritmaların geleneksel optimizasyon yöntemlerinden farklı olan yönleri vardır. Bu farklılıklar aşağıda özetlenmiştir (Tektaş vd., 2002: 5).

- Genetik algoritmalar, kodlanmış diziler ile çalışır.
- Genetik algoritmalar, tüm olası çözümlerin yer aldığı bir popülasyon içinde arama yapar.
- Optimizasyon tekniklerinde kullanılan amaç fonksiyonuna benzeyen bir uygunluk fonksiyonu yardımı ile çözüm uzayında problemin çözümünü arar.
- Genetik algoritmalar, stokastik yöntemleri kullanarak sonuca ulaşır.

Genetik algoritmalar, problemlere bir tane çözüm üretmektense birbirinden bağımsız olası çözümlerden oluşan bir çözüm kümesi üretir. Bu sayede arama uzayında aynı anda birden fazla nokta değerlendirilmekte ve bütünsel çözüme ulaşma olasılığı artmaktadır. Çözüm kümesinde yer alan çözümler birbirinden tamamen farklıdır. Her bir çözüm, çok boyutlu uzay üzerinde bir vektördür (Ayık vd., 2007: 446).

Standart bir genetik algoritmanın çalışma adımları aşağıda gösterilmiştir (Emel ve Taşkın, 2002: 132).

- Arama uzayında yer alan tüm olası çözümler dizi olacak şekilde kodlanır.
- Rastgele bir çözüm kümesi seçilip başlangıç popülasyonu olarak kabul edilir.
- Tüm dizilerin uygunluk değeri, amaç fonksiyonu ile hesaplanır. Uygunluk değerleri ile her bir dizinin çözüm kalitesi ortaya çıkar.
- Rastgele seçilen bir grup dizi, belirli bir olasılık değerine göre seçilir ve seçilen diziler çoğalma işlemine tabi tutulur.
- Yeni oluşan bireyler için de uygunluk değeri hesaplanır. Sonrasında mutasyon ve çaprazlama işlemleri gerçekleştirilir.
- Daha önce belirlenen kuşak sayısı boyunca önceki adımlar tekrarlanır.
- Tekrarlama işlemi, belirlenen kuşak sayısına ulaştığında sonlandırılır ve amaç fonksiyonuna göre en uygun dizi seçilir.

Genetik algoritmalar günümüzde; pazarlama, üretim, finans, yapay zeka uygulamaları, tesis yerleşimi, en uygun şekle sokma problemleri ve hat dengeleme alanlarında uygulanmaktadır.

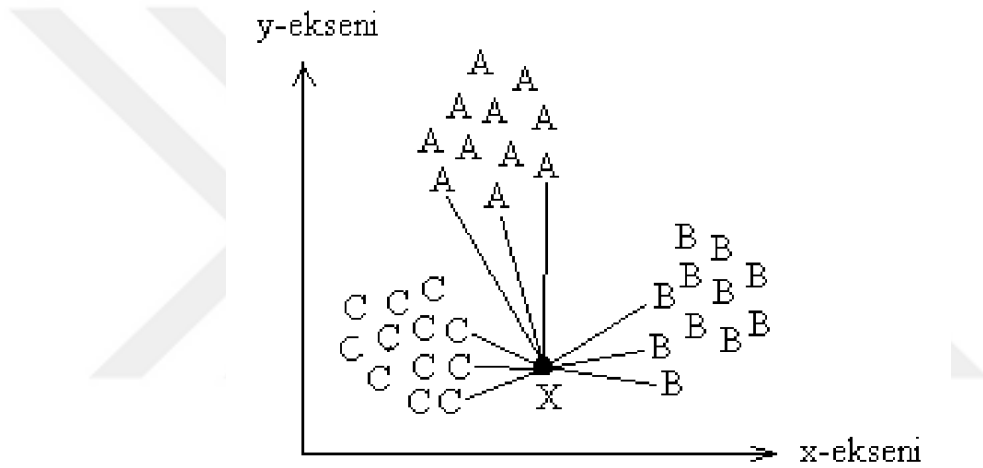
3.1.4. K-En Yakın Komşu

K-en yakın komşu tekniği, denetimli öğrenme teknikleri arasında yer alan ve sınıflandırma problemlerini çözen bir yöntemdir (Çalışkan ve Soğukpınar, 2008: 122). Bu yöntem, yeni bir veri geldiği zaman verinin en yakın “k” komşusuna bakar ve onun öğrenme verisi üzerinde hangi sınıfa atanacağına karar verir. Yeni gelen verinin daha önceki verilerden “k” tanesine yakınlığına bakılır ve en yakın mesafe hangi sınıfta ise yeni gelen veri oraya atanır. K-en yakın komşu tekniğinde her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmesi çok önemlidir. Veriler arasındaki mesafe; Öklid uzaklık fonksiyonu, Minkowski uzaklık fonksiyonu,

Manhattan uzaklık fonksiyonu gibi çeşitli uzaklık fonksiyonlarıyla hesaplanır (Daş ve Türkoğlu, 214: 192).

Şekil 3.1.4.1’de, $K=3$ olmak üzere K-en yakın komşu yöntemi gösterilmiştir. Burada A, B ve C sınıfları temsil etmektedir. Bu durumda X verisinin A, B ve C sınıflarına en yakın 3 tane verinin uzaklıkları hesaplanır. Her bir sınıfın en yakın uzaklıkları toplamı, ilgili sınıfın X verisine olan uzaklığını verir. Böyle bir durumda X verisi, C sınıfına atanır (Aşlıyan ve Günel, 2010: 532).

Şekil 3.1.4.1: K-En Yakın Komşu Yöntemi



Kaynak: (Aşlıyan ve Günel, 2010: 532)

K-en yakın komşu yönteminin uygulaması basittir ve gürültülü veriye karşı dirençlidir. Ancak bu yöntemin hesaplama maliyetleri yüksektir. Algoritma bir “k” parametresine ihtiyaç duyar ve en iyi sonucu alabilmek için hangi ölçümün uygulanacağı, hangi özelliklerin alınacağı bilgisi açık değildir (Gökgöz, 2015: 63).

Günümüzde k-en yakın komşu yöntemi; pazarlama, fen bilimleri, sigortacılık, şehir planlama, sağlık, deprem çalışmaları ve eğitim gibi alanlarda uygulanmaktadır.

3.1.5. K-Ortalamlar Kümeleme Algoritması

K-ortalamlar kümeleme algoritması, 1957 yılında Cox tarafından geliştirilmiştir. Veri madenciliği çalışmalarında verilerin ön hazırlığı ve veri setindeki gizli örüntülerin keşfi için en çok tercih edilen algoritmalarından biridir (Akın: 2008: 79). K-ortalamlar kümeleme algoritması, “n” sayıda birimi “k” sayıda kümeye ayırır ve birimlerin en uygun kümelerde olmalarını sağlar (Koyuncugil, 2007: 10).

K-ortalamlar kümeleme algoritmasının çalışma adımları şöyledir (Demiralay ve Çamurcu, 2005: 4-5).

1. İlk önce küme merkezleri tespit edilir. Bu işlem için iki yol vardır. İlk yolda nesneler içinden küme sayısı olan “k” tane rastgele nokta belirlenir. İkinci yolda ise küme merkezleri, tüm nesnelerin ortalaması alınıp tespit edilir.
2. Tüm nesnelerin seçilen merkez noktalarına olan uzaklıkları hesaplanır. Bulunan sonuçlara göre nesneler, “k” tane küme içinden kendilerine en yakın olan kümeye atanır.
3. Oluşan kümelere ait yeni merkez noktaları ilgili kümedeki tüm nesnelerin ortalama değeri ile değiştirilir.
4. Küme merkez noktalarında değişme olmayana kadar 2. ve 3. işlem basamakları tekrarlanır.

K-ortalamlar yönteminin uygulamasının basit olması, büyük veri kümelerinde hızlı çalışması araştırmacıya avantaj sağlarken gürültülü verilere karşı duyarlılığı ve küme sayısını belirleyemediği için en uygun “k” küme sayısını bulana kadar deneme yanılma işlemi yapması yöntemin dezavantajlarıdır (Yıldırım, 2016: 24).

3.1.6. Naive Bayes Sınıflandırıcısı

Basit bir olasılıksal sınıflandırıcı olan Naive Bayes, ismini İngiliz Matematikçi Thomas Bayes’ten almıştır. Naive Bayes sınıflandırıcısı, Bayes karar teorisini esas alır. Tüm örneklerin sınıflara, hangi olasılıkla ait olduğunu hesaplar. Basit bir

yapısının olması, kayıp verilere karşı duyarsızlığı ve hızlı hesaplama yapması sebeplerinden veri madenciliği çalışmalarında yaygın olarak kullanılmaktadır (Liu ve Lei, 200: 414-415). Bayes teoremi, aşağıdaki eşitlik ile ifade edilir (Çalış vd., 2013: 3).

$$P(A/B) = [P(B/A) * P(A)] / P(B) \quad (3.3)$$

Bu eşitlikte;

$P(A)$: A olayının bağımsız olasılığını,

$P(B)$: B olayının bağımsız olasılığını,

$P(B/A)$: A olayının gerçekleştiği bilindiğinde B olayının gerçekleşme olasılığını,

$P(A/B)$: B olayının gerçekleştiği bilindiğinde A olayının gerçekleşme olasılığını ifade etmektedir.

Naive Bayes algoritması uygulanırken niteliklerin bağımsız olduğu kabul edilmektedir. Nitelikler birbirini etkilerse olasılık hesaplaması oldukça zorlaşır. Bu nedenle niteliklerin birbirinden bağımsız ve hepsinin aynı derecede önemli olduğu kabul edilir (Çalış, 2013: 27).

Naive Bayes sınıflandırıcısı, kesikli veriler ile çalışır. Bu sebeple sürekli değer almış bağımlı ya da bağımsız değişkenler kategorik hale getirilmelidir. Örneğin; sürekli değerler almış yaş değişkeni, “<18” , “18-36” , “36-54” gibi yaş aralıklarına dönüştürülmelidir. Naive Bayes, model öğrenilirken tüm çıktıların öğrenme kümesinde kaç kere gerçekleştiğini hesaplar. Hesaplanan değere, öncelikli olasılık denir. Bağımsız değişken / bağımlı değişken kombinasyonun meydana gelme sıklığı da Naive Bayes ile bulunur. Bulunan sıklıklar, öncelikli olasılıklar ile birleştirilerek tahminlemede kullanılır (Akbulut, 2006: 21-22).

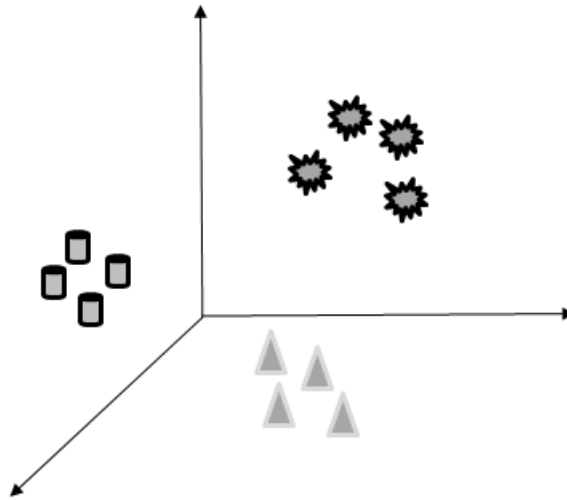
Naive Bayes algoritması, gerçek dünya problemlerine uygulandığında çok iyi çözümler üretebilmektedir. Algoritmanın eğitim ve değerlendirme adımları çok hızlı

işlemesine rağmen karmaşık sınıflama problemlerinde yetersiz kalmaktadır (Uzun, 2007: 51).

3.2. Kümeleme

Veri kümesinde var olan bilgileri, çeşitli uzaklık ölçütlerine göre gruplara bölme işlemine kümeleme analizi adı verilir. Analiz sonunda oluşan grupların her birine ise küme adı verilir. (Dinçer, 2006: 24). Kümeleme analizi, veri kümesindeki örneklerin birbirine çok benzediği kümeler oluşturur. Ancak bu kümeler, özellik bakımından birbirlerinden farklıdır ve örnekler bu farklı kümelere bölünür. Kümeleme analizinde kümeler hakkında ön bilgi olmayabilir ve oluşacak küme sayısı belirsizdir. (Tiryaki, 2006: 17). Başarılı bir kümeleme analizi sonucu, Şekil 3.2.1’de olduğu gibi küme içindeki birimlerin uzaklıkları minimum, kümeler arası ise uzaklık maksimum olmalıdır.

Şekil 3.2.1: Kümeleme Analizinde Kümelerin Konumları



Kümeleme analizi, birimleri homojen gruplara bölmek için kullanılır. Kümeleme analizinin amaçları aşağıdaki gibi sıralanabilir (Murat ve Şekerler, 2009: 4761);

- “n” sayıda birim ya da nesneyi, “k” değişkene göre belirlenen özellikleri temel alarak homojen alt gruplara bölmek. Bölünen grupların ise kendi aralarında mümkün olduğu kadar heterojen yapıda olmasını sağlamak.
- “k” sayıda değişkeni, “n” sayıda birimde belirlenen değerlere göre ortak özellikleri açıkladığı kabul edilen alt kümelerle ayırmak ve ortak faktör yapılarını göstermek.
- Birimleri ve değişkenleri beraber ele alıp ortak “n” sayıda birimi, “k” değişkene göre ortak özellikli alt kümelerle bölmek.
- Birimleri, “k” değişkene göre belirlenen değerler için takip ettikleri tipolojik ve biyolojik sınıflamayı ortaya koymak.

Kümeleme modellerinde genellikle oluşacak küme sayısı önceden bilinmez. Verilerin mevcut durumundan hareketle kümeler oluşur. Bu sebeple elde edilen sonuçlar tahmin amacıyla kullanılmaz. Kümeleme analizinde veriler sınıflara ayrılmazlar. Herhangi bir sınıfı olmayan veriler kümelendir. Bazen kümeleme analizi, sınıflama modeli kurulmadan yapılması gereken ön işlem olarak uygulanabilir (Tatlídil, 1992: 329).

3.3. Birliktelik Kuralları

Birliktelik kuralları, eş zamanlı gerçekleşen ilişkileri tanımlayan veri madenciliği yöntemidir. (Gürbüz, 2009: 47). İşletme ve organizasyonlar, kayıt altına alınan verilerin her geçen gün daha da büyümesi yüzünden veri tabanlarındaki birliktelik kurallarını keşfetmek istemektedirler. Birliktelik kuralları yardımıyla daha verimli kararlar alınmaktadır (Ergün, 2012: 158).

Birliktelik kuralı, aşağıdaki gibi tanımlanabilir (Zhu, 1998: 14).

$$X_1, X_2, X_3, \dots, X_m \Rightarrow Y_1, Y_2, Y_3, \dots, Y_n \quad (3.4)$$

Yukarıda yer alan ifade de X_i ve Y_j nesne ya da yapılan işi göstermektedir. Bu kural genellikle, " $X_1, X_2, X_3, \dots, X_m$ " nesneleri ya da işi gerçekleştiğinde, sık olarak " $Y_1, Y_2, Y_3, \dots, Y_n$ " nesne ya da olayların gerçekleşeceğini ifade eder.

Veri tabanlarına birliktelik kuralları uygulanırken ilk önce sık tekrarlanan nesneler tespit edilir. Tespit edilen nesnelerden güçlü birliktelik kuralları oluşturulur. Tekrarlanan nesnelerin her biri en az, önceden belirlenmiş minimum destek sayısı kadar tekrarı olmalıdır. Oluşturulan birliktelik kuralları ise minimum güven ve minimum destek ölçütlerine uygun olmalıdır (Zaki: 1999: 16).

Birliktelik kuralları, en çok pazar sepet analizinde kullanılmaktadır. Pazar sepet analizinde temel amaç, marketlerde hangi ürünlerin birlikte satıldığını ortaya çıkartmaktır (Birant vd., 2010: 260). Birliktelik kuralları, aşağıdaki gibi ilişkileri tanımlayabilir.

- Tüketiciler, bebek bezi satın aldığında % 80 ihtimalle bebek maması da alırlar.
- Yağsız yoğurt alan müşteriler, % 75 ihtimalle diyet içecek de alırlar.
- Diz üstü bilgisayar alan tüketiciler, % 85 ihtimalle diz üstü bilgisayar çantası da alırlar.

Sepet analizi, müşterilerin tüketim alışkanlıklarını analiz ederek hangi ürünleri birlikte aldıklarını keşfeder. Elde edilen bu bilgiler; kampanya, promosyon ve marketlerde raf düzenleme gibi satış stratejilerinde kullanılabilir.

3.4. Ardışık Zaman Örüntüler

Ardışık zaman örüntüler analizi, birliktelik kurallarına benzemektedir. Bu analizin en önemli farkı olayların zaman sıralarıyla ilgilenmesidir. Bir tüketici X ürünü alsın, ertesi gün ya da sonra ki günlerde Y ürününü daha sonraki günlerde ise Z ürününü alması, zaman içinde bir örüntü oluşturacaktır (Silahtaroglu, 2004: 14). Ardışık zaman örüntüler analizi, aşağıdaki örneklerde görüldüğü gibi aralarında ilişki olan ve

birbirini izleyen zaman dilimlerinde gerçekleşen örüntülerin tanımlanmasında kullanılmaktadır.

- İkinci el araç alan bir tüketici % 95 ihtimalle on gün içinde araç sigortası yaptıracaktır.
- Cep telefonu alan bir tüketici % 75 ihtimalle bir ay içinde mobil internet paketi alacaktır.
- X ameliyatını olan bir hasta % 60 ihtimalle iki gün içinde Y enfeksiyonunu kapabilir.

3.5. Veri Madenciliği Çalışmalarında Kullanılan İstatistiksel Teknikler

İstatistiksel teknikler, veri madenciliği sürecinin her adımında kullanılmaktadır. Veri madenciliği çalışmalarında başlıca kullanılan istatistiksel yöntemler aşağıda ele alınmıştır.

3.5.1 Regresyon Analizi

Regresyon, en az iki değişken arasındaki ilişkinin denklem ile ifadesidir. Değişkenler arasındaki ilişki, denklem ile ifade edilebilirse bilinen değişken değerleri yardımı ile bilinmeyen değişken değerleri tahmin edilir. Burada amaç, bir serpm diyagramındaki noktalara en yakın yerden geçen çizgiyi cebirsel bir fonksiyon ile sağlayan denklemi bulmaktır. Bu çizgiye regresyon çizgisi, denkleme ise regresyon denklemi denir (Çil, 2005: 282).

Doğrusal regresyon denklemi, Formül 3.5'te verilmiştir.

$$Y_i = a + b X_i \quad (3.5)$$

Bu denklemdeki a ve b sabit birer değerdir. Regresyon denklemi, bağımsız değişken “X_i” deki bir birimlik değişmeye karşı, bağımlı değişkenden “Y_i” de meydana gelecek ortalama değişikliği izah eder. İki değişken arasındaki regresyon, doğrusal

olacağı gibi eğrisel de olabilir. İkiiden fazla değişkenin olduğu regresyon modeline ise çoklu regresyon denir (Çil, 2005: 282).

Örneğin, günde bir paket sigara içen kişinin akciğer kanserine yakalanma olasılığı % 61 iken hiç sigara kullanmayan bir kişinin akciğer kanseri olma olasılığı % 13' tür sonucu bir regresyon ilişkisidir.

3.5.2. Lojistik Regresyon

Lojistik regresyon, ileri düzey bir regresyon tekniğidir. Lojistik regresyon yönteminin temel amacı, bir ya da daha fazla bağımlı değişken ve bağımsız değişken arasındaki ilişkiyi modellemektir (Atan ve Çatalbaş, 2004: 12). Sınıflama yapma özelliği de olan lojistik regresyon analizinde bağımsız değişken kategorik yapıda olup kesikli değerler almaktadır. Bağımsız değişkenlerin ise tamamının ya da bazılarının kategorik veya sürekli yapıda olma zorunluluğu yoktur (Işığık, 2003:3). Lojistik regresyon analizi, bağımsız değişkenlerin çok değişkenli normal dağılıma uygun olması ve sahip olduğu grupların kovaryans ve varyanslarının homojen olması gibi önemli varsayımlara ihtiyaç duymaz. Bu özellikleri sebebiyle çalışmalarda sıklıkla tercih edilmektedir (Akın ve Şentürk, 2012: 184). Lojistik regresyon, aşağıdaki gibi formüle edilir (Özdamar, 2012: 475).

$$P = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}} \quad (3.6)$$

Bu eşitlikte;

P: İncelenen olayın gözlenme olasılığını,

k: Bağımsız değişken sayısını,

e: 2.71 sayısını,

$\beta_1, \beta_2, \dots, \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

$X_1, X_2, \dots, X_k$: Bağımsız değişkenleri ifade etmektedir.

Lojistik regresyon ile doğrusal regresyon arasındaki önemli farklar vardır. Bu farklar aşağıdaki gibi sıralanabilir (Coşkun vd., 2004: 43);

- Lojistik regresyon analizinde tahmin edilecek bağımlı değişken kesikli yapıdayken doğrusal regresyon analizinde sürekli yapıdadır.
- Lojistik regresyon analizinde bağımlı değişkenin alabileceği değerlerden bir tanesinin gerçekleşme olasılığı tahmin edilirken doğrusal regresyon analizinde bağımlı değişkenin değeri tahmin edilir.
- Lojistik regresyon analizinde bağımsız değişkenlerin dağılımı ile ilgili ön koşul aranmazken doğrusal regresyon analizinde bağımsız değişkenlerin çoklu normal dağılım göstermesi koşulu aranır.

3.5.3. Diskriminat Analizi

Günümüzde hem akademisyenlerin hem de iş dünyasının sıklıkla kullandığı diskriminant analizinin temelleri 1930'lu yıllarda Fisher tarafından atılmıştır. Çok değişkenli istatistik yöntemlerinden biri olan diskriminant analizi; finans, biyoloji, davranış bilimleri gibi alanlarda yaygın olarak kullanılmaktadır (Öz, 2005: 47).

Diskriminant analizi, veri setindeki birimlerin araştırılan özellikleri yönünden minimum hata ile ait oldukları gruplara atanması için yapılan işlemler bütünüdür. Diskriminant analizinin temelinde, ele alınan birimlerin ait oldukları grubun saptanmasını sağlayacak bir ayırma fonksiyonunun bulunması yer alır. Bu ayırma fonksiyonun belirlenmesinde grupların ortalamaları arasındaki farkın en yüksek olması istenmektedir. Diskriminant analizinde diskriminant fonksiyonu bulunur ve bulunan fonksiyonun anlamlılığı test edilir. Son olarak hatalı gruplandırma olasılıkları hesaplanarak birimler ait oldukları gruplara atanır (Filiz, 2005: 79).

Diskriminant analizi ile bulunan diskriminant fonksiyonları, tahmin değişkenlerinin doğrusal bileşenlerinden meydana gelir. Bu fonksiyonlar, hangi tahmin değişkenlerinin gruplar arası farklılığa sebep olduğunu ortaya çıkarır. Gruplar arası farklılığın oluşmasına sebep olan değişkenlere ise diskriminant değişkenler adı

verilir. Diskriminant analizi aynı zamanda analiz esnasında oluşan gruplardan herhangi birine ait olan ancak hangi gruptan geldiği bilinmeyen bir birimin ait olduğu grubu minimum hata ile tespit eder (Ünsal, 2000: 19).

Diskriminant fonksiyonunun matematiksel modeli aşağıdaki gibidir (Çelik, 2010: 136).

$$Z = \alpha + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (3.7)$$

Bu ifadede;

Z: Diskriminant skorunu,

α : Sabit terimi,

$b_1, b_2, \dots, b_n$: Diskriminant katsayılarını,

$x_1, x_2, \dots, x_n$: Bağımsız değişkenleri ifade etmektedir.

Diskriminant analizi ile kümeleme analizinin aynı amaçta çözümlerler yaptıkları düşünülmemelidir. İki yöntemin de bazı benzerlikleri olmasına rağmen kümeleme analizinde küme sayısı tam olarak bilinmez ve gelecekte kullanılabilirlik özelliği yoktur. Kümeleme analizinde kümeler oluşturulduktan sonra her bir kümenin doğru bir şekilde meydana geldiğinin değerlendirilmesi gerekmektedir. Bu işlemi yapabilmek için kümeleme analizinin sonunda oluşan kümelerin, analiz öncesi olarak ayırma analizine sokulması yeterlidir. Kümeleme analizi, bilgiyi özetlediğinden bu bilgiler diskriminant analizi için veri olarak kullanılır (Filiz, 2005: 79).

3.5.4. Faktör Analizi

Faktör analizi, değişkenler arasındaki ilişkilere dayanarak çok sayıdaki değişkeni daha az sayıda değişkene indirebilen çok değişkenli bir analiz tekniğidir. Faktör analizi, çok sayıda değişken kullanılarak yapılan ölçümlerle altta yatan gizli değişken yapısının ve boyutlarının belirlenmesi için kullanılır. Bunu yaparken de çok büyük miktarda veriyi mümkün olan en az bilgi kaybıyla sadeleştirerek daha az sayıda değişkenle göstermeye olanak sağlar (İslamoğlu ve Alınacı, 2016: 395).

Faktör analizi, aynı yapıyı ya da niteliği ölçen değişkenleri bir araya toplayarak ölçmeyi az sayıda faktör ile açıklamayı amaçlayan bir istatistiksel tekniktir. Faktör analizi, bir faktörleştirme ya da ortak faktör adı verilen yeni değişkenleri ortaya çıkartma ya da maddelerin faktör yük değerlerini kullanarak kavramların işlevsel tanımlarını elde etme süreci olarak da tanımlanmaktadır. Açımlayıcı ve doğrulayıcı olmak üzere iki tür faktör analizi yaklaşımı vardır. Açımlayıcı faktör analizinde değişkenler arasındaki ilişkiden hareketle faktör bulmaya yönelik bir işlem; doğrulayıcı faktör analizinde ise değişkenler arasındaki ilişkiye dair daha önce saptanan bir hipotezin ya da kuramın test edilmesi söz konusudur (Büyüköztürk, 2015: 133).

Faktör analizinin en önemli noktası, oluşturulan faktörlerin adlandırılıp anlamlandırılmasıdır. Bu işlem esnasında faktörlerden yoğun bir şekilde etkilenen gözlemsel değişkenleri göz önünde bulundurmalı ve onları neyin etkilediği sorulmalıdır (Işık vd., 2004: 55).

3.5.5. Ki-Kare Testi

Ki-kare testleri en çok kullanılan prosedürler arasında yer alır. Bu testlerin temeli 1900 yılında Karl Pearson tarafından geliştirilen bir tekniğe dayanır. Örneklem yolu ile elde edilen rakamların uygun olup olmadığını, bir başka ifadeyle gözlenen değerlerin beklenen değerlere uygunluk gösterip göstermediği ki-kare testi ile tespit edilir (Başar ve Oktay, 2007: 1).

Ki-kare testi (Güngör ve Bulut, 2008: 85),

- İki ya da daha çok grup arasında fark olup olmadığının testinde,
- İki değişken arasında bağ olup olmadığının testinde,
- Gruplar arasındaki homojenlik testinde,
- Örneklemden elde edilen dağılımın istenilen bir teorik dağılıma uygunluğunun testinde,

- Varyans için ki-kare testinde,
- Kontenjans sayısının hesaplanmasında kullanılmaktadır.

3.5.6. Zaman Serileri Analizi

Zaman serisi, zamanın periyodik noktalarında bir cevap değişkeninin gözlenip verilerin kaydedilmesiyle oluşur (Özek vd., 2010: 101). Zaman serileri analizi, zaman serilerinden özellikler çıkarabilmek için kullanılan matematiksel yöntemlerdir.

Zaman serisi analizi yapılırken verilerin zamana bağlı değişimleri incelenir. Örneğin bir süpermarkete gün içinde gelen müşteri sayısı günün belli saatlerinde fazla olurken belli saatlerinde ise az olmaktadır yani zamana bağlı olarak değişmektedir (Şeker, 2015: 24).

Zaman serileri, lineer zaman serileri ve lineer olmayan zaman serileri olmak üzere ikiye ayrılır. Lineer zaman serileri analiz yöntemiyle bir zaman serinin analiz edilebilmesi için zaman serisinin durağanlaştırılması gerekir. Lineer olmayan zaman serileri analizi, gömülme boyutu ve kaos gibi tekniklere dayanır ve durağanlaştırılma gereksinimi yoktur (Aydın, 2006: 64).

4. GEREÇ VE YÖNTEM

Çalışmanın bu bölümünde TÜİK tarafından 2014 yılında yapılan TSA'ya ait veriler üzerinde karar ağacı algoritmaları uygulanmıştır. Bu doğrultuda TÜİK'ten 2014 yılına ait TSA B grubu mikro verileri talep edilmiştir. Veri madenciliği uygulaması, IBM SPSS Modeler 18.0 programı ile gerçekleştirilmiştir.

4.1. Türkiye Sağlık Araştırması

Hane halklarına uygulamak üzere geliştirilen TSA, ilk kez 2008 yılının Nisan ayında TÜİK tarafından gerçekleştirilmiştir. Bu araştırma ile sağlık alanında fertlerin genel sağlık sorunları ve bu sorunların yarattığı günlük kısıtlamaları, fertlerin gün içindeki bedensel faaliyetleri, son bir yıl içinde alınan sağlık hizmetleri, alkol ve tütün tüketimi gibi konulardaki bilgi eksikliklerini gidermek hedeflenmiştir. TSA, ülke genelinde bu kapsamda yapılan ilk araştırmadır (TÜİK, 2014: 3). Türkiye tahmini verecek şekilde tasarlanan araştırma; 2008, 2010, 2012, 2014 ve 2016 yıllarında uygulanmıştır.

TSA soru formu;

- Kapak sayfası,
- Hane halkı soru formu,
- 0-6 yaş grubu soru formu,
- 7-14 yaş grubu soru formu,
- 15+ yaş grubu soru formu olmak üzere beş ana kısımdan oluşmaktadır.

Hane halkı soru formunda fertlerin; cinsiyet, yaş, medeni durum, çalışma durumu, gelir ve öğrenim durumları ile ilgili sorular sorulmuştur. 0-6 ve 7-14 yaş gruplarında yer alan fertlere; başlarından geçen kazalar, hastalıklar ve sağlık hizmetlerinden faydalanma durumu ile ilgili bilgiler alınmıştır. 15+ yaş grubundaki fertlerden ise genel sağlık durumu, sağlık problemlerine dayalı günlük yaşamsal faaliyetlerini gerçekleştirmedeki kısıtlamalar, günlük bedensel aktiviteler, tütün kullanımı ve son bir yıl içinde almış oldukları sağlık hizmetleri konularında bilgiler alınmıştır.

TSA’da kapsanan kitle, Türkiye Cumhuriyeti sınırları içindeki tüm yerleşim yerlerinde bulunan hane halklarıdır. Kurumsal olarak nitelendirilen nüfus, (hastanede sürekli olarak kalan, yurtlarda, hapishanede, huzurevinde kalan, askerler vb.) kapsam dışı tutulmuştur.

Türkiye toplam tahmini verebilecek şekilde tasarlanan araştırmanın örnek hacmi, 9740 hane olarak belirlenmiştir. Örneğe seçilen 9740 hanenin 8634’ü ile anket çalışması yapılmıştır. Metodoloji gereği Ağustos-Ekim dönemini kapsayan 3 aylık süreçte gerçekleşen çalışmada toplam 26075 fert ile anket yapılmıştır (<http://www.tuik.gov.tr/PreHaberBultenleri.do?id=18854> , Erişim tarihi: 02 Mart 2017).

Araştırma kapsamında anket yapılan fertlere;

- Hane halkı soru formunda 28,
- 0-6 yaş grubu soru formunda 22,
- 7-14 grubu soru formunda 10,
- 15+ yaş grubu soru formunda ise 103 soru sorulmuştur.

4.2. Problemin Tanımlanması

Çalışmanın problemi; demografik, çevresel, ekonomik vb. faktörlerin genel sağlık durumu üzerinde ne kadar etkiye sahip olduğunun belirlenmesidir olarak tanımlanmıştır. Problem tanımlandıktan sonra çözüme yönelik analize girecek bağımlı değişkenin; “genel olarak sağlık durumunuz nasıldır?” sorusuna verilen cevapla oluşturulan “GENELSAGLIKDURUM” değişkeninin olmasına karar verilmiştir.

4.3. Analiz Öncesi Veri Seti

TÜİK’ten talep edilen B grubu mikro veriler, virgülle ayrılmış (.csv) excel dosyası formatında alınmıştır. 26075 fert ile yapılan anketten oluşan veri seti, 27076 satır ve

471 sütundan oluşmaktadır. Veri setinde 465 tane değişken yer almaktadır. Değişkenlerin 226 tanesi 15 yaş ve üzeri fertlere sorulan sorular ile oluşturulmuştur (EK-1).

Veri ön işleme adımından önce, IBM SPSS Modeler yazılımı ile veri setinin kalitesi incelenmiştir. Bunun için Sources sekmesinin altında yer alan Var. File işlemcisi ile veri seti okutulmuştur. Var. File işlemcisine, Output sekmesinin altında yer alan Data Audit işlemcisi bağlanmış ve sonuçlar incelenmiştir.

Şekil 4.3.1: Veri Kalitesinin İncelenmesi

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value
YAS_GRP	Nominal	--	--		Never	Fixed	100	26075	
CINSİYET	Flag	--	--		Never	Fixed	100	26075	
DOĞUM_YER	Nominal	--	--		Never	Fixed	100	26075	
VATANDASLIK	Nominal	--	--		Never	Fixed	100	26075	
TEDAVI_MAS...	Flag	--	--		Never	Fixed	100	26075	
TEDAVI_MAS...	Flag	--	--		Never	Fixed	100	26075	
TEDAVI_MAS...	Flag	--	--		Never	Fixed	100	26075	
TEDAVI_MAS...	Flag	--	--		Never	Fixed	100	26075	
TEDAVI_MAS...	Flag	--	--		Never	Fixed	100	26075	
TEDAVI_MAS...	Flag	--	--		Never	Fixed	100	26075	
TEDAVI_MAS...	Flag	--	--		Never	Fixed	100	26075	
EĞİTİM	Nominal	--	--		Never	Fixed	90,159	23509	2
MEDENİ_DU...	Nominal	--	--		Never	Fixed	73,361	19129	6
ÇALIŞMA_D...	Nominal	--	--		Never	Fixed	73,361	19129	6
ÇALIŞMAMA...	Nominal	--	--		Never	Fixed	44,291	11549	14
ÖNCEDEKİ...	Flag	--	--		Never	Fixed	44,291	11549	14
İSTEKİ_DUR...	Nominal	--	--		Never	Fixed	49,085	12799	13

Quality sekmesini seçip veri kalitesi incelediğimizde aşırı uç değerler gözlenmemiştir ancak veri setinde çok sayıda boş hücre yer almaktadır.

4.4. Veri Ön İşleme

Veri madenciliği sürecinde en çok zaman alan adım, veri ön işleme adıdır. Bu adımda sağlıklı sonuçlara ulaşabilmek için elde edilen veriler üzerinde temizleme, bütünleştirme, dönüştürme ve veri indirgeme işlemlerinden gerekli olanları yapılır.

4.4.1. Veri Temizleme

Veri madenciliği çalışmalarında kullanılan veri seti çok fazla gözlem ve değişkenden oluşabilir. Güçlü bir model kurabilmek için sonuca etki eden değişkenleri seçip doğru miktarda veri ile çalışılmalıdır. Eksik veriler, doğru modelin kurulmasını engelleyebileceği gibi fazla değişken ve veri, algoritmayı yorup doğru sonuçtan uzaklaşmaya sebep olabilir.

Anketin uygulanması esnasında bağımlı değişkene ait “genel olarak sağlık durumunuz nasıldır?” sorusu, 15 yaşından büyük fertlere sorulmuştur. 0-6 ve 7-14 yaş grubunda yer alan fertlerin bağımlı değişkene ait hücreleri boştur. Bu nedenle 0-6 ve 7-14 yaş grubunda yer alan 6946 fert, veri setinden çıkartılmıştır.

4.4.2. Veri Dönüştürme

Çalışmada bağımlı değişken olarak belirlediğimiz “GENELSAGLIKDURUM” değişkenine 5 kategoride cevap alınmıştır. “Çok iyi, iyi, orta, kötü ve çok kötü” cevapları, sınıflama performansını arttırabilmek amacıyla “iyi”, “orta” ve “kötü” olmak üzere 3 kategoride toplanmıştır. Bağımlı değişken üzerinde yapılan veri dönüştürme işlemi Tablo 4.4.2.1’de gösterilmiştir.

Tablo 4.4.2.1: Veri Dönüştürme İşlemi

Dönüştürme İşlemi Uygulanan Değişken	Değişken Değerinin İlk Hali	Değişken Değerinin Son Hali
GENELSAGLIKDURUM	Çok iyi	İyi
	İyi	
	Orta	Orta
	Kötü	Kötü
	Çok Kötü	

4.4.3. Veri İndirgeme

Veri setinde yer alan tüm değişkenlerin veri madenciliği uygulamasına dâhil edilmesi anlamlı sonuçlar vermediği için 45 tane bağımsız değişken belirlenmiştir. Analize girecek bağımsız değişkenler, TÜİK’te çalışan uzman kişilerin görüşleri ve IBM SPSS Modeler programında yer alan Feature Selection işlemcisiyle belirlenmiştir.

Veri setinde yer alan çok sayıda değişkenden hangilerinin bağımlı değişkenin tahmini için önemli etkiye sahip olduğunu tespit etmek için Feature Selection tekniğinden faydalanılmıştır.

Bu teknikte bağımsız değişkenler 0 ile 1 arasında önem değerleri almaktadır.

- Önem değeri $> 0,95$ ise değişken, bağımlı değişken üzerinde önemli etkiye sahiptir,
- $0,95 \geq \text{önem değeri} > 0,9$ ise değişken, bağımlı değişken üzerinde az önemli etkiye sahiptir,
- Önem değeri $< 0,9$ ise değişken, bağımlı değişken üzerinde önemsiz etkiye sahiptir şeklinde yorumlanmaktadır.

Bağımlı değişken olarak belirlenen “GENELSAGLIKDURUM” değişkeni ile diğer değişkenler arasındaki ilişki sonuçları Tablo 4.4.3.1’de gösterilmiştir.

Tablo 4.4.3.1: Bağımlı Değişken ile Diğer Değişkenler Arasındaki İlişkinin İncelenmesi

Sıra	Değişken	Önem	Önem Değeri
1	YASAMFAAL_KISITLANMADURUM	Önemli	1.0
2	DESTEKSİZ12BASAMAK	Önemli	1.0
3	HASTALIK_SAGLIKDURUM	Önemli	1.0
4	BEDENSELAGRI	Önemli	1.0
5	DESTEKSİZ500MYURUYEBİLME	Önemli	1.0
6	ZORLUKCEKME_AGIREVIS	Önemli	1.0
7	YAS_GRUP	Önemli	1.0
8	ZORLUKCEKME_HAFIFEVIS	Önemli	1.0
9	EGITIM	Önemli	1.0
10	OGRENME	Önemli	1.0
11	ZORLUKCEKME_IDARIMALIKONU	Önemli	1.0
12	HATIRLAMA	Önemli	1.0
13	KRONIKHAST_HIPERTANSİYON	Önemli	1.0
14	YORGUNLUK	Önemli	1.0
15	RECETEEDİLEN_ILACKULLANIM	Önemli	1.0
16	HUZURSUZLUK	Önemli	1.0
17	MEDENİ_DURUM	Önemli	1.0
18	UYKUSORUN	Önemli	1.0
19	KRONIKHAST_BELBOLGE	Önemli	1.0
20	BUNALIM	Önemli	1.0
21	DEGERSİZHISSETME	Önemli	1.0
22	ZEVKDUYMA	Önemli	1.0
23	KONSANTRASYONPRB	Önemli	1.0
24	KRONIKHAST_BOYUNBOLGE	Önemli	1.0
25	KRONIKHAST_SEKERHAST	Önemli	1.0
26	ISTAHSIZLIK	Önemli	1.0
27	KRONIKHAST_DEPRESYON	Önemli	1.0
28	CALISMA_DURUM	Önemli	1.0
29	GUNLUK_FAALİYETDURUM	Önemli	1.0
30	SON12AY_YATAKLIHİZMET	Önemli	1.0
31	HANE_GELİR	Önemli	1.0
32	YURUME_10DAK_GUN	Önemli	1.0
33	GOZLUKKULLANMA	Önemli	1.0
34	SON12AY_GUNUBİRLİKHİZMET	Önemli	1.0
35	CİNSİYET	Önemli	1.0
36	KRONIKHAST_ALERJİ	Önemli	1.0
37	GUVENİLİR_YAKIN	Önemli	1.0
38	CEVRE_ILGI	Önemli	1.0
39	ALKOLKULLANIM_DURUM	Önemli	1.0
40	TUTUNKULLANIM_DURUM	Önemli	1.0
41	TEDAVİ_MASRAF_SGK	Önemli	1.0
42	MEYVE_YEMESİKLİK	Önemli	1.0
43	KOMSU_YARDIM	Önemli	1.0
44	KENDİ_ILAC	Önemli	1.0
45	SEBZE_SALATA_YEMESİKLİK	Önemli	1.0

Feature Selection tekniği ile yapılan işlem sonucunda, belirlenen tüm değişkenlerin yapılacak analiz hakkında önemli bilgiler taşıdığı gözlenmiştir. Bu nedenle önceden tespit edilen 45 değişken de analize dâhil edilmiştir.

Veri ön işleme adımının sonunda Data Audit işlemcisi ile veri kalitesi tekrar incelenmiş. İşlemin sonuçları Şekil 4.4.3.1’de yer almaktadır.

Şekil: 4.4.3.1: Veri Ön İşlemi Adımından Sonra Veri Setinin Kalitesi

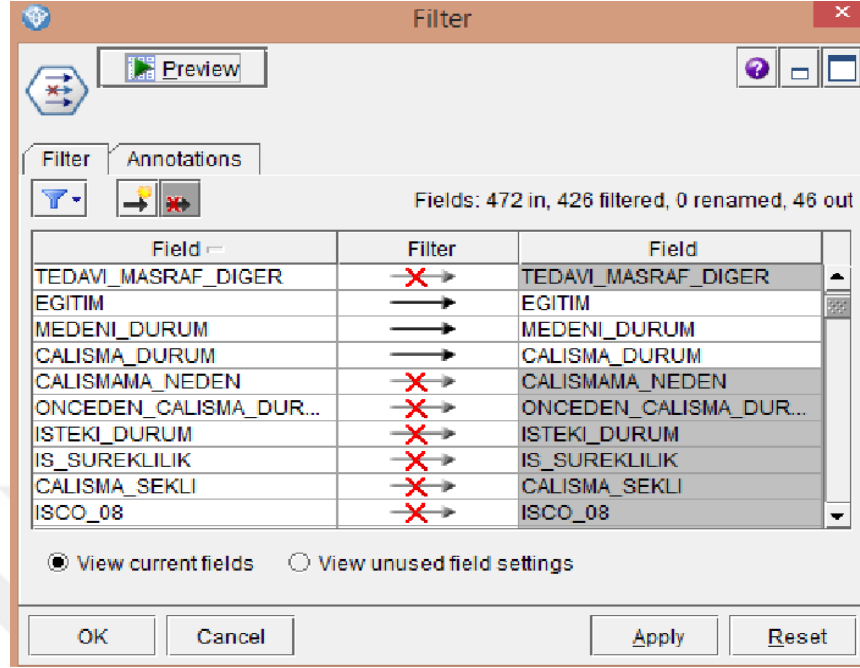
Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
YASAMFAAL	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
DESTEKSİZ1	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
HASTALIK_S	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
BEDENSEL2	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
DESTEKSİZ2	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
ZORLUKCE	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
YAS_GRP	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
ZORLUKCE	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
EGİTİM	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
OGRENİME	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
ZORLUKCE	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
HATIRLAMA	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
KRONİKHAŞ	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
YORGUNLUK	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
RECİTEEDİ	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
HUZURSİZ	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
MEDENİ_DÜ	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
UNKUSORUN	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
KRONİKHAŞ	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
BUNALIM	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
DEĞERSİZLİ	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
ZEKİDÜYMA	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
KONSANTRA	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
KRONİKHAŞ	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
KRONİKHAŞ	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
İSTAHSİZLİK	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
KRONİKHAŞ	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
CALISMA_D	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
GÜNLÜK_FA	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
SÖNİŞİY_Y	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
HANE_GELİR	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
YÜRÜME_10	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
GOZLUKKUL	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
SÖNİŞİY_G	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
CİNSİYET	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
KRONİKHAŞ	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
GÜVENİLİR	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
CEVRE_İLGİ	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
ALKOLKULL	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
TUTURKULL	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
TEDAT_MAS	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
MEYVE_YEM	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
KOMSU_YAR	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
KENDİ_İLAÇ	Flag	---	---	---	Never	Fixed	100	19129	0	0	0	0
SEBZE_SAL	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0
GENELSAĞL	Nominal	---	---	---	Never	Fixed	100	19129	0	0	0	0

Veri setinin son halinde aykırı değerler ve kayıp gözlem bulunmamaktadır. Analize girecek veri seti % 100 doluluk oranına sahiptir.

4.5. Veri Madenciliği Sürecinde Kullanılacak Değişkenlerin Özellikleri ve Tanımlayıcı İstatistikler

Çalışmanın bundan sonraki bölümleri ön işlemeden geçen veri seti üzerinden devam edecektir. Veri madenciliği analizlerine uygun hale gelen veri setinde, 19129 fert ve 46 değişken bulunmaktadır. Analize girecek değişkenler Şekil 4.5.1’de görüldüğü gibi Filter işlemcisiyle seçilmiştir.

Şekil 4.5.1: Değişkenlerin Seçilmesi



Tablo 4.5.1’de veri ön işleme sonucu elde edilen değişkenlerin açıklamaları, değerleri ve frekansları gösterilmiştir.

Tablo 4.5.1: Değişken Tablosu

Değişken Adı	Değişken Açıklaması	Değişken Değerleri	Yüzde (%)	Frekans
YAS_GRUP	Doğum tarihiniz nedir?	15-24	17,71	3388
		25-34	19,14	3661
		35-44	19,7	3768
		45-54	17,42	3332
		55-64	13,36	2555
		64-74	7,83	1498
		75+	4,85	927
CINSIYET	Cinsiyetiniz nedir?	1: Erkek	45,59	8721
		2: Kadın	54,41	10408
TEDAVI_MASRAFI_SGK	Tedavi masraflarınız Sosyal Güvenlik Kurumu tarafından karşılanıyor mu?	1: Evet	82,01	15688
		2: Hayır	17,99	3441

Tablo 4.5.1: Değişken Tablosu (Devamı)

Değişken Adı	Değişken Açıklaması	Değişken Değerleri	Yüzde (%)	Frekans
EGITIM	En son hangi okulu bitirdiniz?	0: Bir okul bitirmedi	5,01	959
		1: İlkokul	37,19	7114
		21: Genel ortaokul	8,58	1641
		22: Mesleki veya teknik ortaokul	0,44	85
		23: İlköğretim	7,72	1477
		3: Lise veya mesleki lise	17,58	3362
		4: 2 veya 3 yıllık yüksekokul	4,63	885
		5: 4 yıllık yüksekokul veya fakülte	7,71	1474
		6: Yüksek lisans (5 veya 6 yıllık fakülteler dahil)	1,05	201
		7: Doktora	0,21	41
		8: Okuma yazma bilmiyor	9,88	1890
MEDENI_DURUM	Medeni durumunuz nedir?	1: Hiç evlenmedi	21,71	4153
		2: Evli	68,8	13161
		3: Boşandı	2,71	518
		4: Eşi öldü	6,78	1297
CALISMA_DURUM	Son bir hafta içerisinde aynı (mal) veya nakdi (para) bir gelir elde etmek amacıyla (ev kadını, öğrenci veya emekli de olsanız) bir saat bile olsa bir işte	1: Çalıştı	38,76	7415
		2: Çalışmadı fakat işi ile ilgisi devam	0,86	165
		3: Çalışmadı	60,37	11549
HANE_GELIR	Hanenizin ortalama aylık geliri aşağıdaki gelir gruplarının hangisinde yer almaktadır?	1: 0-1080 TL	30,81	5894
		2: 1081-1550 TL	20,17	3859
		3: 1551-2170 TL	16,28	3115
		4: 2171-3180 TL	17,12	3274
		5: 3181+ TL	15,62	2987
GENELSAGLIK DURUM	Genel olarak sağlık durumunuz nasıldır?	1: İyi	58,33	11157
		2: Orta	12,16	2326
		3: Kötü	29,52	5646

Tablo 4.5.1: Değişken Tablosu (Devamı)

Değişken Adı	Değişken Açıklaması	Değişken Değerleri	Yüzde (%)	Frekan s
HASTALIK_SAG LIKDURUM	6 ay ya da daha uzun süren/sürmesi beklenen hastalığınız/sağlık probleminiz var mı?	1: Evet	51,03	9762
		2: Hayır	48,97	9367
YASAMSALFAA L_KISITLANMA DURUM	Herhangi bir sağlık problemi nedeniyle günlük faaliyetlerinizin 6 ay ve daha uzun süre kısıtlanma durumu nedir?	1: Ciddi ölçüde kısıtlandı	15,23	2913
		2: Kısıtlandı fakat ciddi ölçüde değil	26,07	4987
		3: Kısıtlanmadı	58,7	11229
KRONIKHAST_H IPERTANSİYON	Son 12 ay içinde hipertansiyon hastalığınızı yaşadınız mı?	1: Evet	18,48	3535
		2: Hayır	81,52	15594
KRONIKHAST_B ELBOLGE	Son 12 ay içinde bel bölgesi problemleri (bel ağrısı, bel fıtığı ve diğer bel defektleri) yaşadınız mı?	1: Evet	35,18	6729
		2: Hayır	64,82	12400
KRONIKHAST_B OYUNBOLGE	Son 12 ay içinde boyun bölgesi problemleri (boyun ağrısı, boyun fıtığı, diğer boyun defektleri) yaşadınız mı?	1: Evet	23,48	4492
		2: Hayır	76,52	14637
KRONIKHAST_S EKERHAST	Son 12 ay içinde şeker hastalığı (diyabet) yaşadınız mı?	1: Evet	10,43	1996
		2: Hayır	89,57	17133
KRONIKHAST_A LERJİ	Son 12 ay içinde alerji (alerjik rinit, dermatit, yiyecek vb. alerjisi) (alerjik astım hariç) yaşadınız mı?	1: Evet	13,23	2531
		2: Hayır	86,77	16598
KRONIKHAST_D EPRESYON	Son 12 ay içinde depresyon durumunu yaşadınız mı?	1: Evet	11,72	2242
		2: Hayır	88,28	16887
GOZLUKKULLA NMA	Gözlük ya da lens kullanıyor musunuz?	1: Evet	38,99	7459
		2: Hayır	60,88	11645
		3: Hiç göremiyorum	0,13	25
DESTEKSİZ500M YURUYEBİLME	Düz bir arazide baston veya diğer bir yürüme aracı kullanmadan ya da destek almadan 500 metre yürümekte zorluk çekiyor musunuz?	1: Hiç zorlanmıyorum	80,39	15378
		2: Biraz zorlanıyorum	11,52	2203
		3: Çok zorlanıyorum	6,98	1336
		4: Hiç yürüyemiyorum	1,11	212

Tablo 4.5.1: Değişken Tablosu (Devamı)

Değişken Adı	Değişken Açıklaması	Değişken Değerleri	Yüzde (%)	Frekans
DESTEKSİZ1 2BASAMAK	12 Basamak merdiven inip çıkmada zorluk çekiyor musunuz?	1: Hiç zorlanmıyorum	73,7	14099
		2: Biraz zorlanıyorum	16,33	3123
		3: Çok zorlanıyorum	8,86	1694
		4: Hiç merdiven inip çıkamıyorum	1,11	213
OGRENME	Yaştlarınıza göre yeni bilgi beceri öğrenmede (okuma-yazma, hesaplama, basit dört işlem yapmada vb.) güçlük çekiyor musunuz?	1: Hayır, güçlük çekmiyorum	84,14	16095
		2: Biraz güçlük çekiyorum	10,59	2025
		3: Çok güçlük çekiyorum	3,62	692
		4: Hiç öğrenemiyorum	1,66	317
HATIRLAM A	Yakın zamanda yaşanan olayları (gün içindeki olaylar, yakınlarınızın adları gibi) hatırlamakta güçlük çekiyor musunuz?	1: Hayır, güçlük çekmiyorum	77,2	14767
		2: Biraz güçlük çekiyorum	18,81	3598
		3: Çok güçlük çekiyorum	3,73	714
		4: Hiç hatırlamıyorum	0,26	50
ZORLUKCE KME_HAFIF EVIS	Hafif ev işleri faaliyetini yardım almadan gerçekleştirirken zorluk çekiyor musunuz?	1: Hiç zorlanmıyorum	87,07	16656
		2: Biraz zorlanıyorum	5,2	995
		3: Çok zorlanıyorum	1,69	324
		4: Kendi kendime yapamıyorum	2,34	448
		5: Hiçbir zaman denemedim/Yapmaya ihtiyaç duymadım	3,69	706
ZORLUKCE KME_AGIRE VIS	Ağır ev işleri faaliyetini yardım almadan gerçekleştirirken zorluk çekiyor musunuz?	1: Hiç zorlanmıyorum	69,84	13360
		2: Biraz zorlanıyorum	12,52	2395
		3: Çok zorlanıyorum	7,1	1359
		4: Kendi kendime yapamıyorum	6,05	1158
		5: Hiçbir zaman denemedim/Yapmaya ihtiyaç duymadım	4,48	857

Tablo 4.5.1: Değişken Tablosu (Devamı)

Değişken Adı	Değişken Açıklaması	Değişken Değerleri	Yüzde (%)	Frekans
ZORLUKCEKME_I DARIMALIKONU	Günlük idari sorumluluklar ve mali konularla ilgilenme faaliyetini yardım almadan gerçekleştirirken zorluk çekiyor musunuz?	1: Hiç zorlanmıyorum	87,47	16732
		2: Biraz zorlanıyorum	6,25	1196
		3: Çok zorlanıyorum	2,06	395
		4: Kendi kendime yapamıyorum	2,56	490
		5: Hiçbir zaman denemedim/Yapmaya ihtiyaç duymadım	1,65	316
BEDENSELAGRI	Son 4 hafta içerisinde ne kadar bedensel ağrınız oldu?	1: Hiç	45,17	8640
		2: Çok az	12,87	2462
		3: Az	11,5	2199
		4: Orta	15,22	2911
		5: Fazla	10,51	2011
		6: Çok fazla	4,74	906
ZEVKDUYMA	Son 2 hafta içerisinde bir şeyler yaparken çok az ilgi duyma ve keyif alma problemi sizi ne sıklıkla rahatsız etti?	1: Hiç	62,41	11938
		2: Bazı günler	31,94	6110
		3: Bir haftadan fazla	1,64	314
		4: Neredeyse her gün	4,01	767
BUNALIM	Son 2 hafta içerisinde morali bozuk, bunalımlı veya umutsuz hissetme problemi sizi ne sıklıkla rahatsız etti?	1: Hiç	53,59	10251
		2: Bazı günler	38,84	7429
		3: Bir haftadan fazla	2,34	448
		4: Neredeyse her gün	5,23	1001
UYKUSORUN	Son 2 hafta içerisinde uykuya dalmakta veya uyumakta zorluk çekme/çok fazla uyuma problemi sizi ne sıklıkla rahatsız etti?	1: Hiç	63,63	12172
		2: Bazı günler	27,27	5217
		3: Bir haftadan fazla	2,26	433
		4: Neredeyse her gün	6,83	1307
YORGUNLUK	Son 2 hafta içerisinde yorgun/enerjisiz hissetme problemi sizi ne sıklıkla rahatsız etti?	1: Hiç	49,03	9379
		2: Bazı günler	41,55	7949
		3: Bir haftadan fazla	2,52	483
		4: Neredeyse her gün	6,89	1318
ISTAHSIZLIK	Son 2 hafta içerisinde iştahsızlık/aşırı yemek yeme problemi sizi ne sıklıkla rahatsız etti?	1: Hiç	74,43	14238
		2: Bazı günler	20,52	3925
		3: Bir haftadan fazla	1,31	250
		4: Neredeyse her gün	3,74	716
DEGERSIZHISSET ME	Son 2 hafta içerisinde kendini değersiz ve kötü hissetme problemi sizi ne sıklıkla rahatsız etti?	1: Hiç	74,13	14180
		2: Bazı günler	22,01	4211
		3: Bir haftadan fazla	0,98	188
		4: Neredeyse her gün	2,88	550

Tablo 4.5.1: Değişken Tablosu (Devamı)

Değişken Adı	Değişken Açıklaması	Değişken Değerleri	Yüzde (%)	Frekans
KONSANTRASYO NPRB	Son 2 hafta içerisinde konsantrasyon problemi (gazete okurken, TV izlerken vb.) sizi ne sıklıkla rahatsız etti?	1: Hiç	78,09	14937
		2: Bazı günler	18,65	3567
		3: Bir haftadan fazla	0,88	168
		4: Neredeyse her gün	2,39	457
HUZURSUZLUK	Son 2 hafta içerisinde başkalarının fark edeceği derecede ağır hareket etme/konuşma ya da huzursuz/fazla hareket etme problemi sizi ne sıklıkla rahatsız etti?	1: Hiç	85	16259
		2: Bazı günler	12,68	2426
		3: Bir haftadan fazla	0,57	109
		4: Neredeyse her gün	1,75	335
SON12AY_YATAK LIHIZMET	Son 12 ay içerisinde en az bir gece hastanede yattınız mı?	1: Evet	12,19	2332
		2: Hayır	87,81	16797
SON12AY_GUNUB IRLIKHIZMET	Son 12 ay içerisinde hastanede gününbirlik (24 saatten daha az süreyle) tedavi hizmeti aldınız mı?	1: Evet	64,62	12362
		2: Hayır	35,38	6767
RECETEEDILEN_I LACKULLANIM	Son 2 hafta içerisinde bir doktor tarafından size reçete edilen bir ilaç kullandınız mı?	1: Evet	37,25	7125
		2: Hayır	62,75	12004
KENDI_ILAC	Son 2 hafta içerisinde kendi kararınız/tercihinizle bir ilaç kullandınız mı?	1: Evet	30,1	5757
		2: Hayır	69,9	13372
GUNLUK_FAALIY ETDURUM	Çalışırken/günlük faaliyetlerinizi gerçekleştirirken aşağıdakilerden hangisi durumunuzu en iyi tanımlar?	1: Çoğunlukla oturan veya duran	39,86	7625
		2: Çoğunlukla yürüyüş veya orta derecede fiziksel güç gerektiren işler	52,78	10096
		3: Çoğunlukla ağır iş veya fiziksel güç gerektiren işler	7,36	1408

Tablo 4.5.1: Değişken Tablosu (Devamı)

Değişken Adı	Değişken Açıklaması	Değişken Değerleri	Yüzde (%)	Frekans
YURUME_10D AK_GUN	Normal bir hafta içerisinde bir yerden bir yere giderken, kaç gün en az 10 dakika yürürsünüz?	0: Hiçbir zaman böyle bir fiziksel aktivite gerçekleştirmedim	24,65	4715
		1: 1 gün	5,53	1057
		2: 2 gün	5,71	1093
		3: 3 gün	7,03	1344
		4: 4 gün	3,87	741
		5: 5gün	8,23	1575
		6: 6 gün	3,26	624
		7: 7 gün	41,72	7980
MEYVE_YEME SIKLIK	Ne sıklıkla meyve yersiniz?	1: Günde bir kere ya da daha çok	49,29	9429
		2: Haftada 4-6 kere	17,05	3262
		3: Haftada 1-3 kere	24,03	4597
		4: Haftada bir kereden az	7,32	1401
		5: Hiç	2,3	440
SEBZE_SALAT A_YEMESIKLI K	Ne sıklıkla sebze ya da salata yersiniz (patates hariç)?	1: Günde bir kere ya da daha çok	61,35	11736
		2: Haftada 4-6 kere	19,13	3659
		3: Haftada 1-3 kere	16,12	3083
		4: Haftada bir kereden az	2,68	512
		5: Hiç	0,73	139
TUTUNKULLA NIM_DURUM	Tütün mamulü kullanıyor musunuz? (bir kere dahi olsa)	1: Evet, her gün	26,08	4988
		2: Evet, ara sıra	5,19	993
		3: Hiçbir zaman	49,89	9543
		4: Bıraktım	18,85	3605
ALKOLKULLA NIM_DURUM	Hiç alkollü içecek kullandınız mı? (Bir yudum dahi olsa)	1: Evet	33,84	6474
		2: Hayır	66,16	12655
GUVENILIR_Y AKIN	Ciddi bir kişisel probleminiz olursa güvенеbileceğiniz, size çok yakın kaç kişi vardır?	1: Hiç	6,12	1171
		2: 1-2 kişi	34,17	6536
		3: 3-5 kişi	37,44	7161
		4: 6 veya daha fazla kişi	22,28	4261

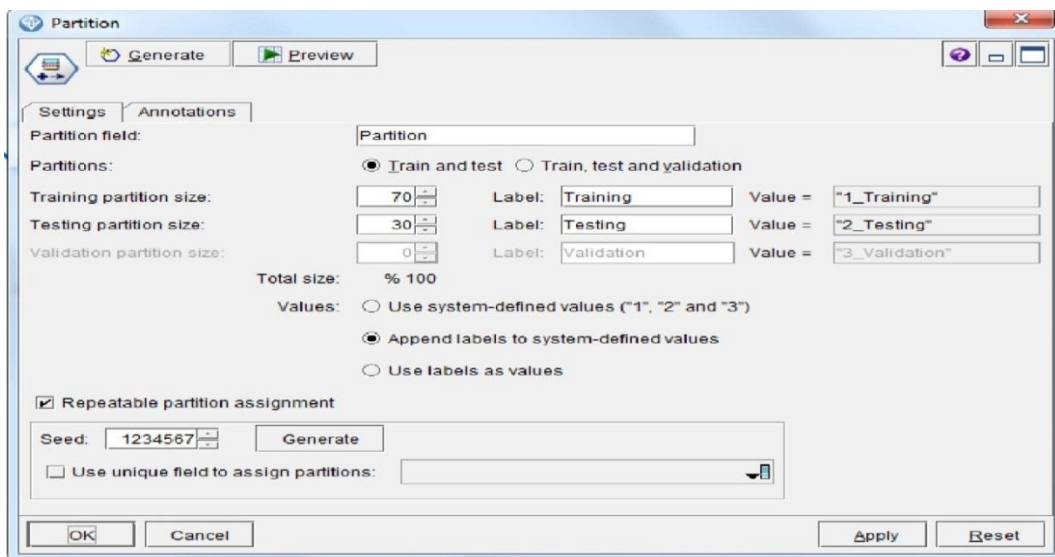
Tablo 4.5.1: Değişken Tablosu (Devamı)

Değişken Adı	Değişken Açıklaması	Değişken Değerleri	Yüzde (%)	Frekans
CEVRE_ILGI	Çevrenizdeki insanlar ne yaptığınıza ne kadar ilgi gösterir?	1: Çok fazla	16,15	3090
		2: Fazla	47,11	9012
		3: Emin değilim	23,43	4481
		4: Az	9,82	1879
		5: Hiç	3,49	667
KOMSU_YARDIM	İhtiyacınız olduğunda komşularınızdan yardım alabilme durumunuz nedir?	1: Çok kolay	21,11	4039
		2: Kolay	49,86	9537
		3: Orta	14,68	2808
		4: Zor	8,88	1698
		5: Çok zor	5,47	1047

4.6. Veri Setinin Bölünmesi

Karar ağacı algoritmaları uygulanmadan önce veri seti, eğitim ve test olmak üzere iki parçaya bölünmüştür. Algoritmanın uygulandığı sırada eğitim kümesi üzerinde model çalıştırılır ve sınıflandırıcının öğrenmesi sağlanır. Test kümesinde ise model kontrol edilir. Veri setinin bölünmesi konusunda literatürde en çok kullanılan yöntem “% ayırım” yöntemi olduğu için veri setinin % 70’i eğitim için ayrılmış, kalan % 30’luk kısmı ise modelin testi için bırakılmıştır.

Şekil 4.6.1: Veri Setinin Bölünmesi

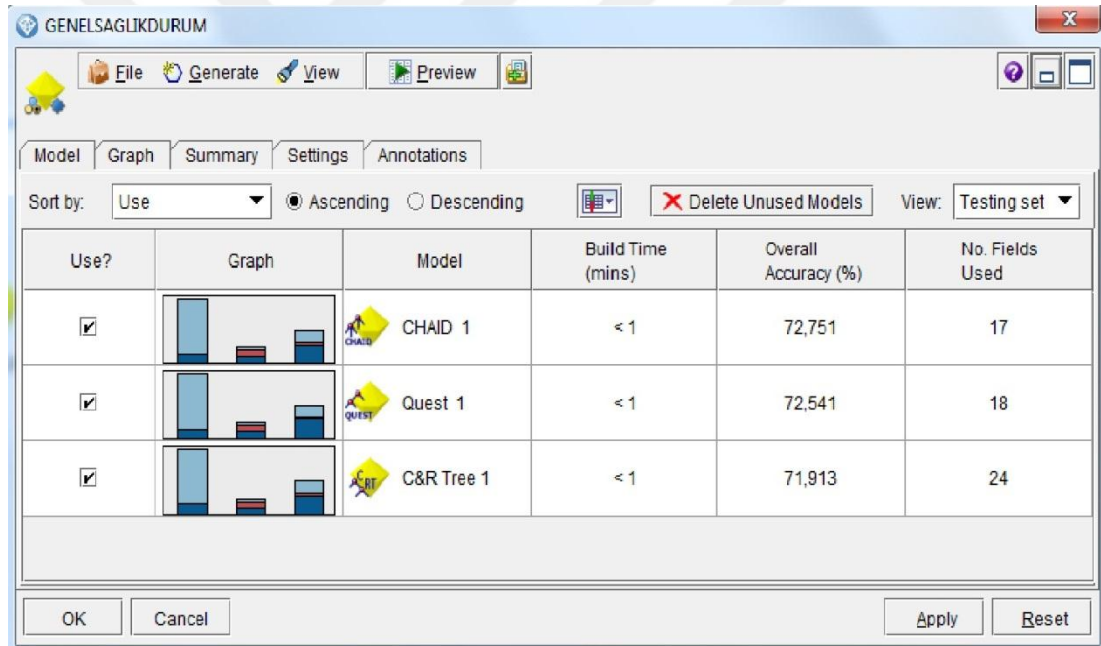


IBM SPSS Modeler programında bölme işlemi, partition işlemcisi ile yapılır. Şekil 4.6.1’deki gibi veri seti ikiye bölünüp analiz işlemi devam etmiştir.

4.7. Veri Seti Üzerinde Karar Ağacı Algoritmalarının Uygulanması

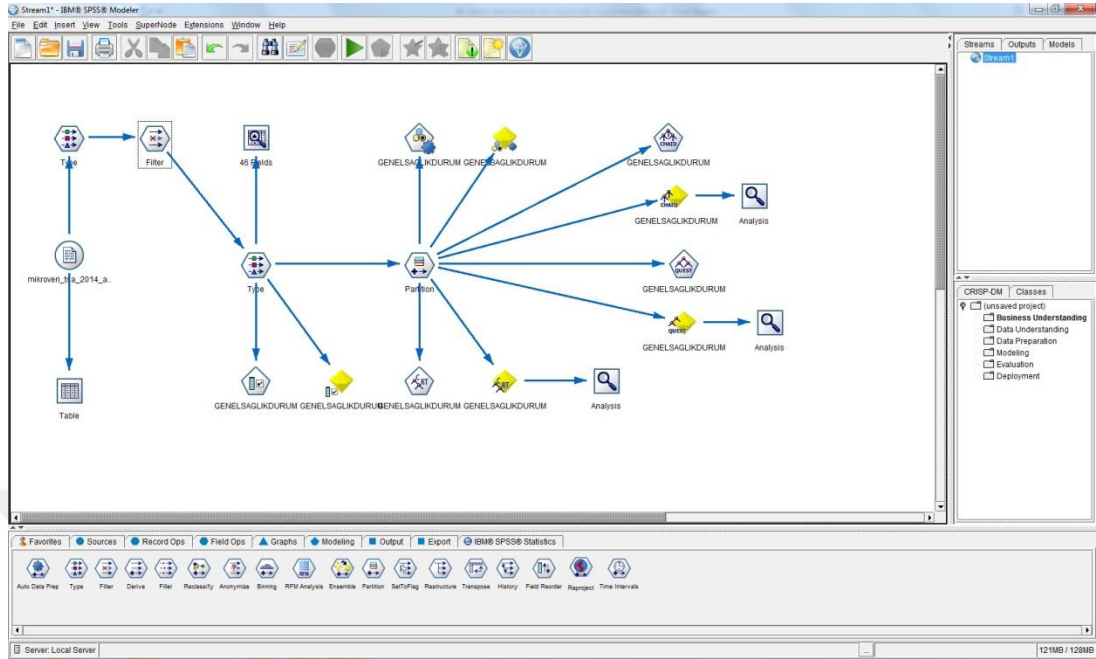
IBM SPSS Modeler programında otomatik sınıflama yapabilen Auto Classifier işlemcisi yer almaktadır. Bu işlemci, programda bulunan sınıflama algoritmalarını birbirleriyle karşılaştırarak veri seti için en ideal algoritmaları kullanıcıya sunmaktadır. Programda yer alan C5.0, Random Trees, Tree-AS, CHAID, Quest ve C&R Tree karar ağacı algoritmaları seçilmiş ve Auto Classifier işlemcisi çalıştırılmıştır.

Şekil 4.7.1: Auto Classifier İşlemcisi ile Karar Ağacı Algoritmalarının Belirlenmesi



Şekil 4.7.1 incelendiğinde veri setine en uygun karar ağacı algoritmalarının CHAID, Quest ve C&R Tree algoritmaları olduğu gözlenmiştir.

Şekil 4.7.2: Kurulan Modelin Ekran Görüntüsü



Auto Classifier işlemcisinin sonuçlarına göre, IBM SPSS Modeler programında üç farklı model kurulmuştur. Bu modellerin ekran görüntüsü Şekil 4.7.2'deki gibidir.

4.7.1. CHAID Algoritması

IBM SPSS Modeler programı, analiz esnasında değişken önemi (variable importance) grafiğini göstermektedir. Bu sayede kurulan model için üzerinde durulması gereken değişkenler bilinebilir. Değişken önemi grafiği ile verilen değerler görecelidir ve algoritmanın modele dâhil ettiği değişkenlerin toplam değeri 1.0'dır. Değişken öneminin modelin doğruluğu ile ilgisi yoktur. Tahmin yaparken her bir değişkenin önemini gösterir (Kuzey, 2012: 212).

CHAID algoritması ile elde edilen değişkenlerin önem değerleri Tablo 4.7.1.1'de gösterilmiştir.

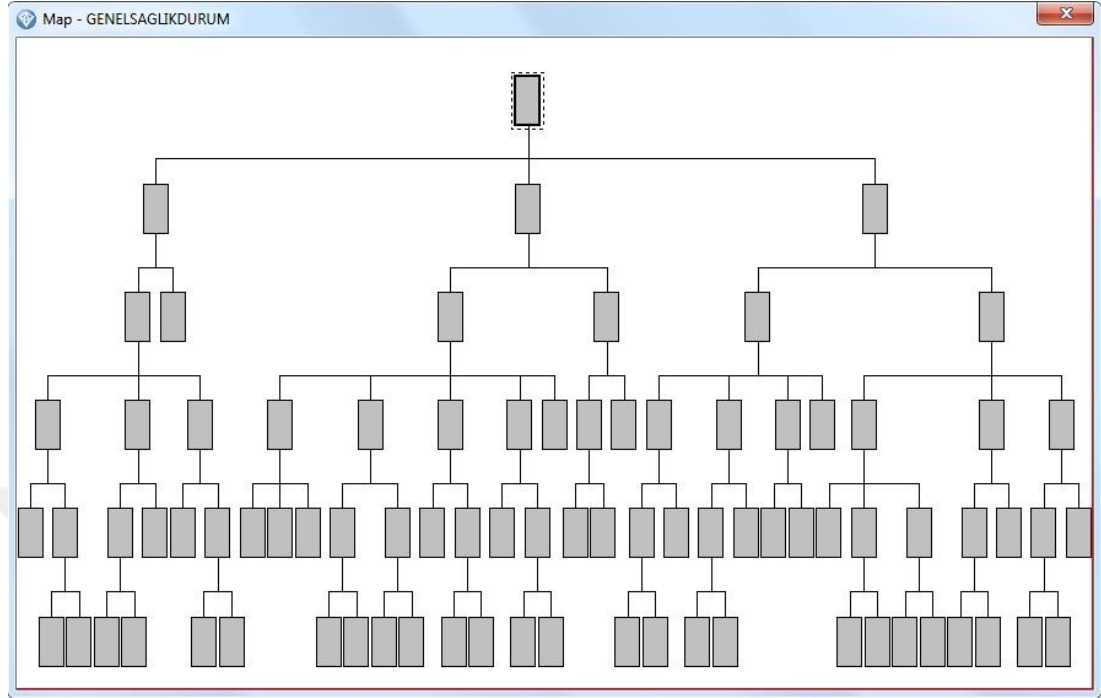
Tablo 4.7.1.1: CHAID Algoritmasında Kullanılan Değişkenlerin Önem Değerleri

Değişken	Değişken Önem Değeri
HASTALIK_SAGLIKDURUM	0,2601
YASAMFAAL_KISITLANMADURUM	0,2003
BEDENSELAGRI	0,1713
EGITIM	0,0577
KRONIKHAST_BELBOLGE	0,0405
KRONIKHAST_HIPERTANSİYON	0,0398
DESTEKSİZ12BASAMAK	0,0346
YAS_GRUP	0,0317
RECETEEDİLEN_ILACKULLANIM	0,0241
KRONIKHAST_BOYUNBOLGE	0,0214
UYKUSORUN	0,0198
ZORLUKCEKME_AGIREVIS	0,0182
YORGUNLUK	0,0178
ZORLUKCEKME_IDARIMALIKONU	0,0178
GUNLUK_FAALİYETDURUM	0,0178
KOMSU_YARDIM	0,0149
SEBZE_SALATA_YEMESIKLIK	0,012

CHAID algoritması ile kurulan modelde en önemli değişken, 0,2601 önem değeri ile “HASTALIK_SAGLIKDURUM” değişkeni olmuştur. Bu değişkeni 0,2003 önem değeri ile “YASAMFAAL_KISITLANMADURUM” değişkeni, 0,1713 önem değeri ile “BEDENSELAGRI” değişkeni takip etmiştir.

CHAID algoritması ile oluşturulan karar ağacının genel görünümü Şekil 4.7.1.1’de gösterilmiştir.

Şekil 4.7.1.1: CHAID Algoritması İle Oluşturulan Karar Ağacının Görünümü



Karar ağacının ilk düğümünde bağımlı değişken olan genel sağlık durumu değişkeni yer almaktadır. Ağaç derinliği 5 olan karar ağacı, toplam 83 düğümden oluşmuştur. Karar ağacının başlangıç düğümü Şekil 4.7.1.2’de gösterilmiştir.

Şekil 4.7.1.2: CHAID algoritması ile Oluşan Karar Ağacının Başlangıç Düğümü

GENELSAGLIKDURUM

Node 0		
Category	%	n
iyi	58.221	7804
kötü	12.086	1620
orta	29.693	3980
Total	100.000	13404

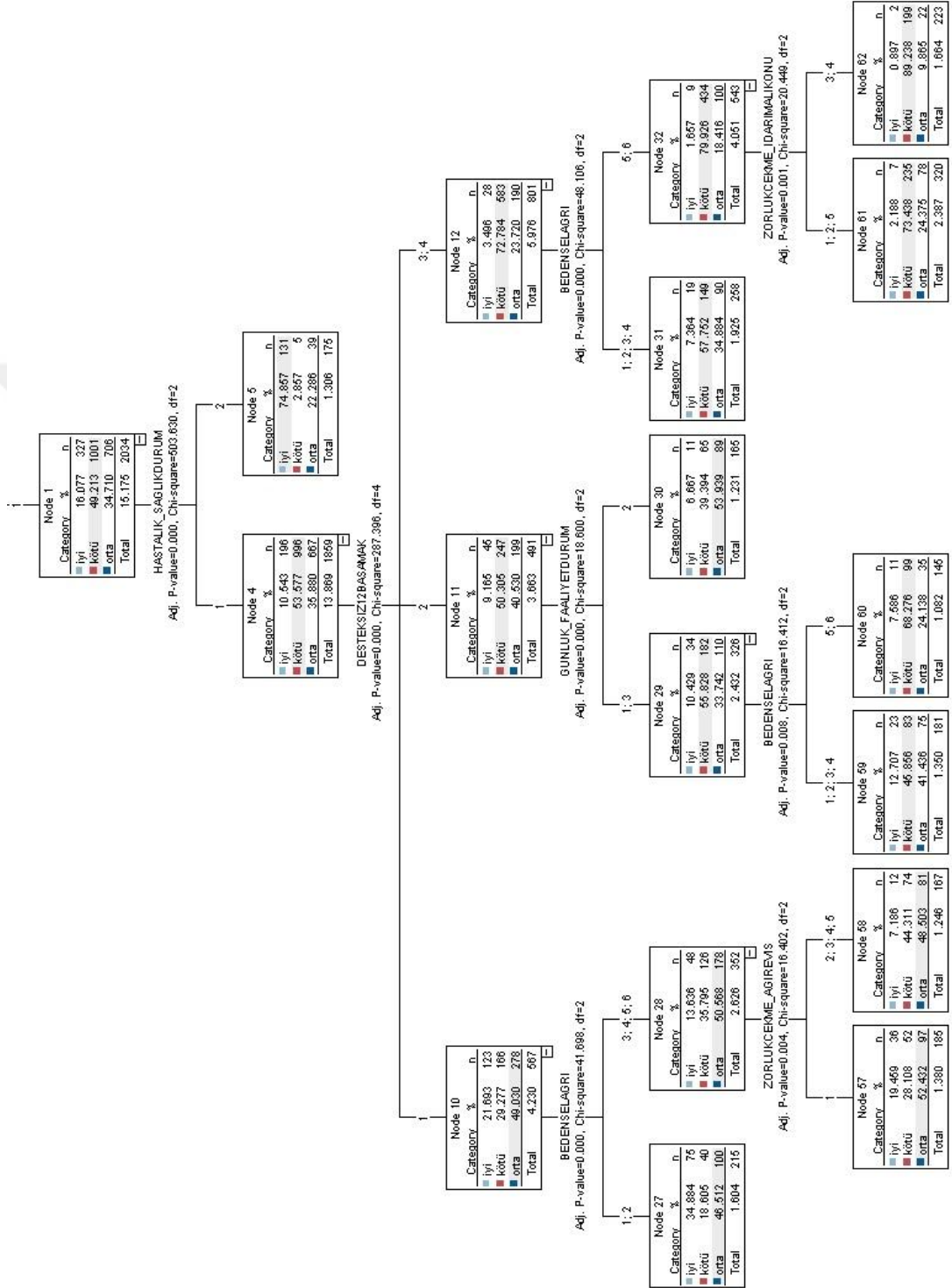
YASAMFAAL_KISITLANMADURUM
Adj. P-value=0.000, Chi-square=5482.491, df=4

Başlangıç düğümünde yer alan 13404 ferdin % 58,221'i genel sağlık durumunu iyi olarak nitelendirirken % 29,693'ü orta, % 12,086'sı ise kötü olarak nitelendirmiştir.

CHAID algoritmasına ile oluşturulan karar ağacı dallanmaya “YASAMFAAL_KISITLANMADURUM” değişkeni ile 3 alt sınıf oluşturarak başlamıştır. “Herhangi bir sağlık problemi nedeniyle günlük faaliyetlerinizin 6 ay ve daha uzun süre kısıtlanma durumu nedir?” sorusuna verilen “ciddi ölçüde kısıtlandı” cevabı ile 1. düğüm, “kısıtlandı fakat ciddi ölçüde değil” cevabı ile 2. düğüm, “kısıtlanmadı” cevabı ile de 3. düğüm oluşmuştur. İlk düğüme göre oluşan karar ağacı Şekil 4.7.1.3'te gösterilmiştir.



Şekil 4.7.1.3: CHAID Algoritmasının 1. Düzgüne G6re Dallanması



Karar ağacının 1. düğümü incelendiğinde genel sağlık durumunu iyi olarak nitelendiren fertlerin oranının % 16,077, kötü olarak nitelendirenlerin oranının ise % 49,213 olduğu gözlenmektedir.

1. düğüm, “HASTALIK_SAGLIKDURUM” değişkeni ile dallanmıştır. “6 ay ya da daha uzun süren/sürmesi beklenen hastalığınız/sağlık probleminiz var mı?” sorusuna verilen “evet” cevabı ile 4. düğüm, “hayır” cevabı ile de 5. düğüm oluşmuştur. 4. düğümde sağlık durumunun iyi olduğunu söyleyen fertlerin oranı %10,543, kötü olduğunu söyleyen fertlerin oranı ise % 53,577’dir. 5. düğümde sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 74,857’ye yükselmiş, kötü olarak nitelendirenlerin oranı ise % 2,857’ye düşmüştür.

5. düğümde dallanma meydana gelmemiştir fakat 4. düğüm, “DESTEKSİZ12BASAMAK” değişkenine göre dallanmıştır. “12 basamak merdiven inip çıkmada zorluk çekiyor musunuz?” sorusuna verilen “hiç zorlanmıyorum” cevabı ile 10. düğüm meydana gelmiştir ve sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 21,693 olmuştur. “Biraz zorlanıyorum” cevabı ile 11. düğüm oluşmuştur ve sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 9,165’e düşmüştür. “Çok zorlanıyorum ve hiç merdiven inip çıkamıyorum” cevapları ile de 12. düğüm oluşmuştur. 12. düğümde yer alan fertlerin % 3,496’sı sağlık durumunu iyi olarak nitelendirmiştir.

10. düğüm, “BEDENSELAGRI” değişkenine göre dallanmıştır. “Son 4 hafta içerisinde ne kadar bedensel ağrınız oldu?” sorusuna verilen “hiç ve çok az” cevapları ile 27. düğüm oluşmuştur. Bu düğümde sağlık durumunun iyi olduğunu söyleyen fertlerin oranı % 34,884, kötü olduğunu söyleyen fertlerin oranı ise % 18,605 olmuştur. “Az, orta, fazla ve çok fazla” cevapları ile 28. düğüm oluşmuştur ve bu düğümde sağlık durumunun iyi olduğunu söyleyen fertlerin oranı % 13,636’ya düşmüş, kötü olduğunu söyleyen fertlerin oranı ise % 35,795’e yükselmiştir.

27. düğümde dallanma meydana gelmemiştir fakat 28. düğüm, “ZORLUKCEKME_AGIR EVIS” değişkenine göre dallanmıştır. “Ağır ev işleri faaliyetini yardım almadan gerçekleştirirken zorluk çekiyor musunuz?” sorusuna verilen “hiç zorlanmıyorum” cevabı ile 57. düğüm, “biraz zorlanıyorum, çok zorlanıyorum, kendi kendime yapamıyorum ve hiçbir zaman denemedim/yapmaya ihtiyaç duymadım” cevapları ile de 58. düğüm oluşmuştur. 57. düğümde sağlık durumunu iyi olarak nitelendirenlerin oranı % 19,459, kötü olarak nitelendirenlerin oranının ise % 28,108 olduğu gözlenmiştir. 58. düğüm incelendiğinde sağlık durumu iyi olan fertlerin oranı % 7,186’ya düşmüş, kötü olan fertlerin oranı ise % 44,311’e yükselmiştir.

11. düğümde yer alan fertlerin % 9,165’i genel sağlık durumunu iyi olarak nitelendirirken % 50,305’i ise kötü olarak nitelendirmiştir. Bu düğüm, “GUNLUK_FAALIYETDURUM” değişkenine göre dallanmıştır. “Çalışırken/günlük faaliyetlerinizi gerçekleştirirken aşağıdakilerden hangisi durumunuzu en iyi tanımlar?” sorusuna, “çoğunlukla oturan veya duran ve çoğunlukla ağır iş veya fiziksel güç gerektiren işler” cevaplarını veren fertler ile 29. düğüm oluşmuştur. 29. düğümde yer alanların % 10,429’u sağlık durumunun iyi olduğunu söylemiştir. “Çoğunlukla yürüyüş veya orta derecede fiziksel güç gerektiren işler” cevabını veren fertlerden oluşan 30. düğümde ise sağlık durumunu iyi olarak nitelendirenlerin oranının % 6,667 olduğu gözlenmiştir.

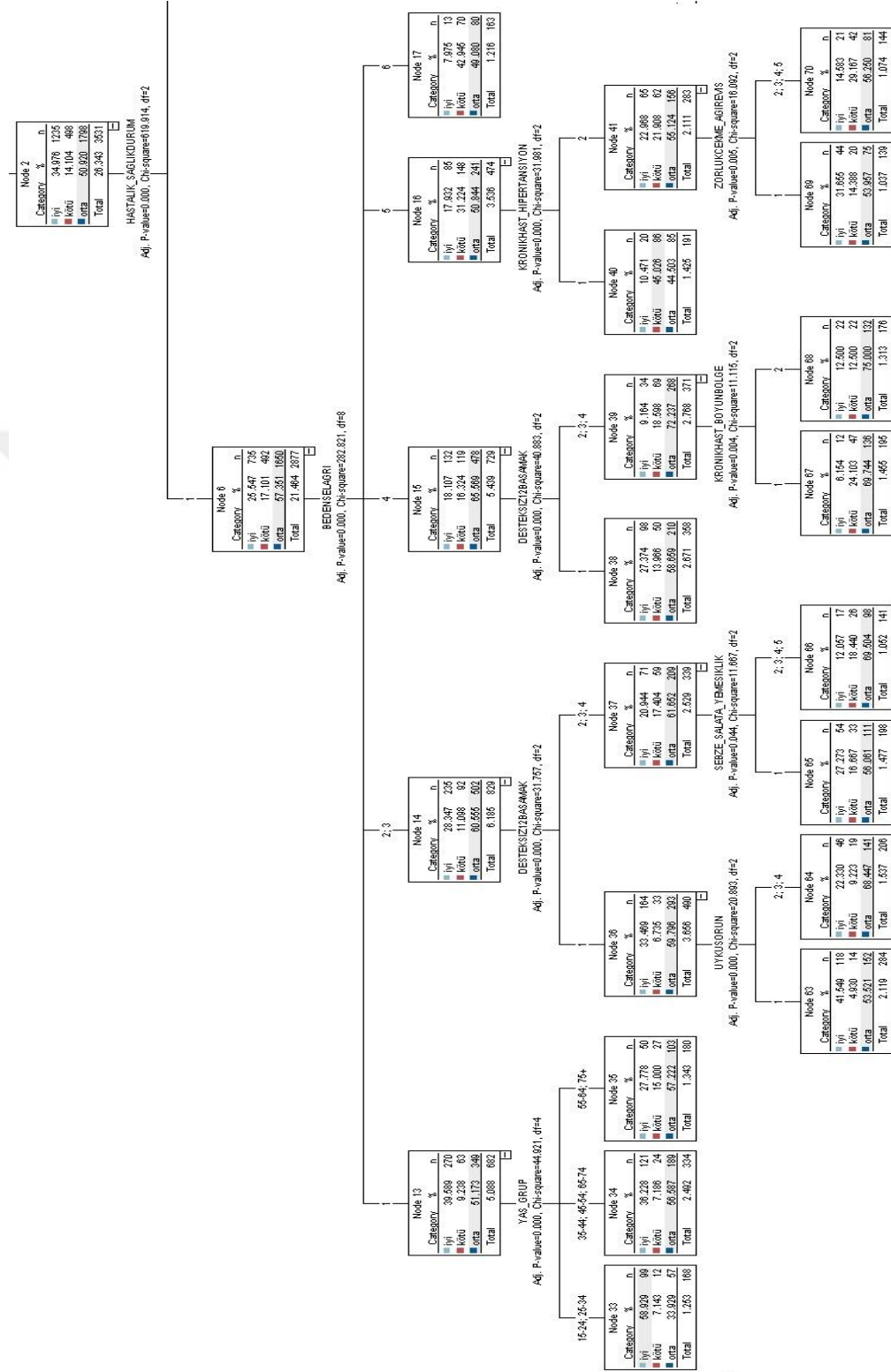
29. düğüm, “BEDENSELAGRI” değişkenine göre dallanmıştır. Son 4 hafta içinde “hiç, çok az, az ve orta” bedensel ağrısı olan fertlerin % 12,707’si sağlık durumunu iyi olarak nitelendirirken “çok ve çok fazla” bedensel ağrısı olan fertlerin % 7,586’sı sağlık durumunu iyi olarak nitelendirmiştir.

12. düğüm de “BEDENSELAGRI” değişkenine göre dallanmıştır. Son 4 hafta içinde “hiç, çok az, az ve orta” ağrısı olan fertlerin % 7,364’ü sağlık durumunu iyi olarak nitelendirirken “çok ve çok fazla” bedensel ağrısı olan fertlerin % 1,657’si sağlık durumunu iyi olarak nitelendirmiştir.

Karar ağacının 32. düğümü ise “ZORLUKCEKME_IDARIMALIKONU” değişkenine göre dallanmıştır. “Günlük idari sorumluluklar ve mali konularla ilgilenme faaliyetini yardım almadan gerçekleştirirken zorluk çekiyor musunuz?” sorusuna, “hiç zorlanmıyorum, biraz zorlanıyorum ve hiçbir zaman denemedim/yapmaya ihtiyaç duymadım” cevaplarını veren fertlerin % 2,188’i sağlık durumunu iyi, % 73,438’i kötü olarak nitelendirmiştir. “Çok zorlanıyorum ve kendi kendime yapamıyorum” cevaplarını veren fertlerin ise % 0,897’si sağlık durumunu iyi olarak nitelendirirken % 89,238’i kötü olarak nitelendirmiştir.



Şekil 4.7.1.4: CHAID Algoritmasının 6. Düzgüne Gre Dallanması



2. düğümde yer alan fertlerin % 34,976'sı sağlık durumunu iyi olarak nitelendirirken % 14,404'ü kötü olarak nitelendirmiştir. 2. düğüm, "HASTALIK_SAGLIK DURUM" değişkenine göre 2 alt sınıf oluşturarak dallanmıştır. Şekil 4.7.1.4'te görüldüğü gibi "6 ay ya da daha uzun süren/sürmesi beklenen hastalığınız/sağlık probleminiz var mı?" sorusuna "evet" cevabını fertler 6. düğümü oluşturmuştur.

6. düğümde yer alan fertlerin % 25,547'si sağlık durumunun iyi olduğunu söylerken % 17,101'i kötü olduğunu söylemiştir. 6. düğüm, "BEDENSELAGRI" değişkenine göre 5 alt sınıf oluşturarak dallanmıştır. "Son 4 hafta içerisinde ne kadar bedensel ağrınız oldu?" sorusuna "hiç" cevabını fertlerden oluşan 13. düğümde sağlık durumunu iyi olarak nitelendirenlerin oranının % 39,589, kötü olarak nitelendirenlerin oranının ise % 9,238 olduğu gözlenmiştir. "Çok az ve az" cevabını veren fertlerden oluşan 14. düğümde sağlık durumu iyi olan fertlerin oranı % 28,347, kötü olan fertlerin oranı ise % 11,098 olmuştur. "Orta" cevabını veren fertlerin oluşturduğu 15. düğümde ise sağlık durumu iyi olan fertlerin oranı % 18,107'ye düşmüş, sağlık durumu kötü olan fertlerin oranı ise % 16,324'e yükselmiştir. "Fazla" cevabını veren fertlerden oluşan 16. düğümde sağlık durumu iyi olan fertlerin oranı düşmeye devam edip % 17,932 olmuş, kötü olan fertlerin oranı ise yükselerek % 31,224 olmuştur. "Çok fazla" bedensel ağrısının söyleyen fertlerin oluşturduğu 17. düğümde ise sağlık durumu iyi olan fertler % 7,975 oranı ile diğer 4 düğüme göre en düşük, sağlık durumu kötü olan fertler ise % 42,945 oranı ile en yüksek değere sahip olduğu gözlenmiştir.

13. düğüm, "YAS_GRUP" değişkeni ile dallanmıştır. "15-24 ve 25-34" yaş grubundaki fertler ile 33. düğüm, "35-44, 45-54 ve 64-74" yaş grubundaki fertler ile 34. düğüm, "55-64 ve 75+" yaş grubundaki fertler ile de 35. düğüm oluşmuştur. 33. düğümde sağlık durumunu iyi olan fertlerin oranı % 58,929, kötü olan fertlerin oranı ise % 7,143 olmuştur. 34. düğümde yer alan fertlerin yaş ortalaması bir önceki düğümdeki fertlere göre yükselmiş ve sağlık durumu iyi olan fertlerin oranı % 36,228'e düşerken sağlık durumu kötü olan fertlerin oranı % 7,186 olmuştur. 35.

düğümde ise sağlık durumunu iyi olarak nitelendirenlerin oranı % 27,778, kötü olarak nitelendirenlerin oranı ise % 15 olmuştur.

14. düğüm, "DESTEKSİZ12BASAMAK" değişkeni ile dallanmıştır. "12 basamak merdiven inip çıkmada zorluk çekiyor musunuz?" sorusuna "hiç zorlanmıyorum" cevabını veren fertler 36. düğümü oluşturmuş ve bu düğümde sağlık durumu iyi olanların oranı % 33,469, kötü olanların oranı ise % 6,735 olmuştur. "Biraz zorlanıyorum, çok zorlanıyorum ve hiç merdiven inip çıkamıyorum" cevabını veren fertler ise 37. düğümü oluşturmuştur. Bu düğümde sağlık durumu iyi olan fertlerin oranı % 20,944'e düşerken sağlık durumu kötü olan fertlerin oranı % 17,404'e yükselmiştir.

36. düğümdeki dallanma, "UYKUSORUN" değişkeni ile gerçekleşmiştir. "Son 2 hafta içerisinde uykuya dalmakta veya uyumakta zorluk çekme/çok fazla uyuma problemi sizi ne sıklıkla rahatsız etti?" sorusuna, "hiç" cevabını fertler 63. düğümü oluşturmuştur ve sağlık durumunu iyi olarak nitelendirenlerin oranı % 41,549 olmuştur. "Bazı günler, bir haftadan fazla ve neredeyse her gün" cevaplarını veren fertler ise 64. düğümü oluşturmuş ve sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 22,33'e düşmüştür.

37. düğüm, "SEBZE_SALATA_YEMESIKLIK" değişkenine göre dallanmıştır. "Ne sıklıkla sebze ya da salata yersiniz (patates hariç)?" sorusuna verilen "günde bir kere ya da daha çok" cevabı ile 65. düğüm oluşmuştur ve bu düğümde sağlık durumu iyi olan fertlerin oranı % 27,273, kötü olan fertlerin oranı ise % 16,667 olmuştur. "Haftada 4-6 kere, haftada 1-3 kere, haftada bir kereden az ve hiç" cevaplarını fertlerden oluşan 66. düğümde ise sağlık durumu iyi olan fertlerin oranı % 12,057'ye düşerken sağlık durumu kötü olan fertlerin oranı % 18,44'e yükselmiştir.

15. düğümde gerçekleşen dallanma, "DESTEKSİZ12BASAMAK" değişkeni ile olmuştur. "12 basamak merdiven inip çıkmada zorluk çekiyor musunuz?" sorusuna "hiç zorlanmıyorum" cevabını veren fertler 38. düğümü oluşturmuş ve bu düğümde

sağlık durumu iyi olanların oranı % 27,334, kötü olanların oranı ise % 13,966 olmuştur. “Biraz zorlanıyorum, çok zorlanıyorum ve hiç merdiven inip çıkamıyorum” cevabını veren fertler ise 39. düğümü oluşturmuştur. Bu düğümde sağlık durumu iyi olan fertlerin oranı % 9,164’e düşerken sağlık durumu kötü olan fertlerin oranı % 18,859’a yükselmiştir.

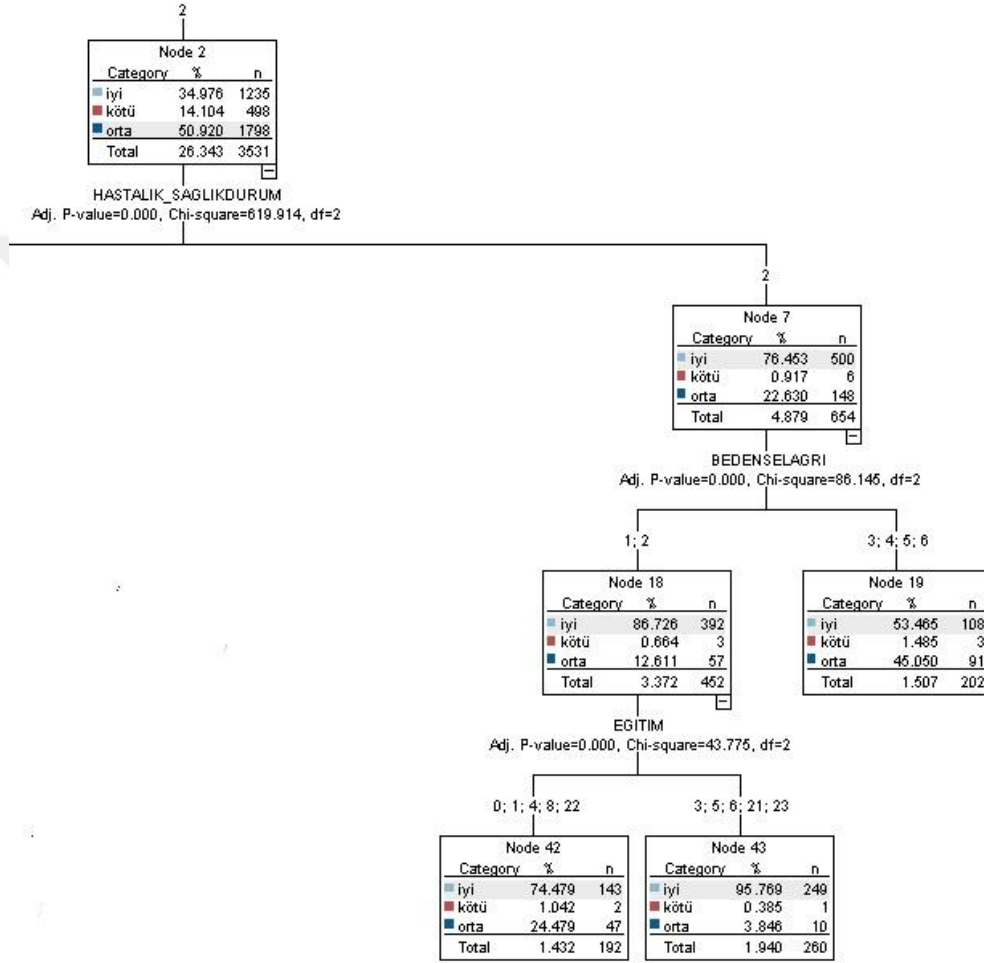
38. düğümde dallanma meydana gelmemiştir fakat 39. düğüm “KRONIKHAŞT_BOYUNBOLGE” değişkenine göre dallanmıştır. “Son 12 ay içinde boyun bölgesi problemleri (boyun ağrısı, boyun fıtığı, diğer boyun defektleri) yaşadınız mı?” sorusuna “evet” cevabını veren fertler 67. düğümü oluşturmuştur ve bu düğümde sağlık durumunun iyi olduğunu söyleyen fertlerin oranı % 6,154, kötü olduğunu söyleyen fertlerin oranı ise % 24,103 olmuştur. “Hayır” cevabını veren fertler ise 68. düğümü oluşturmuştur. Bu düğümde sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 12,5’e yükselirken kötü olarak nitelendiren fertlerin oranı % 12,5’e düşmüştür.

5. düğüm, “KRONIKHAŞT_HIPERTANSİYON” değişkenine göre dallanmıştır. “Son 12 ay içinde hipertansiyon hastalığını yaşadınız mı?” sorusuna “evet” cevabını veren fertler 40. düğümü oluşturmuş ve sağlık durumu iyi olan fertlerin oranı % 10,471, kötü olan fertlerin oranı ise % 45,026 olmuştur. “Hayır” cevabını veren fertlerden oluşan 41. düğümde ise sağlık durumu iyi olan fertlerin oranı % 22,968’e yükselmiş, kötü olan fertlerin oranı ise % 21,908’e düşmüştür.

40. düğümde dallanma olmazken 41. düğüm, “ZORLUKCEKME_AGIREVIS” değişkenine göre dallanmıştır. “Ağır ev işleri faaliyetini yardım almadan gerçekleştirirken zorluk çekiyor musunuz?” sorusuna “hiç zorlanmıyorum” cevabını verenlerden oluşan 69. düğümde sağlık durumu iyi olan fertlerin oranı % 31,655 olduğu gözlenmiştir. “Biraz zorlanıyorum, çok zorlanıyorum, kendi kendime yapamıyorum ve hiçbir zaman denemedim/yapmaya ihtiyaç duymadım” cevaplarını verenlerden oluşan 70. düğümde ise sağlık durumu iyi olan fertlerin oranının % 14,583’e düştüğü gözlenmiştir.

“HASTALIK_SAGLIKDURUM” değişkenine göre 2 alt sınıf oluşturarak dallanan 2. düğüm, “6 ay ya da daha uzun süren/sürmesi beklenen hastalığınız/sağlık probleminiz var mı?” sorusuna “hayır” cevabını fertler ile 7. düğümü oluşturmuştur. 7. düğüme göre oluşan karar ağacı Şekil 4.7.1.5’de gösterilmiştir.

Şekil 4.7.1.5: CHAID Algoritmasının 7. Düğüme Göre Dallanması



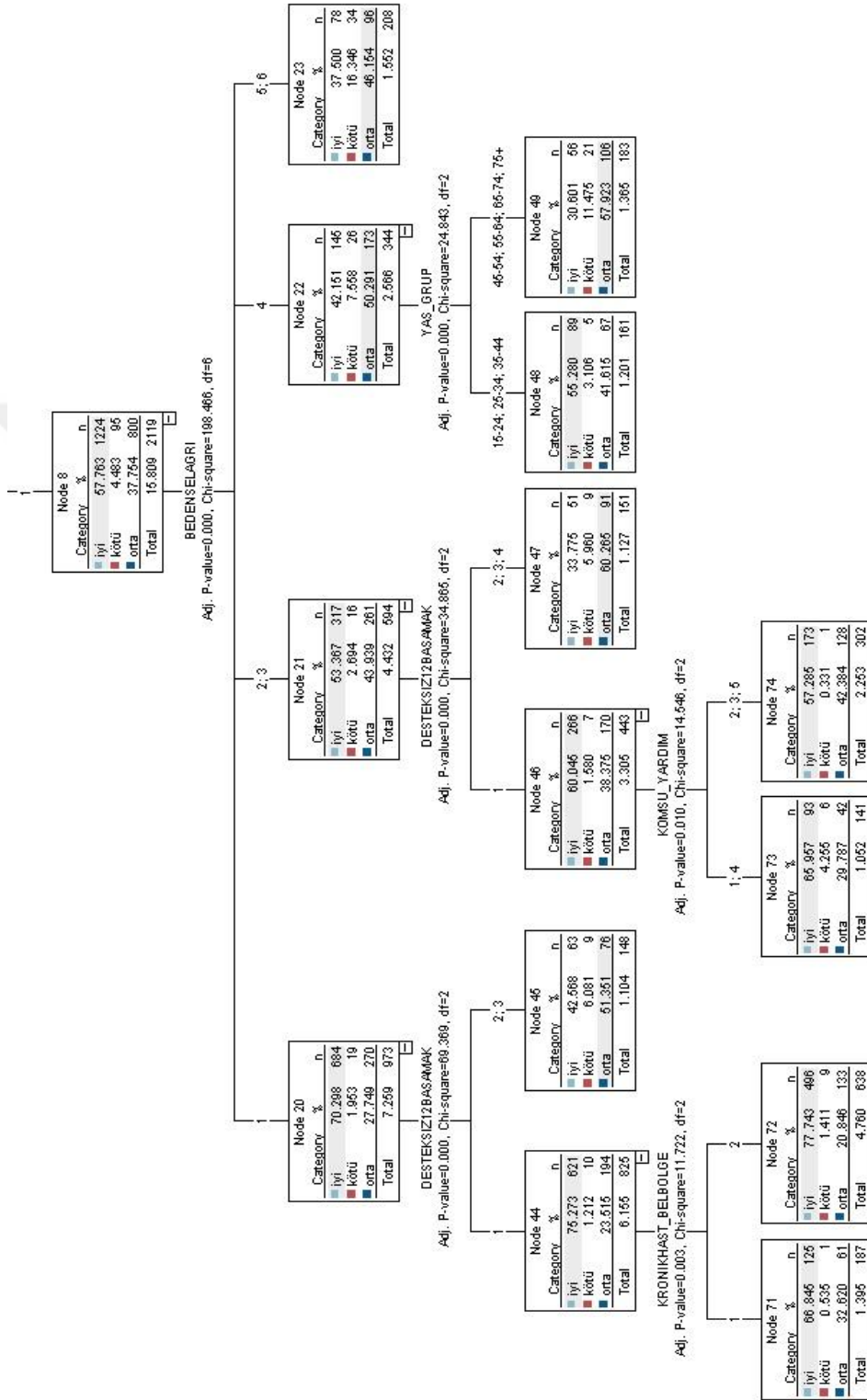
7. düğüm, “BEDENSELAGRI” değişkenine göre dallanmıştır. “Son 4 hafta içerisinde ne kadar bedensel ağrınız oldu?” sorusuna “hiç ve çok az” cevaplarını veren fertlerin oluşturduğu 18. düğümde sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 86,726 olurken kötü olarak nitelendiren fertlerin oranı % 0,664 olmuştur. “Az, orta, fazla ve çok fazla” cevaplarını veren fertlerden oluşan 19.

düğümde ise sağlık durumu iyi olan fertlerin oranı % 53,465'e düşerken sağlık durumu kötü olan fertlerin oranı % 1,485'e yükselmiştir.

18. düğüm, "EGITIM" değişkenine göre dallanırken 19. düğümde dallanma meydana gelmemiştir. Eğitim durumu, "bir okul bitirmedi, ilkokul, 2 veya 3 yıllık yüksekokul ve okuma yazma bilmiyor" olan fertler 42. düğümü, "lise, fakülte, yüksek lisans, genel ortaokul ve ilköğretim" olan fertler ise 43. düğümü oluşturmuştur. 42. düğümde sağlık durumunun iyi olduğunu söyleyen fertlerin oranı % 74,479 iken 43. düğümde sağlık durumunun iyi olduğunu söyleyen fertlerin oranı, % 95,769 olmuştur.

CHAID algoritmasıyla oluşan karar ağacının 3. düğümü "HASTALIK_SAGLIK DURUM" değişkenine göre 2 alt sınıf oluşturarak dallanmıştır. "6 ay ya da daha uzun süren/sürmesi beklenen hastalığınız/sağlık probleminiz var mı?" sorusuna "evet" cevabını veren fertler 8. düğümü, "hayır" cevabını veren fertler ise 9. düğümü oluşturmuştur. 8. düğüme göre oluşan karar ağacı Şekil 4.7.1.6'da gösterilmiştir.

Şekil 4.7.1.6: CHAID Algoritmasının 8. Düzüme Göre Dallanması



8. düğümde yer alan fertlerin % 57,763'ü sağlık durumunu iyi olarak nitelendirirken % 4,483'ü kötü olarak nitelendirmiştir. 8. düğüm, "BEDENSELAGRI" değişkenine göre 4 alt sınıf oluşturarak dallanmıştır. "Son 4 hafta içerisinde ne kadar bedensel ağrınız oldu?" sorusuna "hiç" cevabını veren fertler 20. düğümü oluşturmuş ve sağlık durumunun iyi olduğunu söyleyenlerin oranının % 70,289, kötü olduğunu söyleyenlerin oranının ise % 1,983 olduğu gözlenmiştir. "Çok az ve az" cevabını veren fertlerden oluşan 21. düğümde ise sağlık durumu iyi olan fertlerin oranı % 53,367'ye düşerken kötü olan fertlerin oranı % 2,694'e yükselmiştir. "Orta" cevabını veren fertlerden oluşan 22. düğümde sağlık durumu iyi olan fertlerin oranı düşmeye devam ederek % 42,151 olmuş, kötü olan fertlerin oranı da yükselişini sürdürmüş ve % 7,558 olmuştur. "Fazla ve çok fazla" bedensel ağrısı olan fertlerden oluşan 23. düğümde ise sağlık durumu iyi olan fertler % 37,5 değeri ile diğer 3 düğüme göre en düşük oranda, sağlık durumu kötü olan fertler ise % 16,346 değeri ile en yüksek oranda olduğu gözlenmiştir.

20. düğümde "DESTEKSIZ12BASAMAK" değişkenine göre dallanma meydana gelmiştir. "12 basamak merdiven inip çıkmada zorluk çekiyor musunuz?" sorusuna "hiç zorlanmıyorum" cevabını veren fertler 44. düğümü oluşturmuş ve bu düğümde sağlık durumu iyi olan fertlerin oranı % 75,273, kötü olan fertlerin oranı ise % 1,212 olmuştur. "Biraz zorlanıyorum ve çok zorlanıyorum" cevapları ile oluşan 45. düğümde ise sağlık durumu iyi olan fertlerin oranı % 42,568'e düşerken kötü olan fertlerin oranı % 6,081'e yükselmiştir.

45. düğümde dallanma olmamıştır fakat 46. düğüm, "KRONIKHAST_BELBOLGE" değişkenine göre dallanmıştır. "Son 12 ay içinde bel bölgesi problemleri (bel ağrısı, bel fıtığı ve diğer bel defektleri) yaşadınız mı?" sorusuna "evet" cevabını verenlerin oluşturduğu 71. düğümdeki fertlerin % 66,845'i sağlık durumunu iyi olarak nitelendirirken "hayır" cevabını veren fertlerin oluşturduğu 72. düğümdeki fertlerin % 77,743'ü sağlık durumunu iyi olarak nitelendirmiştir.

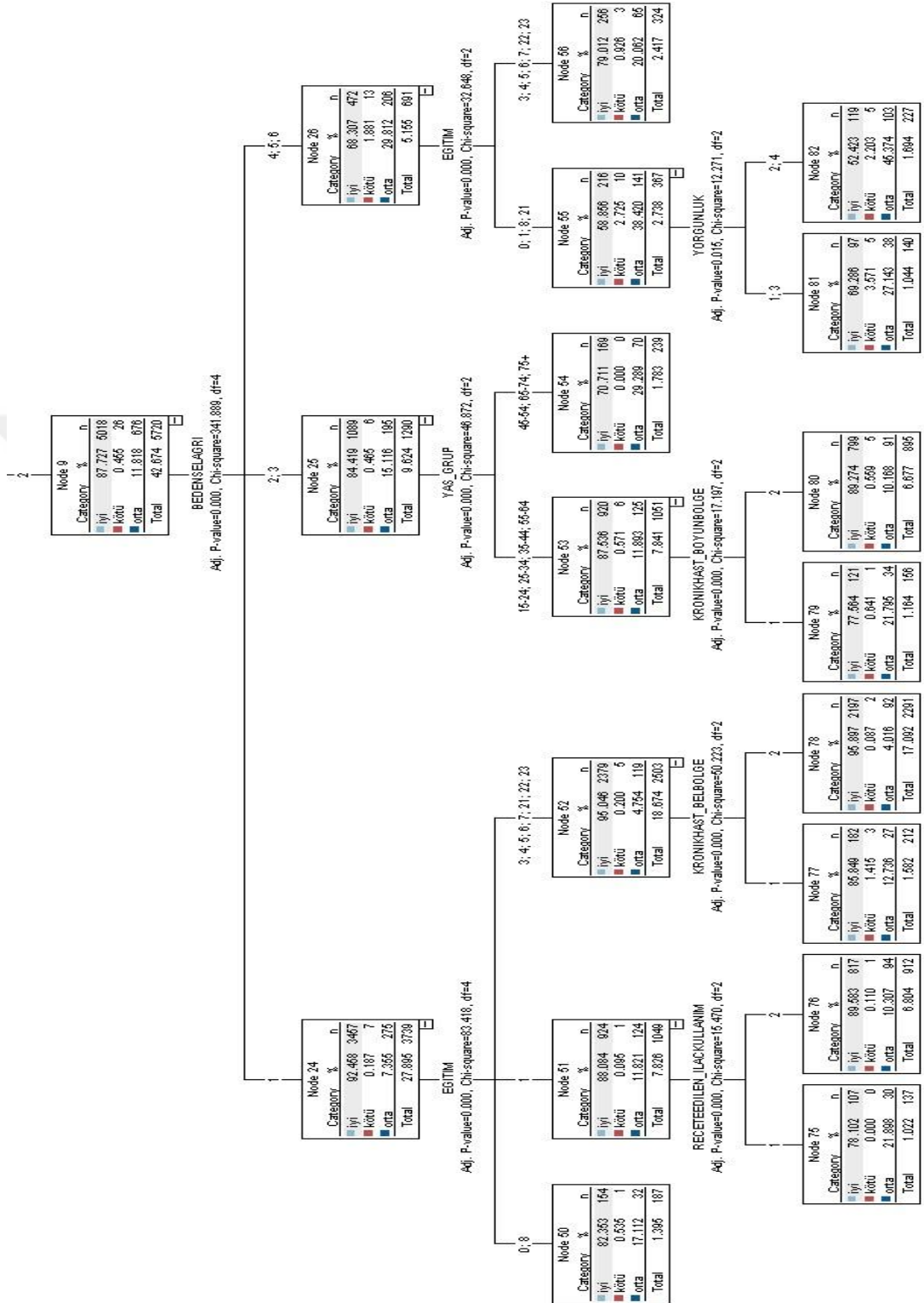
21. düğümde “DESTEKSİZ_12_ BASAMAK” değişkenine göre dallanma meydana gelmiştir. “12 basamak merdiven inip çıkmada zorluk çekiyor musunuz?” sorusuna “hiç zorlanmıyorum” cevabını veren fertler 46. düğümü oluşturmuş ve bu düğümde sağlık durumu iyi olan fertlerin oranı % 60,045, kötü olan fertlerin oranı ise % 1,58 olmuştur. “Biraz zorlanıyorum, çok zorlanıyorum ve hiç merdiven inip çıkamıyorum” cevapları ile oluşan 47. düğümde ise sağlık durumu iyi olan fertlerin oranı % 33,775’e düşerken kötü olan fertlerin oranı ise % 5,96’ya yükselmiştir.

47. düğümde dallanma meydana gelmezken 46. düğüm, ”KOMSU_YARDIM” değişkenine göre dallanmıştır. “İhtiyacınız olduğunda komşularınızdan yardım alabilme durumunuz nedir?” sorusuna “çok kolay ve zor” cevabını verenlerden oluşan 73. düğümdeki fertlerin % 65,957’si sağlık durumunu iyi olarak nitelendirirken “kolay, orta ve çok zor” cevabını veren fertlerin oluşturduğu 74. düğümdeki fertlerin ise % 57,285’i sağlık durumunu iyi olarak nitelendirmiştir

22. düğüm, “YAS_GRUP” değişkenine göre dallanmıştır. “15-24, 25-34 ve 35-44” yaş gruplarında yer alan fertlerin oluşturduğu 48. düğümde sağlık durumunu iyi olarak nitelendirenlerin oranı % 55,28, kötü olarak nitelendirenlerin oranının ise % 3,106 olduğu gözlenmiştir. “45-54, 55-64, 65-74 ve 75+” yaş grubunda yer alan fertlerden oluşan 49. düğümde ise sağlık durumu iyi olan fertlerin oranı % 30,601’e düşerken kötü olan fertlerin ise oranı % 11,475’e yükselmiştir.

CHAID algoritmasıyla oluşan karar ağacının 9. düğüme göre yaptığı dallanma Şekil 4.7.1.7’de yer almaktadır.

Şekil 4.7.1.7: CHAID Algoritmasının 9. Düzüme Göre Dallanması



9. düğüm, “BEDENSELAGRI” değişkenine göre 3 alt sınıf oluşturarak dallanmıştır. “Son 4 hafta içerisinde ne kadar bedensel ağrınız oldu?” sorusuna “hiç” cevabını verenler ile 24. düğüm oluşmuş ve sağlık durumunun iyi olduğunu söyleyen fertlerin oranı % 92,458, kötü olduğunu söyleyen fertlerin oranının ise % 0,187 olmuştur. “Çok az ve az” cevaplarını veren fertlerden oluşan 25. düğümde ise sağlık durumu iyi olan fertlerin oranı % 84,419’a düşerken sağlık durumu kötü olan fertlerin oranı ise % 0,465’e yükselmiştir. “Orta, fazla ve çok fazla” bedensel ağrısının olduğunu söyleyen fertlerin oluşturduğu 26. düğümde sağlık durumu iyi olan fertlerin oranı düşmeye devam ederek % 68,307, sağlık durumu kötü olan fertlerin oranı ise diğer 2 düğüme göre yükselerek % 1,881 olmuştur.

24. düğüm, “EGITIM” değişkenine göre dallanmıştır. Mezuniyet durumu “bir okul bitirmedim ve hiç okuma yazması yok” olan fertlerden oluşan 50. düğümde sağlık durumu iyi olan fertlerin oranı % 82,353 olmuştur. Mezuniyet durumu “ilkokul” olan fertlerin oluşturduğu 51. düğümde sağlık durumu iyi olan fertlerin oranı % 88,084’e yükselmiştir. Mezuniyet durumu “lise, yüksekokul, fakülte, yüksek lisans, doktora, genel ortaokul, teknik ortaokul ve ilköğretim” olan fertlerin oluşturduğu 52. düğümde ise sağlık durumunu iyi olarak nitelendirenlerin oranı yükselmeye devam ederek % 95,046 olduğu gözlenmiştir.

50. düğümde dallanma meydana gelmemiştir ancak 51. düğüm, “RECETE_EDILEN_ILAC_KULLANIM” değişkenine, 52. düğüm ise “KRONIKHAŞT_BELBOLGE” değişkenine göre dallanmıştır.

51. düğümden sonra, “son 2 hafta içerisinde bir doktor tarafından size reçete edilen bir ilaç kullandınız mı?” sorusuna evet cevabını fertler ile oluşan 75. düğümde sağlık durumu iyi olan fertlerin oranı % 78,102 olmuştur. “Hayır” cevabını veren fertlerin oluşturduğu 76. düğümde ise sağlık durumu iyi olanların oranı % 89,583 olmuştur.

52. düğümden sonra, “son 12 ay içinde bel bölgesi problemleri (bel ağrısı, bel fıtığı ve diğer bel defektleri) yaşadınız mı?” sorusuna evet cevabını fertler ile oluşan 77.

düğümde sağlık durumu iyi olan fertlerin oranı % 85,849 olmuştur. “Hayır” cevabını veren fertlerin oluşturduğu 78. düğümde ise sağlık durumu iyi olanların oranı % 95,897’ye yükselmiştir.

25. düğüm, “YASGRUP” değişkenine göre dallanmıştır. “15-24, 25-34, 35-44 ve 55-64” yaş grubunda yer alan fertlerden oluşan 53. düğümde sağlık durumu iyi olan fertlerin oranı, % 87,536 iken “45-54, 65-74 ve 75+” yaş grubunda yer alan fertlerin oluşturduğu 54. düğümde sağlık durumu iyi olan fertlerin oranı, % 70,711’e düşmüştür.

53. düğüm, “KRONIKHAST_ BOYUNBOLGE” değişkenine göre dallanırken 54. düğümde dallanma meydana gelmemiştir. “Son 12 ay içinde boyun bölgesi problemleri (boyun ağrısı, boyun fıtığı, diğer boyun defektleri) yaşadınız mı?” sorusuna “evet” cevabını veren fertlerin %77,564’ü, “hayır” cevabını veren fertlerin ise % 89,274’ü sağlık durumunu iyi olarak nitelendirmiştir.

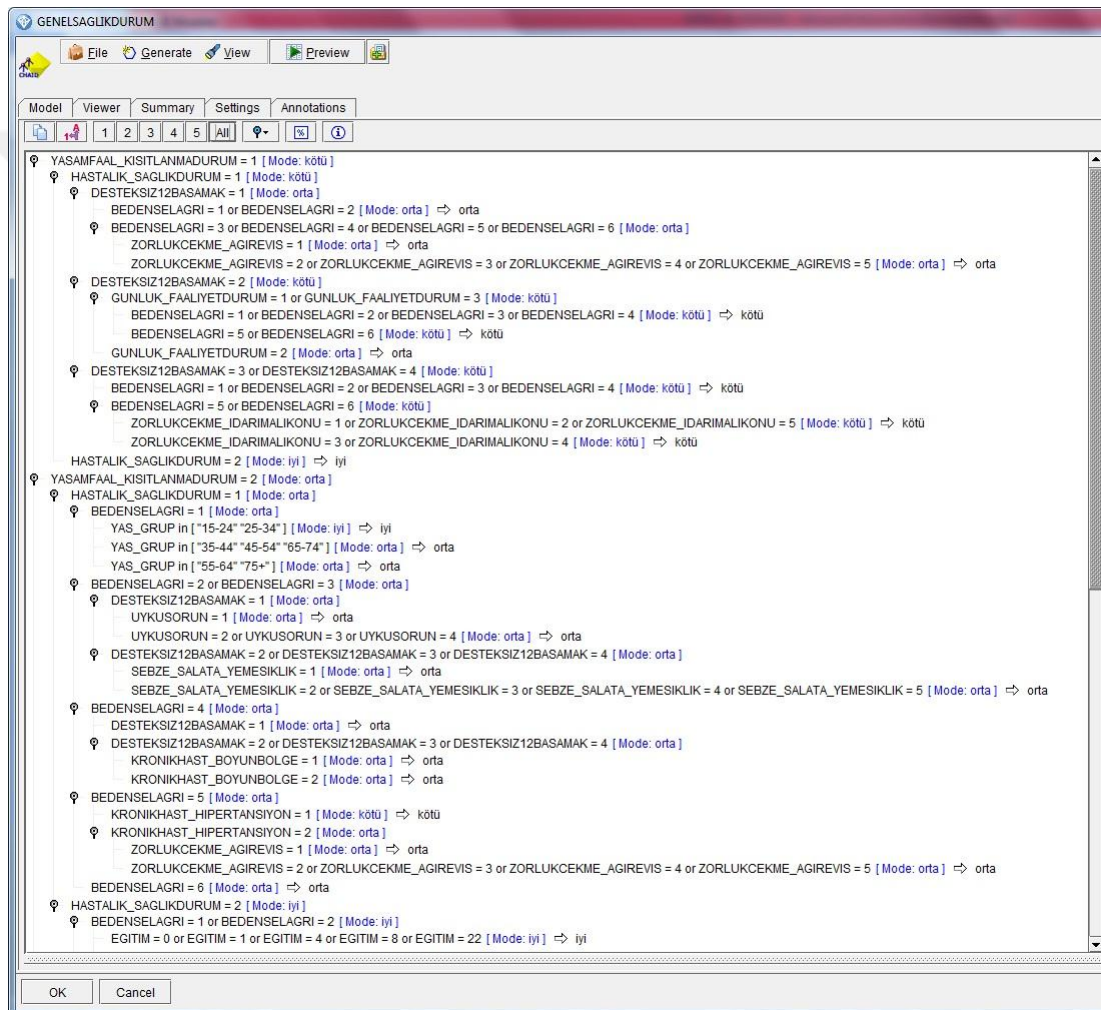
26. düğümdeki dallanma, “EGITIM” değişkenine göre olmuştur. Mezuniyet durumu “bir okul bitirmedi, ilkokul, okuma yazma bilmiyor ve genel ortaokul” olan fertlerden oluşan 55. düğümde sağlık durumu iyi olan fertlerin oranı % 58,856, kötü olan fertlerin oranının ise % 2,725 olduğu gözlenmiştir. Mezuniyet durumu, “lise, yüksekokul, fakülte, yüksek lisans, mesleki veya teknik ortaokul ve ilköğretim” olan fertlerin oluşturduğu 56. düğümde ise sağlık durumu iyi olan fertlerin oranı % 79,012’ye yükselirken sağlık durumu kötü olan fertlerin oranı ise % 0,926’ya düşmüştür.

55. düğüm, “YORGUNLUK” değişkenine göre dallanırken 56. düğümde dallanma meydana gelmemiştir. “Son 2 hafta içerisinde yorgun/enerjisiz hissetme problemi sizi ne sıklıkla rahatsız etti?” sorusuna “hiç ve bir haftadan fazla” cevaplarını veren fertlerden oluşan 81. düğümde sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 69,286 olmuştur. “Bazı günler ve neredeyse her gün” cevaplarını veren fertlerden

oluşan 82. düğümde ise sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 52,423'e düşmüştür.

Analiz için model kurulup algoritma çalışmaya başladığında program tarafından bir kural seti üretilir. CHAID algoritması 47 tane kural üretmiştir. Kural setinin bir kısmı Şekil 4.7.1.8'de gösterilmektedir.

Şekil 4.7.1.8: CHAID Algoritması İle Üretilen Kural Setinin Bir Kısımı



CHAID algoritması ile oluşturulan karar ağacında; genel sağlık durumu iyi olan fertler en yüksek oranla 78. düğümde, orta olan fertler en yüksek oranla 68. düğümde

ve kötü olan fertler en yüksek oranla 62. düğümde yer almıştır. Bu düğümlere ait kurallar aşağıdaki gibi ifade edilebilir.

- Eğer bir kişinin herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre kısıtlanmadıysa ve 6 ay ya da daha uzun süren/sürmesi beklenen hastalığı/sağlık problemi yoksa ve son 4 hafta içerisinde hiç bedensel ağrısı olmadıysa ve eğitim durumu lise veya mesleki lise veya 2 veya 3 yıllık yüksekokul veya 4 yıllık yüksekokul veya fakülte veya yüksek lisans (5 veya 6 yıllık fakülteler dâhil) veya doktora veya genel ortaokul veya mesleki veya teknik ortaokul veya ilköğretim ise ve son 12 ay içinde bel bölgesi problemleri (bel ağrısı, bel fıtığı ve diğer bel defektleri) yaşamadıysa genel sağlık durumu iyidir.
- Eğer bir kişinin herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre kısıtlandı fakat ciddi ölçüde kısıtlanmadıysa ve 6 ay ya da daha uzun süren/sürmesi beklenen hastalığı/sağlık problemi varsa ve son 4 hafta içerisinde bedensel ağrısı orta seviyede olduysa ve 12 basamak merdiven inip çıkmada biraz zorlanıyor veya çok zorlanıyor veya hiç merdiven inip çıkamıyorsa ve son 12 ay içinde boyun bölgesi problemleri (boyun ağrısı, boyun fıtığı, diğer boyun defektleri) yaşamadıysa genel sağlık durumu ortadır.
- Eğer bir kişinin herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre ciddi ölçüde kısıtlandıysa ve 6 ay ya da daha uzun süren/sürmesi beklenen hastalığı/sağlık problemi varsa ve 12 basamak merdiven inip çıkmada çok zorlanıyor veya hiç merdiven inip çıkamıyorsa ve son 4 hafta içerisinde bedensel ağrısı fazla veya çok fazla ise ve günlük idari sorumluluklar ve mali konularla ilgilenme faaliyetini yardım almadan gerçekleştirirken çok zorlanıyor veya kendi kendine yapamıyorsa genel sağlık durumu kötüdür.

CHAID algoritmasının uygulanması sırasında önceden bölünmüş eğitim kümesi ile model kurulmuş ve test kümesi ile de kurulan modelin güvenilirliği ve doğruluğu test edilmiştir. Modelin performans testini gerçekleştirebilmek için IBM SPSS Modeler programında yer alan analiz işlemcisi, CHAID algoritmasına bağlanarak çalıştırılmıştır. CHAID algoritmasının eğitim ve test kümesi için sınıflandırma performans ölçüm sonuçları Tablo 4.7.1.2’de verilmiştir.

Tablo 4.7.1.2: CHAID Algoritmasının Performans Ölçüm Sonuçları

Sınıflandırma	Eğitim Kümesi		Test Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	9819	% 73,25	4165	% 72,75
Yanlış	3585	% 26,75	1560	% 27,25
Toplam	13404	% 100	5725	% 100

Eğitim kümesinde yer alan 13404 adet veri eğitildikten sonra test kümesinde yer alan 5725 adet veri, test edilerek modelin doğruluk oranı belirlenmiştir. CHAID algoritmasının modeli doğru tahmin etme oranının % 72,75 olduğu söylenebilir.

4.7.2. QUEST Algoritması

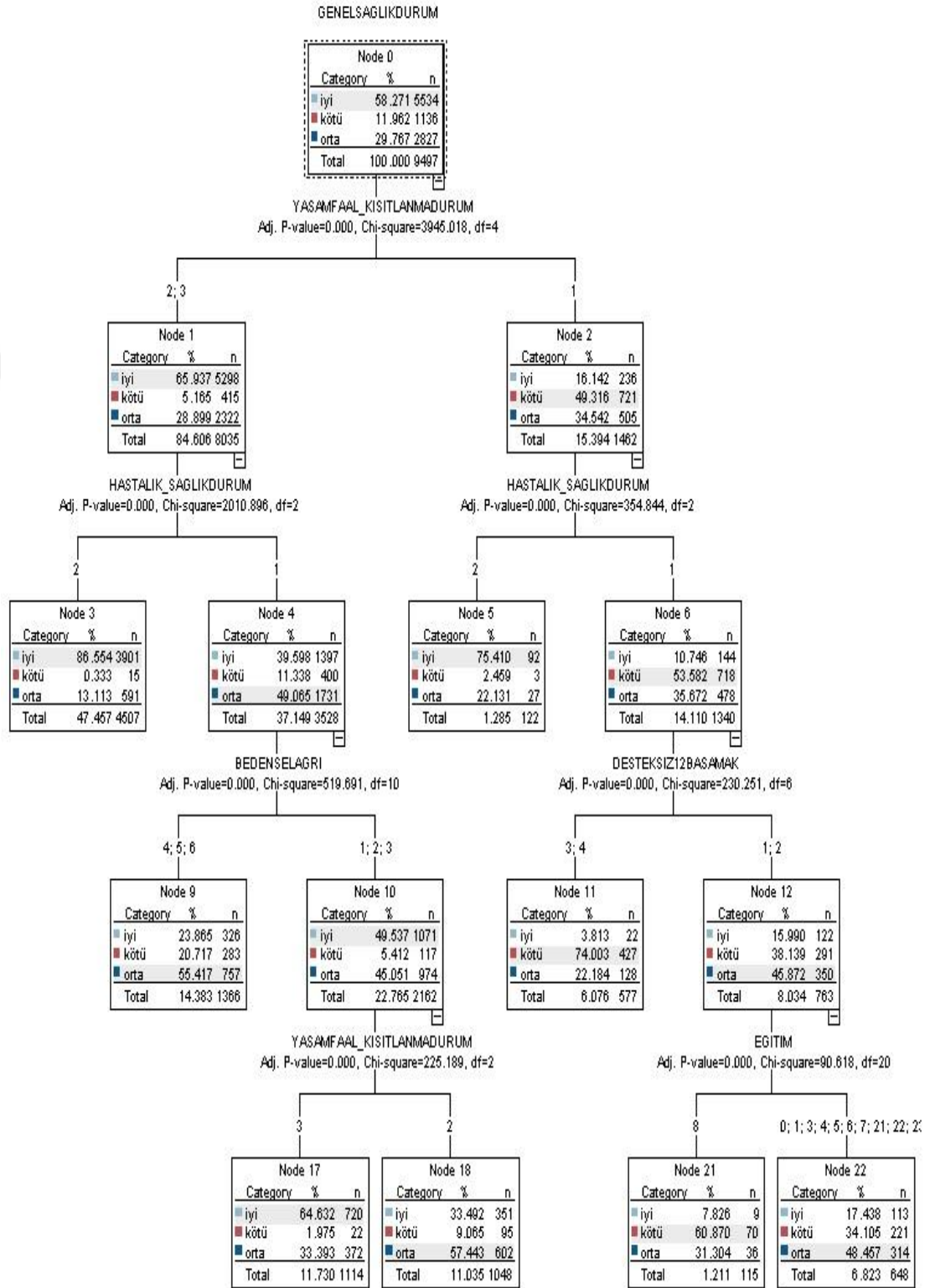
Quest algoritması ile oluşturulan karar ağacında en önemli değişken, 0,4825 önem değeri ile “HASTALIK_SAGLIK DURUM” değişkeni olmuştur. Bu değişkeni 0,1746 önem değeri ile “YASAMFAAL_KISITLANMADURUM” değişkeni ve 0,0414 önem değeri ile “DESTEKSİZ12BASAMAK” değişkenleri takip etmiştir. QUEST algoritması ile elde edilen değişkenlerin önem değerleri Tablo 4.7.2.1.’de gösterilmiştir.

Tablo 4.7.2.1: QUEST Algoritmasında Kullanılan Değişkenlerin Önem Değerleri

Değişken	Değişken Önem Değeri
HASTALIK_SAGLIKDURUM	0,4825
YASAMFAAL_KISITLANMADURUM	0,1746
DESTEKSIZ12BASAMAK	0,0414
BEDENSELAGRI	0,0362
EGITIM	0,019
OGRENME	0,019
YORGUNLUK	0,019
HATIRLAMA	0,019
KRONIKHAST_BELBOLGE	0,019
ZORLUKCEKME_AGIREVIS	0,019
KRONIKHAST_BOYUNBOLGE	0,019
RECETEEDILEN_ILACKULLANIM	0,019
YAS_GRUP	0,019
KRONIKHAST_HIPERTANSİYON	0,019
HUZURSUZLUK	0,019
ZORLUKCEKME_IDARIMALIKONU	0,019
DESTEKSIZ500MYURUYEBİLME	0,019
ZORLUKCEKME_HAFİFEVIS	0,019

QUEST algoritması ile oluşturulan karar ağacının ilk düğümünde bağımlı değişken olarak belirlenen genel sağlık durumu değişkeni yer almaktadır. Ağaç derinliği 4 olan karar ağacı, toplam 15 düğümden oluşmuştur. Karar ağacının yapısı Şekil 4.7.2.1’de gösterilmiştir.

Şekil: 4.7.2.1: Quest Algoritması ile Oluşturulan Karar Ağacı



Başlangıç düğümünde yer alan 9497 ferдин % 58,271'i genel sağlık durumunu iyi olarak nitelendirirken % 29,762'si orta, % 11,962'si ise kötü olarak nitelemiştir.

Karar ağacı ilk dallanmayı “YASAMSALFAAL_KISITLANMADURUM” değişkeni ile yapmıştır. “Herhangi bir sağlık problemi nedeniyle günlük faaliyetlerinizin 6 ay ve daha uzun süre kısıtlanma durumu nedir?” sorusuna verilen “kısıtlandı fakat ciddi ölçüde değil ve kısıtlanmadı” cevaplarıyla 1. düğüm, “ciddi ölçüde kısıtlandı” cevabını verenlerle 2. düğüm oluşmuştur.

1. düğümde genel sağlık durumunun iyi olduğunu söyleyen fertlerin oranı % 65,937 iken kötü olduğunu söyleyen fertlerin oranı ise % 5,165 olmuştur. Daha sonra 1. düğüm, “HASTALIK_SAGLIK DURUMU” değişkenine göre dallanmıştır. “6 ay ya da daha uzun süren/sürmesi beklenen hastalığınız/sağlık problemiz var mı?” sorusuna “hayır” cevabını veren fertlerle 3. düğüm, “evet” cevabını veren fertlerle 4. düğüm oluşmuştur. 3. düğümde yer alan fertlerin % 86,554'ü sağlık durumunu iyi, % 0,333'ü ise kötü olarak nitelendirmiştir. 4. düğümde sağlık durumuna iyi cevabını fertlerin oranı % 39,598'e düşerken kötü cevabını veren fertlerin oranı % 11,338'e yükselmiştir. 3. düğümde dallanma olmamıştır ancak 4. düğüm, “BEDENSELAGRI” değişkenine göre dallanmıştır. “Son 4 hafta içerisinde ne kadar bedensel ağrınız oldu?” sorusuna “orta, fazla ve çok fazla” cevabını veren fertlerle 9. düğüm, “hiç, çok az ve az” cevabını veren fertlerle ise 10. düğüm oluşmuştur.

9. düğümde bulunan fertlerin % 23,865'i sağlık durumunun iyi olduğunu söylerken % 20,717'si kötü olduğunu söylemiştir. 10. düğümde sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 49,537'ye yükselirken kötü olarak nitelendiren fertlerin oranı ise % 5,412'ye düşmüştür. 9. düğümde dallanma olmazken 10. düğüm “YASAMSALFAAL_KISITLANMADURUM” değişkeni ile dallanmıştır. Herhangi bir sağlık problemi nedeniyle günlük faaliyetlerinizin 6 ay ve daha uzun süre kısıtlanma durumu nedir?” sorusuna “kısıtlanmadı” cevabını veren fertler 17. düğümü, “Kısıtlandı fakat ciddi ölçüde değil” cevabını veren fertler ise 18. düğümü oluşturmuştur.

17. düğümde yer alan fertlerin % 64,632'si sağlık durumunu iyi olarak nitelendirirken % 1,975'i kötü olarak nitelemiştir. 18. düğümde ise sağlık durumunu iyi olarak nitelendirenlerin oranı % 33,492'ye düşmüş, kötü olarak nitelendirenlerin oranı ise % 9,065'e yükselmiştir.

Karar ağacının 2. düğümünde genel sağlık durumunun iyi olduğunu söyleyen fertlerin oranının % 16,142, kötü olduğunu söyleyen fertlerin oranının ise % 49,316 olduğu gözlenmiştir. Bu düğümdeki dallanma "HASTALIK_SAGLIKDURUM" değişkenine göre gerçekleşmiştir. "6 ay ya da daha uzun süren/sürmesi beklenen hastalığınız/sağlık problemiz var mı?" sorusuna "hayır" cevabını veren fertlerle 5. düğüm, "evet" cevabını veren fertlerle ise 6. düğüm oluşmuştur. 5. düğümde yer alan fertlerin % 75,41'i sağlık durumunu iyi olarak nitelendirirken % 2,459'u kötü olarak nitelendirmiştir. 6. düğümde yer alan fertlerin % 10,746'sı sağlık durumunun iyi, % 53,582'si kötü olduğunu söylemiştir.

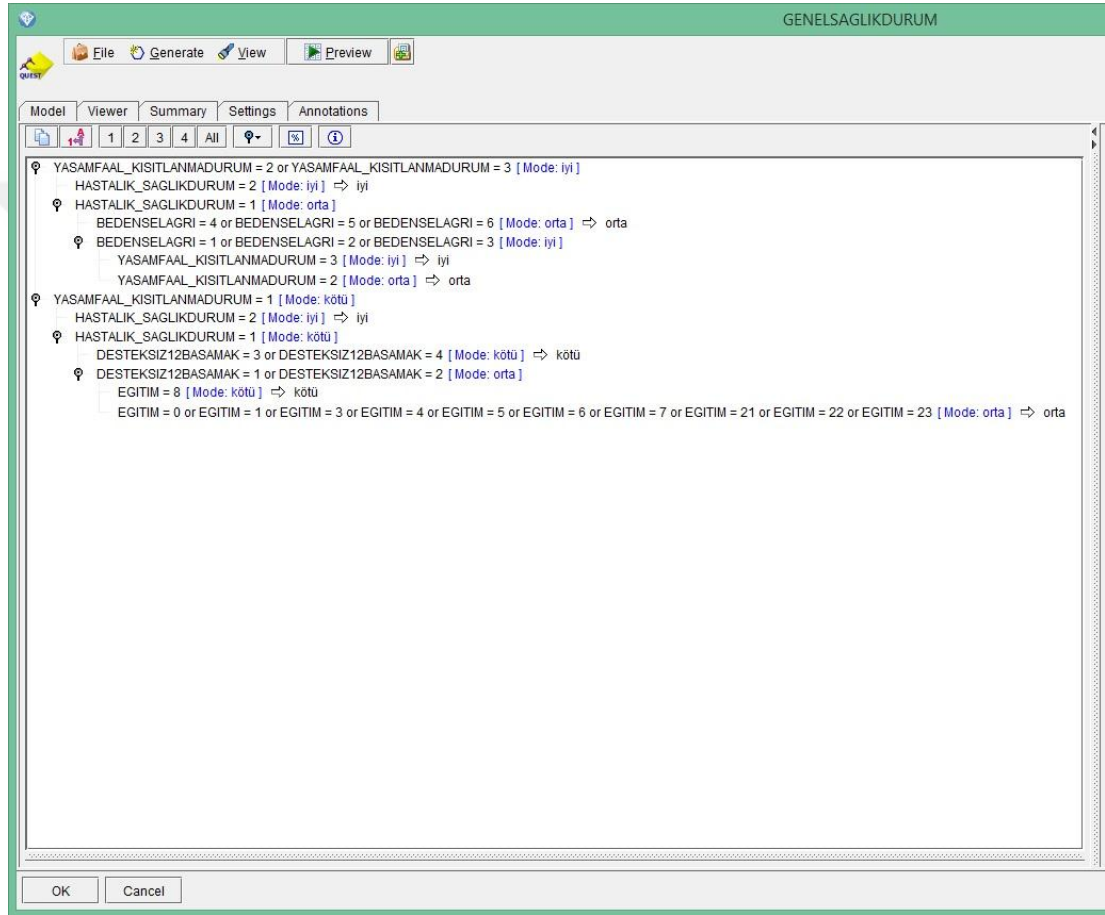
5. düğümde dallanma gerçekleşmezken 6. düğüm, "DESTEKSIZ12BASAMAK" değişkenine göre dallanmıştır. "12 basamak merdiven inip çıkmada zorluk çekiyor musunuz?" sorusuna "çok zorlanıyorum ve hiç merdiven inip çıkamıyorum" cevabını veren fertlerle 11. düğüm, "hiç zorlanmıyorum ve biraz zorlanıyorum" cevabını veren fertlerle 12. düğüm oluşmuştur. 11. düğümdeki fertlerin % 3,813'ü sağlık durumunu iyi olarak nitelendirirken % 74,003'ü kötü olarak nitelendirmiştir. 12. düğümde yer alıp sağlık durumunun iyi olduğunu söyleyenlerin oranı % 15,99'a yükselmiş, kötü olduğunu söyleyenlerin oranı ise % 38,139'a düşmüştür.

11. düğümde dallanma meydana gelmemiştir ancak 12. düğüm, "EGITIM" değişkenine göre dallanmıştır. Hiç okuma yazma bilmediğini söyleyen fertler 21. düğümü oluşturmuştur. "Bir okul bitirmedi, ilkokul, genel ortaokul, mesleki (teknik) ortaokul, ilköğretim, lise, yüksekokul, fakülte, yüksek lisans ve doktora" mezuniyet durumu olan fertler ise 22. düğümü oluşturmuştur. 21. düğümde yer alan fertlerin % 7,826'sı sağlık durumunun iyi olduğunu söylerken % 60,87'si sağlık durumunun

kötü olduğunu söylemiştir. 22. düğümde yer alan fertlerin ise % 17,438'i sağlık durumunu iyi olarak nitelendirirken % 34,105'i ise kötü olarak nitelendirmiştir.

Quest algoritması toplam 8 tane kural üretmiştir. Algoritmanın ürettiği kural seti Şekil 4.7.2.2'de gösterilmiştir.

Şekil 4.7.2.2: QUEST Algoritması İle Üretilen Kural Seti



Genel sağlık durumu iyi olan fertler en yüksek oranla 3. düğümde, orta olan fertler en yüksek oranla 18. düğümde ve kötü olan fertler en yüksek oranla 11. düğümde yer almaktadır. Bu düğümlere ait kurallar aşağıdaki gibi ifade edilebilir.

- Eğer bir kişinin herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre kısıtlanma durumu, kısıtlandı fakat ciddi ölçüde değil

veya hiç kısıtlanmadı ise ve 6 ay ya da daha uzun süren/sürmesi beklenen hastalığı/sağlık problemi yoksa genel sağlık durumu iyidir.

- Eğer bir kişinin herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre kısıtlanma durumu, kısıtlandı fakat ciddi ölçüde değil veya hiç kısıtlanmadı ise ve 6 ay ya da daha uzun süren/sürmesi beklenen hastalığı/sağlık problemi varsa ve son 4 hafta içerisinde hiç veya çok az veya az bedensel ağrısı olduysa ve herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre kısıtlanma durumu kısıtlandı fakat ciddi ölçüde değilse genel sağlık durumu ortadır.
- Eğer bir kişinin herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre kısıtlanma durumu, ciddi ölçüde kısıtlandı ise ve 6 ay ya da daha uzun süren/sürmesi beklenen hastalığı/sağlık problemi varsa ve 12 basamak merdiven inip çıkmada çok zorlanıyor veya hiç merdiven inip çıkamıyorsa genel sağlık durumu kötüdür.

QUEST algoritmasının uygulanması sırasında önceden bölünmüş eğitim kümesi ile model kurulmuş ve test kümesi ile kurulan modelin güvenilirliği ve doğruluğu test edilmiştir. QUEST algoritmasının eğitim ve test kümesi için sınıflandırma performans ölçüm sonuçları Tablo 3.4.6.2’de verilmiştir.

Tablo 4.7.2.2: QUEST Algoritmasının Performans Ölçüm Sonuçları

Sınıflandırma	Eğitim Kümesi		Test Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	9668	% 72,13	4153	% 72,54
Yanlış	3736	% 27,87	1572	% 27,46
Toplam	13404	% 100	5725	% 100

Eđitim kümesinde yer alan 13404 adet veri eđitildikten sonra test kümesinde yer alan 5725 adet veri test edilerek modelin dođruluk oranı belirlenmiştir. QUEST algoritmasının modeli dođru tahmin etme oranının % 72,54 olduđu söylenebilir.

4.7.3. C&R Tree Algoritması

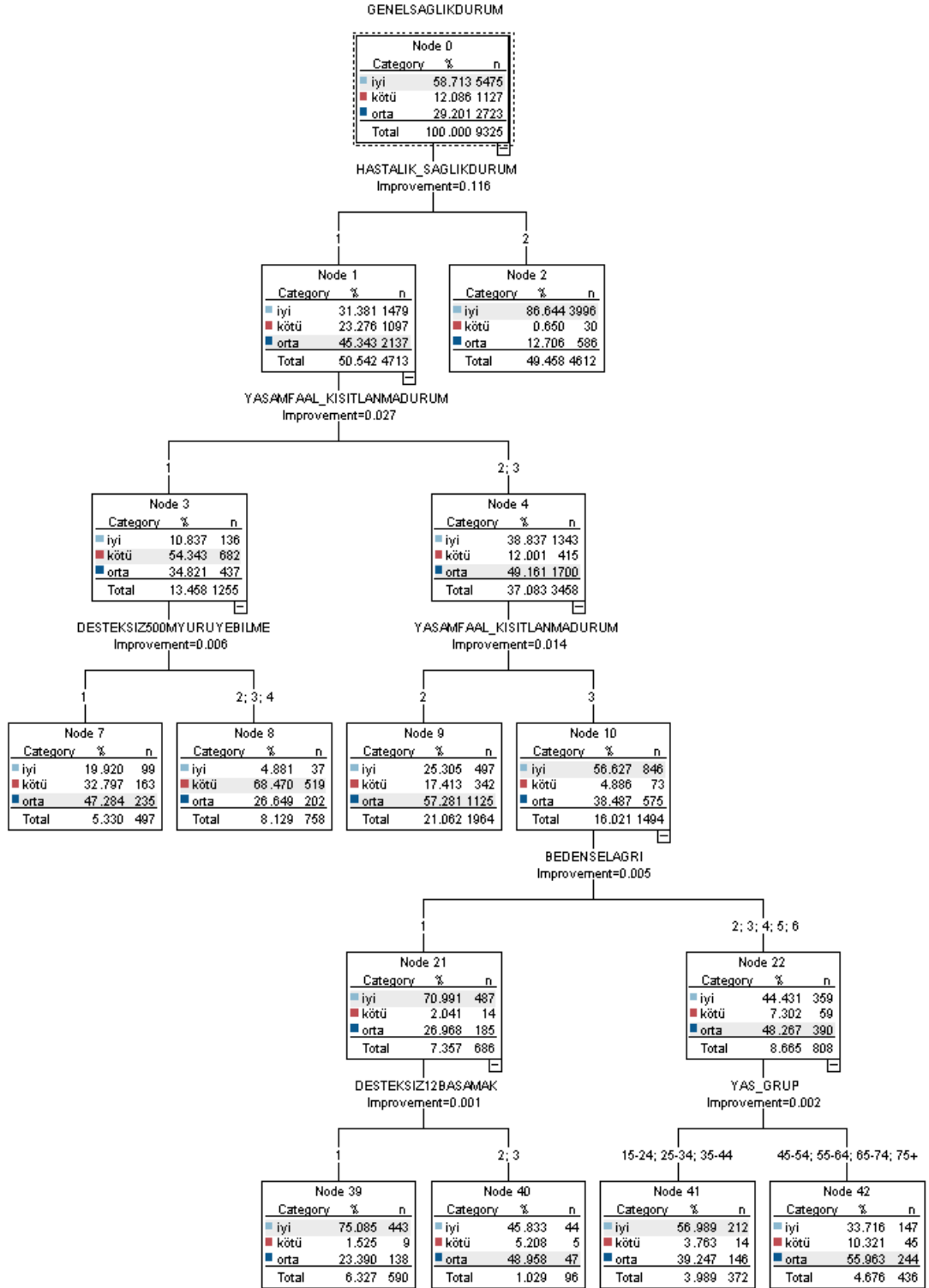
C&R Tree algoritması ile kurulan modelde en önemli deđişken, 0,4281 önem deđeri ile “HASTALIK_SAGLIKDURUM” deđişkeni olmuştur. Bu deđişkeni 0,152 önem deđeri ile “YASAMFAAL_KISITLANMADURUM” deđişkeni ve 0,0453 önem deđeri ile “DESTEKSIZ500MYURUYEBILME” deđişkenleri takip etmiştir. C&R Tree algoritması ile elde edilen deđişkenlerin önem deđerleri Tablo 4.7.3.1’de gösterilmiştir.

Tablo 4.7.3.1: C&R Tree Algoritmasında Kullanılan Değişkenlerin Önem Değerleri

Değişken	Değişken Önem Değeri
HASTALIK_SAGLIKDURUM	0,4281
YASAMFAAL_KISITLANMADURUM	0,152
DESTEKSIZ500MYURUYEBILME	0,0453
YAS_GRUP	0,0356
DESTEKSIZ12BASAMAK	0,0225
BEDENSELAGRI	0,0203
CALISMA_DURUM	0,0165
MEDENI_DURUM	0,0165
KRONIKHAST_HIPERTANSIYON	0,0165
GOZLUKKULLANMA	0,0165
HATIRLAMA	0,0165
KRONIKHAST_BOYUNBOLGE	0,0165
ZEVKDUYMA	0,0165
BUNALIM	0,0165
KRONIKHAST_BELBOLGE	0,0165
CINSIYET	0,0165
EGITIM	0,0165
YORGUNLUK	0,0165
OGRENME	0,0165
GUNLUK_FAALİYETDURUM	0,0165
ZORLUKCEKME_AGIREVIS	0,0165
ZORLUKCEKME_IDARIMALIKONU	0,0165
ZORLUKCEKME_HAFIFEVIS	0,0165
RECETEEDILEN_ILACKULLANIM	0,0165

C&R Tree algoritması ile oluşturulan karar ağacının ilk düğümünde bağımlı değişken olarak belirlenen genel sağlık durumu değişkeni yer almaktadır. Ağaç derinliği 5 olan karar ağacı, toplam 15 düğümden oluşmuştur. Karar ağacının yapısı Şekil 4.7.3.1’de gösterilmiştir.

Şekil 4.7.3.1: C&R Tree Algoritması ile Oluşturulan Karar Ağacı



Başlangıç düğümünde yer alan 9325 ferдин % 58,713'ü genel sağlık durumunu iyi olarak nitelendirirken % 29,201'i orta, % 12,086'sı ise kötü olarak nitelendirmiştir.

Karar ağacı dallanmaya "HASTALIK_SAGLIK DURUMU" değişkeni ile 2 alt sınıf oluşturarak başlamıştır. "6 ay ya da daha uzun süren/sürmesi beklenen hastalığınız/sağlık probleminiz var mı?" sorusuna verilen "evet ve hayır" cevaplarıyla dallanma gerçekleşmiştir.

1. düğüm, "evet" cevabını veren fertlerden oluşmuştur. Bu düğümde genel sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 31,381 iken kötü olarak nitelendirenlerin oranı % 23,276 olmuştur. "Hayır" cevabını veren fertlerden oluşan 2. düğümde ise sağlık durumunu iyi olarak nitelendiren fertlerin oranı % 86,644'e yükselmiş kötü olarak nitelendirenlerin oranı ise % 0,65'e düşmüştür.

2. düğümde dallanma meydana gelmemiştir ancak 1. düğüm, "YASAMSALFAAL_KISITLANMADURUM" değişkenine göre dallanmıştır. "Herhangi bir sağlık problemi nedeniyle günlük faaliyetlerinizin 6 ay ve daha uzun süre kısıtlanma durumu nedir?" sorusuna, "ciddi ölçüde kısıtlandı" cevabını veren fertler ile 3. düğüm, "kısıtlandı fakat ciddi ölçüde değil ve kısıtlanmadı" cevabını veren fertlerle 4. düğüm oluşmuştur.

3. düğümde yer alan fertlerin % 10,837'si genel sağlık durumunu iyi olarak nitelendirirken % 54,343'ü kötü olarak nitelendirmiştir. 3. düğüm, "DESTEKSİZ500MYURUYEBİLME" değişkeni ile tekrar dallanmıştır. "Düz bir arazide baston veya diğer bir yürüme aracı kullanmadan ya da destek almadan 500 metre yürümekte zorluk çekiyor musunuz?" sorusuna "hiç zorlanmıyorum" cevabını veren fertler ile 7. düğüm, "biraz zorlanıyorum, çok zorlanıyorum ve hiç yürüyemiyorum" cevaplarını veren fertler ile de 8. düğüm oluşmuştur. 7. düğümde yer alan fertlerin % 19,92'si genel sağlık durumunun iyi olduğunu söylerken % 32,797'si kötü olduğunu söylemiştir. 8. düğümde sağlık durumunu iyi olarak

nitelendiren fertlerin oranı diğer düğüme göre % 4,881'e düşmüş, kötü olarak nitelendirenlerin oranı % 68,47'ye yükselmiştir.

4. düğümden yer alan fertlerin % 38,837'si genel sağlık durumunu iyi olarak nitelendirirken % 12,001'i kötü olarak nitelendirmiştir. 4. düğüm, "YASAMSALFAAL_KISITLANMADURUM" değişkenine göre dallanmıştır. "Herhangi bir sağlık problemi nedeniyle günlük faaliyetlerinizin 6 ay ve daha uzun süre kısıtlanma durumu nedir?" sorusuna verilen "kısıtlandı fakat ciddi ölçüde değil" cevabı ile 9. düğüm, "kısıtlanmadı" cevabı ile de 10. düğüm oluşmuştur. 9. düğümden bulunan fertlerin % 25,305'i sağlık durumunun iyi olduğunu söylerken % 17,413'ü kötü olduğunu söylemiştir. 10. düğümden ise sağlık durumunun iyi olduğunu söyleyen fertlerin oranı diğer düğüme göre % 56,627'ye yükselmiş, kötü olduğunu söyleyenlerin ise oranı % 4,886'ya düşmüştür.

10. düğüm, "BEDENSELAGRI" değişkenine göre dallanmıştır. "Son 4 hafta içerisinde ne kadar bedensel ağrınız oldu?" sorusuna verilen "hiç" cevabı ile 21. düğüm, "çok az, az, orta, fazla ve çok fazla" cevapları ile 22. düğüm oluşmuştur.

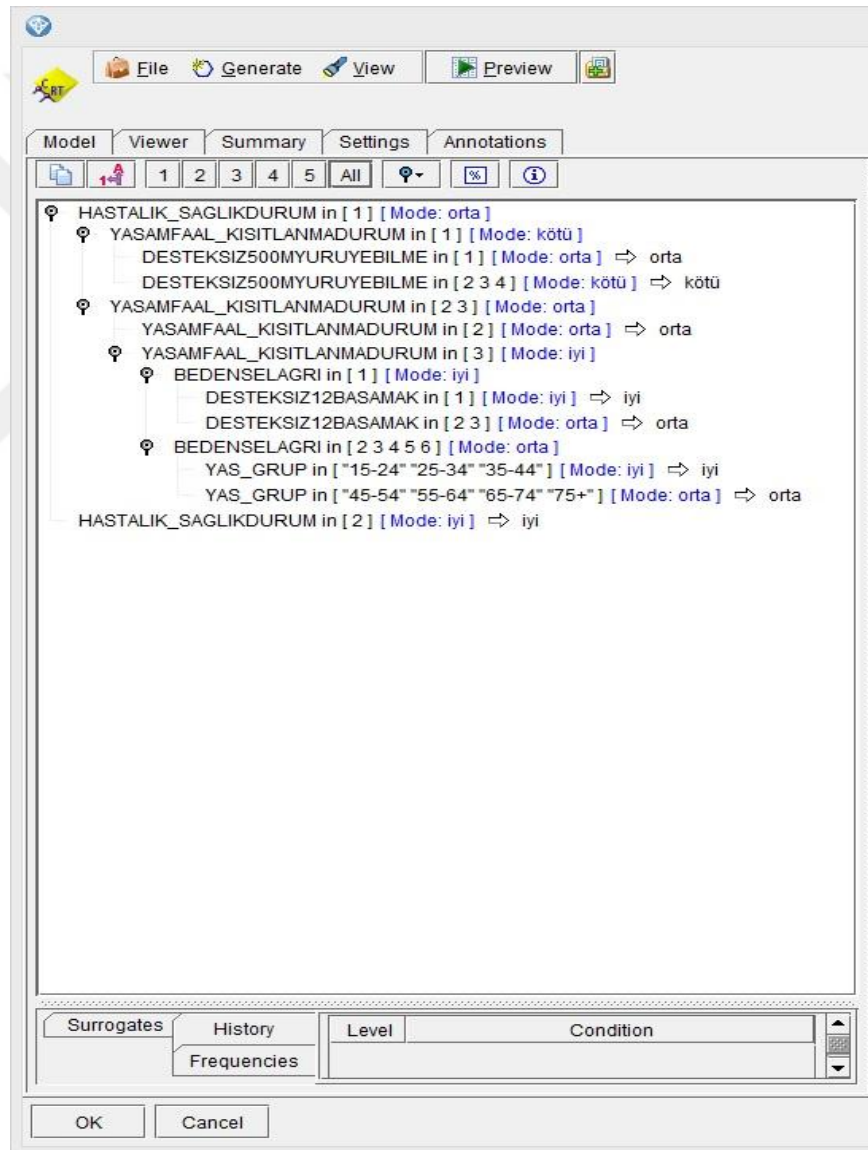
21. düğümden yer alan fertlerin % 70,991'i genel sağlık durumunu iyi olarak nitelendirirken % 2,041'i kötü olarak nitelendirmiştir. 21. düğüm, "DESTEKSIZ12BASAMAK" değişkeni ile dallanmıştır. "12 Basamak merdiven inip çıkmada zorluk çekiyor musunuz?" sorusuna verilen "hiç zorlanmıyorum" cevabı ile 39. düğüm, "biraz zorlanıyorum ve çok zorlanıyorum" cevapları ile de 40. düğüm oluşmuştur. 12 basamak inip çıkmada zorluk çekmeyen fertlerin % 75,085'i sağlık durumunu iyi olarak nitelendirirken % 1,525'i kötü olarak nitelendirmiştir. 12 basamak inip çıkmada biraz veya çok zorlanan fertlerde ise sağlık durumu iyi olarak nitelendirenlerin oranı % 45,8332'e düşmüş, kötü olarak nitelendirenlerin oranı % 5,208'e yükselmiştir.

22. düğümden yer alan fertlerin % 44,431'i genel sağlık durumunun iyi olduğunu söylerken % 7,302'si kötü olduğunu söylemiştir. Bu düğüm, "YAS_GRUP" değişkenine göre

dallanmıştır. “15-24, 25-34 ve 35-44” yaş gruplarında yer alan fertlerin % 56,989’u genel sağlık durumunun iyi olduğunu, % 3,763’ü ise kötü olduğunu söylemiştir. “45-54, 55-64, 65-74 ve 75+” yaş gruplarında yer alan fertlerden ise % 33,716’sı genel sağlık durumunun iyi olduğunu söylerken % 10,321’i kötü olduğunu söylemiştir.

C&R Tree algoritması toplam 8 tane kural üretmiştir. Algoritmanın ürettiği kural seti Şekil 4.7.3.2’de gösterilmiştir.

Şekil 4.7.3.2: C&R Tree Algoritması İle Üretilen Kural Seti



C&R Tree algoritmasıyla oluşan karar ağacında genel sağlık durumu iyi olan fertler en yüksek oranla 2. düğümde, orta olan fertler en yüksek oranla 9. düğümde ve kötü olan fertler en yüksek oranla 8. düğümde yer almaktadır. Bu düğümlere ait kurallar aşağıdaki gibi ifade edilebilir.

- Eğer bir kişinin 6 ay ya da daha uzun süren/sürmesi beklenen hastalığı/sağlık problemi yoksa genel sağlık durumu iyidir.
- Eğer bir kişinin 6 ay ya da daha uzun süren/sürmesi beklenen hastalığı/sağlık problemi varsa ve herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre ile kısıtlandı fakat ciddi ölçüde değil veya kısıtlanmadı ise ve herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre ile kısıtlandı fakat ciddi ölçüde değilse genel sağlık durumu ortadır.
- Eğer bir kişinin 6 ay ya da daha uzun süren/sürmesi beklenen hastalığı/sağlık problemi varsa ve herhangi bir sağlık problemi nedeniyle günlük faaliyetleri 6 ay ve daha uzun süre ciddi ölçüde kısıtlandıysa ve düz bir arazide baston veya diğer bir yürüme aracı kullanmadan ya da destek almadan 500 metre yürümekte biraz zorlanıyor veya çok zorlanıyor veya hiç yürüyemiyorsa genel sağlık durumu kötüdür.

C&R Tree algoritmasının uygulanması sırasında önceden bölünmüş eğitim kümesi ile model kurulmuş ve test kümesi ile de kurulan modelin güvenilirliği ve doğruluğu test edilmiştir. C&R Tree algoritmasının eğitim ve test kümesi için sınıflandırma performans ölçüm sonuçları Tablo 3.4.7.2’de verilmiştir.

Tablo 4.7.3.2: C&R Tree Algoritmasının Performans Ölçüm Sonuçları

Sınıflandırma	Eğitim Kümesi		Test Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	9753	% 72,26	4117	% 71,91
Yanlış	3651	% 26,24	1608	% 28,09
Toplam	13404	% 100	5725	% 100

Eğitim kümesinde yer alan 13404 adet veri eğitildikten sonra test kümesinde yer alan 5725 adet veri test edilerek modelin doğruluk oranı belirlenmiştir. C&R Tree algoritmasının modeli doğru tahmin etme oranı % 71,91 olduğu söylenebilir.

5. SONUÇ

Bu tez çalışmasında amaç, TÜİK tarafından yapılan TSA'ya ait B grubu mikro veri seti kullanılarak genel sağlık durumu üzerinde etkili olabilecek faktörlerin karar ağacı algoritmaları ile belirlenmesidir.

TÜİK'ten temin edilen B grubu mikro veri seti; veri temizleme, veri dönüştürme ve veri indirgeme yöntemleriyle veri madenciliği analizlerine uygun hale getirilmiştir. Daha sonra IBM SPSS Modeler programı ile modeller kurulmuştur. Kurulan modellere ait analiz sonuçları, tablolar ve şekiller halinde elde edilmiştir.

Çalışmanın amacı doğrultusunda genel sağlık durumu değişkeni, bağımlı değişken olarak belirlenmiştir. TSA'nın kapsamı, Türkiye Cumhuriyeti sınırları içinde ikamet eden hanelerde bulunan tüm fertlerdir. Ancak “genel olarak sağlık durumunuz nasıldır?” sorusu, anketin uygulanması esnasında 15 yaşından büyük fertlere sorulmuştur. Bu nedenle analize 15 yaşından büyük fertler dâhil edilmiştir.

Veri setinin ham halinde 465 değişken bulunmaktadır. IBM SPSS Modeler programında yer alan feature selection işlemcisi yardımı ile değişkenler indirgenmiştir. Analiz, 45 bağımsız ve 1 bağımlı değişken olmak üzere 46 değişken ile gerçekleştirilmiştir.

Anketin uygulanması esnasında genel sağlık durumu değişkeni, 5 kategoride alınan cevaplar ile oluşturulmuştur. Sınıflandırma performansını arttırabilmek amacıyla 5 kategoriden oluşan bağımlı değişken 3 kategoriye dönüştürülmüştür. Bağımlı değişken üzerinde uygulanan dönüşüm sayesinde;

- CHAID algoritmasının modeli doğru tahmin etme oranı, % 60,761'den % 72,751'e,
- Quest algoritmasının modeli doğru etme oranı, % 60,082'den 72,541'e,

- C&R Tree algoritmasının modeli doğru etme oranı ise % 55,486'dan % 71,913'e yükselmiştir.

Belirlenen bağımlı değişken ile çok sayıdaki bağımsız değişken arasındaki ilişkileri tespit etmek için IBM SPSS Modeler programında yer alan Auto Classifier işlemcisi kullanılmıştır. Karar ağacı algoritmaları arasında otomatik karşılaştırma yapabilen bu işlemci, veri setine uygun ve en yüksek modeli tahmin etme oranına sahip algoritmaların CHAID, QUEST ve C&R Tree algoritmaları olduğu sonucuna varmıştır.

CHAID algoritması, 13404 fert ve 17 değişken ile derinliği 5 olan karar ağacı oluşturmuştur. Genel sağlık durumunu etkileyen en önemli 3 değişkeni;

- 0,2601 önem değeri ile HASTALIK_SAGLIKDURUM,
- 0,2003 önem değeri ile YASAMSALFAAL_KISITLANMADURUM,
- 0,1713 önem değeri ile BEDENSELAGRI olarak tespit etmiştir.

QUEST algoritması, 9497 fert ve 6 değişken ile derinliği 4 olan karar ağacı oluşturmuştur. Genel sağlık durumunu etkileyen en önemli 3 değişkeni;

- 0,4825 önem değeri ile HASTALIK_SAGLIKDURUM,
- 0,1746 önem değeri ile YASAMSALFAAL_KISITLANMADURUM,
- 0,0414 önem değeri ile DESTEKSIZ12BASAMAK olarak tespit etmiştir.

C&R Tree algoritması ise 9325 fert ve 7 değişken ile derinliği 5 olan karar ağacı oluşturmuştur. Genel sağlık durumunu etkileyen en önemli 3 değişkeni;

- 0,4281 önem değeri ile HASTALIK_SAGLIKDURUM,
- 0,152 önem değeri ile YASAMSALFAAL_KISITLANMADURUM,
- 0,0453 önem değeri ile DESTEKSIZ500MYURUYEBILME olarak tespit etmiştir.

Daha açık bir ifade ile analiz için kullanılan 3 algoritma, 6 ay ya da daha uzun süren veya sürmesi beklenen sağlık problemi faktörünün, kişinin genel sağlık durumu etkileyen en önemli faktör olduğunu belirlemiştir. Bireyler, 6 ay ya da daha uzun süren veya sürmesi beklenen sağlık problemi varsa genel sağlık durumunu kötü, yoksa genel sağlık durumunu iyi olarak nitelendirmiştir.

Analiz için kullanılan algoritmalar, herhangi bir sağlık problemi nedeniyle, günlük faaliyetlerin 6 ay ve daha uzun süre kısıtlanma durumunun da kişinin genel sağlık durumunu en çok etkileyen ikinci faktör olduğunu tespit etmiştir. Bireyler, herhangi bir sağlık problemi nedeniyle, günlük faaliyetlerinde 6 ay ve daha uzun süre kısıtlanma yaşamadıysa genel sağlık durumlarını iyi, aksi bir durumda ise genel sağlık durumlarını kötü olarak nitelendirmiştir.

Modeli doğru tahmin etme oranları incelendiğinde kullanılan karar ağacı algoritmalarının birbirlerine çok yakın performans gösterdiği söylenebilir. Analiz sonuçlarına göre en yüksek modeli doğru tahmin etme oranına sahip olan CHAID algoritmasını sırasıyla QUEST ve C&R Tree algoritmaları takip etmiştir. Ancak CHAID algoritması, karar ağacını diğer algoritmalarla göre daha fazla veri ve değişken ile oluşturmuştur. Ayrıca CHAID algoritması ile kurulan karar ağacı, diğer algoritmaların kurduğu ağaçlara göre sadece ikili dallanma yapmamıştır. Veri yapısına göre üçlü, dördü hatta beşli dallanmalar yapan CHAID algoritması, daha fazla alt gruplar oluşturarak ayrıntılı değerlendirmeler yapmaya imkân tanımıştır.

Sonuç olarak maksimum değişken ve fert ile karar ağacı oluşturması, fertleri çok sayıda kural ile sınıflandırması, diğer algoritmalarla göre daha fazla alt sınıflar oluşturarak homojen düğümlerde değerlendirme yapmaya imkân tanınması gibi güçlü yönleriyle CHAID algoritmasının daha verimli sınıflama yaptığı söylenebilir.

KAYNAKLAR

- Akgöbek, Ö. ve Çakır, F. (2009). “Veri Madenciliğinde Bir Uzman Sistem Tasarımı” [Bildiri]. *Akademik Bilişim’09*, 11-13 Şubat 2009, Harran Üniversitesi, Şanlıurfa.
- Akgöbek, Ö. ve Öztemel, E. (2006). “Endüktif Öğrenme Algoritmalarının Kural Öğretme Yöntemleri ve Performanslarının Karşılaştırılması”. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 10(1): 1-9.
- Akın, H. B. ve Şentürk, E. (2012). “Bireylerin Mutluluk Düzeylerinin Ordinal Lojistik Regresyon Analizi ile İncelenmesi”. *Öneri Dergisi*, 10(37): 183-193.
- Akın, Y. K. (2008). *Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi*. Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, İstanbul.
- Akkaya, G. C., Demireli, E. ve Yakut, Ü. H. (2009). “İşletmelerde Finansal Başarısızlık Tahminlemesi: Yapay Sinir Ağları Modeli ile İMKB Üzerine Bir Uygulama”. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 10(2): 187-216.
- Akman, M. (2010). *Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama*. Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Akman, M., Genç, Y. ve Ankaralı, H. (2011). “Random Forest Yöntemi ve Sağlık Alanında Bir Uygulama”. *Türkiye Klinikleri Journal of Biostatistics*, 3(1): 36-48.
- Akpınar, H. (2014). *DATA Veri Madenciliği Veri Analizi*. İstanbul: Papatya Yayıncılık Eğitim.
- Aksoy, E. (2014). *Matematik Alanında Üstün Yetenekli ve Zekâlı Öğrencilerin Bazı Değişkenler Açısından Veri Madenciliği ile Belirlenmesi*. Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü, Yüksek Lisans Tezi, İzmir.
- Aktürk, H. (2008). *Borsa ve Döviz Verileri Üzerinde Veri Madenciliği Teknolojisini Kullanarak Zarar Riskini Azaltan Bir Uygulama Geliştirimi*. Ege Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İzmir.

- Albayrak, A. S. ve Yılmaz, K. (2009). “Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama”. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1): 31-52.
- Anane, R. (2001). “Data Mining and Serial Documents”. *Computers and the Humanities*, 35(3): 299-314.
- Argüden, Y., Erşahin, B. (2008). “*Veri Madenciliği: Veriden Bilgiye, Masraftan Değere*”. İstanbul: ARGE Danışmanlık Yayınları.
- Asilkan, Ö. (2008). *Veri Madenciliği Kullanılarak İkinci El Otomobil Pazarında Fiyat Tahmini*. Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, Antalya.
- Aşlıyan, R. ve Günel, K. (2010). “Metin İçerikli Dokümanların Sınıflandırılması” [Bildiri]. *Akademik Bilişim’10 – Akademik Bilişim Konferansı Bildirileri*, 10–12 Şubat 2010, Muğla Üniversitesi, Muğla.
- Atan, M. ve Çatalbaş, E. (2004). “Çok Değişkenli Analiz Yöntemleri ile Türk Bankacılık Sektöründe Çok Boyutlu Mali Başarısızlık Tahmin Modelleri Oluşturulması” [Bildiri]. 4. *İstatistik Günleri Sempozyumu*, 21-22 Mayıs 2004, Dokuz Eylül Üniversitesi, İzmir.
- Atasoy, Y. (2015). *Veri Madenciliği Yöntemleri ile Ankilozan Spondilit Hastalığında Radyografik Progresyona Etkili Faktörlerin Analizi*. İstanbul Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Aydın, İ. (2006). *Arıza Teşhisinde Veri Madenciliği ve Yumuşak Hesaplama Teknikleri*. Fırat Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Elazığ.
- Aydın, S. (2007). *Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama*. Eskişehir Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Eskişehir.
- Ayık, Y. Z., Özdemir, A. ve Yavuz, U. (2007). “Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte ile İlişkisinin Veri Madenciliği Tekniği İle Analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2): 441-454.

- Aytaç, E. B. (2014). *ATM’lerde Bulundurulacak Günlük Para Miktarının Veri Madenciliği Teknikleri Kullanılarak Optimize Edilmesi*. Maltepe Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Aytaç, M.B. ve Bilge, H. Ş. (2013). “Tele Pazarlama Verilerinin Birliktelik Kurallarıyla ve CRISP–DM Yöntemiyle Analiz Edilmesi”. *Aksaray Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 5(2): 25-39.
- Ballou, D. P. ve Tayı, G. K. (1999). “Enhancing Data Quality in Data Warehouse Environments”. *Communications of the ACM*, 42(1): 73-78.
- Başar, A. ve Oktay, E. (2014). *Uygulamalı İstatistik 2 Kısa Teorik Bilgiler ve Çözülmüş Problemler*. Erzurum: Erzurum Kültür ve Eğitim Vakfı.
- Baykal, A. (2006). “Veri Madenciliği Uygulama Alanları”. *D.Ü. Ziya Gökalp Eğitim Fakültesi Dergisi*, 7: 95-107.
- Baykasoğlu, A. (2005). “Veri Madenciliği ve Çimento Sektöründe Bir Uygulama”. *Akademik Bilişim*, 2-4.
- Bilgin, Ş. (2013). *Banka Müşterilerinin İnternet Bankacılığına İlişkin Yaklaşımlarının Veri Madenciliği Teknikleri ile İncelenmesi*. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Birant, D., Kut, A., Ventura, M., Altınok, H., Altınok, B., Altınok, E., ve İhlamur, M. (2010). “İş Zekası Çözümleri için Çok Boyutlu Birliktelik Kuralları Analizi”, [Elektronik Sürüm]. *Akademik Bilişim*, 10: 256-262.
- Bolat, B. (2006). *Asansör Kontrol Sistemlerinin Genetik Algoritma ile Simülasyonu*. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, İstanbul.
- Büyüköztürk, Ş. (2015). *Sosyal Bilimler İçin Veri Analizi El Kitabı, İstatistik, Araştırma Deseni SPSS Uygulamaları ve Yorum*. Ankara: Pegem Akademi Yayıncılık.
- Canbazoğlu, T. (2000). “Yararlı Bilgi Yönetiminde İnsan Faktörü”. Türkiye Bilişim Vakfı Seminerleri, 2000, İstanbul.
- Chen, Q., Wu, X. ve Zhu, X. (2004). “OIDM: Online Interactive Data Mining”. *Innovations in Applied Artificial Intelligence*, 66-76.
- Collier, K., Carey, B., Grusy, E., Marjaniemi, C. ve Sautter, D. (1998). “A Perspective on Data Mining”. *Centre for Data Insight*, 2-4.

- Coşkun, S., Kartal, M. Coşkun, A ve Bircan, H. (2004). “Lojistik Regresyon Analizinin İncelenmesi ve Dişhekimliğinde Bir Uygulaması”. *Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi*, 7(1): 41-50.
- Çakar, T., Türker, A. K. ve Toraman, A. (1996). “İmalat Sistemlerinin Tasarlanmasında Yapay Sinirsel Ağların Kullanılması” [Bildiri]. *Birinci Ulusal Zeki İmalat Sistemleri Sempozyumu*, Sakarya.
- Çalış, A. (2013). *Veri Madenciliği Yaklaşımı ile Bireysel Müşterilerin Kredi Ödeme Performanslarının Değerlendirilmesi*. Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kocaeli.
- Çalış, K., Gazdağı, O. ve Yıldız, O. (2013). “Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti”. *Bilişim Teknolojileri Dergisi*, 6(1): 1-7.
- Çalışkan, S. K. ve Soğukpınar, İ. (2008). “KxKNN: K-Means ve K En yakın Komşu Yöntemleri ile Ağlarda Nüfuz Tespiti”. *EMO Yayınları*, 120-124.
- Çelik, M. K. (2010). “Bankaların Finansal Başarısızlıklarının Geleneksel ve Yeni Yöntemlerle Öngörüsü”. *Celal Bayar Üniversitesi Yönetim ve Ekonomi Dergisi*, 17(2): 129-143.
- Çoban, A. (2006). *İmalat Sanayinde Veri Madenciliği Destekli Tedarikçi Seçimi Uygulaması*. Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Sakarya.
- Daş, B. ve Türkoğlu, İ. (2014). “DNA Dizilimlerindeki Nükleotit Çiftlerinin Frekans Değerlerine Göre Farklı Sınıflandırma Yöntemleri ile Karşılaştırılması” [Bildiri]. *Tıp Teknolojileri Ulusal Kongresi (TIPTEKNO'2014)*, 25-27 Eylül 2014, Perrissia Hotel & Convention Center, Nevşehir.
- Deconinck, E., Hancock, T., Coomans, D., Massart, D. L. ve Vander Heyden, Y. (2005). Classification of Drugs in Absorption Classes Using the Classification and Regression Trees (CART) Methodology. *Journal of Pharmaceutical and Biomedical Analysis*, 39(1): 91-103.
- Demiralay, M. ve Çamurcu, A. Y. (2005). “Cure, Agne ve K-Means Algoritmalarındaki Kümeleme Yeteneklerinin Karşılaştırılması”. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 8: 1-18.

- Dinçer, E. (2006). *Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması*. Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kocaeli.
- Doğan, S. ve Diri, B. (2010). Türkçe Dokümanlar İçin N-gram Tabanlı Yeni Bir Sınıflandırma (Ng-ind): Yazar, Tür ve Cinsiyet. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 3(1): 11-19.
- Dolgun, M. Ö. (2006). *Büyük Alışveriş Merkezleri İçin Veri Madenciliği Uygulamaları*. Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Dolgun, M. Ö. (2014). *Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının; Bağımlı Değişken Prevelansı, Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası İlişki Yapısına Göre Karşılaştırılması*. Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, Doktora Tezi, Ankara.
- Döşlü, A. (2008). *Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi*. İstanbul Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Dunham, M. H. (2006). *Data Mining: Introductory and Advanced Topics*. India: Pearson Education.
- Düzgünoğlu, S. (2006). *Veri Ambarı ve OLAP Teknolojilerinden Yararlanılarak Karar Destek Amaçlı Raporlama Aracı Gerçekleştirimi*. Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Efe, Ö. ve Kaynak, O. (2004). *Yapay Sinir Ağları ve Uygulamaları*. İstanbul: Boğaziçi Üniversitesi Yayınları.
- Elbiad, Z. (2013). *Web Tabanlı Anket Sistemi ile Verilerin Veri Madenciliği Yöntemi ile Analizi*. İstanbul Aydın Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Elmas, Ç. (2003). *Yapay Sinir Ağları: (Kuram, Mimari, Eğitim, Uygulama)*. Ankara: Seçkin Yayıncılık.
- Elmas, F. (2014). *Kalp Krizi Riskinin Bir Veri Madenciliği Uygulaması ile Analizi*. Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Muğla.

- Emek, G. G. ve Taşkın, Ç. (2002). “Genetik Algoritmalar ve Uygulama Alanları”. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 21(1): 129-152.
- Ergün, C. (2012). “ÜNAK’09 Bilgi Çağında Varoluş: Fırsatlar ve Tehditler Sempozyumu” [Bildiri]. Yıldızeli, A., Arıkan, A ve Çakmak, T. (Ed.). *Bildiriler Kitabı, 01-02 Ekim 2009*, (ss. 01-245). İstanbul: Yeditepe Üniversitesi.
- Erol, B. (2013). *Müşteri İlişkileri Yönetimi İçin Veri Madenciliği Kullanılması ve Sigortacılık Üzerine Bir Uygulama*. Marmara Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Filiz, Z. (2005). “İllerin Sosyo-Ekonomik Gelişmişlik Düzeylerine Göre Gruplandırılmasında Farklı Yaklaşımlar”. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(1): 77-100.
- Giudici, P. (2005). *Applied Data Mining: Statistical Methods for Business and Industry*. England: John Wiley & Sons.
- Govindarajan, M. (2007). “Text Mining Technique for Data Mining Application”. *Engineering and Technology International Journal of Computer*. 1(11): 3460-3465.
- Gökgöz, K. B. (2015). *Veri Madenciliği ile Tıbbi Cihaz Bakım Karar Modeli*. Başkent Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Gökmen, Ş. (2014). *Müşteri İlişkileri Yönetiminde Bir Araç Olarak Veri Madenciliği ve Perakende Sektöründe Bir Uygulama*. Yıldız Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Göral, M. A. (2007). *Kredi Kartı Başvuru Aşamasında Sahtecilik Tespiti İçin Bir Veri Madenciliği Modeli*. İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Gülçe, G. (2010). *Veri Ambarı ve Veri Madenciliği Teknikleri Kullanılarak Öğrenci Karar Destek Sistemi Oluşturma*. Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Denizli.

- Güner, Z. B. (2014). “Veri Madenciliğinde CART ve Lojistik Regresyon Analizinin Yeri: İlaç Provizyon Sistemi Verileri Üzerinde Örnek Bir Uygulama”. *Sosyal Güvençe Dergisi*, 6: 53-99.
- Güngör, M. ve Bulut, Y. (2008). Ki-Kare Üzerine, *Doğu Anadolu Bölgesi Araştırmaları*, 88-89.
- Gürbüz, F. (2009). *Havacılık Sektöründe Veri Madenciliği İle Farklı Sınıflandırma Tekniklerinin Karşılaştırmalı Olarak Uygulanması*. Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Kayseri.
- Gürsoy Şimşek, U. T. (2009). *Veri Madenciliği ve Bilgi Keşfi*. Ankara: Pegem Akademi Yayıncılık.
- Han, J. ve Kamber, M. (2006). *Data Mining Concepts & Techniques*. USA: Elsevier.
- Hand, D. J. (1998). “Data Mining: Statistics and More?”. *The American Statistician*, 52(2): 112-118.
- Hand, J., Pei, J. ve Kamber, M. (2011). *Data Mining: Concepts and Techniques*. USA: Elsevier.
- Işığçık, E. (2003). “Bebeklerin Doğum Ağırlıklarını ve Boylarını Etkileyen Faktörlerin Lojistik Regresyon Analizi ile Karşılaştırılması” [Bildiri]. VI. *Ulusal Ekonometri ve İstatistik Sempozyumu*, Gazi Üniversitesi, Ankara.
- Işık, S., Duman, K. ve Korkmaz, A. (2004). “Türkiye Ekonomisinde Finansal Krizler: Bir Faktör Analizi Uygulaması”. *D.E.Ü. İ.İ.B.F Dergisi*, 19(1): 45-69.
- İslamoğlu, A. H. ve Alnıaçık, Ü. (2016). *Sosyal Bilimlerde Araştırma Yöntemleri*. İstanbul: Beta Basım Yayım.
- Jacobs, P. (1999). “Data Mining: What General Managers Need to Know”. *Harvard Management Update*, 4(10): 8.
- Karaca, E. (2010). *Uçak Kaza ve Olaylarının Veri Madenciliği Yöntemiyle Analiz Edilmesi*. Hava Harp Okulu Havacılık ve Uzay Teknolojileri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Karaibrahimoğlu, İ. (2014). *Veri Madenciliğinden Birliktelik Kuralı İle Onkoloji Verilerinin Analiz Edilmesi: Meram Tıp Fakültesi Onkoloji Örneği*. Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Konya.

- Kartum, G. (2014). *Kurumsal Kaynak Yönetiminde, Veri Madenciliği, Müşteri İlişkileri Yönetimi ve Birebir Pazarlama Stratejisi Durum Analizi; Netsis Yazılım Memnuniyet Araştırması*. Gediz Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İzmir.
- Kavzaoglu, T., Şahin, E. K. ve Çölkesen, İ. (2012). “Heyelan Duyarlılığının İncelenmesinde Regresyon Ağaçlarının Kullanımı: Trabzon Örneği”. *Harita Dergisi*, 147: 21-33.
- Khoonsari, P. E. ve Motie, A. (2012). “A Comparasion of Efficiency and Robustness of ID3 and C4.5 Algorithms Using Dynamic Test and Training Data Sets”. *International Journal of Machine Learning and Computing*. 2(5): 540.
- Kılınç, Ç. (2015). *Üniversite Öğrenci Başarısı Üzerine Etki Eden Faktörlerin Veri Madenciliği Yöntemleri İle İncelenmesi*. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Eskişehir.
- Kiremitçi, B. (2005). *Veri Ambarlarında Veri Madenciliği ve Ulaştırma – Lojistik Sektöründe Bir Uygulama*. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Koyuncugil, A. S. (2006). *Bulanık Veri Madenciliği ve Sermaye Piyasalarına Uygulanması*. Ankara Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Ankara.
- Koyuncugil, A. S. (2007). “Borsa Şirketlerinin Sektörel Risk Profillerinin Veri Madenciliğiyle Belirlenmesi”. *Sermaye Piyasası Kurulu Araştırma Raporu*, Ankara: Araştırma Dairesi.
- Kulluk, S., Özbakır, L. ve Baykasoğlu, A. (2011). “Yapay Sinir Ağlarının Eşzamanlı Eğitimi ve Kural Çıkarımı: Difaconn-Miner” [Bildiri]. *Endüstri Mühendisliği Yazılımları ve Uygulama Kongresi*, 30 Eylül-01.02 Ekim 2011, Tepekule Kongre ve Sergi Merkezi, İzmir.
- Kuzey, C. (2012). *Veri Madenciliğinde Destek Vektör Makinaları ve Karar Ağaçları Yöntemlerini Kullanarak Bilgi Çalışanlarının Kurum Performansı Üzerine Etkisinin Ölçülmesi ve Bir Uygulama*. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, İstanbul.

- Larose D. T. ve Larose D. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons.
- Liu, P. ve Lei, L. (2005). "A Review of Missing Data Treatment Methods". *International Journal on Intelling Information Systems and Tech*, 1(3): 412-419.
- Mackinnon, M. J. ve Ned, G. (1999). "Applications: Data Mining and Knowledge Discovery in Databases - An Overview". *Austrilian & New Zealand Journal of Statistics*, 41(3): 255-275.
- Murat, Y. Ş. ve Şekerler, A. (2009). "Trafik Kaza Verilerinin Kümeleme Analizi Yöntemi ile Modellenmesi". *İMO Teknik Dergi*, 4759-4777.
- Oğuzlar, A. (2004). "CART Analizi İle Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi". *İktisadi ve İdari Bilimler Dergisi*, 18(3-4): 79-90.
- Olson, D. L. ve Delen, D. (2008). *Advanced Data Mining Techniques*, Berlin: Springer Science & Business Media.
- Onan, A. (2015). "Şirket İflaslarının Tahmin Edilmesinde Karar Ağacı Algoritmalarının Karşılaştırmalı Başarım Analizi". *Bilişim Teknolojileri Dergisi*, 8(1): 9-19.
- Öğüt, S. (2005). "Veri Madenciliği Kavramı ve Gelişim Süreci". Veri Madenciliği Paneli, 2005, İstanbul.
- Öz, B. (2005). *Türkiye’de Ticari Bankaların Başarısızlığında Etkili Olan Faktörlerin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi*. İstanbul Üniversitesi Sosyal Bilimler Entitüsü, Doktora Tezi, İstanbul
- Özbay, E. (2007). *Finans Sektöründe Veri Madenciliği ile Dolandırıcılık Tespiti*. Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kayseri.
- Özdamar, K. (2012). *Paket Programlar ile İstatistiksel Veri Analizi*. Eskişehir: Kaan Kitapevi.
- Özek, M. B., Daş, B. ve Akpolat, Z. H. (2010). "Zaman Serisi Analizinde Tip-2 Bulanık Mantık Tabanlı Veri Madenciliği Uygulaması". *Technological Applied Sciences*, 5(2). 100-109.
- Özekeş, S. (2003). "Veri Madenciliği Modelleri ve Uygulama Alanları". *İstanbul Ticaret Üniversitesi Dergisi*, 2(3): 65-82.

- Özer, M. A. (2013). “Bilgi Yönetimi Aracılığıyla Bilginin Etkin Kullanımı”. *Mukaddime*, (7): 69-99.
- Özmen, Ş. (2001). “İş Hayatı ve Veri Madenciliği ile İstatistik Uygulamalarını Yeniden Keşfediyor”. V. Ulusal Ekonometri ve İstatistik Sempozyumu, 2001, Adana.
- Pala, T. (2013). *Tıbbi Karar Destek Sistemlerinin Veri Madenciliği Yöntemleriyle Gerçekleştirilmesi*. Marmara Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Pektaş, A. O. (2013). *SPSS ile Veri Madenciliği*. İstanbul: Ecem Basım Yayın.
- Refaat, M. (2007). *Data Preparation for Data Mining Using SAS*. USA: Morgan Kaufmann.
- Savaş, S., Topaloğlu, N. ve Yılmaz, M. (2012). “Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri”. *İstanbul Üniversitesi Fen Bilimleri Dergisi*, 21: 1-23.
- Sever, H. ve Oğuz, B. (2003). “Veri Tabanlarında Bilgi Keşfine Formel Bir Yaklaşım, Kısım I: Eşleştirme Sorguları ve Algoritmalar”. *Bilgi Dünyası*, 3(2): 173-204.
- Seyrek, İ. H. ve Ata, H. A. (2010). “Veri Zarflama Analizi ve Veri Madenciliği ile Mevduat Bankalarında Etkinlik Ölçümü”. *BDDK Bankacılık ve Finansal Piyasalar*, 4(2): 67-84.
- Sezgin, E. ve Çelik, Y. (2013). “Veri Madenciliğinde Kayıp Veriler İçin Kullanılan Yöntemlerin Karşılaştırılması” [Bildiri]. *Akademik Bilişim Konferansı*, 23-25 Ocak 2013, Akdeniz Üniversitesi, Antalya.
- Silahtaroglu, G. (2004). *Veri Madenciliğinde Kümeleme Analizi ve Öğretim Başarısının Değerlemesine İlişkin Bir Uygulama*. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, İstanbul.
- Silahtaroglu, G. (2008). *Kavram ve Algoritmalarıyla Temel Veri Madenciliği*. İstanbul: Papatya Yayıncılık.
- SPSS. (1999). *AnswerTree Algorithm Summary*. U.S.A: SPSS Inc.
- SPSS. (2002). *AnswerTree 3.1 User’s Guide*, U.S.A: SPSS Inc.
- Sumati, S. ve Sivanandam S. N. (2006). *Introduction to Data Mining and its Applications*. New York: Springer.

- Sun, J. ve Li, H. (2008). "Data Mining Method for Listen Companies Financial Distress Prediction". *Konowledge-Based Systems*, 21(1): 1-5.
- Şeker, S. E. (2015). "Zaman Serisi Analizi (Time Series Analysis)". *YBS Ansiklopedi*, 2(4): 23-31.
- Şeker, S.E. (2015). "Veri Ambarı (Data Warehouse)". *YBS Ansiklopedi*, 2(4): 6-13.
- Şengür, D. (2013). *Öğrencilerin Akademik Başarılarının Veri Madenciliği Metotları ile Tahmini*. Fırat Üniversitesi Eğitim Bilimleri Enstitüsü, Yüksek Lisans Tezi, Elazığ.
- Şık, M. Ş. (2014). *Veri Madenciliği ve Kansere Erken Teşhisinde Kullanımı*. İnönü Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Malatya.
- Tantuğ, A. C. (2002). *Veri Madenciliği ve Demetleme*. İstanbul Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Tapkan, P., Özbakır, L. ve Baykasoğlu, A. (2011). WEKA ile Veri Madenciliği Süreci ve Örnek Uygulama. *Endüstri Mühendisliği Yazılımları ve Uygulama Kongresi*, 30 Eylül-01/02 Ekim 2011, İzmir.
- Tatlıdil, H. (1992). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Ankara: Engin Yayınları.
- Tektaş, M., Akbaş, A. ve Topuz, V. (2002). "Yapay Zeka Tekniklerinin Trafik Kontrolünde Kullanılması Üzerine Bir İnceleme" [Bildiri]. *Uluslararası Trafik ve Yol Güvenliği Kongresi*, Gazi Üniversitesi, Ankara.
- Tiryaki, S. (2006). *Lojistik Alanında Bir Veri Madenciliği Uygulaması*. İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Ture, M., Tokatlı, F. ve Kurt, I. (2009). "Using Kaplan-Meier Analysis Together with Decision Tree Methods (C&RT, CHAID, QUEST, C4.5 and ID3) in Determining Recurrence-Free Survival of Breast Cancer Patient". *Expert Systems with Applications*, 36(2): 2017-2026.
- Turna, F. (2011). *Veri Madenciliği Teknikleriyle Tramvay Arıza Kayıtlarından Kural Çıkarımı*. Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kayseri.
- TÜİK. (2014). *İstatistiksel Veri Analizi*. Ankara: Türkiye İstatistik Kurumu Matbaası.

- TÜİK. (2014). Resmi İstatistikle Dersi *Sağlık İstatistikleri*. Ankara: Türkiye İstatistik Kurumu Matbaası.
- Tüzüntürk, S. (2010). “Veri Madenciliği ve İstatistik”. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19(1): 65-90.
- Ukuş, S. G. (2014). *Veri Madenciliğinin Satış Tahminleri Açısından Önemi ve Bir Araştırma*. Galatasaray Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Uzar, C. (2013). *Finansal Bilgi Sisteminde Veri Madenciliği Teknolojisinin Kullanılması: Borsa İstanbul Üzerine Bir Uygulama*. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İzmir.
- Uzun, E. (2007). *İnternet Tabanlı Bilgi Erişimi Destekli Bir Otomatik Öğrenme Sistemi*. Trakya Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Edirne.
- Ünsal, A. (2000). “Diskriminant Analizi ve Uygulaması Üzerine Bir Örnek”. *Gazi Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, 3: 19-36.
- Ünveren, E. ve Doğan, H. (2010). “Beş Yıldızlı Konaklama İşletmelerinde Çalışanların İş Tatmin Düzeylerinin CHAID Analiz Yöntemiyle Değerlendirilmesi”. *C.Ü. İktisadi ve İdari Bilimler Dergisi*, 11(2): 39-52.
- Vatansever, M. (2008). *Görsel Veri Madenciliği Tekniklerinin Kümeleme Analizlerinde Kullanımı ve Uygulanması*. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2008.
- Velickov, S. ve Solomatine, D. (2000). “Predictive Data Mining: Practical Examples”. *2nd Joint Workshop on Applied AI in Civil Engineering*, Germany.
- Web_1: <http://en.wikipedia.org/wiki/Data>, Erişim Tarihi: 04 Ağustos 2016.
- Web_2: <https://www.ibm.com/us-en/marketplace/learning-lab/courses/coursera-data-mining>, Erişim Tarihi: 09 Ağustos 2016.
- Web_3: http://www.sas.com/en_us/software/enterprise-miner.html, Erişim Tarihi: 09 Ağustos 2016.
- Web_4: <http://crisp-dm.eu/reference-model/>, Erişim Tarihi: 26 Aralık 2016
- Web_5: <https://tr.wikipedia.org/wiki/SPSS>, Erişim tarihi: 12 Aralık 2016

Web_6: <http://www.tuik.gov.tr/PreHaberBultenleri.do?id=18854>, Eriřim tarihi: 02 Mart 2017

- Yakut, E. (2012). *Veri Madencilięi Tekniklerinden C5.0 Algoritması ve Destek Vektör Makineleri İle Yapay Sinir Ağlarının Sınıflandırma Başarılarının Karşılaştırılması: İmalat Sektöründe Bir Uygulama*, Doktora Tezi, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Erzurum.
- Yener, F. (2014). *Veri Madencilięi ve Optimizasyon Teknikleri Kullanılarak Bir Depo Tasarımı: Perakende Sektöründe Uygulaması*. Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Sakarya.
- Yıldırım, M. (2016). *İldeki Kurumlar Arası Çalışma Performansının Arttırılmasında Veri Madencilięi Tekniklerinin Kullanılması*. Fırat Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Elazığ.
- Yurdakul, S. (2015). *Veri Madencilięi ile Lise Öğrencilerinin Performanslarının Deęerlendirilmesi*. Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kırıkkale.
- Zaki, M. J. (1999). "Parallel and Distributed Association Mining: A Survey". IEEE Concurrency, 7(4): 14-25.
- Zhu, H. (1998). *On-Line Analytical Mining of Association Rules*. Simon Fraser University, Doktora Tezi, Kanada.

EKLER

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu

Sütun adı	Soru
YAS_GRUP	
CINSIYET	3. Sayın \$AD_SOYAD\$ cinsiyetiniz nedir?
DOGUM_YER	6. Sayın \$AD_SOYAD\$ nerede doğdunuz?
VATANDASLIK	7. Sayın \$AD_SOYAD\$ hangi ülke vatandaşısınız?
TEDAVI_MASRAF_SGK	9.1. Sayın \$AD_SOYAD\$ tedavi masraflarınız SGK tarafından mı karşılanıyor?
TEDAVI_MASRAF_OZELSAGSIG	9.2. Sayın \$AD_SOYAD\$ tedavi masraflarınız OZEL SAGLIK SIGORTASI tarafından mı karşılanıyor?
TEDAVI_MASRAF_OZELSANDIK	9.3. Sayın \$AD_SOYAD\$ tedavi masraflarınız OZEL SANDIK tarafından mı karşılanıyor?
TEDAVI_MASRAF_GSS	9.4. Sayın \$AD_SOYAD\$ tedavi masraflarınız GSS tarafından mı karşılanıyor?
TEDAVI_MASRAF_YESILKART	9.5. Sayın \$AD_SOYAD\$ tedavi masraflarınız YESILKART tarafından mı karşılanıyor?
TEDAVI_MASRAF_KENDI	9.6. Sayın \$AD_SOYAD\$ tedavi masraflarınızı KENDİNİZ mi karşılıyorsunuz?
TEDAVI_MASRAF_SAGSIGORTAYOK	9.7. Sayın \$AD_SOYAD\$ tedavi masraflarınız için sağlık sigortanız yok mu?
TEDAVI_MASRAF_DIGER	9.98. Sayın \$AD_SOYAD\$ tedavi masraflarınız diğer kaynaklardan mı karşılanıyor?
EGITIM	10. Sayın \$AD_SOYAD\$ en son hangi okulu bitirdiniz?
MEDENI_DURUM	15. Sayın \$AD_SOYAD\$ medeni durumunuz nedir?
CALISMA_DURUM	16. Sayın \$AD_SOYAD\$ son 1 hafta içerisinde aynı(mal) veya nakdi(para) bir gelir elde etmek amacıyla(ev kadını,öğrenci veya emekli de olsanız) bir saat bile olsa bir işte çalıştınız mı?
CALISMAMA_NEDEN	17. Sayın \$AD_SOYAD\$ çalışmama nedeniniz nedir?
ONCEDEN_CALISMA_DURUM	18. Sayın \$AD_SOYAD\$ daha önce hiç aynı(mal) ya da nakdi(para) gelir elde etmek amacıyla bir işte çalıştınız mı?
ISTEKI_DURUM	19. Sayın \$AD_SOYAD\$ şu anda veya en son çalıştığınız işyerindeki işteki durumunuz nedir/neydi?
IS_SUREKLILIK	20. Sayın \$AD_SOYAD\$ şu anda veya en son çalıştığınız bu işin süreklilik durumu nedir?
CALISMA_SEKLI	21. Sayın \$AD_SOYAD\$ bu işinizde çalışma şekliniz nedir/neydi?
ISCO_08	23. Sayın \$AD_SOYAD\$ işyerinizdeki yaptığınız iş(görev ve sorumluluğunuz) nedir/neydi?
NACE_REV2	24. Sayın \$AD_SOYAD\$ işyerinin ana faaliyeti?
GELIR_KAYNAK_UCRET	25.1 Sayın \$AD_SOYAD\$ ücret, maaş veya yevmiye geliriniz var mı?
GELIR_KAYNAK_MUTESEBBIS	25.2. Sayın \$AD_SOYAD\$ müteşebbis geliriniz var mı?
GELIR_KAYNAK_GAYRIMENK	25.3. Sayın \$AD_SOYAD\$ gayrimenkul kira geliriniz(ev, tarla vb. kira geliri) var mı?

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu (Devamı)

Sütun adı	Soru
GELIR_KAYNAK_MENKUL	25.4. Sayın \$AD_SOYAD\$ menkul kıymet faiz geliniz(banka faizi, nema, hisse senedi faizi vb.) var mı?
GELIR_KAYNAK_EMEKLIMAAS	25.51. Sayın \$AD_SOYAD\$ emekli ve dul yetim aylığınız var mı?
GELIR_KAYNAK_DIGERSOSYAL_TRNS	25.52. Sayın \$AD_SOYAD\$ diğer sosyal transferler(sosyal yardımlar, işsizlik geliri, karşılıksız burs vb.) var mı?
GELIR_KAYNAK__HANE_TRNS	25.6. Sayın \$AD_SOYAD\$ hanelerarası transferler(alınan nafaka, yurt dışından gelen işçi döviz vb.) var mı?
HANE_GELIR_BILME	26. Sayın \$AD_SOYAD\$ sizin ya da diğer hanehalkı üyeleri için daha önce ifade ettiğiniz gelir türlerini düşünerek hanehalkının aylık net gelirinin ne olduğunu biliyor musunuz?
HANE_GELIR	28. Sayın \$AD_SOYAD\$ hanenizin ortalama aylık geliri aşağıdaki gelir gruplarından hangisinde yer almaktadır?
GENELSAGLIKDURUM	1. Sayın \$AD_SOYAD\$ genel olarak sağlık durumunuz nasıldır?
HASTALIK_SAGLIKDURUM	2. Sayın \$AD_SOYAD\$ 6 ay ya da uzun süren/sürmesi beklenen hastalığınız/sağlık probleminiz var mı?
YASAMFAAL_KISITLANMADURUM	3. Sayın \$AD_SOYAD\$ son 6 ay içinde herhangi bir sağlık problemi nedeniyle günlük yaşam faaliyetlerinizin kısıtlanma durumu nedir?
KRONIKHAST_ASTIM	4.A. Sayın \$AD_SOYAD\$ son 12 ay içerisinde ASTIM(ALERJİK ASTİM DAHİL) yaşadınız mı?
KRONIKHAST_KRONIKBRONSIT	4.B. Sayın \$AD_SOYAD\$ son 12 ay içerisinde KRONİK BRONSİT, KRONİK OBSTRÜKTİF AKCİGER HASTALIĞI, AMFİZEM yaşadınız mı?
KRONIKHAST_ENFARKTUS	4.C. Sayın \$AD_SOYAD\$ son 12 ay içerisinde MİYOKARDİYAL ENFARKTUS(KALP KRİZİ) yaşadınız mı?
KRONIKHAST_KORONERKALPHAST	4.D. Sayın \$AD_SOYAD\$ son 12 ay içerisinde KORONER KALP HASTALIĞI(ANJİNA, GÖĞÜS AĞRISI, SPAZM) yaşadınız mı?
KRONIKHAST_HİPERTANSİYON	4.E. Sayın \$AD_SOYAD\$ son 12 ay içerisinde HİPERTANSİYON yaşadınız mı?
KRONIKHAST_INMEFELC	4.F. Sayın \$AD_SOYAD\$ son 12 ay içerisinde İNME-FELC(BEYİN KANAMASI, SEREBRAL TROMBOZ) yaşadınız mı?
KRONIKHAST_ARTHRISIS	4.G. Sayın \$AD_SOYAD\$ son 12 ay içerisinde ARTHROSİS(YAŞLANMAYA BAĞLI KIKIRDAK DEJENERASYONU, EKLEM İLTİHABİ HARİC) yaşadınız mı?
KRONIKHAST_BELBOLGE	4.H. Sayın \$AD_SOYAD\$ son 12 ay içerisinde BEL BÖLGESİ PROBLEMLERİ(BEL AĞRISI, BEL FİTİĞİ VE DİĞER BEL DEFEKTLERİ) yaşadınız mı?
KRONIKHAST_BOYUNBOLGE	4.I. Sayın \$AD_SOYAD\$ son 12 ay içerisinde BOYUN BÖLGESİ PROBLEMLERİ(BOYUN AĞRISI, BOYUN FİTİĞİ, DİĞER BOYUN DEFEKTLERİ) yaşadınız mı?
KRONIKHAST_SEKERHAST	4.J. Sayın \$AD_SOYAD\$ son 12 ay içerisinde SEKER HASTALIĞI(DİYABET) yaşadınız mı?
KRONIKHAST_ALERJİ	4.K. Sayın \$AD_SOYAD\$ son 12 ay içerisinde ALERJİ(ALERJİK RİNİT, DERMATİT, YIYECEK VB. ALERJİSİ, ALERJİK ASTİM HARİC) yaşadınız mı?
KRONIKHAST_KARACİGERYETMZ	4.L. Sayın \$AD_SOYAD\$ son 12 ay içerisinde KARACİGER SİROZU, KARACİGER YETMEZLİĞİ yaşadınız mı?

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu (Devamı)

Sütun adı	Soru
KRONIKHAST_IDRARKACIRMA	4.M. Sayın \$AD_SOYAD\$ son 12 ay içerisinde IDRAR KACIRMA, IDRARI TUTAMAMA yaşadınız mı?
KRONIKHAST_BOBREK	4.N. Sayın \$AD_SOYAD\$ son 12 ay içerisinde BOBREK PROBLEMLERİ yaşadınız mı?
KRONIKHAST_DEPRESYON	4.O. Sayın \$AD_SOYAD\$ son 12 ay içerisinde DEPRESYON yaşadınız mı?
KRONIKHAST_ALZHEIMER	4.P. Sayın \$AD_SOYAD\$ son 12 ay içerisinde ALZHEIMER yaşadınız mı?
KRONIKHAST_COLYAK	4.R. Sayın \$AD_SOYAD\$ son 12 ay içerisinde COLYAK yaşadınız mı?
KRONIKHAST_MADDEKUL	4.S. Sayın \$AD_SOYAD\$ son 12 ay içerisinde MADDE KULLANIMINA BAĞLI HASTALIK yaşadınız mı?
KRONIKHAST_DOWN	4.T. Sayın \$AD_SOYAD\$ son 12 ay içerisinde DOWN SENDROMU yaşadınız mı?
YARALANMA_TRAFIKKAZA	5.1. Sayın \$AD_SOYAD\$ son 12 ay içerisinde yaralanma ile sonuçlanan TRAFİK KAZASI geçirdiniz mi?
TEDAVI_TRAFIKKAZA	6.1. Sayın \$AD_SOYAD\$ TRAFİK KAZASI sonucunda aşağıdakilerden hangisine ihtiyacınız oldu?
YARALANMA_EVKAZA	5.2. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, yaralanma ile sonuçlanan EV KAZASI geçirdiniz mi?
TEDAVI_EVKAZA	6.2. Sayın \$AD_SOYAD\$ EV KAZASI sonucunda aşağıdakilerden hangisine ihtiyacınız oldu?
YARALANMA_GUNDELIKKAZA	5.3. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, yaralanma ile sonuçlanan GUNDELİK KAZA geçirdiniz mi?
TEDAVI_GUNDELIKKAZA	6.3. Sayın \$AD_SOYAD\$ GUNDELİK KAZA sonucunda aşağıdakilerden hangisine ihtiyacınız oldu?
YARALANMA_ISKAZA	5.4. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, yaralanma ile sonuçlanan İS KAZASI geçirdiniz mi?
TEDAVI_ISKAZA	6.4. Sayın \$AD_SOYAD\$ İS KAZASI sonucunda aşağıdakilerden hangisine ihtiyacınız oldu?
SON12AY_ISEARAVERME	7. Sayın \$AD_SOYAD\$ son 12 ay içerisinde sağlık sorunları nedeniyle işe ara verdiniz mi?
SON12AY_ISEARAVERME_TOPGUN	8. Sayın \$AD_SOYAD\$ son 12 ay içerisinde sağlık sorunları nedeniyle toplam kaç gün işe ara verdiniz?
GOZLUKKULLANMA	9. Sayın \$AD_SOYAD\$ GOZLUK YA DA LENS kullanıyor musunuz?
GORMEDEZORLUK	10. Sayın \$AD_SOYAD\$ NORMAL METİN YAZILARINI OKURKEN GORMEDE zorluk yaşıyor musunuz?
ISITMECIHAZIKULLANMA	11. Sayın \$AD_SOYAD\$ ISITME CİHAZI kullanıyor musunuz?
DUYMA_SESSİZLİK	12. Sayın \$AD_SOYAD\$ SESSİZ BİR ODADA BİR KISIYLA SOHBET EDERKEN DUYMADA zorluk çekiyor musunuz?
DUYMA_GURULTU	13. Sayın \$AD_SOYAD\$ GURULTULU BİR ODADA BİR KISIYLA SOHBET EDERKEN DUYMADA zorluk çekiyor musunuz?
DESTEKSİZ500MYURUYEBİLME	14. Sayın \$AD_SOYAD\$ DUZ BİR ARAZİDE BASTON VEYA DİĞER YÜRÜME ARACI KULLANMADAN YA DA DESTEK ALMADAN 500 METRE YÜRÜMEKTE zorluk çekiyor musunuz?

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu (Devamı)

Sütun adı	Soru
DESTEKSİZ12BASAMAK	15. Sayın \$AD_SOYAD\$ 12 BASAMAK MERDİVEN İNİP ÇIKMAKTA zorluk çekiyor musunuz?
OGRENME	16. Sayın \$AD_SOYAD\$ YASITLARINIZA GÖRE YENİ BİLGİ BECERİ ÖĞRENMEDE(OKUMA-YAZMA, HESAPLAMA, BASİT DÖRT İŞLEM YAPMADA VB.) güçlük çekiyor musunuz?
HATIRLAMA	17. Sayın \$AD_SOYAD\$ YAKIN ZAMANDA YASANAN OLAYLARI (GÜN İÇİNDEKİ OLAYLAR, YAKINLARINIZIN ADLARI GİBİ) HATIRLAMAKTA güçlük çekiyor musunuz?
ZORLUKÇEKME_BESLENME	18.1. Sayın \$AD_SOYAD\$ YARDIM ALMADAN KENDİ KENDİNİZE BESLENİRKEN zorluk çekiyor musunuz?
YARDIMALMA_BESLENME	19.1. Sayın \$AD_SOYAD\$ KENDİ KENDİNİZE BESLENİRKEN genellikle yardım alıyor musunuz?
FAZLAYARD_BESLENME	20.1. Sayın \$AD_SOYAD\$ KENDİ KENDİNİZE BESLENİRKEN yardıma ihtiyacınız var mı?
ZORLUKÇEKME_YATMOTURMKALKM	18.2. Sayın \$AD_SOYAD\$ yardım almadan YATAGA GİRME/KALKMA, SANDALYEYE OTURMA/KALKMADA zorluk çekiyor musunuz?
YARDIMALMA_YATMOTURMKALKM	19.2. Sayın \$AD_SOYAD\$ YATAGA GİRME/KALKMA, SANDALYEYE OTURMA/KALKMADA genellikle yardım alıyor musunuz?
FAZLAYARD_YATMOTURMKALKM	20.2. Sayın \$AD_SOYAD\$ YATAGA GİRME/KALKMA, SANDALYEYE OTURMA/KALKMADA yardıma ihtiyacınız var mı?
ZORLUKÇEKME_GIYINMESOYUNMA	18.3. Sayın \$AD_SOYAD\$ yardım almadan GIYINME VE ELBİSELERİNİ ÇIKARMADA zorluk çekiyor musunuz?
YARDIMALMA_GIYINMESOYUNMA	19.3. Sayın \$AD_SOYAD\$ GIYINME VE ELBİSELERİNİ ÇIKARMADA genellikle yardım alıyor musunuz?
FAZLAYARD_GIYINMESOYUNMA	20.3. Sayın \$AD_SOYAD\$ GIYINME VE ELBİSELERİNİ ÇIKARMADA yardıma ihtiyacınız var mı?
ZORLUKÇEKME_TUVALETKLLANIM	18.4. Sayın \$AD_SOYAD\$ yardım almadan TUVALET KULLANIMINDA zorluk çekiyor musunuz?
YARDIMALMA_TUVALETKLLANIM	19.4. Sayın \$AD_SOYAD\$ TUVALET KULLANIMINDA genellikle yardım alıyor musunuz?
FAZLAYARD_TUVALETKLLANIM	20.4. Sayın \$AD_SOYAD\$ TUVALET KULLANIMINDA yardıma ihtiyacınız var mı?
ZORLUKÇEKME_BANYOYAPMA	18.5. Sayın \$AD_SOYAD\$ yardım almadan BANYO YAPMADA zorluk çekiyor musunuz?
YARDIMALMA_BANYOYAPMA	19.5. Sayın \$AD_SOYAD\$ BANYO YAPMADA genellikle yardım alıyor musunuz?
FAZLAYARD_BANYOYAPMA	20.5. Sayın \$AD_SOYAD\$ BANYO YAPMADA yardıma ihtiyacınız var mı?
ZORLUKÇEKME_YEMEHAZIRLAM	21.1. Sayın \$AD_SOYAD\$ yardım almadan YEMEK HAZIRLAMADA zorluk çekiyor musunuz?
YARDIMALMA_YEMEHAZIRLAM	22.1. Sayın \$AD_SOYAD\$ YEMEK HAZIRLAMADA genellikle yardım alıyor musunuz?
FAZLAYARD_YEMEHAZIRLAM	23.1. Sayın \$AD_SOYAD\$ YEMEK HAZIRLAMADA yardıma ihtiyacınız var mı?

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu (Devamı)

Sütun adı	Soru
ZORLUKCEKME_TELEFONKULLANM	21.2. Sayın \$AD_SOYAD\$ yardım almadan TELEFON KULLANIMINDA zorluk çekiyor musunuz?
YARDIMALMA_TELEFONKULLANM	22.2. Sayın \$AD_SOYAD\$ TELEFON KULLANIMINDA genellikle yardım alıyor musunuz?
FAZLAYARD_TELEFONKULLANM	23.2. Sayın \$AD_SOYAD\$ TELEFON KULLANIMINDA yardıma ihtiyacınız var mı?
ZORLUKCEKME_ALISVERIS	21.3. Sayın \$AD_SOYAD\$ yardım almadan ALISVERIS YAPMADA zorluk çekiyor musunuz?
YARDIMALMA_ALISVERIS	22.3. Sayın \$AD_SOYAD\$ ALISVERIS YAPMADA genellikle yardım alıyor musunuz?
FAZLAYARD_ALISVERIS	23.3. Sayın \$AD_SOYAD\$ ALISVERIS YAPMADA yardıma ihtiyacınız var mı?
ZORLUKCEKME_ILACKULLANIM	21.4. Sayın \$AD_SOYAD\$ yardım almadan ILAC KULLANIMINDA zorluk çekiyor musunuz?
YARDIMALMA_ILACKULLANIM	22.4. Sayın \$AD_SOYAD\$ ILAC KULLANIMINDA genellikle yardım alıyor musunuz?
FAZLAYARD_ILACKULLANIM	23.4. Sayın \$AD_SOYAD\$ ILAC KULLANIMINDA yardıma ihtiyacınız var mı?
ZORLUKCEKME_HAFIFEVIS	21.5. Sayın \$AD_SOYAD\$ yardım almadan HAFIF EV ISLERINI YAPMADA zorluk çekiyor musunuz?
YARDIMALMA_HAFIFEVIS	22.5. Sayın \$AD_SOYAD\$ HAFIF EV ISLERINI YAPMADA genellikle yardım alıyor musunuz?
FAZLAYARD_HAFIFEVIS	23.5. Sayın \$AD_SOYAD\$ HAFIF EV ISLERINI YAPMADA yardıma ihtiyacınız var mı?
ZORLUKCEKME_AGIREVIS	21.6. Sayın \$AD_SOYAD\$ yardım almadan AGIR EV ISLERINI YAPMADA zorluk çekiyor musunuz?
YARDIMALMA_AGIREVIS	22.6. Sayın \$AD_SOYAD\$ AGIR EV ISLERINI YAPMADA genellikle yardım alıyor musunuz?
FAZLAYARD_AGIREVIS	23.6. Sayın \$AD_SOYAD\$ AGIR EV ISLERINI YAPMADA yardıma ihtiyacınız var mı?
ZORLUKCEKME_IDARIMALIKONU	21.7. Sayın \$AD_SOYAD\$ yardım almadan GUNLUK SORUMLULUKLAR VE MALI KONULARLA ILGILENMEDE zorluk çekiyor musunuz?
YARDIMALMA_IDARIMALIKONU	22.7. Sayın \$AD_SOYAD\$ GUNLUK SORUMLULUKLAR VE MALI KONULARLA ILGILENMEDE genellikle yardım alıyor musunuz?
FAZLAYARD_IDARIMALIKONU	23.7. Sayın \$AD_SOYAD\$ GUNLUK SORUMLULUKLAR VE MALI KONULARLA ILGILENMEDE yardıma ihtiyacınız var mı?
BEDENSELAGRI	24. Sayın \$AD_SOYAD\$ son 4 hafta içerisinde ne kadar bedensel ağrınız oldu?
AGRI_HAYATENGELLEME	25. Sayın \$AD_SOYAD\$ son 4 hafta içerisinde ağrı normal hayatınızı ne derece engelledi?
ZEVKDUYMA	26.A. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde BİRSEYLER YAPARKEN ÇOK AZ ILGI DUYMA VE KEYIF ALMA ne sıklıkla sizi rahatsız etti?
BUNALIM	26.B Sayın \$AD_SOYAD\$ son 2 hafta içerisinde MORALI BOZUK, BUNALIMLI VEYA UMUTSUZ HİSSETME ne sıklıkla sizi rahatsız etti?
UYKUSORUN	26.C. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde UYKUYA DALMAKTA VEYA UYUMAKTA ZORLUK ÇEKME/ ÇOK FAZLA UYUMA ne sıklıkla sizi rahatsız etti?
YORGUNLUK	26.D. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde YORGUN/ENERJISİZ HİSSETME ne sıklıkla sizi rahatsız etti?

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu (Devamı)

Sütun adı	Soru
ISTAHSIZLIK	26.E. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde ISTAHSIZLIK/ASIRI YEMEK YEME ne sıklıkla sizi rahatsız etti?
DEGERSIZHISSETME	26.F. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde KENDİNİ DEGERSİZ VE KÖTÜ HİSSETME ne sıklıkla sizi rahatsız etti?
KONSANTRASYONPRB	26.G. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde KONSANTRASYON PROBLEMİ(GAZETE OKURKEN, TV İZLERKEN VB.) ne sıklıkla sizi rahatsız etti?
HUZURSUZLUK	26.H. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde BASKALARININ FARK EDECEĞİ DERECEDE AĞIR HAREKET ETME/KONUSMA YA DA HUZURSUZ/FAZLA HAREKET ETME ne sıklıkla sizi rahatsız etti?
SON12AY_YATAKLIHİZMET	27. Sayın \$AD_SOYAD\$ son 12 ay içerisinde en az bir gece hastanede YATTINIZ MI?
YATAKLIHİZMET_TOPGECE	28. Sayın \$AD_SOYAD\$ YATAKLI TEDAVİ HİZMETİ/HİZMETLERİ DÜŞÜNDÜĞÜNÜZDE, hastanede toplam kaç gece geçirdiniz?
SON12AY_GÜNÜBİRLİKHİZMET	29. Sayın \$AD_SOYAD\$ son 12 ay içerisinde hastanede GÜNÜBİRLİK (24 SAATTEN DAHA AZ SUREYLE) TEDAVİ HİZMETİ aldınız mı?
GÜNÜBİRLİK_TOPGÜN	30. Sayın \$AD_SOYAD\$ son 12 ay içerisinde kaç kere hastanede GÜNÜBİRLİK (24 SAATTEN DAHA AZ SUREYLE) TEDAVİ HİZMETİ aldınız?
SAGKURUM_TERCİH	31. Sayın \$AD_SOYAD\$ en son gittiğiniz sağlık kuruluşunu tercih etme sebebiniz nedir?
DIS_HEK_HİZALMAZAMAN	32. Sayın \$AD_SOYAD\$ kendiniz için en son ne zaman bir DİS HEKİMDEN hizmet aldınız?
AİLE_PRTS_HEK_HİZALMADURUM	33. Sayın \$AD_SOYAD\$ kendiniz için, PRATİSYEN HEKİMDEN VEYA AİLE HEKİMDEN en son ne zaman hizmet aldınız?
AİLE_PRTS_HEK_HİZALMA_KERE	34. Sayın \$AD_SOYAD\$ son 4 hafta içerisinde bir PRATİSYEN HEKİM YA DA AİLE HEKİMDEN kaç kere hizmet aldınız?
UZMAN_HEK_HİZALMADURUM	35. Sayın \$AD_SOYAD\$ kendiniz için, en son ne zaman UZMAN HEKİMDEN hizmet aldınız?
UZMAN_HEK_HİZALMA_KERE	36. Sayın \$AD_SOYAD\$ son 4 hafta içerisinde bir UZMAN HEKİMDEN kaç kere hizmet aldınız?
SON12AY_FİZİYOTERAPİST	37.1. Sayın \$AD_SOYAD\$ son 12 ay içerisinde kendiniz için FİZİYOTERAPİSTE gittiniz mi?
SON12AY_FİZİKTEDVUZMAN	37.2. Sayın \$AD_SOYAD\$ son 12 ay içerisinde kendiniz için FİZİK TEDAVİ UZMANINA gittiniz mi?
SON12AY_PSIKOLOG	37.3. Sayın \$AD_SOYAD\$ son 12 ay içerisinde kendiniz için PSİKOLOGA gittiniz mi?
SON12AY_PSIKOTERAPİST	37.4. Sayın \$AD_SOYAD\$ son 12 ay içerisinde kendiniz için PSİKOTERAPİSTE gittiniz mi?
SON12AY_PSIKIYATRİST	37.5. Sayın \$AD_SOYAD\$ son 12 ay içerisinde kendiniz için PSIKIYATRİSTE gittiniz mi?
SOSYALYARDIMHİZMETİ	38. Sayın \$AD_SOYAD\$ son 12 ay içerisinde evde sağlık ya da sosyal yardım hizmeti aldınız mı?
EVDESAGLIKHİZMETİ	39. Sayın \$AD_SOYAD\$ evde sağlık ya da sosyal yardım hizmeti için ilk müracatı nereye yaptınız?

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu (Devamı)

Sütun adı	Soru
RECETEEDİLEN_ILACKULLANIM	40. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde bir doktor tarafından size reçete edilen bir ilaç kullandınız mı?
KENDI_ILAC	41. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde kendi kararınız/tercihinizle ILAC kullandınız mı?
KENDI_GIDATAKVIYE	42. Sayın \$AD_SOYAD\$ son 2 hafta içerisinde kendi kararınız/tercihinizle GIDA TAKVIYESİ VEYA VİTAMİN kullandınız mı?
GRIPASISOLMA_ZAMAN	43. Sayın \$AD_SOYAD\$ en son ne zaman grip aşısı oldunuz?
GRIPASISOLMA_TARİH	43.1. Sayın \$AD_SOYAD\$ en son grip aşısı olma tarihini giriniz!!!! (Günü tam olarak hatırlamayanlar için "01" giriniz. Örn; 02/2011 için 01/02/2011 giriniz.)
TANSİYONOLCULME_DURUM	44. Sayın \$AD_SOYAD\$ TANSİYONUNUZ en son ne zaman bir sağlık personeli tarafından ölçüldü?
KOLESTROLOLCULME_DURUM	45. Sayın \$AD_SOYAD\$ KOLESTEROLUNUZ en son ne zaman bir sağlık personeli tarafından ölçüldü?
KANSEKERIOLCULME_DURUM	46. Sayın \$AD_SOYAD\$ KAN SEKERİNİZ en son ne zaman bir sağlık personeli tarafından ölçüldü?
GAITA_GIZLIKANTESTİ_DURUM	47. Sayın \$AD_SOYAD\$ DISKI(GAİTA) İLE İLGİLİ GIZLI KAN TESTİNİ en son ne zaman yaptırdınız?
KOLONOSKOPI_DURUM	48. Sayın \$AD_SOYAD\$ en son ne zaman KOLONOSKOPI yaptırdınız?
KENDIMUAYENE_DURUM	49. Sayın \$AD_SOYAD\$ kendi kendinize meme muayenesini ne sıklıkla yaparsınız?
MAMMOGRAFİCEKTİRME_ZAMAN	50. Sayın \$AD_SOYAD\$ en son ne zaman mammografi/meme filmi çektiniz?
SMEARYAPTIRMA_ZAMAN	51. Sayın \$AD_SOYAD\$ en son ne zaman rahim ağzı sürüntü testi(servikal smear, pap testi) yaptırdınız?
GECİKME_RANDEVU	52. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, RANDEVU ALMA SURESİ ÇOK UZUN OLDUGU İÇİN sağlık hizmeti almada gecikme yaşadınız mı?
GECİKME_ULASIM	53. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, UZAKLIK YA DA ULASIM PROBLEMLERİ NEDENİYLE sağlık hizmeti almada gecikme yaşadınız mı?
ODEMEGUCLUGU_TIBBIBAKIM	54.A. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, TIBBİ BAKIMA İHTİYACINIZ OLDUGU HALDE ODEME GUCLUGU NEDENİYLE karşılayamadığınız oldu mu?
ODEMEGUCLUGU_DISBAKIM	54.B. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, DIS BAKIMINA İHTİYACINIZ OLDUGU HALDE ODEME GUCLUGU NEDENİYLE karşılayamadığınız oldu mu?
ODEMEGUCLUGU_ILAC	54.C. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, RECETE EDİLEN İLACA İHTİYACINIZ OLDUGU HALDE ODEME GUCLUGU NEDENİYLE karşılayamadığınız oldu mu?
ODEMEGUCLUGU_RUHTEDV	54.D. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, RUHSAL TEDAVİYE(PSİKİYATRİST VB. UZMANLAR TARAFINDAN) İHTİYACINIZ OLDUGU HALDE ODEME GUCLUGU NEDENİYLE karşılayamadığınız oldu mu?
BOY	55. Sayın \$AD_SOYAD\$ ayakkabılarınız olmadan boyunuz(cm) ne kadardır?
KİLO	56. Sayın \$AD_SOYAD\$ elbise ve ayakkabılarınız olmadan vücut ağırlığınız (kg) ne kadardır?

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu (Devamı)

Sütun adı	Soru
GUNLUK_FAALİYETDURUM	57. Sayın \$AD_SOYAD\$ çalışırken/günlük faaliyetlerinizi gerçekleştirirken, aşağıdakilerden hangisi durumunuzu en iyi tanımlar?
YURUME_10DAK_GUN	58. Sayın \$AD_SOYAD\$ normal bir hafta içerisinde bir yerden bir yere giderken, kaç gün EN AZ 10 DAKIKA YURURSUNUZ?
YURUME_ZAMAN	59. Sayın \$AD_SOYAD\$ normal bir günde, bir yerden bir yere giderken YURUYUS ICIN ne kadar zaman harcarsınız?
BISIKLET_GUN	60. Sayın \$AD_SOYAD\$ normal bir hafta içerisinde bir yerden bir yere giderken, kaç gün EN AZ 10 DAKIKA BISIKLET KULLANIRSINIZ?
BISIKLET_ZAMAN	61. Sayın \$AD_SOYAD\$ normal bir günde, bir yerden bir yere giderken BISIKLET SURME ICIN ne kadar zaman harcarsınız?
SPOR_GUN	62. Sayın \$AD_SOYAD\$ normal bir hafta içerisinde, kaç gün EN AZ 10 DAKIKA SPOR, FITNESS VEYA BOS ZAMAN AKTİVİTESİ gerçekleştirirsiniz?
SPOR_ZAMAN	63. Sayın \$AD_SOYAD\$ normal bir hafta içerisinde SPOR, FITNESS VEYA BOS ZAMAN AKTİVİTESİ GERCEKLESTIRIRKEN toplamda ne kadar zaman harcarsınız?
KASGUCLENDIRME	64. Sayın \$AD_SOYAD\$ normal bir hafta içerisinde, KASLARINIZI GUCLENDIRMEK ICIN kaç gün zaman harcarsınız?
MEYVE_YEMESIKLIK	65. Sayın \$AD_SOYAD\$ ne sıklıkta MEYVE yersiniz?
MEYVE_PORSIYON	66. Sayın \$AD_SOYAD\$ günde kaç porsiyon MEYVE yersiniz?
SEBZE_SALATA_YEMESIKLIK	67. Sayın \$AD_SOYAD\$ ne sıklıkta SEBZE YA DA SALATA YERSINIZ(PATATES HARIC)?
SEBZE_SALATA_PORSIYON	68. Sayın \$AD_SOYAD\$ günde kaç porsiyon SEBZE YA DA SALATA YERSINIZ?
TUTUNKULLANIM_DURUM	69. Sayın \$AD_SOYAD\$ tütün mamulü kullanıyor musunuz?
TUTUNKULLANMAMA_GUN	70. Sayın \$AD_SOYAD\$ ne kadar süredir kullanmıyor musunuz?
TUTUNKUL_ILKKEZ_YAS	71. Sayın \$AD_SOYAD\$ ilk kez kaç yaşında tütün mamulü kullanmayı denediniz?
TUTUN_BASLAMANEDEN_1	72.1. Sayın \$AD_SOYAD\$ tütün mamulü kullanmaya en önemli başlama nedeniniz nedir?
TUTUNKUL_DUZENLI_DURUM	73. Sayın \$AD_SOYAD\$ düzenli olarak(yılda en az 100 adet olmak üzere) hiç tütün mamulü kullandınız mı?
TUTUNKUL_DUZENLI_YAS	74. Sayın \$AD_SOYAD\$ düzenli olarak kaç yaşında tütün mamulü kullanmaya başladınız?
TUTUN_TUR	75. Sayın \$AD_SOYAD\$ çoğunlukla hangi tütün mamulünü kullanıyorsunuz?

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu (Devamı)

Sütun adı	Soru
SIGARA_ADET	76. Sayın \$AD_SOYAD\$ ortalama olarak her gün kaç adet sigara kullanıyorsunuz?
S_BIRAKMADENEME	77. Sayın \$AD_SOYAD\$ son 12 ayda sigarayı bırakmayı denediniz mi?
S_BIRAKMAYONTEM	78. Sayın \$AD_SOYAD\$ sigarayı bırakmak için hangi yöntemi kullandınız?
TUTUNDUMAN_MARUZ	79. Sayın \$AD_SOYAD\$ kapalı alanda, tütün dumanına ne sıklıkta maruz kalıyorsunuz?
MARUZKALMA_S_LOKANTA	80S.1. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz LOKANTA/RESTORANDA 09:00-17:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_A_LOKANTA	80A.1. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz LOKANTA/RESTORANDA 17:01-02:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_S_PASTANE	80S.2. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz PASTANEDE 09:00-17:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_A_PASTANE	80A.2. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz PASTANEDE 17:01-02:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_S_KAFE	80S.3. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz KAFEDE 09:00-17:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_A_KAFE	80A.3. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz KAFEDE 17:01-02:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_S_KAHVE	80S.4. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz KAHVEHANEDE 09:00-17:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_A_KAHVE	80A.4. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz KAHVEHANEDE 17:01-02:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_S_BAR	80S.5. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz BARDA 09:00-17:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_A_BAR	80A.5. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz BARDA 17:01-02:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_S_DIGER	80S.98. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz DİGER MEKANLARDA 09:00-17:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
MARUZKALMA_A_DIGER	80A.98. Sayın \$AD_SOYAD\$ son 30 gün içerisinde gittiğiniz DİGER MEKANLARDA 17:01-02:00 SAATLERİ ARASINDA tütün kullanan var mıydı?
EVDEKULLANMA_SIKLIK	81. Sayın \$AD_SOYAD\$ evinizin içerisinde herhangi bir kişi ne sıklıkta tütün mamulü kullanır?
EVDEKULLANMA_YER	82. Sayın \$AD_SOYAD\$ evinizde genellikle nerelerde tütün mamulü kullanılır?
ALKOLKULLANIM_DURUM	83. Sayın \$AD_SOYAD\$ hiç alkollü içecek kullandınız mı? (bir yudum dahi olsa)
ALKOLKUL_ILKKEZ_YAS	84. Sayın \$AD_SOYAD\$ ilk kez kaç yaşında alkollü içecek kullandınız ya da kullanmayı denediniz?
ILKDENEEME_ALKOL_TUR	85. Sayın \$AD_SOYAD\$ ilk olarak denediğiniz alkollü içecek hangisidir?
ALKOL_KULLANMANEDEN_1	86.1. Sayın \$AD_SOYAD\$ alkol kullanmanızın ya da kullanmayı denemenizin en önemli nedeni nedir?

Ek-1: Türkiye Sağlık Araştırması Değişkenler Tablosu (Devamı)

Sütun adı	Soru
SON12AY_ALKOLKULLANIM_SIKLIK	87. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, ne sıklıkta alkollü içecek kullandınız?
ALKOLKUL_DUZENLI_YAS	88. Sayın \$AD_SOYAD\$ düzenli olarak kaç yaşında alkollü içecek kullanmaya başladınız?
H_BIRA	89B. Sayın \$AD_SOYAD\$ haftada genellikle kaç şişe BIRA kullanırsınız?
H_RAKI	89R. Sayın \$AD_SOYAD\$ haftada genellikle kaç bardak RAKI kullanırsınız?
H_SARAP	89S. Sayın \$AD_SOYAD\$ haftada genellikle kaç kadeh SARAP kullanırsınız?
H_VOTKA	89VO. Sayın \$AD_SOYAD\$ haftada genellikle kaç bardak VOTKA kullanırsınız?
H_VISKI	89V. Sayın \$AD_SOYAD\$ haftada genellikle kaç bardak VISKI kullanırsınız?
H_CIN	89C. Sayın \$AD_SOYAD\$ haftada genellikle kaç bardak CIN kullanırsınız?
H_LIKOR	89L. Sayın \$AD_SOYAD\$ haftada genellikle kaç bardak LIKOR kullanırsınız?
H_DIGER	89D. Sayın \$AD_SOYAD\$ haftada genellikle kaç bardak DIGER ICECEK kullanırsınız?
SON12AY_6VEUZERI	90. Sayın \$AD_SOYAD\$ son 12 ay içerisinde, (herhangi bir fırsatta) ne sıklıkta 6 adet veya daha fazla alkollü içecek içtiniz?
ALKOLZARAR_SAGLIK	91.1. Sayın \$AD_SOYAD\$ alkol kullanımının SAGLIGA zarar verdiğini düşünüyor musunuz?
ALKOLZARAR_EVLILIK	91.2. Sayın \$AD_SOYAD\$ alkol kullanımının EV YASANTISINA VE/VEYA EVLILIKLERE zarar verdiğini düşünüyor musunuz?
ALKOLZARAR_ARKADASLIK	91.3. Sayın \$AD_SOYAD\$ alkol kullanımının ARKADASLIK ILISKILERI VE/VEYA SOSYAL YASANTIYA zarar verdiğini düşünüyor musunuz?
ALKOLZARAR_IS	91.4. Sayın \$AD_SOYAD\$ alkol kullanımının IS YASAMINA zarar verdiğini düşünüyor musunuz?
ALKOLZARAR_BUTCE	91.5. Sayın \$AD_SOYAD\$ alkol kullanımının AILE BUTCESINE zarar verdiğini düşünüyor musunuz?
ALKOLZARAR_TRAFIK	91.6. Sayın \$AD_SOYAD\$ alkol kullanımının TRAFIK GUVENLIGINE zarar verdiğini düşünüyor musunuz?
GUVENILIR_YAKIN	98. Sayın \$AD_SOYAD\$ ciddi kişisel bir probleminiz olursa güvенеbileceğiniz, size çok yakın kaç kişi vardır?
CEVRE_ILGI	99. Sayın \$AD_SOYAD\$ çevrenizdeki insanlar ne yaptığınıza ne kadar ilgi gösterir?
KOMSU_YARDIM	100. Sayın \$AD_SOYAD\$ ihtiyacınız olduğunda komşularınızdan yardım alabilme durumunuz nedir?
RESMIOLMYIRD_DURUM	101. Sayın \$AD_SOYAD\$ bazı yaşlılık problemleri, kronik sağlık problemleri veya sakatlık yaşayan bir veya daha fazla kişiye, haftada en az bir kez bakım ya da yardım sağlıyor musunuz?
RESMIOLMYIRD_YAKINLIK	102. Sayın \$AD_SOYAD\$ bu kişi veya kişiler.....
RESMIOLMYIRD_SAAAT	103. Sayın \$AD_SOYAD\$ haftada ortalama kaç saat bakım veya yardım sağlıyorsunuz?
FERTFAKTOR	

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	Onur CAN
Doğum Yeri	Ankara
Doğum Tarihi	16 Ağustos 1985

LİSANS EĞİTİM BİLGİLERİ

Üniversite	Samsun Ondokuz Mayıs Üniversitesi
Fakülte	Fen Edebiyat Fakültesi
Bölüm	İstatistik

İŞ DENEYİMİ

Çalıştığı Kurum	Türkiye İstatistik Kurumu
Görevi	İstatistikçi
Tecrübe Süresi	4

İLETİŞİM

Adres	TÜİK Kars Bölge Müdürlüğü Merkez/KARS
E-mail	onur.can@tuik.gov.tr