



**İŞLETME ANABİLİM DALI
TÜRKİYE'DE FERTLERİN ELEKTRONİK TİCARET
KULLANIM SÜRELERİNİN DESTEK VEKTÖR
MAKİNELERİ VE KARAR AĞAÇLARI YÖNTEMLERİ
İLE KARŞILAŞTIRMALI ANALİZİ**

**YÜKSEK LİSANS TEZİ
SAHİB RAMAZANOV
Danışman
PROF. DR. ÖTÜKEN SENER
Kars – 2020**



T.C.

**KAFKAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANA BİLİM DALI**

**TÜRKİYE'DE FERTLERİN ELEKTRONİK TİCARET
KULLANIM SÜRELERİNİN DESTEK VEKTÖR
MAKİNELERİ VE KARAR AĞAÇLARI YÖNTEMLERİ
İLE KARŞILAŞTIRMALI ANALİZİ**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN

SAHİB RAMAZANOV

Danışman ve Jüri Üyeleri

Prof. Dr. Ötüken SENGER

Doç. Dr. Ali Kemal ÇELİK

Dr. Öğr. Üyesi Mehmet Emin KARABAYIR

Kars – 2020

KAFKAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Sahib RAMAZANOV tarafından hazırlanan Türkiye’de Fertlerin Elektronik Ticaret Kullanım Sürelerinin Destek Vektör Makineleri ve Karar Ağaçları Yöntemleri ile Karşılaştırmalı Analizi başlıklı bu çalışma,// 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. Ötügen SENGER

Üye : Dr. Öğr. Üyesi Mehmet Emin KARABAYIR

Üye : Doç. Dr. Ali Kemal ÇELİK

İmza:

İmza:

İmza:

ONAY

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

Doç. Dr. Yaşar KOP
Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİMİ

Yüksek Lisans tezi olarak sunduğum “**Türkiye’de Fertlerin Elektronik Ticaret Kullanım Sürelerinin Destek Vektör Makineleri ve Karar Ağaçları Yöntemleri ile Karşılaştırmalı Analizi**” 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 taahhüt 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 Comparative Analysis Individual Use Time of Electronic Commerce with Support Vector Machines and Decision Trees Methods in Turkey as master’s thesis I prepared, that I obtained all information in master’s thesis 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 master’s thesis in accordance with the writing rules and Works which I used have been shown in the bibliography.

07 / 01 / 2020



Sahib RAMAZANOV

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
ÖNSÖZ	iii
KISALTMALAR	iv
TABLO LİSTESİ	v
ŞEKİLLER LİSTESİ	vi
1.GİRİŞ	1
1.BİLİŞİM VE BİLİŞİM SİSTEMLERİ	3
1.1.Bilişim	3
1.1.1.Bilişim Teknolojileri (BT).....	3
1.1.1.1.Bilişim Teknolojilerinin İşletmelere Sağladığı Yararlar	5
1.1.1.2.Bilişim Teknolojileri ve İşletme Yönetimi İlişkisi.....	6
1.1.2.Bilişim Sistemleri.....	7
1.1.2.1.Bilişim sistemlerinin bileşenleri / kaynakları	8
1.1.3.Bilişim Sistemi Çalışmaları	12
1.1.3.1.Veri Kaynaklarının Girişleri	12
1.1.3.2.Verinin İşlenmesi	13
1.1.3.3.Bilişim Sistemi Çıktıları	13
1.1.3.4.Veri Kaynaklarının Depolanması.....	13
2.KARAR AĞAÇLARI	14
2.1.Karar Ağaçları ile İlgili Literatür Taraması.....	15
2.2.Karar Ağaçlarının Kullanım Alanları.....	18
2.3.Karar Ağaçlarının Avantaj ve Dezavantajları	19
2.4.Sınıflandırma İşleminin Doğruluğu	20
2.4.1.Hold-out.....	21
2.4.2.K-fold cross validation	21
2.4.3.Leave-one-out	22
2.4.4.Bootstrap.....	23
2.5.Sınıflandırma Performanslarının Değerlendirilmesi	23
2.5.1.Karışıklık Tabloları	23
2.5.2.ROC Eğrileri	24
2.6.Karar Ağacı Algoritmaları.....	26

2.6.1.CLS Algoritması	26
2.6.2.CHAID Algoritması (Kass, 1980: 119-127).....	26
2.6.3.ID3 Algoritması (Quinlan, 1986: 81-106).....	27
2.6.4.C4.5 Algoritması.....	27
2.6.5.C5.0 Algoritması.....	27
2.6.6.QUEST Algoritması (Loh ve Shih, 1997: 815-840).....	28
2.6.7.CART (C&R Tree) Algoritması (Breiman vd., 1984: 216-264)	28
3.DESTEK VEKTÖR MAKİNELERİ (DVM).....	31
3.1.Doğrusal Destek Vektör Makineleri ile İlgili Literatür Taraması	32
3.2.Doğrusal Destek Vektör Makineleri	35
3.2.1.Doğrusal Ayrılabilir Durum (Sert Sınır)	35
3.2.1.1.Optimizasyon Problemi	41
3.2.1.2.Lagrange Çarpımları.....	42
3.2.1.3.Krush-Kuhn-Tucker (KKT) koşulları	43
3.2.2.Tümüyle Doğrusal Ayrılamayan Durum (Yumuşak Sınır)	45
3.3.Doğrusal Olmayan Destek Vektör Makineleri	52
3.3.1.Çekirdek Fonksiyonları (Kernel Functions)	56
3.3.1.1.Mercer Koşulları	59
3.3.2.Kernel Fonksiyon Seçimi	59
3.3.3.DVM için Kerneller	60
3.3.3.1.Doğrusal Kernel	60
3.3.3.2.Polinomial Kernel	60
3.3.3.3.Gaussian (Radyal Tabanlı) Kernel	60
4.UYGULAMA	62
4.1.Analiz Öncesi Veri Seti.....	62
4.2.Verdi Ön İşleme.....	63
4.2.1.Verdi Temizleme	64
4.2.2.Verdi Dönüştürme.....	64
4.3.Verdi Setinin Bölünmesi (training ve testing).....	65
4.4.Otomatik sınıflama işlemcisi ile kullanılacak modelin belirlenmesi	65
4.4.1.CHAID Algoritması	69
4.4.2.C&R Tree Algoritması	93
4.4.3.C5.0 Algoritması.....	102

4.5.RBF Kernel Fonksiyonu Kullanılarak Destek Vektör Makineleri Yardımı İle Değişkenlerin İncelenmesi	117
4.6.Doğrusal Kernel Fonksiyonu Kullanılarak Destek Vektör Makineleri Yardımı İle Değişkenlerin İncelenmesi	120
4.7.Sigmoid Kernel Fonksiyonu Kullanarak DVM Modelinin İncelenmesi	122
5.SONUÇ	126
KAYNAKLAR	131
ÖZGEÇMİŞ	140



KAFKAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANA BİLİM DALI
TÜRKİYE'DE FERTLERİN ELEKTRONİK TİCARET KULLANIM
SÜRELERİNİN DESTEK VEKTÖR MAKİNELERİ VE KARAR AĞAÇLARI
YÖNTEMLERİ İLE KARŞILAŞTIRMALI ANALİZİ
YÜKSEK LİSANS TEZİ
Sahib RAMAZANOV
Prof. Dr. Ötüken SENGER
2020 – VII + 140

ÖZET

Bu çalışmanın temel amacı, bireylerin karar ağaçları ve DVM modelleri ile e-ticaret kullanımı üzerinde uygulanarak belirlenmesidir. Bu amaç doğrultusunda, 2017 yılına ait Hanehalkı Bilişim Teknolojileri Kullanım Araştırmasına ait B grubu mikro veri seti Türkiye İstatistik Kurumu aracılığıyla temin edilmiştir. Daha sonra çeşitli manipülasyon yöntemleriyle veri madenciliği analizlerine uygun hale getirilen veri seti üzerinde C5.0, CHAID, C&R Tree algoritmaları ve destek vektör makineleri (DVM) uygulanmıştır.

Çalışmanın amacı doğrultusunda e-ticaret kullanımı değişkeni, bağımlı değişken olarak belirlenmiştir. Hanehalkı bilişim teknolojileri kullanım araştırmasının kapsamı, Türkiye Cumhuriyeti sınırları içinde ikamet eden hanelerde bulunan tüm fertlerdir. “İnternet üzerinden mal veya hizmet satın alma ya da sipariş verme (e-posta, SMS ve MMS ile verilen siparişler hariç) işlemini en son ne zaman gerçekleştirdiniz?” sorusu, anketin uygulanması esnasında 16 yaş ve üzeri fertlere sorulmuştur. Bu nedenle analize 16 yaş ve üzeri fertler dahil edilmiştir.

Karar ağacı algoritmaları ve destek vektör makineleri (DVM) birbirleri ile kıyaslandığında ise DVM makinesinin hem C5.0 hem CHAID hem de C&R Tree algoritmasına göre daha iyi sonuçlar verdiği gözlenmiştir.

Anahtar Kelimeler: E-ticaret Kullanımı, Karar Ağaçları, Destek Vektör Makineler

KAFKAS UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF BUSINES
COMPARATIVE ANALYSIS INDIVIDUAL USE TIME OF ELECTRONIC
COMMERCE WITH SUPPORT VECTOR MACHINES AND DECISION
TREES METHODS IN TURKEY
MASTER'S THESIS
Sahib RAMAZANOV
Prof. Dr. Ötüken SENGER
2020 – VII + 140

ABSTRACT

The main purpose of this study is to determine the decision tree and DVM models by applying the e-commerce usage of the individuals. For this purpose, micro data set of group B from the 2017 Household Information Technology Usage Survey was obtained through Turkey Statistical Institute. Then, C5.0, CHAID, C&R Tree algorithms and support vector machines (DVM) were applied on the data set which was made suitable for data mining analysis with various manipulation methods.

For the purpose of the study, e-commerce usage variable was taken as dependent variable. The scope of this research on household usage of information technology contains all individuals of households residing within the boundaries of Republic of Turkey. The question “When was the last time that you have purchased or ordered goods or services over the Internet (except for orders placed by e-mail, SMS and MMS)?” was asked to individuals aged 16 and over during the survey. Therefore, individuals over 16 years old were included in the analysis.

When decision tree algorithms and support vector machines (DVM) are compared with each other, it is observed that DVM machine gives better results than both C5.0 and CHAID and C&R Tree algorithms.

Keywords: E-commerce Usage, Decision Trees, Support Vector Machines

ÖNSÖZ

Çalışmamın her aşamasında bilgisi ve tecrübesiyle desteklerini esirgemeyen değerli hocam Prof. Dr. Ötüken SENGER'e teşekkürlerimi ve şükranlarımı sunarım. Tez aşamasında yardımlarını esirgemeyen Dr. Öğr. Üyesi Deniz TANIR'a teşekkürlerimi sunarım.

Beni bu günlere getiren ve her zaman yanımda hissettiğim annem Şehla XUDAVERDİYEVA, babam Yagub RAMAZANOV, kardeşim Fatime RAMAZANOVA'ya ve üzerimde emeği olan, isimlerini buraya sığdıramadığım herkese teşekkürü bir borç bilirim.

07/01/2020

Sahib RAMAZANOV

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
DVM	: Destek Vektör Makineleri
SPSS	: Statistical Package for the Social Sciences
TÜİK	: Türkiye İstatistik Kurumu

TABLO LİSTESİ

Tablo No	Sayfa
Tablo 1.1: Bilişim Sistemi Kaynakları.....	12
Tablo 2.1: Karışıklık Matrisi	24
Tablo 3.1: Popolüer Kernel Fonksiyonları.....	58
Tablo 4.1: Veri Dönüştürme İşlemi	64
Tablo 4.2: CHAID Algoritmasında Kullanılan Değişkenlerin Önem Değerleri	70
Tablo 4.3: CHAID Algoritmasının Performans Ölçüm Sonuçları	92
Tablo 4.4: C&R Tree Algoritmasında Kullanılan Değişkenlerin Önem Değerleri	94
Tablo 4.5: C&R Tree Algoritmasının Performans Ölçüm Sonuçları	102
Tablo 4.6: C5.0 Algoritmasında Kullanılan Değişkenlerin Önem Değerleri.....	103
Tablo 4.7: C5.0 Algoritmasının Performans Ölçüm Sonuçları	117
Tablo 4.8: RBF(Radyal Tabanlı)-DVM Değişkenlerin Göreceli Önem Oranları	118
Tablo 4.9: RBF(Radyal Tabanlı)-DVM Değişkenlerin Performans Ölçüm Sonuçları	119
Tablo 4.10: Doğrusal-DVM Değişkenlerin Göreceli Önem Oranları	121
Tablo 4.11: Doğrusal-DVM Değişkenlerin Performans Ölçüm Sonuçları.....	122
Tablo 4.12: Sigmoid DVM Değişkenlerin Göreceli Önem Oranları.....	124
Tablo 4.13: Sigmoid DVM Değişkenlerin Performans Ölçüm Sonuçları	125

ŞEKİLLER LİSTESİ

Şekil No	Sayfa
Şekil 1.1: Bilişim Teknolojilerindeki Temel İşlevler	4
Şekil 1.2: Bilişim Teknolojilerinin İşletmeye Sağladığı Başlıca Yararlar	6
Şekil 1.3: Bilişim Sistemi.....	8
Şekil 1.4: Bilişim Sistemi Bileşenleri/Kaynakları	9
Şekil 2.1: Karar Ağacı Örneği	15
Şekil 2.2: Model Değerlendirmesi (Sınıflama Doğruluğu)	21
Şekil 2.3: ROC Eğrisi	25
Şekil 3.1: Olası Hiper Düzlemler.....	36
Şekil 3.2: Maksimum ayırma başarısına sahip hiper düzlem.....	36
Şekil 3.3: Maksimum Marj Sınıflandırıcısı	37
Şekil 3.4: Destek Vektör Gösterimi	38
Şekil 3.5: İki Boyutlu Hiper Düzlem ve W Vektör Gösterimi	39
Şekil 3.6: Ayrıcı Hiper Düzlem ve Kanonik Hiper Düzlemler	40
Şekil 3.7: Maksimal Marj Sınıflandırıcı İle Ayrılamayan Veri Grafiği.....	45
Şekil 3.8: (a) Grafiğinde Gözlemlerin Çoğu Doğru Tarafa Bulunmaktadır	47
Şekil 3.9: Gevşek Değişken Gösterimi	48
Şekil 3.10: Doğrusal Olmayan Kuadratik DVM Sınıflandırıcısı ve Yanlış Sınıflandırma Yapan Doğrusal Sınıflandırıcı.....	53
Şekil 3.11: Doğrusal Olmayan Bir Veri Setinin Yüksek Boyutlu Bir Uzayda Doğrusal Hiper Düzlem ile Sınıflandırılması.....	54
Şekil 4.1: Veri Kalitesinin İncelenmesi	63
Şekil 4.2: Verilerin Partiton İşlemcisi İle Bölünmesi	65
Şekil 4.3: Auto Classifier İşlemcisi	66
Şekil 4.4: Auto Classifier İşlemcisi Sonuçları.....	67
Şekil 4.5: Kurulan Modelin Ekran Görüntüsü.....	68
Şekil 4.6: CHAID Algoritmasına Göre Önemli Değişkenler.....	69
Şekil 4.7: CHAID Algoritması ile Oluşturulan Karar Ağacının Genel Görünümü....	71
Şekil 4.8: CHAID Algoritması ile Oluşan Karar Ağacının Başlangıç Düzümü	71
Şekil 4.9: CHAID Algoritmasının 1. Düzümü Göre Dallanması	73
Şekil 4.10: CHAID Algoritmasının 2. Düzümü Göre Dallanması	75

Şekil 4.11: CHAID Algoritmasının 5. Düğüme Göre Dallanması	77
Şekil 4.12: CHAID Algoritmasının 6. Düğüme Göre Dallanması	79
Şekil 4.13: Şekil 4.12'nin Devamı.....	80
Şekil 4.14: CHAID Algoritmasının 17. Düğüme Göre Dallanması	82
Şekil 4.15: CHAID Algoritmasının 18. Düğüme Göre Dallanması	85
Şekil 4.16: CHAID Algoritmasının 18. Düğüme Göre Dallanması	86
Şekil 4.17: CHAID Algoritmasının 3. Düğüme Göre Dallanması	90
Şekil 4.18: C&RT Algoritmasına Göre Önemli Değişkenler.....	93
Şekil 4.19: C&RT Algoritması ile Oluşturulan Karar Ağacının Genel Görünümü ...	94
Şekil 4.20: C&R Tree Algoritması ile Oluşan Karar Ağacının Başlangıç Düğümü ..	95
Şekil 4.21: C&R Tree Algoritmasının 1. Düğüme Göre Dallanması	96
Şekil 4.22: C&R Tree Algoritmasının 7. ve 10. Düğüme Göre Dallanması.....	98
Şekil 4.23: C&R Tree Algoritmasının 2. Düğüme Göre Dallanması	100
Şekil 4.24: C5.0 Algoritmasına Göre Önemli Değişkenler.....	103
Şekil 4.25: C5.0 algoritması ile Oluşan Karar Ağacının Başlangıç Düğümü	104
Şekil 4.26: C5.0 Algoritmasının 1. Düğüme Göre Dallanması.....	104
Şekil 4.27: C5.0 Algoritmasının 2. Düğüme Göre Dallanması.....	105
Şekil 4.28: C5.0 Algoritmasının 3. Düğüme Göre Dallanması.....	106
Şekil 4.29: C5.0 Algoritmasının 4. Düğüme Göre Dallanması.....	107
Şekil 4.30: C5.0 Algoritmasının 5. Düğüme Göre Dallanması.....	108
Şekil 4.31: C5.0 Algoritmasının 9. Düğüme Göre Dallanması.....	109
Şekil 4.32: C5.0 Algoritmasının 530. Düğüme Göre Dallanması	110
Şekil 4.33: C5.0 Algoritmasının 686. Düğüme Göre Dallanması	111
Şekil 4.34: C5.0 Algoritmasının 914. Düğüme Göre Dallanması	112
Şekil 4.35: C5.0 Algoritmasının 934. Düğüme Göre Dallanması	113
Şekil 4.36: C5.0 Algoritmasının 980. Düğüme Göre Dallanması	114
Şekil 4.37: C5.0 Algoritmasının 1006. Düğüme Göre Dallanması	115
Şekil 4.38: C5.0 Algoritmasının 1010. Düğüme Göre Dallanması	116
Şekil 4.39: RBF(Radyal Tabanlı)-DVM Değişkenlerin Göreceli Önem Oranları ...	119
Şekil 4.40: Doğrusal-DVM Değişkenlerin Göreceli Önem Oranları	120
Şekil 4.41: Sigmoid -DVM Değişkenlerin Göreceli Önem Oranları	123

GİRİŞ

Yirminci yüzyılın son çeyreğinde bilgi ve iletişim teknolojilerinde yaşanan hızlı gelişmeler, toplumların tüm kesimlerinde, günlük yaşamın her alanında bilgisayar kullanımının çağın gereği olduğu bilincini yaratmıştır.

Birçok yazar bugüne kadar insanlığın karşılaştığı en büyük teknolojik devrimin XX. yüzyılın sonlarına doğru ortaya çıktığına işaret etmektedir. Bu yazarlara göre elektronik ticaret, matbaanın icadı, ya da endüstri devrimi kadar önemli bir dönüşümü simgelemektedir. Bu devrimin, dünya üzerindeki etkilerinin tahmin edilenden çok daha geniş olacağı ileri sürülmekte ve içinde bulunulan bu dönemde, elektronik ticaret devriminin başlarında olduğu özellikle belirtilmektedir.

Elektronik ticaret, örgütsel ve bireysel seviyede tüm ticari faaliyetlerle ilgili işlemleri kapsamaktadır. Bu alanda üretilmiş, işlenmiş ve aktarılmış tüm sayısal veriler, metinler, sesler ve görsel imajların işlenmesi ve aktarılmasına dayanmaktadır. Bu alandaki önemli konular arasında tüketicinin korunması, rekabet, finans ve ödeme sistemleri, vergilendirme, fikri, sınai ve ticari mülkiyet hakları, güvenlik, yasal düzenlemeler, uyumsuzlukların sona erdirilmesi mekanizmaları bulunmaktadır.

Bu çalışma beş bölümden oluşmaktadır. Birinci bölümde bilişim ve bilişim sistemlerine giriş yapılmış ve ön bilgiler verilmiştir. İkinci bölümde, karar ağaçları detaylı bir şekilde anlatılmıştır. Üçüncü bölümde destek vektör makineleri geniş bir şekilde anlatılmıştır. Dördüncü bölümde, gerçek bir veri seti üzerinde karar ağacı algoritmaları ve destek vektör makineleri uygulanmış ve analiz sonuçları ayrıntılı bir şekilde incelenmiştir. Beşinci bölümde, kullanılan karar ağacı algoritmaları ve destek vektör makinelerinin performansları karşılaştırılıp hangi algoritmanın daha iyi sonuç verdiği yorumlanmıştır.

Bu tez çalışmasında, Türkiye İstatistik Kurumunun (TÜİK) 2017 yılında yapmış olduğu Hanehalkı Bilişim Teknolojileri Kullanım Araştırmasına ait mikro veri seti kullanılmıştır. TÜİK Araştırmanın soru kağıdını Avrupa Birliği İstatistik Ofisi (EUROSTAT) tarafından hazırlanan ve önerilen model soru kağıdı esas alarak hazırlamıştır. Bu soru kağıdını Türkiye koşullarına göre adapte etmiş, aynı zamanda Ulusal Bilgi Toplumu Stratejisi'ni yürüten kurum / kuruluşların ihtiyaçları

doğrultusunda düzenlenmiştir. Araştırmanın kapsamı Türkiye sınırları dahilinde bulunan tüm yerleşim yerlerindeki 16-74 yaş grubu yetişkin fertlerden oluşmaktadır. Kurumsal nüfus olarak tanımlanan okul, yurt, otel, çocuk yuvası, huzurevi, hastane ve hapisanede bulunanlar ile kışla ve ordu evlerinde ikamet edenler kapsamamaktadır. Ayrıca nüfusu toplam nüfusun %1'ini geçmeyecek şekilde yeterli (küçük köyler, oba, mezra v.b) örnek hane sayısına ulaşamayacağı düşünülen yerleşim yerleri kapsam dışı bırakılmaktadır. Araştırmanın örneklem tasarımı 2 aşamalı tabakalı sistematik küme örnekleme yöntemiyle yapılmıştır. Birinci aşamada ortalama 100 hane ihtiva eden kümeler (bloklar) PPS yöntemi (büyüklüğü ile orantılı ve olasılıklı) ile örneğe seçilmektedir. İkinci aşamada ise örneğe seçilen bu kümelerden sistematik seçim yöntemi kullanılarak, örnek adresler belirlenmektedir. İstatistiki Bölge Birimleri Sınıflaması 1. Düzey tabakalama kriteri olarak kullanılmaktadır. Analizler yapılırken, modellerin performanslarının kıyaslamaları detaylı olarak tablolar ve şekiller ile gösterilmiştir.

TÜİK'e ait bu veri seti üzerinde karar ağaçları ve DVM modelleri, e-ticaret kullanımı üzerinde uygulanmıştır. Analizler yapılırken, modellerin performanslarının kıyaslamaları detaylı olarak tablolar ve şekiller ile gösterilmiştir. Elde edilen sonuçlara göre, hangi modellerde hangi faktörlerin göreceli olarak etken faktörler olduğu elde edilmiştir.

1. BİLİŞİM VE BİLİŞİM SİSTEMLERİ

1.1. Bilişim

Bilişim kavramı için, bilginin teknoloji ile birlikte kullanılarak, işlenmesinden elde edilen sonuçlardır diyebiliriz. Günümüzde bilgiyi elde etmek ve kullanmak için, bilgi dünyası ile teknoloji arasında ilişki kurularak, elde edilen veri etkin bir şekilde kullanılmaktadır (Dondurmacı, 2011: 228).

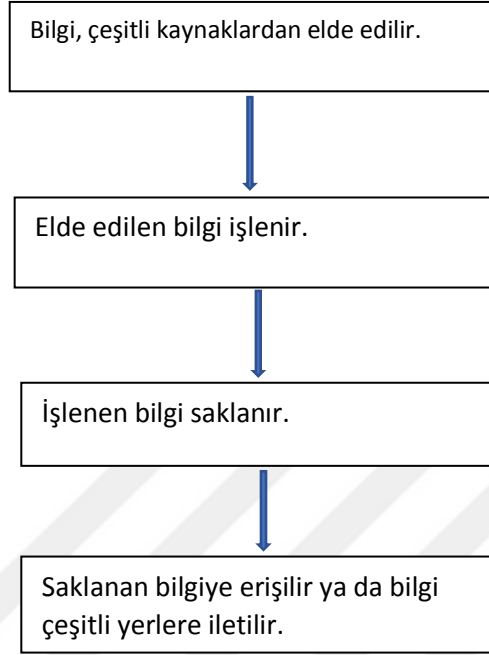
Bilişim, temelde bilgisayar ve bilgisayar yazılımlarının kullanıcıların gereksinimlerine göre düzenlendiği bir platformdur.

Toplumun ihtiyaç duyduğu her türlü bilgi ve verinin bilgisayarlar aracılığı ile işlenmesini ve değerlendirilmesini sağlayan ya da öngören bilime bilişim denilebilir.

1.1.1. Bilişim Teknolojileri (BT)

Bilişim teknolojileri, çeşitli kaynaklardan elde edilen bilginin toplanmasını, bu bilginin işlenmesini, bilgisayarlarda saklanmasını ve çeşitli ağlar aracılığıyla farklı yerlere iletilmesini sağlayarak, insanların kullanımına sunan tüm iletişim ve bilgisayar teknolojilerini içermektedir. Dünya genelinde gerçekleşen konferans, kongre ve bunun gibi tüm etkinlikleri gözlemlediğimizde bilişim teknolojilerinin günümüzde ulaştığı boyutları görebiliriz (Aktaş vd., 2010).

Şekil 1.1: Bilişim Teknolojilerindeki Temel İşlevler



Kaynak: (Dondurmacı, 2011).

20. yüzyılın sonlarından itibaren bilgi ya da bilişim çağı, bilgi toplumu ya da bilişim toplumu kavramları tüm dünyada kabul edilmektedir.

Bilgisayarı kullanarak tüm bilgi gereksinimlerini karşılayabilen topluma, bilgi toplumu diyebiliriz.

Bilginin kullanılabilmesini sağlayan en önemli araç iletişimdir. Çeşitli kaynaklardan toplanan bilgiler, bilgisayarlara girilerek amacına uygun programlar vasıtasıyla işlenerek, bilgisayarlarda saklanır. Bilgisayarlarda depolanan bu bilgiler çeşitli kişi ve kuruluşların yararlanması için kullanıma sunulur. Bu bilgilere erişebilmek için iletişim araçlarından yararlanır. İletişim için internet, intranet ve bunun gibi araçları satabiliriz (Bayoğlu, 2010).

İletişim ağları kullanılarak, bilgiye çok hızlı ve kolay bir şekilde erişilebilmektedir. Bilginin kullanımında iletişim alt yapısı; bilgi bankalarına, birçok araştırma kurumlarına ve bireysel bilgisayarlara bağlıdır. Bu bağlantıyla bilginin üretilmesi ve sunulması, bilişim teknolojileri sisteminin içinde oluşmaktadır (Aktepe vd., 2009).

1.1.1.1. Bilişim Teknolojilerinin İşletmelere Sağladığı Yararlar

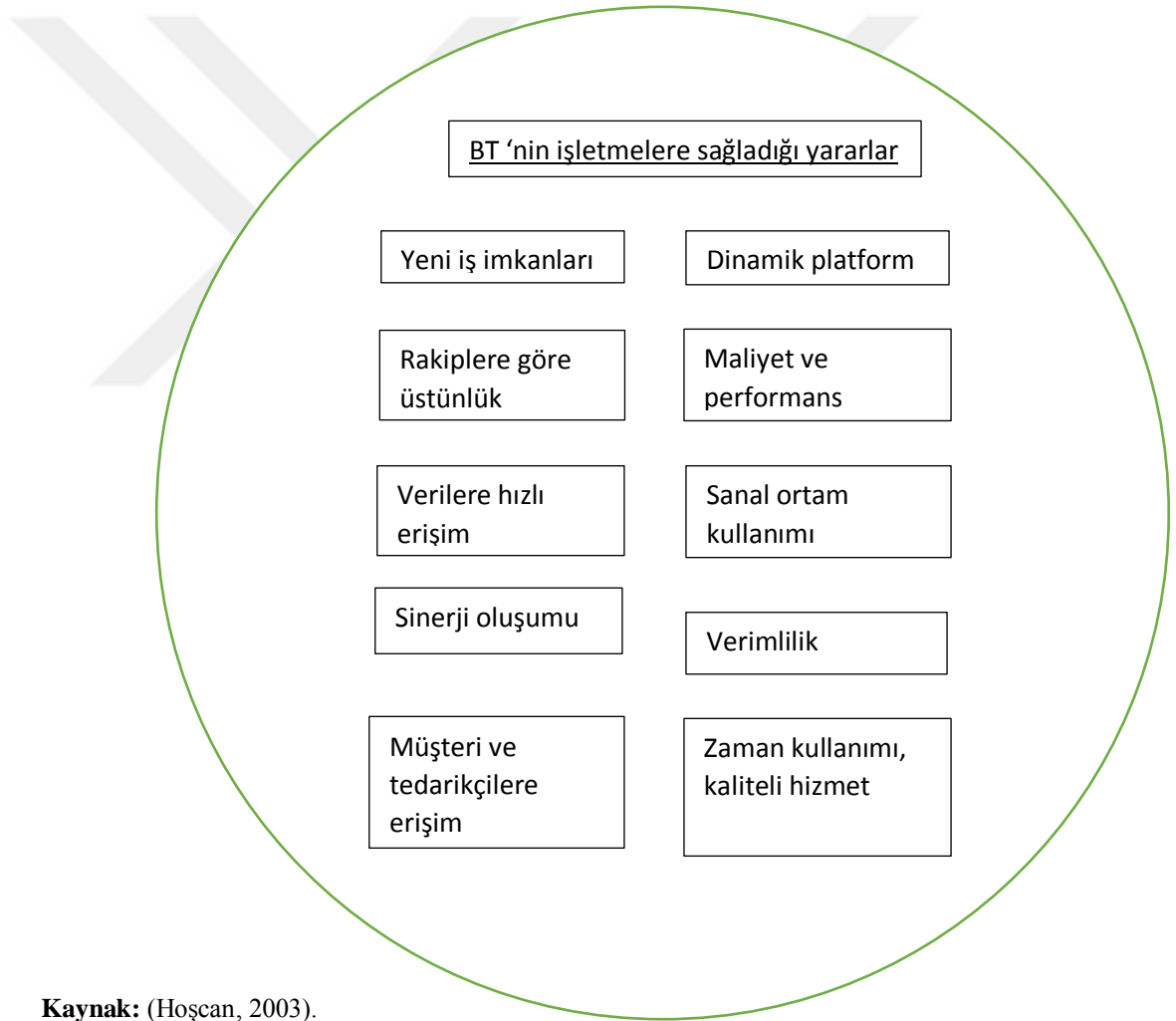
Bilişim Teknolojileri, bilgi toplumunun gelişmesinde çok önemli rolü olmasının yanı sıra işletme örgütlerine sağladığı katkılar ile özellikle örgüt içi ve örgütler arası iletişimde çok önemli değişiklikler sağlamıştır. Bilişim teknolojileri, alt düzeyde çalışanlara da sistem içinde çeşitli yetkiler vererek, tüm çalışanların verilen yetkilere göre işletme bilgilerini verimli bir şekilde kullanmalarını sağlamıştır. Bilişim teknolojileri, işletmelerde tüm kademelerdeki (alt, orta, üst) yönetim faaliyetlerinde, yöneticiler tarafından kullanılarak işletme faaliyetlerinde köklü değişikliklerin yapılmasını ve işletmenin verimlilik ve etkinliğinin artmasını sağlamıştır (Laudon ve Laudon, 2011).

Bilişim teknolojilerinin işletmelere sağladığı başlıca yararları aşağıdaki şekilde sıralayabiliriz:

- Bilişim teknolojilerinde meydana gelen değişimler bir yandan mevcut iş olanaklarının ortadan kalkmasına neden olurken bir yandan da yeni iş olanaklarının ortaya çıkmasını sağlamıştır.
- İşletmeler bilişim teknolojilerinin getirdiği yenilikleri, yeni ürün ve hizmetlerin pazarlamasında kullanarak rakiplerine göre üstünlükler sağlamaktadırlar.
- Bilişim teknolojileri sayesinde işletmeler şirket içi ve şirket dışından gelen veriye hızlı ve kolay bir şekilde erişebilmektedirler.
- Bilişim ağları, iletişim ve işbirliğini önemli ölçüde kolaylaştırmaktadır. Bu sayede işletmelerde sinerji oluşmakta ve girişimler desteklenmektedir.
- Bilişim teknolojileri kullanılarak, işletmeler internet ortamında müşterileri ve tedarikçileri ile hızlı ve sürekli iletişim kurma olanağına sahip olmaktadır.
- Bilişim Teknolojileri sayesinde, işletmenin tüm çalışanları bilgiye hızlı ve kolay erişerek dinamik bir platform içinde bulunmaktadır.
- Bilişim teknolojileri işletmelere, maliyetlerin düşürülmesi ve performansın artması yönünde katkı sağlamaktadır.

- İşletmelerin pazarlama birimleri, bilişim teknolojilerini kullanarak pazarlama ve dağıtım çalışmalarını sanal ortamlarda doğrudan müşteriye erişerek yapmaktadırlar.
- İşletmeler bilişim teknolojilerini kullanarak, yeni pazarlar oluşturmakta, yeni müşteriler kazanmakta ve işletmede verimliliğin artmasını sağlamaktadırlar.
- Bilişim teknolojileri işletmelerin zaman kullanımı, kalite ve müşteri hizmeti faaliyetlerini olumlu yönde etkilemektedir.

Şekil 1.2: Bilişim Teknolojilerinin İşletmeye Sağladığı Başlıca Yararlar



Kaynak: (Hoşcan, 2003).

1.1.1.2. Bilişim Teknolojileri ve İşletme Yönetimi İlişkisi

Bilişim teknolojilerinin bir işletmede, ilgili bilgilerin çeşitli kaynaklardan toplanması, işletme faaliyetlerinin düzenli bir şekilde yürütülmesi, maliyetlerin

düşmesi ve verimliliğin artması yönünde çok önemli katkıları bulunmaktadır (Bagad, 2011). Bilişim teknolojileri geliştikçe bazı sektörlerin tümüyle ortadan kalktığını ve yeni sektörlerin ortaya çıktığını gözlemlemedeyiz. Bilişim teknolojileri günümüzde artık işletmenin rutin işlerinin yanı sıra işletme yönetiminin stratejik karar almalarında önemli bir etken haline gelmiştir. Yöneticiler karar alma aşamalarında bilişim teknolojilerinin sunduğu olanakları yoğun bir şekilde kullanmaya başlamışlardır (Aktaş vd., 2010).

1.1.2. Bilişim Sistemleri

İşletmenin içinden ve dışından olmak üzere çeşitli kaynaklardan gelen veri girdi olarak alınıp, dönüşüm işlemi sürecinden geçirildikten sonra, bilgiyi ürün olarak ortaya çıkaran sisteme bilişim sistemi denilmektedir (Bayoğlu, 2010).

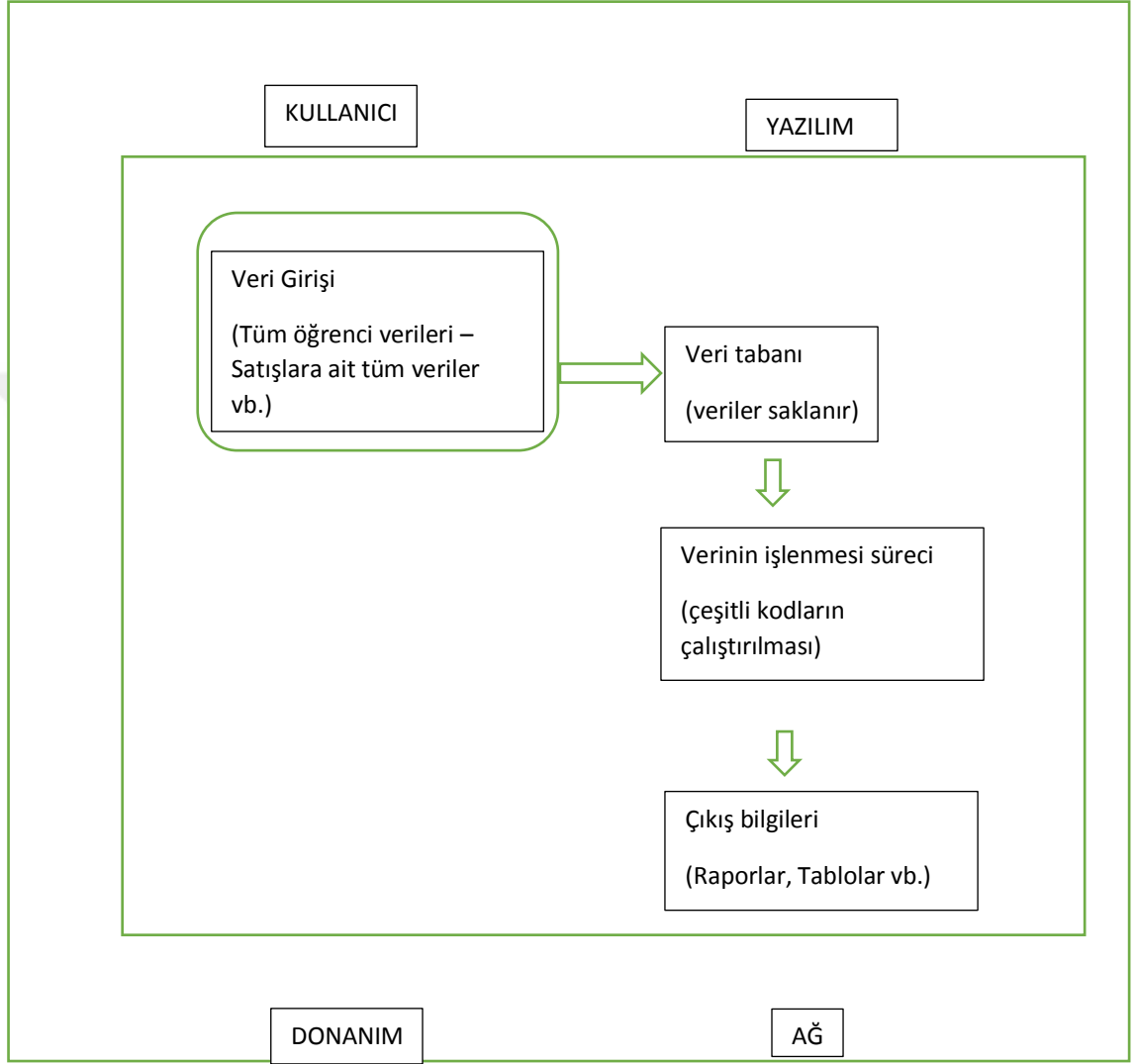
Bir işletmede kısa, orta ve uzun vadeli amaçlara ulaşmak açısından "bilgi" hayati öneme sahip bir olgudur. Yönetim, çeşitli dönemlerde örneğin satışlar ve maliyetlere ilişkin olarak bazı ön bilgilere sahip olmak isteyecektir. Yönetim yüksek satış fiyatları, en düşük maliyetler ve hizmetlerle kazancını en üst düzeye çıkarmak isteyecektir. Bu tür bir amaca ulaşabilmek için de yeterli bilgi ve kullanabileceği karar tekniklerine ihtiyaç duyacaktır. Pek çok yönetici bilgiyi rekabet gücünün kaynağı olarak kabul etmektedir (Dondurmacı, 2011: 209).

Günümüz işletmelerinin başarısını etkileyen faktörlerin arasında kuşkusuz bilgi çok önemli bir yer tutmaktadır. İşletmeler kapalı bir kutu değildir. Sürekli olarak çevresiyle etkileşim halindedir. İşletmenin sağlıklı bir biçimde yönetilebilmesi için, iç ve dış faktörlere ilişkin sağlıklı bir bilgi kaynağına sahip olması gerekiyor. Artan malzeme fiyatları, işçi maliyetleri, dış rekabet sinyalleri ve bunun yanı sıra firmanın ekonomik çevresi, dış faktörler ve iç faktörlerin koordinasyonu "bilişim sistemleri" yardımıyla gerçekleştirilebilir (Bensghir, 1996: 274).

Bilişim sistemleri girdi, süreç ve çıktı işlemlerini yerine getirirken, insan (kullanıcı), veri, iletişim teknolojileri, donanım ve yazılım bileşenlerini kullanmaktadır. Bilişim sistemlerinde yapılan tüm işlemler için insana her zaman ihtiyaç duyulmaktadır. Bilişim sisteminde iki tür kullanıcı görev yapmaktadır. Birincisi veriyi yararlı bilgiye

dönüştüren bilişim uzmanları, ikincisi yararlı bilgiyi kullanan son kullanıcılarıdır (Akpınar,2004).

Şekil 1.3: Bilişim Sistemi

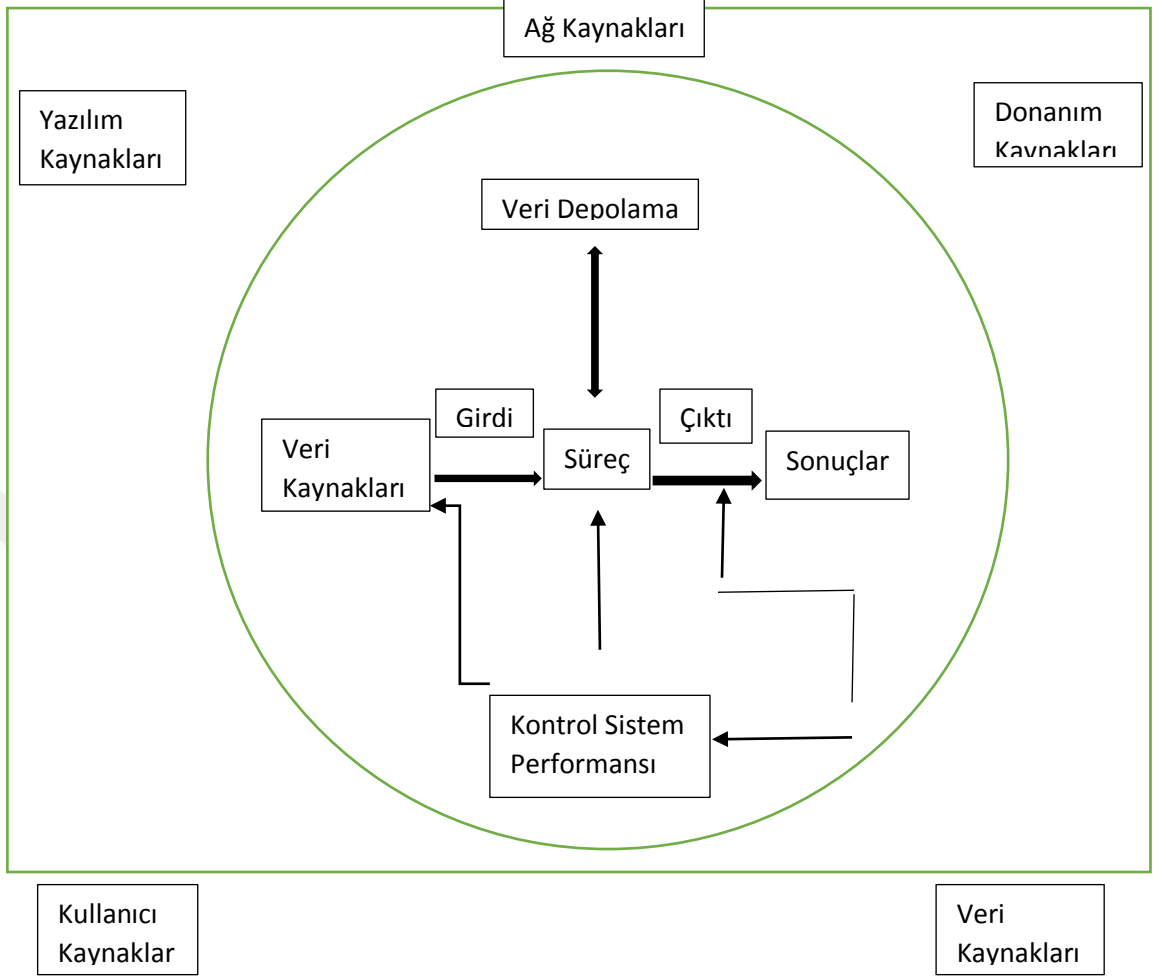


Kaynak: (Aktepe vd., 2009).

1.1.2.1. Bilişim sistemlerinin bileşenleri / kaynakları

Herhangi bir bilişim sistemi modeli beş ana bileşen/kaynak içermektedir. Bunlar Şekil 1.4' de görüldüğü gibi donanım, yazılım, ağ, veri ve kullanıcı kaynağı olarak adlandırılmaktadır.

Şekil 1.4: Bilişim Sistemi Bileşenleri/Kaynakları



Kaynak: (Bensghir, 1996: 274).

a) Donanım Kaynakları

Bu kaynaklar bilişim sistemlerinin fiziksel bölümünü oluşturmaktadır. Kaynaklarla ilgili temel kavramlar ve roller bu bölümde yer almaktadır.

Modern bir bilgisayar sistemi girdiler, işlemler, depolama ve çıktılar olmak üzere dört alt sistemden oluşmaktadır.

Girdi alt sistemi: Karakter tabanlı veri, ses, resim, grafik. İşlemler Alt Sistemi: Bu sistemde merkezi işlem birimi (Central Processing Unit-CPU) bulunmaktadır. Bu birimdeki bileşenler, kontrol birimi, mantık birimi, ana depolama ve kayıt işlemi birimi olarak adlandırılabilir (Elibol ve Kesici, 2004).

Depolama Alt Sistemi: Bu sistemler işlenmiş verinin saklanması için kullanılmaktadır. Çıktı Alt Sistemi: Bilgisayardaki çıktı cihazlarından sonuçlar elde edilir. Çıktı cihazları, metin, sese dayalı, resim bazında ve grafik olarak elde edilen bilgileri içermektedir (Bagad, 2011).

b) Yazılım Kaynakları

Bu kaynaklar bilgisayar programlarıdır. Yazılım genellikle üç ana gruba ayrılmaktadır. Bunlar sistem yazılımları, iletişim yazılımları ve uygulama yazılımlarıdır. (Ay ve Çil, 2010).

Sistem Yazılımları: Uygulama yazılımları ve donanım arasında bir ara yüz görevini üstlenmiştir. Sistem yazılımlarının en önemli türü işletim sistemleridir.

İletişim Yazılımları: Bir ağ üzerinde farklı bilgi işlem cihazları arasında iletişimi sağlamak için kullanılan yazılımlardır.

Uygulama Yazılımları: Uygulama yazılımları, belirli bir görevi yerine getirmek için tasarlanmış bir programlama dili ya da bir araç kullanılarak oluşturulur. Paket yazılımlar, özel yazılımlar, web uygulamaları, açık kaynak kodlu yazılımlar, paylaşımlı yazılımlar ve ücretsiz yazılımlardan uygulama yazılımı olarak söz edilmektedir (Hoşcan, 2003).

c) Veri Kaynağı (Veri tabanları)

Veri tabanı bilişim sistemlerinin en önemli bileşenidir. İletişim ağı üzerinden verinin iletilmesi için, bilgisayar yazılım ve donanımı kullanılarak veri işlenir ve depolanır. Verinin etkili ve düzenli yapılması son kullanıcılar için çok önemlidir. Veri alfa nümerik, metin, resim ve ses olarak veri tabanında yer alır. Veri işlenmemiş gerçekleri ifade etmektedir. Enformasyon ise işlenmiş veridir (Golemanov ve Golemanova, 2004).

d) Ağ kaynakları

Bilgisayar sistemine kurulmuş herhangi bir ağ bazı iletişim teknolojileri tarafından birleştirilir. Bu teknolojiler LAN, WAN, MAN ve ring, star, bus ağları olarak tanımlanabilir. Telekomünikasyon ağı bir siteden diğerine bilgi transfer etmek için

kullanılan telekomünikasyon yapısını sağlar. Telekomünikasyon ağı; bilgisayarlar, işlemciler düğme (switch), iletişim ortamı ve iletişim yazılımlarından oluşmaktadır. Ağ desteği, bilgisayar, modem, işlemci kontrol yazılımları ve işletim sistemi yazılımları kaynaklarını içermektedir (Karahoca ve Karahoca, 1998).

e) Kullanıcı Kaynağı

Bunlar son kullanıcı ya da bilişim sistemi uzmanlarıdır. Son kullanıcılar, genellikle mühendis, doktor, satış elemanı, muhasebeci, yönetici ve benzeri kişilerdir. Son kullanıcılara istemci (client) denilmektedir. Bilişim sistem uzmanları ise bilişim sistemlerini işleten ve geliştiren kişilerdir. Bilişim sistemi, sistem analistleri, programcılar ve yöneticilerden oluşmaktadır. Bu uzmanlar son kullanıcının isteğine göre bilgiyi tasarlarlar. Bilişim ürünleri son kullanıcının ihtiyaçlarını karşılamaya yönelik olmalıdır (Laudon ve Laudon, 2011).

Tablo 1.1: Bilişim Sistemi Kaynakları

Donanım Kaynakları <ul style="list-style-type: none">• Bilgisayar sistemleri• Bilgisayar çevre birimleri	Masaüstü bilgisayarlar, sunucular, dizüstü bilgisayarlar, mini bilgisayarlar Klavye, fare, ekran, yazıcı
Yazılım Kaynakları <ul style="list-style-type: none">• Programlar• Yazılım	İşletim sistemi ve programlar Sistem yazılımları, uygulama yazılımları, iletişim yazılımları
Veri tabanları <ul style="list-style-type: none">• Veri türleri	Alfa nümerik veri, metin verisi, resim verisi, ses verisi
Ağ Kaynaklar <ul style="list-style-type: none">• İletişim cihazları	Fiber optik kablo, mikrodalga, kızıl ötesi, uydu
Kullanıcı Kaynakları <ul style="list-style-type: none">• Son kullanıcı• Bilgi sistemi uzmanları	Bilgi sistemini kullanan kişiler Sistem analist, programcı, sistem yöneticileri

Kaynak: (Şentürk, 2006).

1.1.3. Bilişim Sistemi Çalışmaları

Bilişim sistemi çalışmalarını; veri kaynaklarının girişleri, verinin işlenmesi, bilişim sistemi çıktıları ve veri kaynaklarının depolanması olmak üzere dört ana gruba bölerek inceleyebiliriz (Yozgat, 1998).

1.1.3.1. Veri Kaynaklarının Girişleri

Verinin girilmesi ve kayıt edilmesi süreci genellikle veri kaynakları girişini içermektedir. Veri girişleri bilgisayardan manuel olarak yapılabilir, ya da başka yollarla olabilir. El ile yazılmış kağıt üzerinden veri girişi yapılmasının yanı sıra, bazen veri girişleri ticari işlemler yapılabilir (Şentürk, 2006).

Veri girişlerinin doğru olarak yapılmasını sağlamak için çeşitli işlemler yapılabilir. Kullanıcıların veri girişinde en uygun ve verimli yöntem, görsel yönlendirilmeler, menülerin görüntüleri ve optik taramalardır. Mağazalardaki barkod prensibi ile optik tarama girişleri örnek olarak verilebilir (Özmen, 2013).

1.1.3.2. Verinin İşlenmesi

Veri analiz edilebilecek ya da rapor alınabilecek bilgilere dönüştürülmediği sürece hiçbir değeri yoktur. Bu ilgili bilgiler hesaplama ve özetleme gibi birtakım işlemlere tabi tutulmaktadır. Veri işleme süreci veriyi son kullanıcıların ihtiyaç duyduğu şekle dönüştürür. Bu süreçte veri sürekli düzeltilerek yenilenmektedir (Uwizemungu ve Raymond, 2004).

1.1.3.3. Bilişim Sistemi Çıktıları

Son kullanıcıların çeşitli formlardaki enformasyona (işlenmiş bilgi) ihtiyaçları vardır. İşlenmiş bilgi, çıktı faaliyeti olarak son kullanıcı için hazır duruma getirilir. Bunlar bilgi ürünleri (Information products) olarak adlandırılır. Bu bilgi ürünleri; mesaj, raporlar, formlar, resimler, ses ya da kağıt çıkışı olabilir (Van Der Heijden, 2009).

1.1.3.4. Veri Kaynaklarının Depolanması

Bir işletmede satışların yapılması, ürünlerin elde edilmesi ve verinin talep edilmesi çalışmaları verinin kullanılması faaliyetleridir. Elde edilen veri, işletme faaliyetlerinde yararlı olduğu sürece anlamlı ve değerlidir (Yenişmar ve Gölcü, 2003).

Veri kaynaklarından elde edilen bilgiler veri tabanlarında saklanmaktadır. Bilgiler veri tabanlarındaki veri modeline uygun bir şekilde saklanmaktadır (Şentürk, 2006).

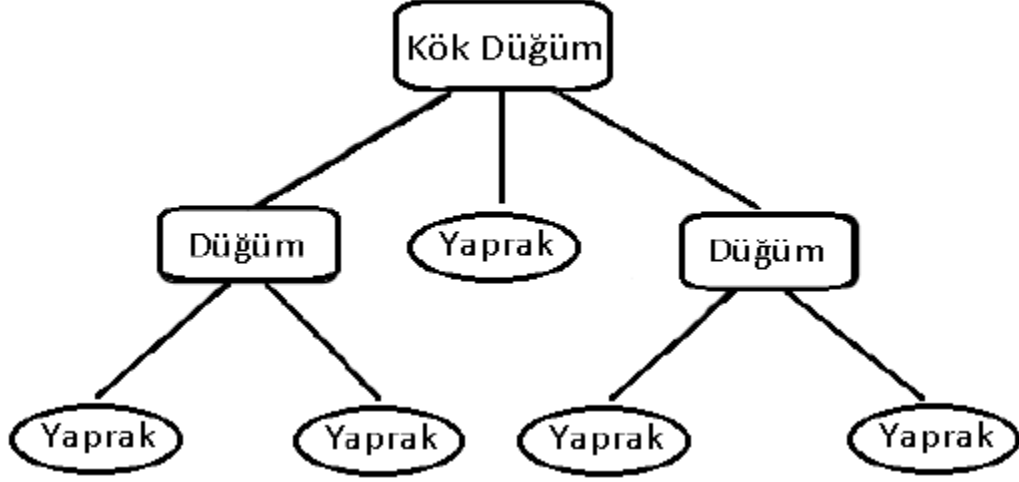
2. KARAR AĞAÇLARI

Karar ağaçları sınıflandırma problemlerinde en çok kullanılan yöntemlerdendir. Diğer yöntemlerle kıyaslandığında karar ağaçlarının yapılandırılması, anlaşılması ve yorumlanması daha kolaydır. Bunların yanında karar ağaçlarının diğer bir avantajı ise başarılı modeller üretmesidir. Karar ağaçları yönteminde sınıflandırma yapmak için elimizdeki verilerden ağaç oluşturulur ve veri setindeki kayıtlar bu ağaca uygulanır, çıkan sonuca göre kayıtların sınıflandırma işlemi gerçekleşir. Yani veritabanından elde edilen karar ağaçlarına göre hangi sınıfa ait olduğunu bilmediğimiz bir veri elimize geldiğinde oluşturulmuş olan kural dizisine göre sınıfı tahmin edilir (Silahtaroglu, 2013).

Sınıflarını bildiğimiz eğitim veri setinden tümevarım yöntemiyle ağaç şekilli karar yapılarının oluşturulması işlemine karar ağaçları denir (Dondurmacı, 2011). Bu ağaçlar, geniş hacimli veri setlerini daha küçük alt gruplara karar verme kuralları ile bölünerek oluşturur. Doğru ayrışma işlemleri ile elde edilen grupların üyeleri birbirleri ile daha benzer olmaktadır. Bir karar ağacı modeli büyük heterojen grupları belirlenmiş hedef değişkene göre yani kökten yaprağa doğru kural seti oluşturarak homojen alt gruplara ayırır (Berry ve Linoff, 2004). Karar ağaçları, hangi faktörlerin etken olduğunu belirlemek ve bu faktörlerin geçmiş ilişkilerini ortaya koymakta yardımcı olan bir yöntemdir (Kayri, 2014).

Karar ağaçları adından anlaşılabilceği üzere ağaç şeklinde gösterilebilmektedir. Bu ağaçlar şekil 2.1'de görülebileceği üzere karar düğümlerinden, dal ve yapraklardan oluşmaktadır. Karar ağaçları ile sınıflandırma süreci kök düğüm ile başlar. Niteliklerin değerine göre düğümler alt dallara ayrılır, bu süreç yaprak elde edilinceye kadar devam eder. Burada düğümler; test işlemine tabi tutulan nitelikleri gösterir ancak bilinmelidir ki tüm değişkenlerin ya da niteliklerin karar ağacında yer alması beklenmez. Dallar; teste tabi tutulan düğümlerin sonuçlarını gösterir. Elde edilen dallar ile tahmin edilecek sınıfa giden yol belirlenir. Dalın sonucunda sınıflama tamamlanmıyorsa tekrar düğüm oluşturulur. Yapraklar ise veri setindeki bir karar sınıfını temsil etmektedir.

Şekil 2.1:Karar Ağacı Örneği



Kaynak: (Silahtaroglu, 2013).

Karar ağaçlarının oluşturulması için önce girdi verisine, yani örneklere gereksinim vardır. Örnekler kategorik ya da nümerik değişkenlerden oluşabilir. Bunun yanı sıra hedef değişkene yani sınıflandırma için kullanılacak verilere ihtiyaç vardır.

2.1. Karar Ağaçları ile İlgili Literatür Taraması

Karar ağacı analizi hem tek başına hem de diğer yöntemler ile kullanılabilen bir yöntem olduğundan literatürde karar ağacı yöntemleri ile yapılmış çok sayıda çalışma mevcuttur. Karar ağacı yöntemleri; personel seçimi, yazılım programı seçimi, web sitesi seçimi, android yazılım seçimi gibi pek çok probleme uygulanmıştır. Bunlardan bazıları şunlardır:

Yıldıztepe ve Kocataş (2018), çalışmasında 2013 yılı TÜİK Hanehalkı İşgücü Araştırması (HİA) verilerini kullanarak Türkiye’de işgücünün yapısı hakkında bilgi verilmiş ve işgücü durumu için karar ağacı modelleri kurulmuştur. Çalışma sonucunda geçmiş değerlere bakıldığında 2013 yılında istihdama katılımın arttığı ve işsizlik oranının azaldığı görülmüştür. 2009 ve 2010 yıllarına baktığımızda kadınların işgücüne katılım oranının erkeklerin üçte biri kadar olduğu belirtilmiştir. Buna göre kadınların işgücüne katılım oranında bir iyileşme olduğu söylenmiştir.

Utku,Dođru ve Akcayol (2018), alıřmasında C4.5 karar ađacı ve Hoeffding ađacı algoritması kullanılarak izin tabanlı bir ktcl yazılım tespit sistemi geliřtirilmiřtir. Yapılan analiz sonuları C4.5 karar ađacı algoritmasının daha iyi sonu verdiđini gstermiřtir.

Kaya, Erkaymaz, Ayar ve zer (2017), alıřmasında ilk olarak 25 giriř, 1 ara katman ve 1 ıkıř ieren bir yapay sinir ađları (YSA) modeli oluřturulmuřtur. Bu topoloji iin veri seti apraz dođrulama yntemi ile eđitim ve test olarak iki kısıma ayrılmıřtır. Ara katmandaki nronların sayısını bulmak iin ortalama karesel hata performansı zerinde ara katman nron sayısının deđiřiminin etkisi arařtırılmıřtır. Bulgular, C4.5 karar ađacı algoritması ile seilen zneliklerin, ayrıık dalgacık dnřm ile ıkarılan zneliklere gre daha iyi sınıflandırma performansı sađladıđını gstermiřtir.

Aksu ve Karaman (2017), alıřmasında Artvin oruh niversitesinin web sitesine ait 2536 adet Google Analitik verisi weka programı kullanarak eřitli sınıflandırma algoritmaları kullanılarak sınıflandırılmıř ve en iyi sınıflandırma iřlemine yapan algoritma kullanılarak 8 adet kural reten bir karar ađacı oluřturulmuřtur. Karar ađacını oluřturduktan sonra verilen linkin ana sayfada olup olmadıđına karar verilebilmiřtir. Bylece karar ađacı retme ve link tespiti iřlemi ok daha hızlı bir Őekilde gerekleřtirilmiřtir.

Ersz, zseven ve Ersz (2017), alıřmasında cep telefonu kullanıcılarının tercihleri model, fiyat, kamera, hafıza ve ekran faktrleri zerinden arařtırılmıřtır. Elde edilen sonular kullanıcıların cep telefonu seiminde en nemli faktrn fiyat olduđunu ve bunu hafızanın takip ettiđini gstermiřtir.

Yılmaz, Durak, zdemir, Kiracı ve Tanrıverdi (2017), alıřmasında Eskiřehir ilinde havayolu ile seyahat eden kesim analiz edilmiř ve mevcut potansiyelin artmasında rol oynayan gruplar belirlenmiřtir. Seyahat amacı ve ulařım btesi ile seyahat araları arasında olan iliřkinin anlamlı olduđu grlmřtr. Veri setinde yer alan kayıtların zerine eklenecek olan yeni bireylerin nasıl bir davranıř sergileyecekleri ve hangi ulařım aracını tercih edecekleri tahmin edilebileceđi grlmřtr. Bu sayede potansiyelin dřk olduđu kesimlere ynelik olarak yapılacak harcamalar

önlenecek, kaynakların potansiyeli arttırmada rol oynayan kesime yönlendirilmesi mümkün olacağı görülmüştür.

Irmak ve Ercan (2017), çalışmasında elde edilen karar ağacı incelendiğinde hanehalkı zeytinyağı tüketimini etkileyen en önemli karakteristiğın gelir düzeyi olduđu ortaya çıkmıştır. Bu bulgular doğrultusunda Türkiye hanehalkı zeytinyağı tüketiminde gelir düzeyinin, hanehalkı reisinin eğitim durumunun, hanede küçük çocuk veya yaşlı birey olup olmamasının ve hanenin büyüklüğünün önemli ve dikkate alınması gereken faktörler olduđu söylenmiştir.

Akbal, Dogan ve Varol (2017), çalışmasında son günlerin popüler araştırma konularından biri olan telefon dolandırıcılığı üzerine bir uygulamayı geliştirilmiştir. Yapılan çalışmada, farklı bireysel özelliklere (öğrenim durumu, meslek grubu, yaş, yaşam yeri ve gelir düzeyi) sahip kişilerden telefon dolandırıcılığı üzerine elde edilen veriler karar ağaçları yöntemlerinden C4.5 algoritması ile sınıflandırılarak sonuçlar değerlendirilmiştir. Bu çalışmada elde edilen bulgular dikkate alındığında insanların dolandırıcılık faaliyetlerine maruz kalmamaları için yeterince bilgilendirilmeleri gerektiği ortaya çıkmıştır.

Yaşar (2016), çalışmasında bağımsız denetim görüşünü tahmin etmek ve etkileyen faktörleri belirlemek amacıyla veri madenciliğinin C5.0 ve CART algoritması karar ağaçları yöntemi ve birliktelik kuralları algoritmalarından GRİ algoritması yöntemi kullanılmak suretiyle, olumlu görüş dışındaki denetim görüşünü tahminine yönelik karar ve birliktelik kuralları çıkartılmıştır. Veri madenciliğinin C5.0 karar ağacı algoritması analizine ilişkin elde edilen bulgular; olumlu görüş dışındaki denetim görüşü tahmininde en etkili değişkenin Önceki Denetim Görüşü (ÖDG) olduğunu, olumlu görüş dışındaki denetim görüşü üzerinde etkili bulunan diğer önemli değişkenlerin ise sırasıyla Finansal Başarısızlık (FD), Cari Oran (DV KY) ve İşletme Büyüklüğü (ln_TV) değişkenleri olduğunu göstermiştir.

Emel ve Taşkın (2005), çalışmasında bir satış hacmi analizi için veri madenciliğinin bir sınıflama işlevi olan karar ağacı yöntemi uygulanmıştır. Yapılan C&RT karar ağacı tekniği uygulaması sonucunda harcama tutarına göre müşteriler sınıflara ayrılabilmiştir.

Sezer, Bozkır, Yağız ve Gökçeoğlu (2010), CART algoritmasının kestirim kapasitesine etkisinin incelenmesi amaçlanmıştır. Ayrıca, kaya mühendisliğinde karar ağaçlarının kestirim amaçlı kullanılabilirliği araştırılmıştır. Burada CART algoritması bir tünel açma makinesinin ilerleme hızının kestirimi amaçlı kullanılmıştır. Yapılan kestirim değerlendirmelerinde, 6 derinlikli karar ağaçları dahi 0.926 gibi son derece yüksek bir performans sağlamıştır. Derinlik sayısına bağlı olarak karar ağaçlarının kestirim performanslarında bir artış görülmüş, 8. derinlikten sonra kestirim performansı sabitlemiştir.

2.2. Karar Ağaçlarının Kullanım Alanları

Karar ağaçları sınıflama, karar teorisi, kümeleme ve tahminsel fonksiyonlarda kullanılmaktadır. Karar ağaçlarını oluşturacak verideki değişkenler (nitelik/özelliik) kategorik veya sürekli olabilirler. Eğer bağımlı değişken sürekli ise karar ağaçları regresyon ağaçları olarak adlandırılır. Eğer bağımlı değişkenler kategorik ise buna sınıflama ağacı denilmektedir. Bu farklılığa rağmen karar ağaçları benzer biçimde kurulmaktadır. Karar ağaçları tıp alanında teşhis için, botanikte sınıflama için, felsefede karar teorisi için, ekonomide ise yatırım alternatiflerini belirlemek için sıklıkla kullanılır. Karar ağaçları kurulma biçimlerine göre birbirlerinden ayrılmaktadırlar. Bazı durumlarda yukarıdan aşağı doğru kurulurken bazı durumlarda soldan sağa doğru kurulabilirler (Omitaomu, 2006).

Karar ağacı temelli analizlerin yaygın olarak kullanıldığı sahalara;

- Belirli bir sınıfın üyesi olacak elemanların belirlenmesi,
- Çeşitli vakaların yüksek, orta, düşük risk grupları biçiminde kategorilere ayrılması,
- Gelecekte gerçekleşebilecek olayların tahmin edilebilmesi için kurallar oluşturulması,
- Parametrik modellerin kurulmasında kullanılacak çok sayıdaki değişken ve veri kümesinden önemli olanlarının seçilmesi,
- Yalnızca belirli alt gruplara özgü ilişkilerin tanımlanması,
- Sürekli değişkenlerin kategorik değişkenlere dönüştürülmesi ve kategorilerin birleştirilmesi olarak sayılabilir.

Karar ağacının kullanıldığı uygulamalardan bazıları ise aşağıdaki biçimde sıralanabilir:

- Hangi demografik grupların mektup aracılığıyla yapılan pazarlama uygulamalarında yüksek cevaplama oranına sahip olduğunun belirlenmesi,
- Kredi geçmişlerinin kullanılmasıyla bireylere ilişkin kredi kararlarının verilmesi,
- İşletmeye en faydalı olan bireylerin özelliklerinin kullanılmasıyla ise alma süreçlerinin belirlenmesi,
- Tıp ile ilgili gözlem verilerinden hareketle en etkin kararların verilmesi,
- Satışları hangi değişkenlerin etkilediğinin belirlenmesi,
- Ürün hatalarına yol açan değişkenlerin belirlenmesi (Oğuzlar, 2004).

2.3. Karar Ağaçlarının Avantaj ve Dezavantajları

Karar ağacı kökenli sınıflandırma yöntemlerinin bazı avantajlarından bahsedecek olursak (Murthy, 1998 ; Mitra 2003);

- Kolay anlaşılabilir ve yorumlanabilir olmasından dolayı daha kullanışlıdır,
- Çok büyük boyutlardaki veri setine kolayca uygulanabilir,
- Ağaç modelleri çıkarımsa modellerin aksine araştırmacıdır,
- Karar ağaçları non-parametrikdir. Diğer tekniklerde olduğu gibi istatistiksel varsayımlarla uğraşmaz,
- Hiyerarşik ayrışması mevcut özelliklerinin ilişki yönünü ve önem sırasını ortaya koyar,
- Ağaç yapısından dolayı karışık bağlantılar ortaya çıkarılabilir,
- Sürekli ve kategorik niteliklerle çalışabilir,
- En çok kullanılan veri madenciliği modeli olduğu için bu alan da yayımlanmış birçok doküman ve bilgisayar programı bulunması,

Karar ağaçlarının bazı dezavantajları ise (Myatt, 2007; Öğüdücü, 2008);

- Hesaplama maliyeti: özellikle büyük veri setlerinde çok sayıda sürekli değişkenin olması durumunda karar ağacı oluşturmak diğer sınıflandırma yöntemlerine nazaran daha maliyetli olabilir,

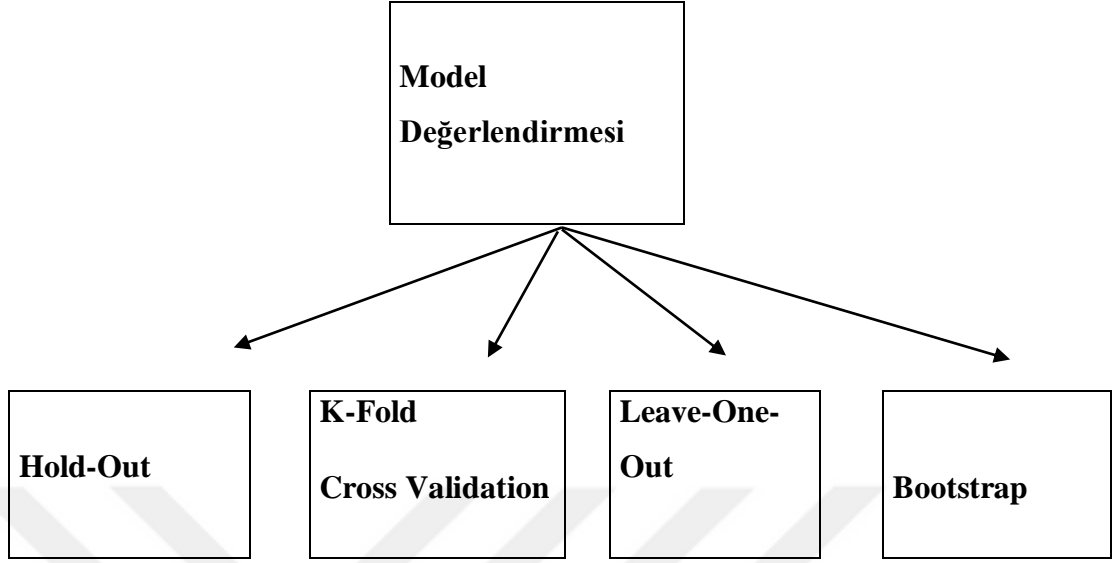
- En uygun ağacı bulmak bazen zor olabilir. Yararlı karar ağacını oluşturmaya çalışılırken birden karmaşık ve büyük ağaçlar üretilebilir,
- Girdi sayısı az olduğunda ağaç yeterli bilgiyi yakalayabilir,
- Bazı algoritmalar sadece kategorik verilerle çalışır ve sürekli değişkenlerin sınıflamasında başarılı olmayabilir,
- Öğrenme kümesinin hacmine ve değişkenlerin sayısına bağlı olarak zaman ihtiyacı ve ağaç karmaşıklıkları değişmektedir.

2.4. Sınıflandırma İşleminin Doğruluğu

Sınıflandırma sürecinde karar ağacının eğitilmesinde kullanılacak olan verilerin seçilmesi gerekmektedir. Veri madenciliği süreci aşamalarında bahsetmiş olduğumuz veri hazırlığı aşaması sonucunda elde edilmiş olan analize tabi tutulacak temizlenmiş verilere bu aşamada ihtiyaç duyulmaktadır. Ancak yöntem kısmında bahsedildiği üzere bütün veriler ağacın eğitilmesinde kullanılmaz. Veriler ikiye bölünür; biri öğrenme ya da eğitim (training, training) diğeri ise deneme ya da test (testing)'dir. Bu ikiye bölme işlemi oluşturulan ağacın doğruluğunu ölçmek için gerçekleştirilir. Bazı durumlarda veri seti bölünebilecek kadar büyük olmayabilir böyle durumlarda doğruluğu test etmek için geliştirilmiş yöntemler vardır. Yapılan işlemlerin sonunda doğruluk oranı tatmin edici düzeyde ise karar ağacı verilerin sınıflandırılmasında kullanılabilir. Aksi takdirde oluşturulan karar ağacı kuralları tahminleme işleminde kullanılamaz (Dondurmacı, 2011).

Sınıflandırma sürecinde eğitim verileri üzerinde kullanılan algoritmanın doğruluğunun belirlenmesi (model değerlendirilmesi) için geliştirilen birçok yöntem bulunmaktadır. Bu kapsamda en yaygın kullanılan yöntemler şekil 2.2'de belirtildiği gibi; Holdout, K-fold Cross-Validation, Bootstrap ve leave-one-out olarak sıralanabilir (Souza ve ark., 2002). Bu yöntemler sırasıyla aşağıda tanımlanmıştır:

Şekil 2.2: Model Değerlendirmesi (Sınıflama Doğruluğu)



Kaynak: (Souza ve ark., 2002).

2.4.1. Hold-out

Veri seti eğitim ve test kümesi olmak üzere iki ayrı parçaya ayrılır. Eğitim verilerinden elde edilen sınıflandırma modelinin performansı, test verilerinde ölçülür. Hold-out yönteminde verilerin genellikle 2/3'ü eğitim, geriye kalan 1/3'ü ise test için ayrılır. Sınıflandırmanın doğruluğu ise test kümesi üzerinden hesaplanır. Hold-out yönteminin iki eksikliği bulunmaktadır; birinci olarak bu yöntem çok fazla veri olmadığı durumlarda test için verileri belirli oranlarda ayırmak mümkün olmayabilir. Diğer eksikliği ise isabetli veri ayırımı yapılamadığında yanıltıcı hata oranının bulunabilmesi ihtimalidir. Yani şöyle bir ikilem ortaya çıkmaktadır iyi bir sınıflandırma yapabilmek için mümkün olduğunca veriyi daha çok öğrenme verileri üzerinde uyguluyoruz ancak iyi bir hata oranı elde etmek için ise verileri daha çok test verilerinde kullanmak isteriz. Bu ikilemden dolayı diğer yöntemler geliştirilmiştir (Witten ve ark., 2011).

2.4.2. K-fold cross validation

Çapraz geçerlilikten türetilen “K-katlı çapraz doğrulama” olarak adlandırılmaktadır. Çapraz geçerlilik tüm verileri kullanır. İlk veri seti üzerinden bir model oluşturulur ikinci model üzerinde bu model uygulanır ve hata oranı tespit edilir. Daha sonra

ikinci veri seti üzerinden model oluşturulur daha sonra ise birinci model üzerinden hata oranı tespit edilir ve bu iki hata oranının ortalaması alınarak hata oranı elde edilir (Babadağ, 2003).

K-katlı çapraz doğrulamada ise veri seti k adet alt kümeye ayrılır. Uygulamalarda genel olarak $k=10$ alınır ve alt kümeler hemen hemen birbirlerine eşittir. Alt kümelerin belirlenmesinin ardından test kümesinin yeri değiştirilmek suretiyle hata oranları sırayla elde edilir. Verilerin 10 gruba bölüdüğü varsayıldığında ilk grup test için kalan dokuz grup hep birlikte model oluşturulmada kullanılır. Model verinin %90'ı tarafından oluşturulur, bu işlem her seferinde test grubunun değiştirilmesiyle 10 kez tekrar edilir. Elde edilen 10 adet bağımsız hata değerinin ortalaması, kurulan modelin hata oranı olmaktadır. Uygulamada ve teoride oluşturulması gereken grup sayısının 10 olması yönünde bir görüş oluşmuştur. 10'lu çapraz geçerlilik standart bir yöntem olmaya başlasa da, yine de oluşturulacak grup sayısı analist tarafından belirlenir (Babadağ, 2003).

Veri tabanlarının birkaç bin ya da daha az satırdan meydana gelmesi durumunda kullanılması uygundur. Bu yöntemin sakıncası ise maliyetin k 'ya bağlı olarak artmasıdır.

2.4.3. Leave-one-out

Leave-one-out yöntemi, K-katlı çapraz geçerlilik yönteminin özel halidir. Bu yöntemdeki K sayısı veri kümesindeki örnek sayısına yani N 'ye eşittir. Model $N-1$ adet örnek eğitim verisi olarak kullanılır, dışarıda kalan 1 örnek ise test için kullanılır. Bu işlemi her örnek bir kere sınama için kullanılacak şekilde gerçekleştirilir (Özbudak 2009). Model başarısı ise yani sınıflamanın doğruluk oranı denemeler sonucunda elde edilen hataların ortalaması alınarak bulunur (Vercellis, 2009). İki sebepten dolayı çok kullanışlıdır. Birincisi; her durumunda mümkün olduğunca çok miktarda veri eğitim için kullanılır bu da büyük olasılıkla doğru bir sınıflama yapma şansını yakalar. İkincisi yöntem deterministiktir: Rastgele örnekleme yoktur, 10 defa tekrarlama veya hepsini tekrarlama gibi bir durum da yoktur. Aynı sonuç her seferinde gözlemlenir (Witten ve ark., 2011).

Ancak örnek sayısı kadar yani N kere test yapılması gerektiği için çok büyük veri setlerinde kullanılması mümkün olmayabilir. Ayrıca çok fazla deneme olabileceği için yüksek hesaplama maliyeti dezavantajları arasında sayılabilir (Witten ve ark., 2011).

2.4.4. Bootstrap

Son olarak bahsedilen model hata tahmin yöntemi olan bootstrap, yerine koyarak istatistiksel örnekleme işlemine dayanır (Witten ve ark., 2011). Örnekleme, anakütle parametrelerinin tahmin edilmesinde kullanılan istatistiksel yöntemdir. Parametrik istatistiksel yöntemlerin uygulanmadığı durumlarda test edilen hipotezdeki istatistiğin anlamlılığını ortaya koymak için kullanılan tekrarlı örnekleme yöntemleri içinde en iyi bilinen yöntemlerden biri Bootstrap yöntemidir (Dondurmacı, 2011). Genellikle çok küçük veri setlerinde uygulanır, model tüm veri setinden oluşturulur. Daha sonra bootstrap olarak adlandırılan çok sayıda örneklem çekilir. Her gözlem örnekleme seçildikten sonra yerine konur ve tüm örneklem seçilene kadar işlem tekrar edilir. Bir kayıt birden fazla veri setinde yer alabilir. Model bu veri setinden oluşturulur ve hata oranı hesaplanır. En az 200 olmak üzere bazen de 1000'i geçecek şekilde birçok bootstrap örnekleme yaratılır (Kavurgacı ve ark., 2011). Model için tüm veri setinden oluşan son hata tahmin değeri bütün bootstrap örneklemlerinin hata tahmin değerinin ortalaması alınarak hesaplanır (Babadağ, 2003).

2.5. Sınıflandırma Performanslarının Değerlendirilmesi

Veri madenciliği yöntemlerinde kullanılmak istendiğinde pek çok sınıflandırma algoritması var olduğundan, hangi algoritmanın seçilmesi için geliştirilmiş birkaç kriter bulunmaktadır. Bunlar karışıklık (confusion) tablolarından hesaplanan; doğruluk (accuracy), hata oranı (error rate), anma (recall), kesinlik (precision) ve F-ölçütü (F-score) ile ROC eğrileri olarak gösterilebilir (Öğüdücü, 2008; Dalkılıç ve Dalkılıç 2015).

2.5.1. Karışıklık Tabloları

Modelin başarısı, doğru ve yanlış sınıflandırma işleminin gerçekleştirilmesi ile alakalı bir durumdur (Coşkun, 2010). Şekil 2.1'de yer alan karışıklık matrisinden

doğruluk (accuracy), hata oranı (error rate), anma (recall), kesinlik (precision) ve F-ölçütü gibi değerler hesaplanmaktadır.

Tablo 2.1: Karışıklık Matrisi

Tahmin Edilen	Pozitif	Negatif
Gerçek Pozitif	TP(Doğru Olumlular)	FN (Yanlış Olumsuzluklar)
Durum Negatif	FP (Yanlış Olumlular)	TN (Doğru Olumsuzluklar)

Kaynak: (Witten ve ark, 2011).

$$\text{Doğruluk} = \frac{TP + TN}{\text{Hepsi}} \quad (2.1)$$

$$\text{Hata Oranı} = \frac{FN + FP}{\text{Hepsi}}$$

Doğruluk ve hata oranları yardımı ile modelin başarı durumu incelenir (Şeker, 2013). Doğruluk en basit değerlendirme ölçütüdür (Öğüdücü, 2008).

$$\text{Kesinlik} = \frac{TP}{TP + FN} \quad (2.2)$$

$$\text{Anma} = \frac{TP}{TP + FN}$$

Modellerin karşılaştırılması durumunda “kesinlik” ve “anma” değerleri yüksek olan model daha iyi bir sınıflandırıcıdır denilebilir (Öğüdücü, 2008). “kesinlik” ile “anma” arasında ters orantı mevcuttur.

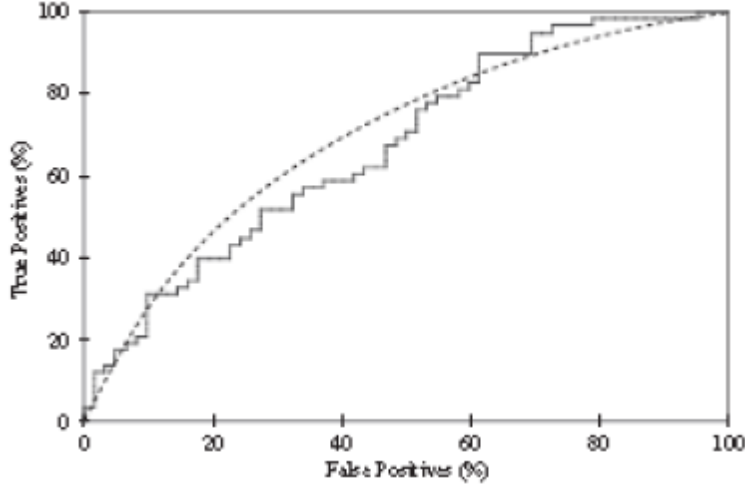
$$F - \text{ölçütü} = \frac{2 * \text{kesinlik} * \text{anma}}{\text{kesinlik} + \text{anma}} \quad (2.3)$$

F ölçütü “anma” ve “kesinlik” değerlerinin harmonik ortalaması ile bulunur (Dalkılıç ve Dalkılıç, 2015).

2.5.2. ROC Eğrileri

ROC (Receiver Operating Characteric) eğrileri farklı sınıflandırma tekniklerinin karşılaştırılması için kullanılır. ROC eğrisi yatay eksenle yanlışlıkla pozitif (FP) düşey eksenle ise doğrulukla pozitif (TP) değerlerini gösterir.

Şekil 2.3: ROC Eğrisi



Kaynak: (Witten ve ark., 2011).

ROC eğrisinin çiziminde “doğru pozitif oranı” (TPR-x eksen) ve “yanlış pozitif oranı” değerlerine ihtiyaç vardır. Bunlar;

$$TPR = \frac{TP}{TP + FN}$$
$$FPR = \frac{FP}{TN + FP} \quad (2.4)$$

Eğri üzerindeki her bir nokta bir sınıflandırıcının oluşturduğu alana karşılık düşer.

TP ve FP değerleri ile çizilen ROC eğrisi grafikte yer alan değerler için;

(0,0): Bütün değerler negatif sınıflandırıldığını

(1,1): Bütün örnekler pozitif sınıflandırıldığını,

(0,1): Bütün negatif ve pozitif değerlerin doğru sınıflandırıldığını (en ideal durum),

(1,0): bütün negatif ve pozitif değerlerin hatalı sınıflandırıldığını ifade eder.

Bu durumlardan da anlaşılacağı üzere ROC eğrisi altındaki alanı 1'e daha yakın olan modelin sınıflandırma işleminde kullanılması tercih edilir (Dondurmacı, 2011).

2.6. Karar Ağacı Algoritmaları

2.6.1. 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).

2.6.2. CHAID Algoritması (Kass, 1980: 119-127)

Yetmişli yılların başında başlayan, karar ağaçlarını oluşturmak için araştırmacılar uygulamalı istatistikte çeşitli yöntemler geliştirdiler, örneğin: AID (Sonquist v.d., 1971), MAID (Gillo, 1972) , ve CHAID (Kass, 1980). CHAID (Chisquare-Automatic – Interaction –Detection) orijinal olarak sadece nominal öznitelikleri (değişkenler) ile işlem yapacak şekilde geliştirildi. Her bir girdi özniteliği olan a_i için, CHAID hedef özniteliğine göre en az anlamlı derecede farklı olan V_i 'de değerler çifti bulur. Anlamlı farklılık, bir istatistik testinden elde edilen p değeri yolu ile ölçülür. Kullanılan istatistik testi, hedef özniteliğin türüne bağlıdır. Eğer hedef nitelik sürekli ise, bir F testi kullanılır.

Eğer değişken (öznitelik) nominal ise, bir Pearson chi-squared testi kullanılır. Eğer değişken (öznitelik) sıralı ise, o zamana olabilirlik-oran (likelihood-ratio test) testi kullanılır.

Seçilen her bir çift için, CHAID elde edilen p -değerinin belli bir birleştirme eşik değerinden büyük olup olmadığını kontrol eder. Eğer cevap pozitif ise, değerleri birleştirir ve ilave potansiyel çiftleri birleştirmek için arama yapar. Bu süreç anlamlı çiftler bulunmayana kadar tekrarlanır. Böylece halihazırdaki düğümü bölmek için kullanılan en iyi girdi özniteliği (değişken) seçilir, öyle ki her bir çocuk (alt) düğüm bir seçilen özniteliğin değerlerinin homojen bir grubundan oluşur. Unutmamak gerekir ki, eğer en iyi girdi özniteliğinin düzeltilmiş p değeri belli bir bölme eşik değerinden küçük değil ise bölme gerçekleşmez. Aşağıdaki durumlardan bir tanesi karşılandığı zaman bu prosedürde durur:

- Maksimum ağaç derinliğine ulaşılır.
 - Üst (ebeveyn) düğüm olmak için düğümdeki minimum sayıda duruma ulaşılır, böylece daha fazla bölünmez.
 - Alt (çocuk) düğüm olmak için düğümdeki minimum durumlar sayısına ulaşılır.
- CHAID eksik değerlerin hepsini tek bir geçerli kategori gibi muamele ederek işlem yapar. CHAID budama işlemi gerçekleştirmez.

2.6.3. ID3 Algoritması (Quinlan, 1986: 81-106)

Bu algoritma çok basit bir karar ağacı algoritması olarak kabul edilir (Quinlan J. , 1986). ID3 bilgi kazancını (information gain) bölme kısıtları (splitting criteria) olarak kullanır. Ağacın büyümesi tüm örneklerin tek bir hedef özelliğın değerine ait olduğu zaman durur ya da en iyi bilgi kazancı sıfırdan büyük olmadığı zaman durur. ID3 herhangi bir budama prosedürünü uygulamaz, ayrıca sayısal öznitelikler veya eksik değerler ile işlem yapmaz.

2.6.4. C4.5 Algoritması

ID3 nin bir evrimidir ve aynı yazar tarafından (Quinlan J. , 1993) sunulmuştur. Bölme kısıtları olarak kazanç oranı (gain ratio) kullanır. Bölme işlemi bölünmesi gereken örneklerin sayısı belli bir eşik değerinden düşük olduğu zaman sonlanır. Ağacın büyüme işlemi gerçekleştikten sonra hata tabanlı budama (Error –based pruning) işlemi başlar. C4.5 sayısal öznitelikleri (değişkenleri) işleyebilir. Düzeltilmiş kazanç oranı ölçütü (corrected gain ratio) kullanarak eksik değerler içeren bir eğitim kümesinden indükleyebilir (Rokach ve Maimon, 1993: 181).

2.6.5. 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.

2.6.6. QUEST Algoritması (Loh ve Shih, 1997: 815-840)

QUEST (Quick, Unbiased, Efficient, Statistical Tree) algoritması tek deęişkenli ve doğrusal bileşim bölmeleri destekler (Loh ve Shih, 1997). Her bir bölme için, ANOVA F-testi yada Levene's test (sıralı ve sürekli öznitelikler için) ya da Pearson's Chi-Square (nominal öznitelikler için) kullanılarak her bir girdi öznitelięi (deęişkeni) ile hedef öznitelięi (baęımlı deęişken) arasındaki baęlantılar hesaplanır. Eęer hedef deęişken (öznitelik) çok terimli ise, iki süper sınıf oluşturmak için iki – ortalamalı kümeleme (two –means clustering) kullanılır. Hedef öznitelikli en yüksek baęlantıyı (association) elde eden öznitelik bölme için seçilir. Kuadratik diskriminant analizi (QDA) girdi öznitelięi için optimal bölme noktasını bulmak için uygulanır. QUEST algoritmasının önemsiz yanlılıęa sahiptir (bias) ve ikili ağaęlar saęlar. Aęaęları budamak için on – katlı çapraz - doęrulama (ten - fold cross - validation) (Rokach ve Maimon, a.g.e. : 182) metodunu kullanır.

QUEST bölünmüş alan seçimi (split - field selection), bölünmüş nokta seçimi (split - point selection) ayrı olarak ilgilenir. QUEST'te ki tek deęişkenli bölme, yaklaşık olarak yansız alan seçimi gerçekleştirir. Yani, eęer tüm belirleyici (predictor) alanlar hedef alana göre aynı derecede bilgilendirici ise, QUEST belirleyici alanlardan herhangi bir tanesini eşit olasılıkla seçer. QUEST algoritması, CART algoritmasının birçok avantajlarını tanımaktadır, fakat CART'ta olduęu gibi ağaęlar hantal olabilir. Aęacın boyutunu küçültmek için otomatik maliyet – karmaşıklık budama (automatic cost –complexity pruning) teknięi, QUEST ağacına uygulanabilir. QUEST eksik deęerleri halletmek için vekil (surrogate) bölme kullanır (SPSS, Clementine 12.0 Algorithms Guide. Integral Solutions Limited, 2007: 56)

2.6.7. CART (C&R Tree) Algoritması (Breiman vd., 1984: 216-264)

CART (Classification and Regression Trees) açılımı Sınıflandırma ve Regresyon Aęaęları olarak tanımlanır. CART, 1984 tarihli bir monograf olup Leo Breiman, Jerome Friedman, Richard Olshen and Charles Stone (BFOS) tarafından beraber yazıldı. Bu algoritma yapay zeka, makine öğrenimi, parametrik olmayan istatistik ve veri madencilięi için çok önemli bir dönüm noktasını temsil eder.

CART karar ağacı tekrarlamalı ve ikili bir bölme prosedürüdür. Baęımlı ve baęımsız deęişkenler nominal ya da sürekli olabilir. Veri ham hali ile işlenir. Kök

düğümünden başlanılarak veri, iki çocuk düğüme (alt düğüme) ve daha sonra her bir çocuk düğüme tekrar iki torun çocuk düğümlerine ayrılır. Böylece ağaçlar durdurma kuralı olmaksızın en yüksek büyüklüğe ulaşır. Aslında, ağaç büyüme süreci veri eksikliğinden dolayı artık bölünme gerçekleşmez. Daha sonra maksimum büyüklükteki ağaç, yeni olan maliyet karmaşıklığı budama metodu yolu ile geriye doğru, yani kök'e doğru budama yapılmaya başlanır. Budama işlemi yapılacak olan sonraki bölünme eğitme verisi üzerindeki ağacın toplam performansına faydası en az olan bölünmedir.

CART algoritmasının amacı, sadece bir ağaç ortaya çıkarmaktan ziyade, sıralı iç içe olan, her biri en uygun (optimal) ağaç adayları olan budanmış ağaçlar oluşturmaktır. "Uygun büyüklükteki" veya "dürüst" bir ağaç bağımsız test verisindeki budama serisinde her bir ağacın tahmin performansının hesaplanması ile belirlenir. C4.5 'e kıyasla, CART ağaç seçiminde dahili (eğitme verilerine dayalı) performans ölçütü kullanmıyor. Bunun yerine, ağaç performansı her zaman ya bağımsız test verileri üzerinde ya da çapraz - sağlama yolu ile ölçülür ve ayrıca ağaç seçimi sadece test veriye dayalı hesaplama sonrasında devam eder. Eğer test veya çapraz - sorgulama uygulanmamış ise, CART serisindeki hangi ağacın en iyi olduğu konusunda agnostik kalır. Bu yöntem eğitme verileri ölçütlerine dayanacak şekilde öncelikli model oluşturan klasik istatistik veya C4.5 gibi metotlara çok zıttır.

CART mekanizması opsiyonel otomatik sınıf dengelemesi ve otomatik eksik değer (missing value) işlemlerini içermektedir. Ayrıca CART, maliyet- duyarlı öğrenmeye, dinamik özellik inşasına ve olasılık ağaç tahminlerine fırsat tanımaktadır. Son raporlar yeni nitelikli önem sıralaması içermektedir. Farklı çapraz - sağlama klasörlerindeki ağaçların terminal düğümlerin sayısı ile hizaya gelemeyebileceğini farz edersek, CART yazarları budama dizisinde her bir ağaç için performans ölçümünde çapraz - sağlamanın nasıl yapıldığının göstererek çığır açtılar (Dan, 2009: 181).

Ağaç diyagramının en tepesindeki olan kök düğüme tüm eğitme verilerini (training data) içermektedir. Bağımlı değişkene ait tüm verileri burada toplanmaktadır. Eksik veri olsun veya olmasın, bağımlı değişkene ait her bir örnek bağımsız değişken nitelikleri üzerinde veriler içermektedir. CART mevcut en iyi ayırıcıyı bulmak için

veriyi arayarak başlar, her bir tahmin deęişkeninin öznitelik – deęer çiftini bu tahmin deęişkeninin bölme iyilięi (goodness of split) için test eder. CART ile elde edilen örneklere baktığımızda, ağaçların dallanması (ayrışması) iki düęümde meydana gelmektedir, bundan dolayı CART iki bileşenli dallandırıcı kullanmaktadır.

Bütün bir ağaç oluşturmak için, CART her iki çocuk düęüm’de tasvir edilen ayrışma ya da dallanma süreçlerini tekrar ederek kök’ün torunlar oluşturur. Torunlar dallanarak torun çocuklar oluşturur ve bu süreç veri yetersizliğinden dolayı ayrışmanın imkansız olduęu ana kadar devam eder. Ayrışmanın son bulduęu en uçtaki düęümlere terminal düęüm adı verilir.



3. DESTEK VEKTÖR MAKİNELERİ (DVM)

DVM temel olarak ikili sınıflandırmalarda kullanılan bir öğrenme tekniğidir. DVM’de asıl amaç çok boyutlu bir veriyi bir hiper düzlem yardımıyla tamamen iki ayrı sınıfa ayırmaktır. Gerçek hayat verilerinde ise genellikle veri doğrusal bir düzlem ile ayrılmaz. Bu durumda veriler kernel (çekirdek) fonksiyonları kullanılarak, verilerin doğrusal olarak ayrılabilirdiği yüksek boyutlu bir uzaya taşınır. Verileri yüksek boyutlu uzaya taşımak genellikle hesaplama hatalarına ve çoklu uyum (overfitting) sorununa neden olmaktadır. DVM yönteminde yüksek boyutlu uzayda elde edilen veriler direk olarak ele alınmaz, yalnızca o uzayda ortaya çıkan iç çarpım formülleri gereklidir. Böylece hem hesaplama hem de çoklu uyum sorunları elimine edilmiş olur (Campbell ve Ying, 2011). Ayrıca, diğer öğrenme metotlarından üstünlüğü DVM’nin açık olarak VC boyutunun hesaplanabilmesidir. VC boyutu, bir sistemin bilinmeyen veri üzerindeki başarı olasılığıdır. Örneğin yapay sinir ağları ile ilgili böyle bir ölçüt yoktur. Bütüne bakıldığında, DVM teorik olarak sağlam temelleri olan, sezgisel olarak güçlü ve pratikte başarılı olan bir öğrenme tekniğidir. DVM aynı zamanda, sistemde öğrenilen veriler sınıflandırılmamış, gerçek nümerik değerler ise regresyon çalışmaları için de kullanılmaktadır (Campbell ve Ying, 2011). DVM’yi etkin bir şekilde kullanmak DVM’nin nasıl çalıştığını iyi bir şekilde anlamayı gerektirir. Bir DVM eğitilirken uygulayıcı birçok konuda doğru kararlar vermelidir. Örneğin en başta veri nasıl işlenmeli, hangi çekirdek fonksiyonu kullanılmalı, DVM’de hangi parametreler kullanılmalı sorularının cevapları uygulayıcının kararına bağlıdır. Doğru yapılmamış seçimler algoritmanın performansının düşük olmasına sebep olmaktadır.

DVM iki anabaşlık altında toplanabilir:

1. Doğrusal DVM
2. Doğrusal olmayan DVM

3.1. Doğrusal Destek Vektör Makineleri ile İlgili Literatür Taraması

Destek Vektör Makineleri analizi hem tek başına hem de diğer yöntemler ile kullanılabilen bir yöntem olduğundan literatürde Destek Vektör Makineleri yöntemleri ile yapılmış çok sayıda çalışma mevcuttur. Destek Vektör Makineleri yöntemleri bir çok çalışmaya uygulanmıştır. Bunlardan bazıları şunlardır:

Yılmaz (2016), çalışmasında kardiyotokogram verisinden fetal iyilik halinin belirlenmesi için Temel Bileşen Analizi ve En Küçük Kareler Destek Vektör Makineleri yöntemleri üzerinde temellendirilmiş bir karar destek sistemi önerilmiştir. Sistemin başarımı 10 katlı Çapraz Doğrulama tekniği kullanılarak incelenmiştir. Yapılan deneyler ışığında; bu çalışmada önerilen Temel Bileşen Analizi ve En Küçük Kareler Destek Vektör Makineleri yöntemlerine dayanan sistemin kardiyotokogram verisinden fetal iyilik halinin tespitinde başarılı bir karar destek sistemi olarak kullanılabilceği sonucuna varılmıştır.

Kavzoğlu ve Çölkesen (2010), çalışmasında DVM'lerin kullanımında kritik öneme sahip kernel fonksiyonlarının sınıflandırma sonuçlarına yani performansına olan etkileri detaylı şekilde analiz edilmiştir. Bu analiz sırasında dört farklı kernel fonksiyonu optimum parametre seçimi yapılarak çalışma sahası olarak belirlenen Kocaeli iline ait Landsat ETM+ görüntüsünün sınıflandırılmasında kullanılmıştır. Sonuç olarak; yapılan bu çalışma, DVM sınıflandırıcısının optimum kernel fonksiyonları ve parametre değerlerinin seçimiyle uydu görüntülerinin sınıflandırılmasında etkin bir yöntem olduğunu göstermiştir.

Ayhan ve Erdoğan (2014), çalışmasında rassal blok deney tasarımıyla veri setlerinin farklılıklarından doğacak DVM performansları üzerindeki etkilerinin yok edilmesi amaçlanmıştır. Sonuç olarak en başarılı performansa sahip çekirdek fonksiyonunun radyal tabanlı fonksiyon olduğu ve polinomial çekirdek fonksiyonu ile performansları arasında bir farklılık olmadığı kanıtlanmıştır.

Güner ve Çomak (2011), çalışmasında mühendislik fakültesinde 2007-2008 eğitim - öğretim yılında matematik I dersini veren beş öğretim üyesi açısından temel matematik bilgi ve becerisi yetersiz olarak gelen öğrenciler Destek Vektör Makineleri yöntemini kullanarak tahmin edilmiştir. 2007 yılında Pamukkale Üniversitesi mühendislik fakültesine başlayan 434 öğrencinin üniversiteye giriş

sınavı sonuçlarına ait verilerin kullanıldığı araştırma sonuçlarına göre; öğrencilerin matematik, fen bilimleri, Türkçe testlerinin sonuçları ile lise mezuniyet başarı puanlarının, matematik I dersindeki başarılarını tahminde önemli rol oynadığı bulunmuştur. 289 öğrencinin verisi ile makine eğitilmiş, 145 öğrencinin verisi ise test için kullanılmış ve matematik I dersinden geçen öğrencilerin % 86'sı doğru olarak tahmin edilmiştir.

Sezer, Erçil ve Keskinöz (2005), çalışmasında zaman ve belleğin bileşimsel (kombinezon) patlaması olmaksızın yüksek dereceden istatistikleri kullanan bir nesne tanıma tekniği önerilmiştir. Önerilen yöntem literatürdeki iki gözde yöntem olan Bağımsız Bileşen Analizi (BBA) ve Destek Vektör Makinesi (DVM)'nin kaynaşımıdır. İmgelerdeki artıklığı gidermek ve her imge için daha düşük boyutlu öznelik vektörleri elde etmek için BBA'yı ve sonrasında BBA adından gelen bu öznelik vektörlerini sınıflandırmak için DVM'nin kullanması önerilmektedir. Coil-20 veritabanı ve kendi ürettikleri bir 2B üretim nesnelere veritabanı için deney sonuçları verilmiştir.

Erişti, Uçar ve Demir (2007), çalışmasında Dalgacık Dönüşümleri (DD'ler), Destek Vektör Makineler (DVM'ler), 6leri Beslemeli Yapay Sinir Ağları (6YSA'lar) ve Olasılıksal Sinir Ağları (OSA'lar) kullanılarak geçici rejim bozulmalarının sınıflandırılması için bir yaklaşım sunulmuştur. DD'ler ile sınıflayıcıların başarımını artırmak ve eğitim süresini azaltmak için özellik çıkarımı yapılarak giriş uzayının boyutu azaltılmıştır. Benzetim sonuçlarından DVM sınıflayıcının en yüksek başarıyı ürettiği gösterilmiştir.

Bilişik (2011), çalışmasında destek vektör makinesi ve çoklu regresyon yöntemlerinin talep tahminlerinin doğruluğu analizinde, çapraz doğrulama tekniğine ait korelasyon katsayısı OMH, HKO, HKOY ve OMYH ölçümleri karşılaştırılmıştır. Dolayısıyla, destek vektör makinesi yöntemi ile elde edilen talep fonksiyonları kestirimleri kullanılarak getiri fonksiyonları elde edilmiş ve elde edilen bu getiri fonksiyonları da kısıtlı ve kısıtlı olmayan kapasite koşullarını altında doğrusal olmayan programlama ile analiz edilmiştir. Sonuç olarak; hem kısıtlı hem de kısıtlı olmayan kapasite koşullarının her biri için getiriye en büyükleyen optimal fiyatlar bulunmuş ve grafik bazlı analizler yapılmıştır.

Köse, Güraksın ve Deperlioğlu (2015), çalışmasında GOA tabanlı bir DVM ile oluşturulan, hibrit bir Yapay Zeka sistemi kullanılarak diyabet tespiti gerçekleştirilmiştir. Kurulan sistem yapısına göre, DVM'nin eğitimi esnasında, RBF çekirdek fonksiyonun parametresi olan, sigma (σ) parametresinin belirlenmesi için GOA işe koşulmuştur. Bu çerçevede kurulan yaklaşımın etkinliği Pima Yerlileri diyabet veri seti üzerinde yapılan sınıflandırma süreci ile değerlendirilmiştir. Süreç ile elde edilen bulgular, önerilen GOA-DVM sisteminin diyabet tespitinde yeteri düzeyde etkinliğe sahip olduğunu göstermiştir.

Demir ve Ertürk (2007), çalışmasında piksel değerleri için dalga boyuna bağlı spektrum bilgisine sahip olan hiperspektral görüntülerin destek vektör makinaları ile sınıflandırılması sonucu elde edilen sınıflandırma başarımı, hiperspektral görüntüye dalgacık uzayında gürültü azaltımı yöntemi uygulanarak önemli ölçüde artırılmıştır. Hiperspektral görüntülerin DVM ile sınıflandırılması öncesi gürültü azaltımı uygulanması yöntemi ilk defa bu çalışma kapsamında incelenmiştir ve sınıflandırma başarımını önemli ölçüde arttırdığı gösterilmiştir.

Çelik, Atalay ve Bayer (2014), çalışmasında sismik darbe verileri kullanılarak depremin önceden tahmin edilmesi için karar sistemleri oluşturulmuştur. Kömür madeninden alınan sismik darbe verileri kullanılarak sınıflandırma algoritmaları ile sınıflandırmaya tabi tutulmuştur. Yapay Sinir Ağları ile erken tespit %83 oranında başarı saptanmıştır. Destek Vektör Makineleri ile erken tespit %91 oranında erken tespit saptanmıştır. Hata paylarının yüksek çıkmasında; başka veri setleri ile karşılaştırmalar yapılamaması ve tehlikeli durum verilerinin az olması sebep olmuştur.

Yabanova ve Yumurtacı (2018), çalışmasında dinamik olarak tartılan yumurtaların Gıda Tarım Hayvancılık Bakanlığı Türk Gıda Kodeksi Yumurta Tebliğinde belirtilen ağırlık sınıflarına göre tasnif edilmesi destek vektör makineleri kullanılarak doğrudan ham veriyle gerçekleştirilmiştir. Bu çalışmada yumurtanın ölçüm platformuna düşmesinden itibaren alınan 11 adet ağırlık verisi bir işleme tabi tutulmadan doğrudan DVM sınıflandırıcısına uygulanmıştır. Çıkışta ise ağırlık sınıflarına göre verilen etiketlerden (1, 2, 3, 4) biri elde edilmiştir. Temelde iki sınıfın ayırımında kullanılan DVM yönteminin çoklu sınıflandırma işleminde kullanılan BKB yöntemi Radyal tabanlı kernel fonksiyonuyla birlikte dört grubun ayırımında kullanılmıştır.

DVM modelinin oluşturulması sırasında eğitim işlemi için 29 adet yumurtanın ağırlık verisi kullanılırken modelin test işleminde 14 adet yumurtanın ağırlık verisi kullanılmıştır. Test işlemi sonucunda tüm yumurtalar verilen ağırlık grup etiketlerine göre doğru olarak sınıflandırılmıştır. Önerilen yöntemle, yumurta ölçüm platformundan çıkmadan ağırlığına uygun sınıf ölçüm verisine bir ön sinyal işleme yöntemi uygulanmadan doğrudan DVM ile tam olarak tespit edilebilmiştir.

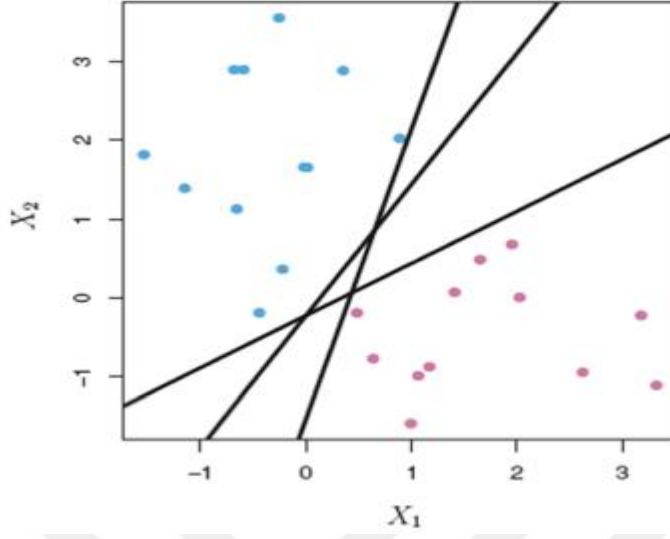
3.2. Doğrusal Destek Vektör Makineleri

DVM yapısını anlamak için en basit YRM yapısını ele alınırsa, bu yapıda iç içe geçmiş alt kümeler bulunmayacaktır. Bu durumda YRM içerisinde özet olarak anlatılmış olan ilk iki adım birleşmiş olur. İlk olarak eğitim hatası olmadan verinin bir hiper düzlem ile iki ayrı gruba ayrılabilirdiği durumu değerlendirmek konuyu daha rahat anlamak için yardımcı olacaktır. Eğer veri böyle bir niteliğe sahipse veri doğrusal olarak ayrılabilir denir ve veriyi ikiye ayıran hiper düzleme ise ayırma hiper düzlemi adı verilir. İki boyutlu uzayda veriyi ayıran düzlem doğru olarak adlandırılırken çok boyutlu uzaylarda hiper düzlem adı verilmektedir (Cristianni ve Shawe-Taylor,2000).

3.2.1. Doğrusal Ayrılabilir Durum (Sert Sınır)

DVM'nin en basit ve ilk olarak tanıtılan modeli doğrusal ayrılabilir destek vektörlerdir. Yalnızca uygun seçilmiş bir kernel fonksiyonuna bağlı özellik uzayında doğrusal olarak ayrılabilir veriler için uygundur. Ancak bu durum gerçek hayat problemleri için çoğu zaman geçerli değildir. Yine de DVM tanımını anlamak için ilk önce en basit DVM'yi anlamak gereklidir. Eğer veri bir hiper düzlem yardımıyla hatasız bir şekilde ayrılabiliriyorsa, burada veriyi ayırabilecek sonsuz sayıda hiper düzlem bulunabilir demektir. Bunun sebebi var olan hiper düzlemin kolaylıkla bir miktar aşağı yukarı hareket ettirilip döndürülerek yine de hiç bir gözlemin üzerinden geçmeyecek şekilde veriyi ayırabilecek olmasıdır (James vd.,2013).

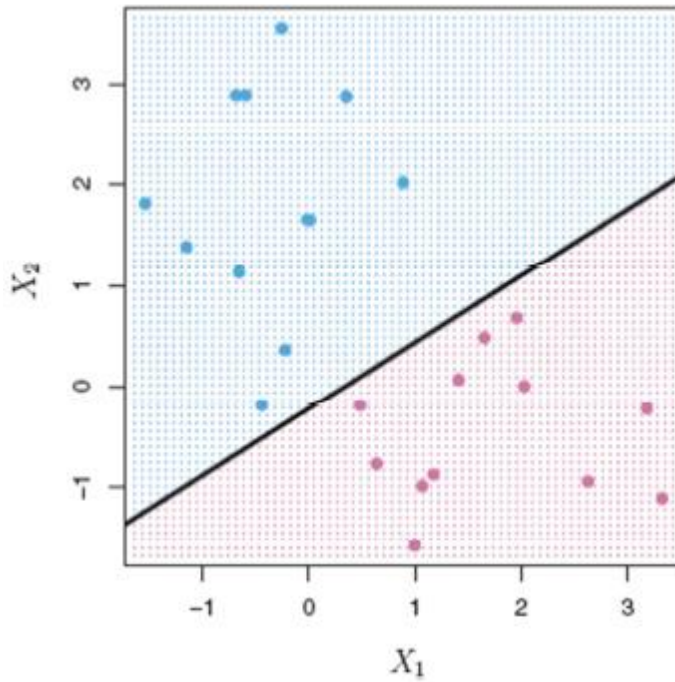
Şekil 3.1: Olası Hiper Düzlemler



Kaynak: (James vd.,2013).

Şekil 3.1’de veriyi ikiye ayıran üç adet hiper düzlem gösterilmiştir. Ancak bunlardan yalnızca bir tanesi maksimum ayırma başarısına sahiptir. Şekil 3.2’de gösterilen ayırma düzlemi maksimum ayırma başarısına sahiptir.

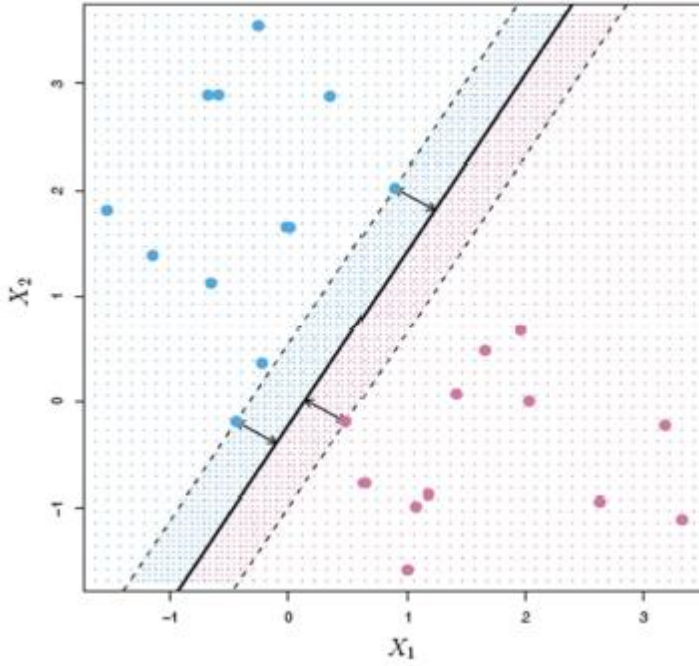
Şekil 3.2: Maksimum ayırma başarısına sahip hiper düzlem



Kaynak: (Cristianni ve Shawe-Taylor,2000).

Hangi hiper düzlemin daha iyi olduğuna nasıl karar verilir? Ayrıcı hiper düzlem veri kümelerinden birine daha yakın veya daha uzak olmamalıdır ki veri doğru bir şekilde ayrılabilir. Bu duruma en uygun olan çözüm ise maksimum marj sınıflandırıcılarıdır (Cristianni ve Shawe-Taylor,2000).

Şekil 3.3: Maksimum Marj Sınıflandırıcısı



Kaynak: (Cristianni ve Shawe-Taylor,2000).

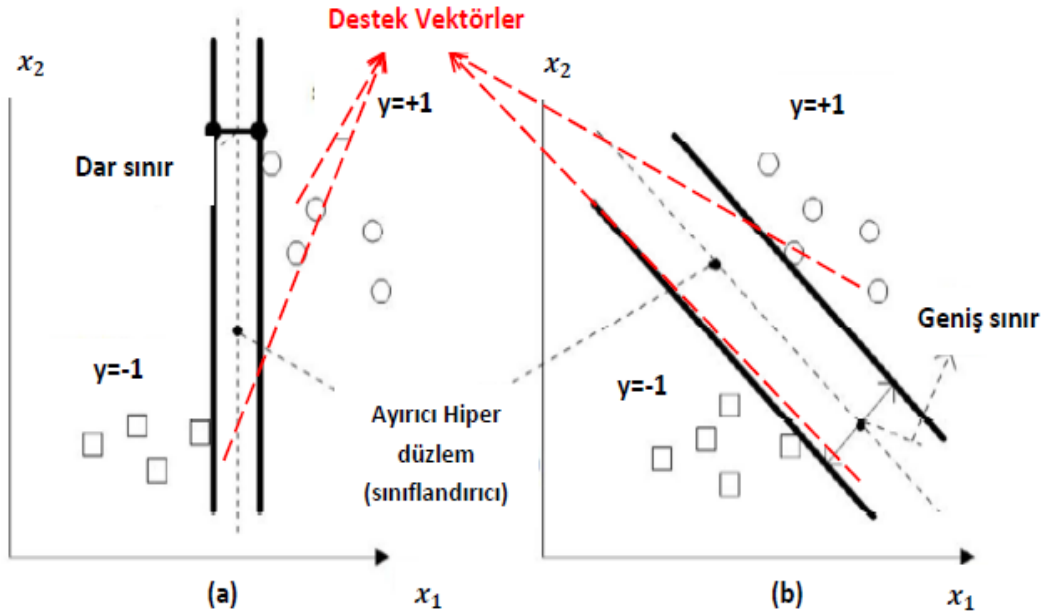
Şekil 3.3'de siyah düz çizgi olarak belirtilen doğru maksimum marj sınıflandırıcısıdır. Kesikli çizgiler arasında kalan kısım ise sınır (marj) olarak adlandırılmaktadır. Kesikli çizgiler ise marj hiper düzlemleri (kanonik hiper düzlemler) olarak adlandırılır. Kanonik hiper düzlemler ayırma hiper düzlemine paraleldir ve maksimal sınır hiper düzlemine en yakın noktalardan geçerler. Şekil 3.3 maksimum marj sınıflandırıcısına örnek olarak gösterilebilir. Aynı zamanda basit doğrusal destek vektör makinası olarak da adlandırılmaktadır. Sorulması gereken diğer bir soru ise neden maksimum marj olması gerektiğidir. Bunun birçok açıklaması bulunmaktadır. En önemlilerinden bir tanesi daha iyi ampirik risk performansı sağlamasıdır. Diğer bir neden ise kümeler birbirinden ne kadar uzak olursa yanlış sınıflandırma o kadar düşük olacaktır, yani daha iyi sınıflandırma performansı sağlanacaktır. Başka bir avantajı ise yerel minimumdan kaçınmayı

sağlamasıdır (Cristianni ve Shawe-Taylor,2000). Maksimum marj hiper düzlemi, sınırın maksimum olduğu ayırıcı hiper düzlemdir.

Bir ikili sınıflandırma problemi üzerinden tanımlar yapılacak olursa, eğitim verileri $X = (x_1, x_2, x_3, \dots, x_n)$ ve $y_i \in \{-1, 1\}$, $i = 1, 2, \dots, n$, $x_i \in R^n$ olarak tanımlanabilir.

Hiper düzlemin bir tarafında kalan etiketler $y_i = +1$, diğer tarafında kalanlar ise $y_i = -1$ olanlar olacaktır. Hiper düzlem ise her iki tarafta bulunan etiketli veri noktalarına eşit maksimum uzaklıkta bulunur. Ayırma düzleminin pozisyonu üzerinde etkisi en büyük olan ve aynı zamanda bu düzleme en yakın olan noktalara ise destek vektör adı verilir (Campbell ve Ying,2011). Destek vektörler marj sınırı üzerinde bulunan noktalardır (Şekil 3.4).

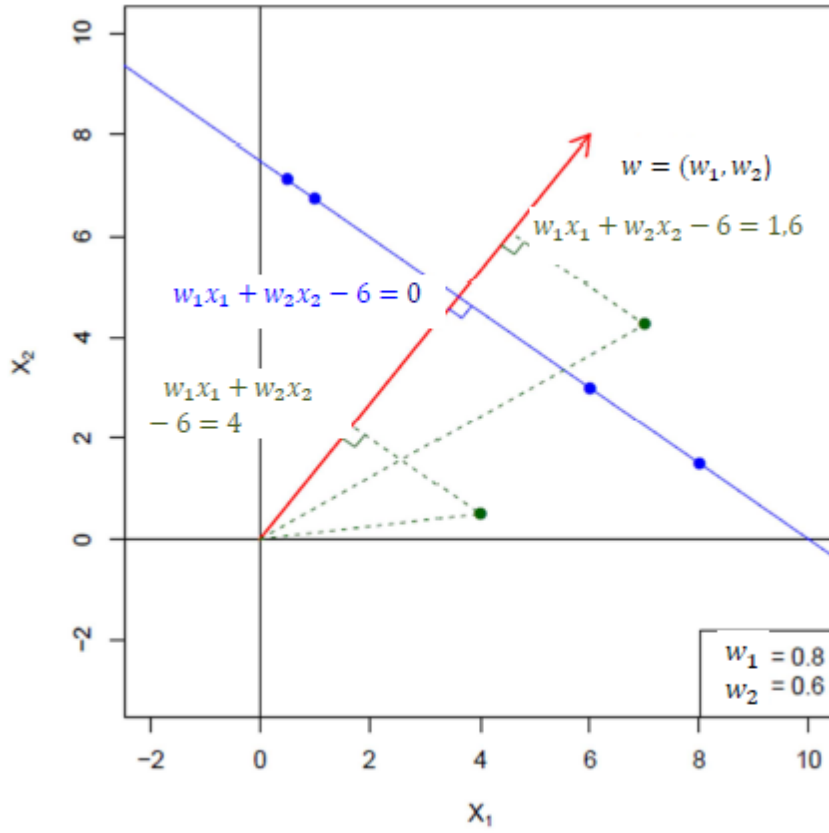
Şekil 3.4: Destek Vektör Gösterimi



Kaynak: (Campbell ve Ying,2011).

Bir hiper düzlemin genel formu $(w' \cdot x) + b = 0$ şeklinde ifade edilir ve karar fonksiyonu, $f(x) \equiv (w' \cdot x) + b$, bir gözleme atanarak, $y = 1$ (pozitif sınıf) ve $y = -1$ (negatif sınıf), sınıflandırma kuralı olarak kullanılabilir. Bu genel form ikili sınıflandırmada bir adım fonksiyonudur (Kecman,2001).

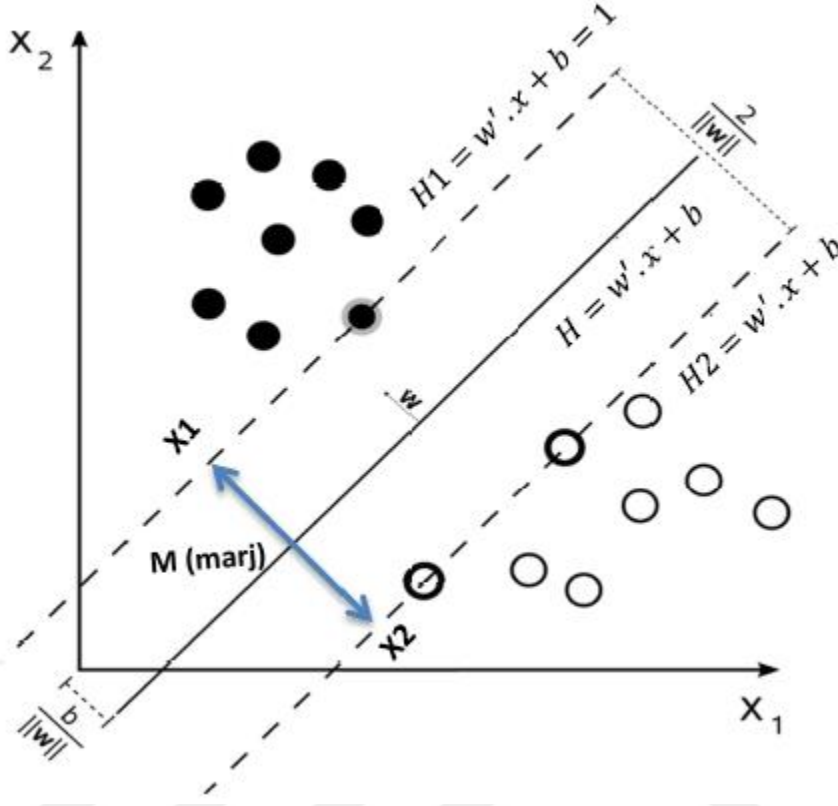
Şekil 3.5: İki Boyutlu Hiper Düzlem ve W Vektör Gösterimi



Kaynak: (Kecman,2001).

Bu eşitlikte x bir vektör noktası, w ağırlık vektörü ve b yan (bias) olmak üzere bir sabit sayıdır. w ağırlık vektörü ayırıcı hiper düzleme dik normal vektördür, b sabiti ise hiper düzlemin orijinden ne kadar sapacağını belirler (w_0). Şekil 3.5’de bu eşitlik ve vektörlerin gösterimine ilişkin bir örnek verilmiştir.

Şekil 3.6: Ayırıcı Hiper Düzlem ve Kanonik Hiper Düzlemler



Kaynak: (Kuhn ve Tucker,1951).

Şekil 3.6'da yuvarlak içine alınan gözlemler destek vektörlerdir. Ayırıcı hiper düzlemi $(w' . x) + b = 0$ aynı zamanda (3.6)'daki şekilde de yazılabilir.

$$\sum_{i=1}^n w_i' x_i + b = 0 \quad (3.1)$$

Burada ağırlık vektörü $w = \{w_1, w_2, \dots, w_n\}$ Şekil 3.5'de de görüldüğü gibi her x için hiper düzleme olan dik uzaklığı belirtmektedir. Kanonik hiper düzlemler ise (3.2) ve (3.3)'te olduğu şekilde gösterilir.

$$H1 = (w' . x) + b = +1 \quad (3.2)$$

$$H2 = (w' . x) + b = -1 \quad (3.3)$$

H1 ve H2 hiper düzlemleri arasındaki uzaklığı yani marjı (M) maksimize etmek amaçlandığı için öncelikle bu uzaklığın ne olduğu bulunmalıdır. H1 ve H2 üzerinde

birer x noktası ele alınırsa, H_1 üzerindeki x_1 , H_2 üzerindeki x_2 olur. Bu durumda Şekil 3.6'da da görüleceği gibi bu uzaklık geometri yardımıyla (3.4) ile (3.5) arasındaki işlemler ile bulunabilir.

$$M = H_1 - H_2 \quad (3.4)$$

$$(w' \cdot (x_1 - x_2)) = 2 \quad (3.5)$$

$$\left(\frac{w'}{\|w\|} \cdot (x_1 - x_2)\right) = \frac{2}{\|w\|} \quad (3.6)$$

$$M = (x_1 - x_2) = \frac{2}{\|w\|} \quad (3.7)$$

Amaç M 'yi yani marjı maksimum yapmaktır. Bu durumda yapılması gereken $2/\|w\|$ ifadesinin maksimize edilmesidir. Bunun için ise $\|w\|$ ifadesinin minimum yapılması gerekmektedir. $\|w\|$ minimum yaparken aynı zamanda veri noktalarının marjın içine düşmesini engellemek için (3.8) ve (3.9)'de gösterilen kısıtlar da bu minimizasyon problemine eklenmelidir.

$$y_i = -1 \text{ için } (w' \cdot x) + b \leq -1 \quad (3.8)$$

$$y_i = +1 \text{ için } (w' \cdot x) + b \geq +1 \quad (3.9)$$

Şekil 3.6'da ayırıcı hiper düzlemin üst tarafında (pozitif tarafında) kalan ifadeler için (3.9) eşitsizliği, alt tarafında kalan ifadeler için ise (3.8) eşitsizliği kullanılabilir.

Bu iki eşitsizlik (3.10)'daki şekilde birleştirilebilir.

$$y_i((w' \cdot x) + b) - 1 \geq 0, \quad \forall i \quad (3.10)$$

3.2.1.1. Optimizasyon Problemi

Yukarıdaki bilgiler ışığında son duruma ilişkin primal form (3.11)'daki şekildedir.

$$\text{Min} : \|w\|$$

$$\text{Kısıtlar: } y_i(w' \cdot x) + b \geq 1, \quad i=1,2,\dots,n \quad (3.11)$$

(2.31)'da görülen optimizasyon problemini çözmek zordur. $\|w\| = \sqrt{w'w} = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$ şeklinde açılır, yani w vektörünün normu, kare kök içermektedir.

Optimizasyon problemini daha kolay bir hale getirebilmek için $\|w\|$ ifadesinin türevi kullanılabilir ki bu $1/2\|w\|^2$ ifadesidir. $1/2\|w\|^2$ ifadesi $\|w\|$ ile aynı minimuma ulaşmaktadır. Optimizasyon probleminin çözümü için w vektörünün normunun türevi hesaplamalarda kolaylık sağladığından, türev formu üzerinden işlemler yapılır (Kecman,2001). Son durumda amaç fonksiyonu kuadratik terim içerdiği için kuadratik optimizasyon problemi olarak karşımıza çıkmaktadır. Bu durumda primal form (3.12)'deki duruma dönüşür.

$$\text{Min: } \frac{1}{2}\|w\|^2$$

$$\text{Kısıtlar: } y_i(w'x) + b \geq 1, \quad i=1,2,\dots,n \quad (3.12)$$

Maksimal marj hiper düzlemi yukarıdaki kuadratik programlama probleminin çözülmesiyle elde edilecektir (Kecman,2001).

3.2.1.2. Lagrange Çarpımları

Lagrange metodu bir fonksiyonun eşitlik içeren belirli kısıtlar dahilinde yerel maksimum ve minimum noktalarının bulunması için kullanılan bir stratejidir (Bertsekas,1999).

$$\text{Optimizasyon problemi: } \quad \text{Max: } f(x, y)$$

$$\text{Kısıtlar: } g(x, y) = c$$

Yukarıdaki optimizasyon problemi için lagrange fonksiyonu (3.13)'daki şekilde oluşturulur.

$$L(x, y, a) = f(x, y) - a_i(g(x, y) - c) \quad a_i \geq 0, \quad i=1,2,\dots,n \quad (3.13)$$

Kısıtlarda bulunan her bir eşitlik durumu için bir lagrange çarpanı tanımlanır. Yerel maksimum veya minimumun bulunabilmesi için lagrange fonksiyonunun x ve y değişkenlerine göre kısmi türevlerinin alınıp sifıra eşitlenmesi gerekmektedir.

Sonrasında bulunan denklemlerin çözümleri aranılan optimum noktaları verecektir. Yukarıda belirtilen (3.11) kuadratik programlama probleminin lagrange fonksiyonu (3.14)'teki gibi oluşturulur.

$$L(w, b, a) = \frac{1}{2} w' w - \sum a_i [y_i (w' x_i + b) - 1] \quad (3.14)$$

Ancak (3.13)'daki primal forma bakıldığında görüldüğü gibi kısıtlar yalnızca eşitlik değil eşitsizlik de içermektedir. Bu durum Krush-Kuhn-Tucker koşulları yardımıyla aşılır.

3.2.1.3. Krush-Kuhn-Tucker (KKT) koşulları

Krush-Kuhn-Tucker koşulları doğrusal olmayan programlama problemleri için birinci dereceden gerek olan koşullardır. Daha öncede belirtildiği gibi lagrange yöntemi kısıtlarda eşitlik olduğu durumlarda kullanılır ancak bu durum eşitsizlik olan durumlara belirli kısıtlar eklenerek geliştirilebilir. Bu geliştirmeyi sağlayan koşullara KKT koşulları adı verilir. Eşitsizlik içeren koşullar olmadığı durumlarda KKT koşulları lagrange koşullarına dönüşür ve KKT çarpanları lagrange çarpanları adını alır (Kuhn ve Tucker,1951). Bu geliştirme için $a_i \geq 0, \forall i$ kısıtı optimizasyon problemine eklenmelidir. Bu kısıta KKT tamamlayıcı (complementary) koşulu adı verilir. KKT her zaman yeter koşulları sağlamaz. Yalnızca amaç fonksiyonunun konveks olduğu durumlarda yeter koşullar sağlanır. DVM konveks programlama problemi olduğu için KKT koşulları w için yeter ve gerek şartları sağlar denir (Cristianni ve Shawe-Taylor,2000).

Son durumda Lagrange fonksiyonu (3.15)'deki duruma dönüşür.

$$L(w, b, a) = \frac{1}{2} w' w - \sum a_i [y_i (w' x_i + b) - 1], \quad a_i \geq 0, \quad i = 1, 2, \dots, n \quad (3.15)$$

(3.15)'de belirtilen lagrange fonksiyonunun w ve b 'ye göre birinci dereceden kısmi türevleri alınarak sifira eşitlenir.

$$\frac{\partial L(w, b, a)}{\partial w} = w' - \sum_{i=1}^n y_i a_i x_i = 0 \Rightarrow w = \sum a_i y_i x_i \quad (3.16)$$

$$\frac{\partial L(w, b, a)}{\partial b} = -\sum_{i=1}^n a_i y_i = 0 \Rightarrow \sum a_i y_i = 0 \quad (3.17)$$

(3.16) ve (3.17)'de elde edilen sonuçlara KKT koşulları adı verilir. Elde edilen ifadeleri (3.15)'de yerine konulduğunda ise (3.18)'deki ifade elde edilir.

$$\begin{aligned}
L(w,b,a) &= \frac{1}{2} \langle \sum a_i y_i x_i, \sum a_j y_j x_j \rangle \\
&\quad - \sum a_i \left[y_i (\langle \sum a_j y_j x_j, x_j + b \rangle - 1) \right] \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j \langle x_i, x_j \rangle - \sum_{i=1}^n a_i a_j y_i y_j \langle x_i, x_j \rangle - b \sum_{i=1}^n a_i y_i + \sum_{i=1}^n a_i \\
&= \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{i=1}^n a_i a_j y_i y_j x_i' x_j
\end{aligned} \tag{3.18}$$

(3.16)'da görüldüğü gibi w artık gözlemlerin doğrusal kombinasyonlarından oluşmaktadır ancak lagrange çarpanları bilinmeden w 'yu hesaplamak mümkün değildir. Optimum bir eyer noktası (sırt noktası) bulabilmek için lagrange fonksiyonu w ve b 'ye göre minimize edilmeli ve lagrange çarpanları olan a_i değerleri için maksimize edilmelidir. Bu problem primal uzayda (w ve b 'ye göre) veya dual uzayda (a 'ya göre) çözülebilir (Wang,2005). Primal çözümün kısıtları dual probleme göre daha karmaşık olduğundan pratikte dual kuadratik problem çözülmektedir ve verdikleri sonuç aynıdır. (3.19)'da elde edilen problem Wolfe dual olarak adlandırılmaktadır ve bir kuadratik programlama problemi olarak elde edilmektedir (Hastievd.,2008).

$$L_{dual}(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{i=1}^n a_i a_j y_i y_j x_i' x_j$$

$$\text{Maks. } L_{dual}(a)$$

$$\text{Kısıtlar } a_i \geq 0, \quad i = 1, 2, \dots, n$$

$$\sum a_i y_i = 0 \tag{3.19}$$

Böyle bir standart kuadratik optimizasyon problemi matris notasyonu ile (3.20)'deki şekilde gösterilebilir.

$$\text{Maks. } L_{Dual}(a) = f'a - \frac{1}{2} a'Ha$$

Kısıtlar $a \geq 0$

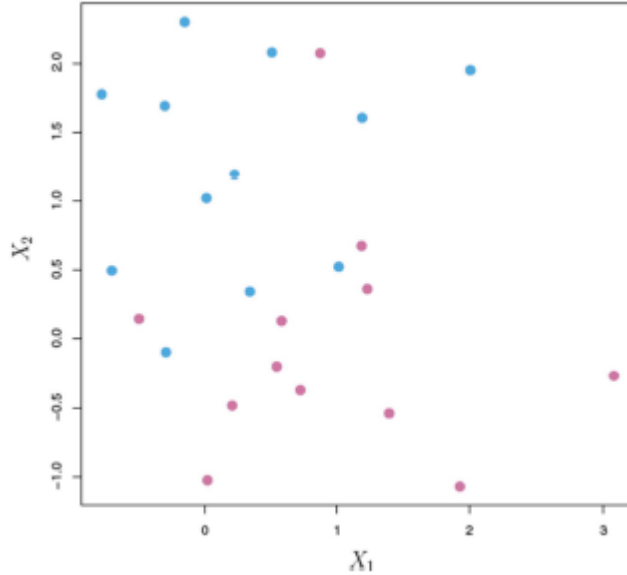
$$y' a = 0 \quad (3.20)$$

3.20'de matris notasyonu ile belirtilen optimizasyon probleminde $a = [a_1, a_2, \dots, a_n]^T$, H bu probleme ait Hessian matrisi ($H_{ij} = y_i y_j (x_i x_j) = y_i y_j x'_i x_j$) ve f vektörü n boyutlu bir birim vektörünü ($f = 1 = [111\dots 1]^T$) göstermektedir (Wang,2005).

3.2.2. Tümüyle Doğrusal Ayrılamayan Durum (Yumuşak Sınır)

Maksimum marj sınıflandırıcısı, eğer bir ayırma hiper düzlemi varsa, uygulama için uygun bir yöntemdir. Ancak, daha önce de belirtildiği gibi, çoğu gerçek hayat probleminde bir ayırıcı hiper düzlem yoktur, dolayısıyla maksimal marj sınıflandırıcısından bahsetmek de mümkün değildir (James vd.,2013). Veride gürültüler veya yanlış veri girişleri olabilir. Bu tür durumlarda ayırıcının hatalar yapmasına izin verilir (Hamel,2009).

Şekil 3.7: Maksimal Marj Sınıflandırıcı İle Ayrılamayan Veri Grafiği

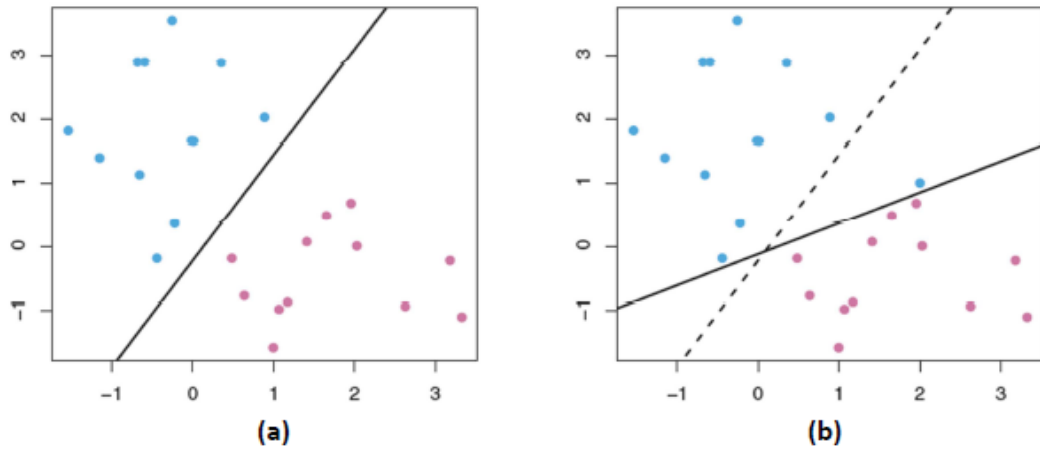


Kaynak: (Kuhn ve Tucker,1951).

Bazen veriyi bir hiper düzlem ile tam olarak ikiye ayırmak mümkün olmayabilir. Ancak, ayırma hiper düzlemi tanımı genişletilip yumuşak marj kullanılarak veriyi

neredeşye iki ayrı sınıfa ayırabilen bir hiper düzlem bulunabilir. Markimal marj sınıflandırıcısının (doğrusal DVM) tam olarak ayrılamayan veriler için genelleştirilmiş haline destek vektör sınıflandırıcısı adı verilir (James vd.,2013). Şekil 3.6'da görölen iki sınıfa ait gözlemler bir hiper düzlemle ayrılamamaktadır. Bazı durumlarda ise bir ayırıcı hiper düzlem olsa bile, bu hiper düzlemin istenilen hiper düzlem olmadığı durumlar olabilir.

Şekil 3.8: Ayırıcı Hiper Düzlemin Tek Gözlemle Yer Değişim Grafiği



Kaynak: (James vd.,2013).

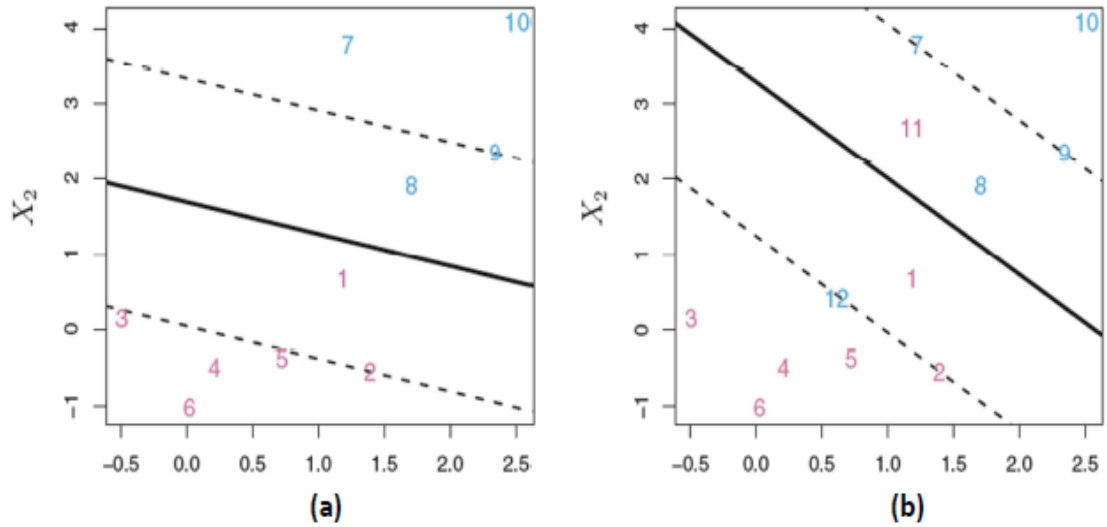
(a) grafiği (b) grafiğine bir gözlem eklendiğinde ayırıcı hiper düzlemin nasıl yer değiştirdiğini göstermektedir. (b) grafiğinde kesikli çizgi (a) grafiğindeki eski hiper düzlemi, siyah düz çizgi ise (b) grafiğine yeni gözlem eklendiğinde yer değiştiren yeni hiper düzlemi temsil etmektedir.

Şekil 3.8'de görüldüğü gibi tek bir gözlem eklenmesi maksimal marj hiper düzleminin yerinde büyük bir değişikliğe neden olabilmektedir. Sonuç olarak ortaya çıkan hiper düzlem ise tatmin edici değildir çünkü dar bir marja sahiptir. Buradan da anlaşılacağı gibi maksimal marj hiper düzlemi tek bir gözlemin değişikliğine bile çok duyarlıdır. Dar marj daha önce de belirtildiği gibi bir problemdir. Gözlemlerin hiper düzlemden uzaklığı aynı zamanda bir verinin ne kadar doğru sınıflandırıldığına ilişkin bir güven ölçüsüdür.

Bu tür durumlarda veriyi tamamen ikiye ayıran bir hiper düzlem seçmek yerine, eğitim verisinin çoğunu doğru sınıflandıran bir hiper düzlem bulmak istenilebilir. Bir

kaç gözlemin yanlış sınıflandırılması, geriye kalan gözlemlerin daha iyi sınıflandırılmasını sağlıyorsa bu sınıflandırıcıyı seçmek tercih edilebilir (James vd.,2013). Destek vektör sınıflandırıcısı tam olarak bunu yapmaktadır. Tüm eğitim verisini marjın doğru tarafında olmasını sağlayacak hiper düzlemi aramak yerine, bazı gözlemleri sınırın yanlış tarafında hatta ayırıcı hiper düzlemin yanlış tarafında olacak şekilde sınıflandıran hiper düzlemleri arar. Marjın yumuşak olarak adlandırılmasının nedeni ise bazı gözlemlerin yanlış tarafta olmasına izin verilmesidir. Şekil 3.8 bu duruma örnek olarak gösterilebilir. Şekil 3.8’de eğitim verisinin çoğu doğru tarafta olmasına rağmen çok küçük bir kısmı ise marjın yanlış tarafında bulunmaktadır.

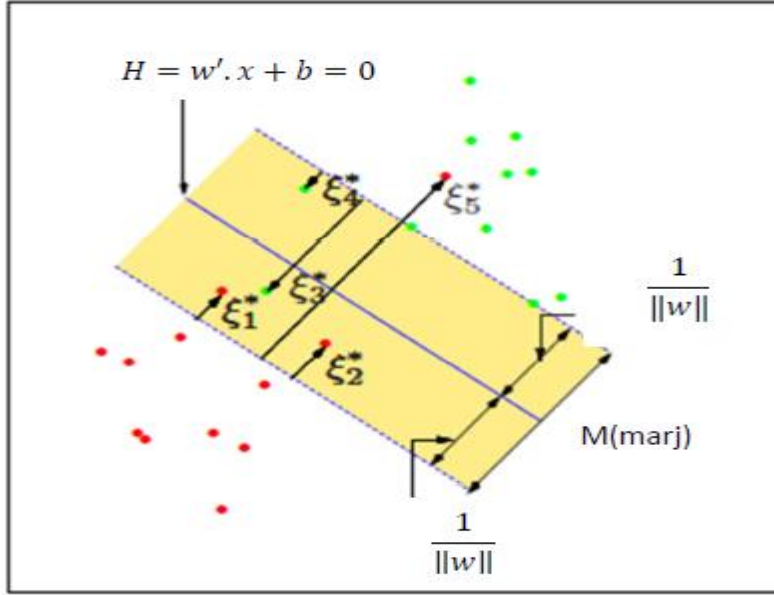
Şekil 3.8: (a) Grafiğinde Gözlemlerin Çoğu Doğru Tarafta Bulunmaktadır



Kaynak: (Wang,2005).

2 numaralı gözlem marjın üzerinde bulunurken, 1 numaralı gözlem marjın yanlış tarafında bulunmaktadır. (b) grafiğinde ise 12 ve 11 numaralı gözlemler eklenmiştir. Buradaki durumda ise 12 yalnızca sınırın yanlış tarafında değil aynı zamanda hiper düzlemin de yanlış tarafında bulunmaktadır. Aynı şekilde 11 numaralı gözlem de sınırın yanlış tarafında yer almaktadır. Anlatılan bu durumlar yumuşak sınırlara örnek olarak gösterilebilir.

Şekil 3.9: Gevşek Değişken Gösterimi



Kaynak: (Kecman,2001).

Verilerin doğrusal bir düzlemle ayrılama durumunda negatif olmayan ve hataları ifade eden ξ_j^* gevşek (slack) değişkenlerinin optimizasyon modeline eklenmesi sağlanarak bu soruna çözüm aranır. Şekil 3.9'da ξ_j^* ile belirtilen gözlemler marjın yanlış tarafında bulunan gözlemlerdir. Burada doğru tarafta olan gözlemler için $\xi_j^* = 0$ 'dır. $\xi_j^* > 1$ olan gözlemler (şekilde ξ_3^* ve ξ_5^*) hiper düzlemin yanlış tarafında kalan gözlemleri, $0 < \xi_i^* < 1$ olan gözlemler (şekilde ξ_1^*, ξ_2^* ve ξ_4^*) ise hiper düzlemin doğru yanında yer alan, ancak marj bölgesi içerisinde kalan gözlem değerlerini ifade eder. Bu duruma üst üste binen veriler de denilmektedir (James vd.,2013). Üst üste binen veriler söz konusu olduğunda daha önce maksimal marj sınıflandırıcı bulmak için kullandığımız kuadratik programlama kısıtı $y_i(w' . x) + b \geq 1$, $i = 1, 2, \dots, n$ sağlanmadığı için kullanılamaz. Şekil 3.9'da verilen durum olduğunda, yani sınıflar üst üste bindiğinde, veriler doğru olarak sınıflandırılmaz ve yanlış sınıflandırılan her x_i için elde edilen lagrange çarpanı a_i sonsuza gitme eğilimi gösterir. Bu tür veriler doğru sınıflandırılabilme için karar sınıfı üzerinde çok güçlü bir etki yaratırlar. a_i değerleri maksimum sınıra ulaştığında

ektikisini daha fazla artıramaz ve çarpana ilişkin gözlem yanlış sınıflandırılmış olarak kalır. Böyle bir durumda, daha önce bahsettiğimiz algoritma neredeyse bütün verleri destek vektör olarak seçer (Kecman,2001). Yumuşak marjın genişliği modelin eğitim hatası ve VC boyutu arasında bir denge kurulmasını sağlayan bir C parametresi ile kontrol edilebilir. C parametresi yanlış sınıflandırma için kullanılan bir ölçüdür. Bu durumda optimizasyon probleminin son hali (3.21)'daki şekilde yazılabilir.

$$\text{Min } \frac{1}{2} w'w + C \quad (\text{yanlış tarafta bulunan gözlemlerin ayırıcı düzleme uzaklığı}) \quad (3.21)$$

Bu şekilde maksimal marj sınıflandırıcısı üst üste binen gözlemler için genelleştirilmiş olur. Optimizasyon problemi ise (3.21)'deki formu alır.

$$\text{Min } \frac{1}{2} w'w + C \sum_i^n \xi_i$$

$$\text{Kısıtlar} \quad y_i(w'.x) + b \geq 1 - \xi_i, \quad i = 1, 2, \dots, n, \xi_i \geq 0$$

$$\text{Diğer bir ifadeyle kısıtlar} \quad (w'.x) + b \leq +1 - \xi_i, \quad \xi_i \geq 0,$$

$$(w'.x) + b \leq -1 - \xi_i, \quad \xi_i \geq 0 \quad (3.21)$$

C parametresi bir sabittir ve kullanıcı tarafından seçilmektedir. C'yi büyük seçmek demek hatalara daha yüksek pay atanması anlamına gelir. Bu tür genelleştirilmiş optimal hiper düzlemde, optimize edilecek fonksiyon, doğrusal ayırmayı engelleyen durumlar için bir ilave terime sahip olacaktır. İlave terimin eklenmesi ile birlikte primal lagrange fonksiyonu (3.22)'deki forma dönüşür.

$$L_{\text{primal}}(w, b, \xi, \alpha, \beta) =$$

$$\frac{1}{2} w'w + C \left(\sum_i^n \xi_i \right) - \sum_i^n a_i (y_i [w'w + b] - 1 + \xi_i) - \sum_i^n \beta_i \xi_i \quad (3.22)$$

Burada a_i ve β_i lagrange çarpanlarıdır. Optimizasyon problemine bir kısıt daha eklendiği için ikinci bir lagrange çarpanı primal fonksiyona eklenmiştir. Yine bu optimizasyon probleminde de amaç eğer noktalarını $(w^*, b^*, \xi^*, \alpha^*, \beta^*)$ bulmaktır. Primal optimazasyon problemi w,b ve ξ parametrelerine göre minimize edilirken a_i

ve β_i parametrelerine göre maksimize edimelidir. Burada da çözüm primal veya dual uzayda bulunabilir, kolaylık olması açısından dual uzayda çözümlenir (Kecman,2001).

Primal amaç fonksiyonu konveks olduğu için tek bir eğer noktası vardır ve Dolayısıyla $w^*, b^*, \xi^*, \alpha^*, \beta^*$ çözümleri KKT koşullarını sağlamalıdır (Hamel,2009). Maksimal marj sınıflandırıcısında olduğu gibi, dual problemi çözebilmek için KKT koşulları uygulanıp primal fonksiyonun türevleri alınarak sıfıra eşitlenir ve primal parametreler (3.23)'deki gibi elde edilir.

$$\frac{\partial L_{primal}}{\partial w^*} = w^* - \sum_{i=1}^n a_i y_i x_i = 0$$

$$w^* = \sum_{i=1}^n a_i y_i x_i \quad (3.23)$$

Yumuşak marj sınıflandırıcı için karar düzleminin optimal normal vektörü eğitim verilenin doğrusal kombinasyonundan oluşmaktadır.

$$\frac{\partial L_{primal}}{\partial b^*} = \sum_{i=1}^n a_i y_i = 0 \quad (3.24)$$

(3.24)'de görüldüğü gibi b parametresi de maksimal marj sınıflandırıcıda olduğu gibi dual için aynı kısıtı vermektedir. Son olarak ise gevşek değişkenler için türev uygulanır (3.25).

$$\frac{\partial L_{primal}}{\partial \xi^*} = C - \alpha_i - \beta_i = 0 \quad (3.25)$$

(3.25)'de elde edilen durum ise bize (3.26)'da görülen yeni bir kısıt getirmektedir.

$$a_i = C - \beta_i \quad (3.26)$$

Elde edilen eğer noktalarını ve kısıtları lagrange probleminde (3.27)'deki gibi yerine koyduğumuzda dual problem elde edilmiş olur.

$$L_{dual}(\alpha, \beta) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j x_i' x_j \quad (3.27)$$

Yumuşak marj için elde edilen dual amaç fonksiyonu, sert marj (maksimal marj) sınıflandırıcısı için bulunan amaç fonksiyonu ile aynıdır. Yani, optimizasyon

probleminin ana formu deđişmemiştir, deđişen ise yalnızca kısıtlardır (Hamel,2009). Eklenen kısıt ise gevşek primal parametre olan ξ ait olan kısıttır. Lagrange çarpanları a_i ve β_i için kısıtlar $a_i \geq 0$ ve $\beta_i \geq 0$ ' dır. Kısıtlar daha uygun bir hale getirilerek (3.28)'deki gibi tekrar yazılabilir.

$$0 \leq a_i \leq C \quad (3.28)$$

$$0 \leq a_i \leq C \quad (3.29)$$

Son durumda yumuşak marj için lagrange dual optimizasyon problemi (3.30)'daki formu alır.

$$\text{Max } L_{dual}(\alpha, \beta) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j x'_i x_j$$

Kısıtlar:

$$\sum a_i y_i = 0, \quad i = 1, 2, \dots, n$$

$$c \geq a_i \geq 0, \quad i = 1, 2, \dots, n \quad (3.30)$$

Maksimal marj sınıflandırıcısı ile yumuşak marj sınıflandırıcısı arasındaki tek fark (3.30)'da da görülebileceđi gibi Lagrange çarpanlarını sınırlayan C sabitidir (Hamel,2009). Maksimal marjda olduđu gibi burada da b^* deđişkeni açık olarak hesaplanamamıştır. Bu deđişken hesaplanırken ise dikkat edilmesi gerekmektedir. Sert marj sınıflandırıcısında b^* 'yi hesaplamak için herhangi bir destek vektör kullanılabilirdi. Ancak burada lagrange çarpanı C sabitine eşit olan destek vektörlerden kaçınılması gerekmektedir. Lagrange çarpanı C'ye eşit olan bir destek vektör, sıfır olmayan bir gevşek deđişken ile ilişkilidir. Bu durum, bu destek vektörün marjın yanlış tarafında bulunduđu anlamına gelir ve b^* yanlış hesaplanır.

Sonuç olarak b hesaplanırken gevşek deđişkeni sıfır, lagrange çarpanı $0 < a_{sv} < C$ koşulunu sađlayan gözlemler seçilmelidir. Tamamlayıcı koşuluna bakıldığında (3.31)'deki eşitlik elde edilir.

$$\sum_{i=0}^n a_i^* y_i x_i x_{sv} - b^* - 1 = 0 \quad (3.31)$$

x_{sv} noktası marj sınırı üzerinde bulunur çünkü bu noktaya karşılık gelen lagrange çarpanı sıfırdan büyük ancak C sabitinden küçüktür ve gevşek değişkeni sıfıra eşittir. Son durumda b (3.32)'de gösterildiği gibidir.

$$b^* = \sum_{i=1}^n a_i^* y_i x_i x_{sv^+} - 1 \quad (3.32)$$

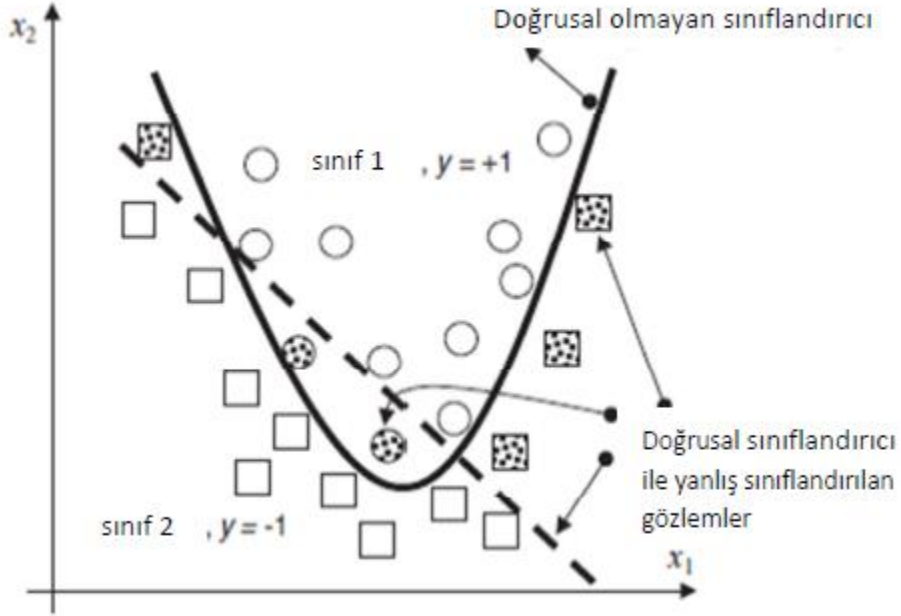
Eğer destek vektör doğru seçilirse yumuşak marj için hesaplanan b^* değeri sert marj sınıflandırıcısıdır. Sonuç olarak elde edilen model sert marj ve yumuşak marjda aynıdır, modelde değişiklik yoktur. Değişik olan tek şey lagrange çarpanlarının bir C sabiti ile sınırlandırılmasıdır ki bu sabit uygulayıcı tarafından belirlenir. Her iki optimizasyon probleminde de sınıflandırıcıları oluşturan terimler iç çarpım halindedir. Bu durum doğrusal olmayan sınıflandırıcılarda özellikle çekirdek fonksiyonları söz konusu olduğunda destek vektör makinelerini değerli kılan en önemli etmendir (Hamel,2009).

3.3. Doğrusal Olmayan Destek Vektör Makineleri

Şimdiye kadar veri kümelerinin doğrusal bir hiper düzlem ile ayrılabilirdiği durumlar irdelendi. Verilerin doğrusal olarak ayrılamadığı durumlarda ise gevşek değişkenler modele eklenerek çözüm üretilmeye çalışıldı. Eğitim verisinin doğrusal olarak ayrılamadığı durumlarda doğrusal bir sınıflandırıcı yerine doğrusal olmayan sınıflandırıcılar da kullanılabilir (Kecman,2001).

Gerçek hayat verilerinin çok azı doğrusal olarak ayrılabilir. Destek vektör makinelerinin en dikkat çekici özelliği, temelinde yer alan doğrusal çerçevesinin kolayca doğrusal olmayan duruma genişletilebilir olmasıdır. Bu genişletmenin arkasındaki temel fikir doğrusal ayrılamayan girdiler uzayının, verilerin doğrusal olarak ayrılabilirdiği daha yüksek boyutlu bir özellikler uzayına dönüştürülmesidir (Hamel,2009). Eğer bu dönüşüm dikkatli ve doğru bir şekilde gerçekleştirilebilirse, özellikler uzayı ile ilgili tüm hesaplamalar gerçek girdi uzayında da yapılabilir. Bu dönüşümleri sağlayan fonksiyonlara çekirdek (kernel) fonksiyonları adı verilmektedir. Doğrusal destek vektör makinelerinden doğrusal olmayan destek vektör makinelerine geçişi sağlayan bu işleme de kernel hilesi adı verilmektedir.

Şekil 3.10: Doğrusal Olmayan Kuadratik DVM Sınıflandırıcısı ve Yanlış Sınıflandırma Yapan Doğrusal Sınıflandırıcı



Kaynak: (Kecman,2001).

Şekil 3.10 doğrusal olmayan sınıflandırıcılara örnek olarak gösterilebilir. Doğrusal sınıflandırıcı ile Şekil 3.10'da altı gözlem yanlış sınıflandırılmıştır. Doğrusal olmayan sınıflandırıcı kullanılarak ise hatasız olarak veri iki sınıfa ayrılabilmiştir. Destek vektör makinelerinin tasarımının arkasındaki ana fikir, girdi uzayını oluşturan $x \in R^n$ vektörlerini $\Phi(x)$ formunda kullanarak daha yüksek boyutlu bir özellikler uzayı F' e taşıyıp $(R^n \rightarrow R^F)$ bu özellikler uzayında doğrusal sınıflandırma problemini çözmektir.

$$x \in R^n \rightarrow \Phi(x) = [\Phi_1(x)\Phi_2(x), \dots, \Phi_n(x)] \in R^F \quad (3.33)$$

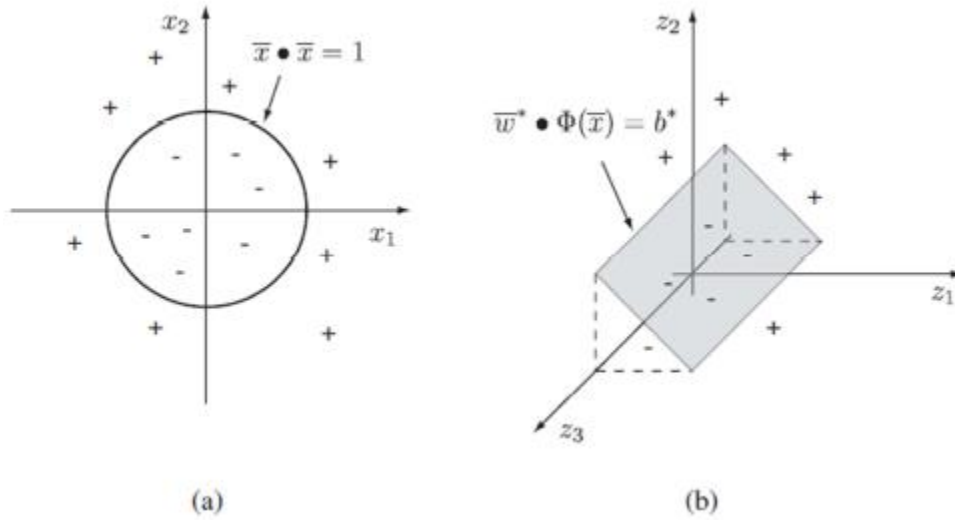
Bir haritalama fonksiyonu, x 'lerin $\Phi(x)$ 'e dönüşümünü sağlayan sabit bir fonksiyondur. Girdi uzayını oluşturan x vektörü x_i gözlemlerinden oluşurken, özellik uzayını oluşturan $\Phi(x)$ vektörü $\phi_i(x)$ 'lerden oluşmaktadır. Böyle bir haritalama yaparak Φ uzayında x görüntülerinin doğrusal olarak sınıflandırılabilirdiği bir öğrenme algoritması bulunması amaçlanmaktadır. Aynı zamanda bu yöntem kullanılarak aynı kısıtlarda bir kuadratik optimizasyon problemi bulunması da

beklenir. Pozitif tanımlı tüm kernel fonksiyonları ile yapılan haritalamalarda özellikler uzayında doğrusal bir ayırıcı bulunur (Bestsekas,1999). Bu beklentilerle elde edilecek karar fonksiyonu (2.54)'teki gibidir.

$$d(x) = \text{sign} \left(\sum_i^n y_i a_i \Phi'(x) \Phi(x) + b \right) = \text{sign} \left(\sum_i^n y_i a_i k(x_i, x) + b \right) \quad (3.34)$$

(3.34)'de belirtilen eşitlikle, özellik uzayında bir doğrusal sınıflandırıcı, girdi uzayında ise doğrusal olmayan bir hiper yüzey elde edilmektedir. $k(x_i, x)$ bir çekirdek fonksiyonunu temsil etmektedir (Wang,2005).

Şekil 3.11: Doğrusal Olmayan Bir Veri Setinin Yüksek Boyutlu Bir Uzayda Doğrusal Hiper Düzlem ile Sınıflandırılması



Kaynak: (Bestsekas,1999).

Şekil 3.11'de verilen eğitim verisi $x = (x_1, x_2)$ şekilde de görüldüğü gibi $(w'x) + b = 0$ formunda bir doğrusal ayırıcı ile hatasız olarak ayrılamaz. Tam tersine doğrusal olmayan $(x \cdot x = 1)$ $x \in R^2$ yüzeyi ile ayrılabilir. Bu durumda etiketleri bulmak için girdi uzayında bir ayırıcı düzlem bulmak yerine, daha yüksek boyutlu, örneğin R^3 , bir uzaydaki iç çarpımlar ile bir hiper yüzey bulmak için (3.35)'deki karar fonksiyonu kullanılır.

$$d(x) = \text{sign}(w \cdot \Phi(x) - b) \quad (3.35)$$

Bu haritalamada girdi uzayı özellik uzayı $R^2 \rightarrow R^3$ olacak şekilde (3.36)'da gösterildiği şekilde tanımlanır.

$$\Phi(X) = \Phi(x_1, x_2) = (x_1^2, x_2^2, \sqrt{2x_1x_2}) = (z_1, z_2, z_3) = z \quad (3.36)$$

Bu durumda $\Phi(x)$ vektörü iki boyutlu girdi uzayını üç boyutlu özellik uzayına haritalamaktadır. Bu haritalamayla doğrusal olmayan karar yüzeyindeki her nokta özellik uzayında (3.37)'de gösterilen formdaki düzleme haritalanmaktadır.

$$w^* \Phi(x) + b^* = 0 \quad (3.37)$$

(3.37)'de gösterilen eşitlikte $w^* = (w_1^*, w_2^*, w_3^*) = (1, 1, 0)$ ve $b^* = 1$ 'dir. Girdi uzayındaki

$q = (1, 0)$ noktası, doğrusal olmayan karar düzeyi $q \cdot q = (1, 0) \cdot (1, 0) = 1^2 + 0^2 = 1$ üzerinde bulunuyordu. Bu noktanın özellik uzayında da bu düzlem üzerinde bulunduğu (3.38)'deki gibi gösterilebilir.

$$\begin{aligned} w^* \Phi(q) &= (1, 1, 0) \cdot (1^2, 0^2, \sqrt{2} \cdot 1 \cdot 0) \\ &= (1, 1, 0) \cdot (1, 0, 0) \\ &= 1^2 + 1 \cdot 0 + 0 \cdot 0 \\ &= 1 = b^* \end{aligned} \quad (3.38)$$

Girdi uzayındaki +1 etiketli her nokta özellik uzayındaki düzeyin üst tarafına, -1 olarak etiketlenen her nokta da özellik uzayındaki düzeyin alt kısmına haritalanmaktadır. Şekil 3.11'de görüldüğü gibi ayırıcı düzlem doğrusaldır. Bu şu anlama gelmektedir, Φ haritalaması girdi uzayında doğrusal olmayan karar problemini özellikler uzayında doğrusal karar problemine dönüştürmüştür. Özellikler uzayında (2.57)'de verilen karar yüzeyi sayesinde karar fonksiyonu (3.39)'daki şekilde tanımlanabilir.

$$d(x) = \text{sign}(w^* \cdot \Phi(x) - b^*) \quad (3.39)$$

İstendiği gibi, girdi uzayında verilen herhangi bir $x \in R^2$, bu karar fonksiyonu ilkönce özellikler uzayı $z = \Phi(x) \in R^3$ 'ya haritalanmış ve sonrasında özellikler uzayındaki doğrusal olan karar yüzeyi ile etiketler hesaplanmıştır. (3.39)'da belirtilen karar fonksiyonunu (3.36)'ya göre açılacak olursa (3.40)'daki denklem elde edilir.

$$\begin{aligned}
d(x) &= \text{sign}(w^* \cdot \Phi(x) - b^*) \\
&= \text{sign}(w_1^* x_1^2 + w_2^* x_2^2 + w_3^* \sqrt{2} x_1 x_2 - b^*) \\
&= \text{sign}(w^* \cdot z - b^*) \\
&= \text{sign}\left(\sum_{i=1}^3 w_i^* z_i - b^*\right)
\end{aligned} \tag{3.40}$$

(3.40) denklemini göstermektedir ki karar fonksiyonunun karmaşıklığı direkt olarak özellik uzayının boyutuna bağlıdır. (3.40)'daki denklemin genelleştirilecek olursa (3.41)'teki fonksiyon elde edilmektedir.

$$d(x) = \text{sign}\left(\sum_{i=1}^m w_i z_i - b\right) \tag{3.41}$$

(3.41)'de m özellikler uzayının boyutunu temsil etmektedir (Hamel,2009).

(3.36)'da belirtilen dönüşüm (3.42)'de görüldüğü şekilde gerçekleşir.

$\Phi: R^2 \rightarrow R^3$, $\Phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$, R^2 'de iç çarpımı $\Phi(x) \cdot \Phi(y)$ (2.62)'deki şekilde bulunur.

$$\begin{aligned}
\Phi(x) \cdot \Phi(y) &= (x_1^2, x_2^2, \sqrt{2}x_1x_2) \cdot (y_1^2, y_2^2, \sqrt{2}y_1y_2) \\
&= x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1 x_2 y_1 y_2 \\
&= (x_1 y_1 + x_2 y_2)(x_1 y_1 + x_2 y_2) \\
&= (x \cdot y)(x \cdot y) \\
&= (x \cdot y)^2
\end{aligned} \tag{3.42}$$

3.3.1. Çekirdek Fonksiyonları (Kernel Functions)

Dönüşümlere daha yakından bakılacak olursa, uygun bir haritalama $\Phi: R^n \rightarrow R^m$, $m > n$ olmak üzere dönüşüm fonksiyonu (3.43)'teki şekilde gösterilir.

$$k(x, y) = \Phi(x) \cdot \Phi(y) \quad x, y \in R^n \tag{3.43}$$

Bu fonksiyonlara çekirdekler veya çekirdek fonksiyonları adı verilir. Çekirdek fonksiyonları özellikler uzayında bir iç çarpım oluştururlar. Bir çekirdeğin belirleyici olan özelliği ise iç çarpımın girdiler uzayında hesaplanabiliyor olmasıdır (Hamel,2009).

(3.35)'de belirtilen karar fonksiyonunun normal vektörü w^* 'nin dual gösterimi (3.44)'teki gibidir.

$$w^* = \sum_{i=1}^n a_i^* y_i \Phi(x_i) \quad (3.44)$$

Dual gösterimi karar fonksiyonunda yerine konulduğunda ise (3.45)'deki eşitlik elde edilmektedir.

$$\begin{aligned} d(x) &= \text{sign}(w^* \cdot \Phi(x) - b^*) \\ &= \text{sign}\left(\sum_{i=1}^n a_i^* y_i \Phi(x_i) \cdot \Phi(x) - b^*\right) \end{aligned} \quad (3.45)$$

(2.65)'te elde edilen karar fonksiyonu ile daha karmaşık bir hesaplamayla karşılaştırılması beklenilmektedir. Ancak (3.42)'deki çıkarım dikkate alındığında (3.45)'deki durum oluşur.

Genelleştirilecek olursa karar fonksiyonu (3.46)'daki formu alacaktır.

$$d(x) = \text{sign}\left(\sum_{i=1}^n a_i^* y_i k(x_i, x) - b^*\right) \quad (3.46)$$

Bu durum çekirdek hilesi (kernel trick) olarak bilinmektedir. Herhangi bir uygun kernel fonksiyonu kullanılırsa özellikler uzayında açıkça fonksiyonu hesaplama zorunluluğu ortadan kalkar. Onun yerine özellikler uzayındaki her hesaplama girdi uzayındaki hesaplamaları kolaylaştırır.

Kernel fonksiyonlardan bir tanesi açıklanmakta $\Phi: R^2 \rightarrow R^3, \Phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$ olacak şekilde (3.42)'deki kullanılmıştır. Kernel fonksiyonu ise (3.47)'de gösterildiği gibi ifade edilir.

$$k(x, y) = \Phi(x) \cdot \Phi(y) = (x \cdot y)^2 \quad (3.47)$$

(3.47)'de belirtilen kernel fonksiyonunda $x, y \in R^2$ 'dir ve ikinci dereceden homojen polynomial kernel olarak adlandırılır. Bu kernel kolaylıkla $x, y \in R^n$ formuna çevrilebilir. Aşağıda verilen Tablo 3.1'de en yaygın olarak kullanılan kerneller listelenmiştir.

Tablo 3.1: Popolüer Kernel Fonksiyonları

$$x, y \in R^n$$

Kernelin Adı	Kernel fonksiyonu	Serbest Parametreler
Doğrusal kernel	$k(x, y) = x' \cdot y$	yok
Polinomial kernel	$k(x, y) = (yx' \cdot y + r)^d$	$d \geq 2, y > 0, r > 0$
Gaussian (Radyal Tabanlı) kernel	$k(x, y) = \exp(-y x - y ^2)$	$y > 0$
Sigmoid kernel	$k(x, y) = \tanh(ax' \cdot y + c)$	$a < 0, c > 0$

Kaynak: (Burges,1998).

b^* teriminin dual gösteriminin kerneller ile ifade ise (3.48)'deki gibidir.

$$\begin{aligned} b^* &= w^* \cdot \Phi(x_{SV^+}) - 1 \\ &= \sum_{i=1}^n a_i^* y_i \Phi(x_i) \cdot \Phi(x_{SV^+}) - 1 \\ &= \sum_{i=1}^n a_i^* y_i k(x_i, x_{SV^+}) - 1 \end{aligned} \quad (3.48)$$

Özellikler uzayında destek vektörleri bulmak için öğrenme algoritmasını oluşturan destek vektör makinesi modelleri üzerinde de çekirdek hilesi uygulanmalıdır.

$$\begin{aligned} \text{Max} \quad L_{dual}(a) &= \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j k(x_i, x_j) \\ \text{Kısıtlar} \quad a_i &\geq 0, \quad i = 1, 2, \dots, n \\ \sum a_i y_i &= 0 \end{aligned} \quad (3.49)$$

a^* terimlerini bulmak için $k(x_i, x_j)$ ifadesi yerine uygun bir kernel fonksiyonu yerleştirilmelidir. Öğrenme algoritması ve modelde kullanılan kernel fonksiyonları birbiriyle aynı olmalıdır ki yapılan işlemler anlamlı olsun. Belirtilen kısıtlar ise kernel uygulamasından etkilenmemektedir (Hamel,2009).

Çekirdek fonksiyonları sınıflandırma probleminin gereksinimine göre seçilir. Örneğin, eğer sınıflandırma problemi doğrusal olarak ayrılabilir bir eğitim verisine sahipse doğrusal kernel $k(x, y) = x.y$ kullanılabilir. Eğer sınıflandırma problemi girdi uzayında bir kuadratik sınıflandırma yüzeyine sahipse ikinci dereceden bir polinomial kernel $k(x, y) = (x.y)^2$ kullanılabilir. Eğer bu kernel fonksiyonu kullanılırsa özellikler uzayında sınıflandırma yüzeyi doğrusal olacaktır. Girdi uzayındaki daha karmaşık sınırlandırma yüzeylerindeyse daha yüksek dereceli polinomial kerneller veya daha karmaşık olan kerneller, örneğin gaussian kernel, özellikler uzayında doğrusal bir ayırıcı yüzey bulmak için denenilebilir. Her fonksiyon özellikler uzayında iç çarpıma sahip değildir. Bir kernel fonksiyonunun özellikler uzayında iç çarpımının tanımlı olması için Mercer koşullarını sağlaması gerekmektedir (Burges,1998).

3.3.1.1. Mercer Koşulları

Belirli koşullar altında özellikler uzayındaki bir iç çarpım girdi uzayında da aynı çekirdek fonksiyona sahiptir (İvanciuc,2007).

$$k(x, y) = \Phi(x).\Phi(y) \quad (3.50)$$

Eğer K simetrik pozitif tanımlı bir fonksiyon ise, (3.51) ve (3.52)'de belirtilen Mercer koşullarını sağlıyor denir (İvanciuc,2007).

$$k(x, y) = \sum_k a_k \Phi_k(x).\Phi_k(y), a_k \geq 0 \quad (3.51)$$

$$\iint k(x, y)g(x)g(y)dxdy > 0 \quad (3.52)$$

(3.51) ve (3.52) fonksiyonları sağlanıyor ise çekirdek fonksiyonu özellikler uzayında iç çarpıma sahiptir denir. Mercer koşulları ile ilgili ayrıntılı bilgi (Mercer,1909)'den edinilebilir.

3.3.2. Kernel Fonksiyon Seçimi

DVM modeli tasarlanırken en önemli olan konulardan birisi kullanılacak olan kernel fonksiyonu ve seçilen kernelde kullanılan parametrelerin belirlenmesidir. Tüm kernel fonksiyonlarını ve parametrelerini eğitim verisi üzerinde uygulamak çok doğru bir davranış değildir çünkü bu durum aşırı uyum sorununa sebep olabilir. Kernel seçimi

için en uygun yöntem önceki çalışmalar hakkında bilgi edinmektir. Örneğin geçmiş çalışmalarda ikinci dereceden bir polinomial kernel kullanılmışsa DVM tasarımına buradan başlamak en doğru yol olacaktır (Boswell,2002). Modelin doğruluğu ve karmaşıklığı açısından parametre seçimi çok önemlidir. Birçok paket program parametre seçimini otomatik olarak yapar. Otomatik olarak parametre seçiminin yapılması ise hücre arama (grid search) yöntemi ile sağlanır (Burges,1998).

3.3.3. DVM için Kerneller

DVM problemlerinde en sık kullanılan kerneller aşağıdaki şekilde verilebilir.

3.3.3.1. Doğrusal Kernel

x ve y 'nin iç çarpımları doğrusal kerneli tanımlamaktadır.

$$k(x, y) = x' \cdot y \quad (3.53)$$

Bu bir doğrusal sınıflandırıcıdır ve eğitim verisinin doğrusal olup olmadığının testi için kullanılır (İvanciuc,2007).

3.3.3.2. Polinomial Kernel

Polinomial kernel doğrusal olmayan ilişkiyi modellemek için kolay ve verimli yöntemlerden bir tanesidir. Rakam tanıma modellerinde kullanımları yaygındır (Mercer,1909).

$$k(x, y) = (yx' \cdot y + r)^d \quad (3.54)$$

Polinomun derecesi arttıkça aşırı uyum ile karşılaşma olasılığı artmaktadır. Ayrıca derece yükseldikçe sınıflandırma yüzeyi daha karmaşık bir hale gelmektedir (İvanciuc,2007). r değeri 0 olduğunda bu kernel fonksiyonu homojen polinomial kernel olarak adlandırılırken, r değeri 0'dan büyük bir değer aldığında homojen olmayan polinomial kernel adını almaktadır.

3.3.3.3. Gaussian (Radyal Tabanlı) Kernel

Radyal tabanlı kernel olarak da bilinir (Steinwart ve Chsirtmann,2008). En çok kullanılan kernel fonksiyonudur. σ parametresi ayırma hiper düzleminin şeklini kontrol eder. σ arttıkça DVM sayısı artar ve bu da genelleştirmenin azalmasına

sebepe olur. Dolayısıyla bu kernelin kullanımında σ seçimi büyük önem taşımaktadır. Çapraz doğrulama prosedürüyle doğru σ aranmalıdır (İvanciuc,2007). Bu kernel bazı paket programlarında ve kitaplarda (3.55)'deki gibi de gösterilmektedir.

$$k(x, y) = e^{-\frac{|x-y|^2}{2\sigma^2}} \quad (3.55)$$
$$k(x, y) = e^{-\gamma|x-y|^2}$$

$\gamma = \frac{1}{2\sigma^2}$ 'dir ve bu durumda aranması gereken doğru parametre γ parametresidir.

Örneğin Libsvm programı bu fonksiyonu kullanmaktadır.

4. UYGULAMA

Bu bölümde karar ağacı algoritmaları ve destek vektör makineleri (DVM) ile uygulama yapılmıştır. Uygulamada kullanılan veri seti Türkiye İstatistik Kurumu (TÜİK) tarafından yapılan hanehalkı bilişim teknolojileri kullanımı 2017 yılına ait B grubu mikro verilerden oluşmaktadır. Uygulamanın yapılması amacı ile kullanılan program IBM SPSS Modeler 18.0'dir.

4.1. Analiz Öncesi Veri Seti

TÜİK'ten talep edilen B grubu mikro veriler, virgülle ayrılmış (.csv) excel dosyası formatında alınmıştır. 45591 fert ile yapılan anketten oluşan veri seti, 45591 satır ve 115 sütundan oluşmaktadır. Veri setinde 109 tane değişken yer almaktadır.

Hanehalkı Bilişim Teknolojileri Kullanım Araştırması, hanelerde ve bireylerde sahip olunan bilgi ve iletişim teknolojileri ile bunların kullanımları hakkında bilgi derlemek amacıyla 2004 yılından itibaren uygulanmakta olup, söz konusu teknolojilerin kullanımı hakkında bilgi veren temel veri kaynağıdır. (Araştırma 2006 yılında gerçekleştirilmemiştir.)

Bu araştırma ile aşağıdaki alanlarda veri derlenmektedir:

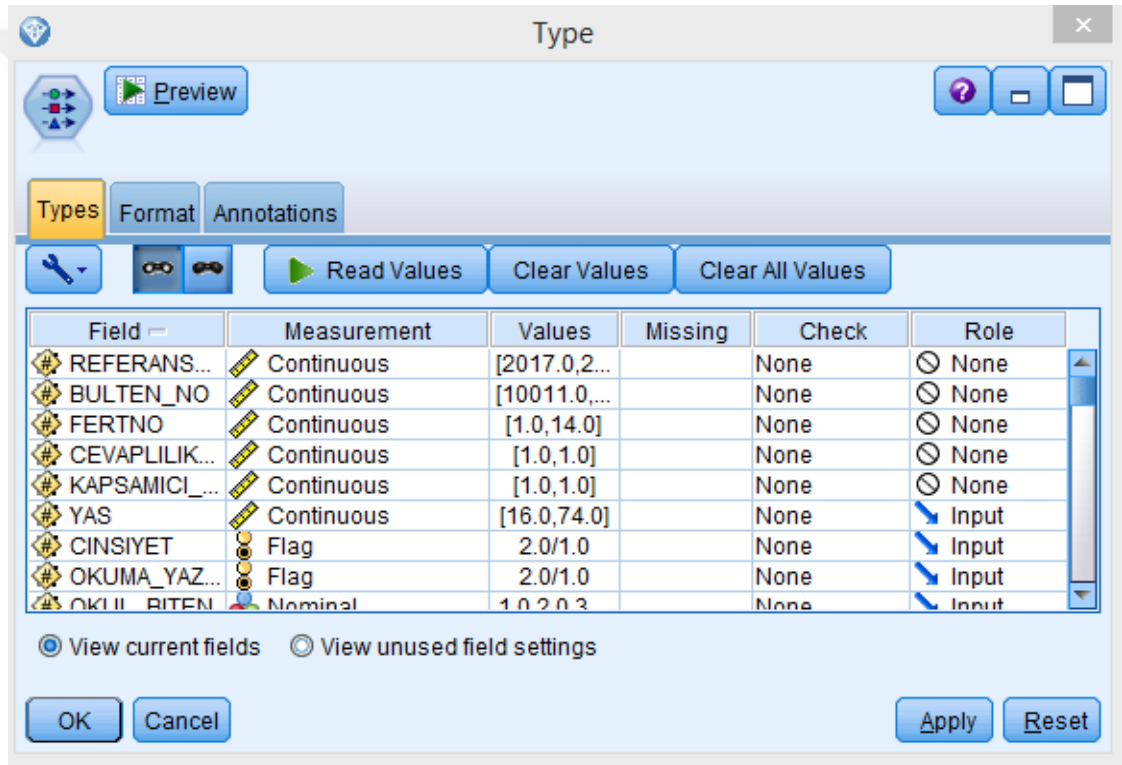
- Hanelerde bulunan bilgi ve iletişim teknolojileri
- Bilgisayar (Bireylerin bilgisayar kullanımı, sıklığı v.b.)
- İnternet (Hanelerde İnternet erişimi sahipliği, bağlantı tipleri, bağlı araçlar, bireylerin İnternet kullanım sıklığı, kullanım amaçları v.b)
- e-Ticaret
- e-Devlet uygulamaları
- Bilişim güvenliği

Veriler, belirlenen örnekleme yöntemine göre seçilen hanehalklarından derlenmektedir. Hanehalkı Bilişim Teknolojileri Kullanım Araştırmalarında kullanılan istatistiksel birim "hanehalkları"dır. Hanehalkında bulunan tüm fertlere ilişkin demografik bilgiler (yaş, cinsiyet) alınmaktadır. Eğitim durumu, işgücü durumu ve bilişim teknolojilerinin kullanımına yönelik sorular ise 16-74 yaş

aralığındaki fertlere sorulmaktadır. 6-15 yaş grubu çocuklara ise bilgisayar, internet ve cep telefonu kullanımları (sıklık, kullanım süreleri, başlangıç yaşları, amaçlar vb.) ile gazete, dergi ve televizyon alışkanlıklarına ilişkin sorular sorulmaktadır.

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 4.1: Veri Kalitesinin İncelenmesi



Quality sekmesini seçip veri kalitesi incelediğimizde aşırı uç değerler gözlenmemiştir.

4.2. Veri Ön İşleme

Hanehalkı bilişim teknolojileri sürecinde en çok zaman alan adım, veri ön işleme adımdır. Bu adımda e-ticaret kullanımına ilişkin sonuçlara ulaşabilmek için elde edilen veriler üzerinde temizleme, bütünleştirme, dönüştürme ve veri indirgeme

işlemlerinden gerekli olanları yapılır. Veri ön işleme işleminden sonra 45.591 olan satır sayısı 11.926 satıra ve 109 olan değişken sayısı da 56'ya düşmüştür.

4.2.1. Veri Temizleme

Hanehalkı bilişim teknolojileri 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 “son kullanım eticaret” sorusu, 16 yaşından büyük fertlere sorulmuştur.

4.2.2. Veri Dönüştürme

Çalışmada bağımlı değişken olarak belirlediğimiz “SON KULLANIM ETİCARET” değişkenine 4 kategoride cevap alınmıştır. “Son üç ay içinde, üç ay ile bir yıl arasında, bir yıldan çok ve hiç kullanmadı” cevapları, sınıflama performansını arttırmak amacıyla “1 yıldan az”, “1 yıldan fazla” ve “hiç kullanmadı” olmak üzere 3 kategoride toplanmıştır. Bağımlı değişken üzerinde yapılan veri dönüştürme işlemi Tablo 4.1’de gösterilmiştir.

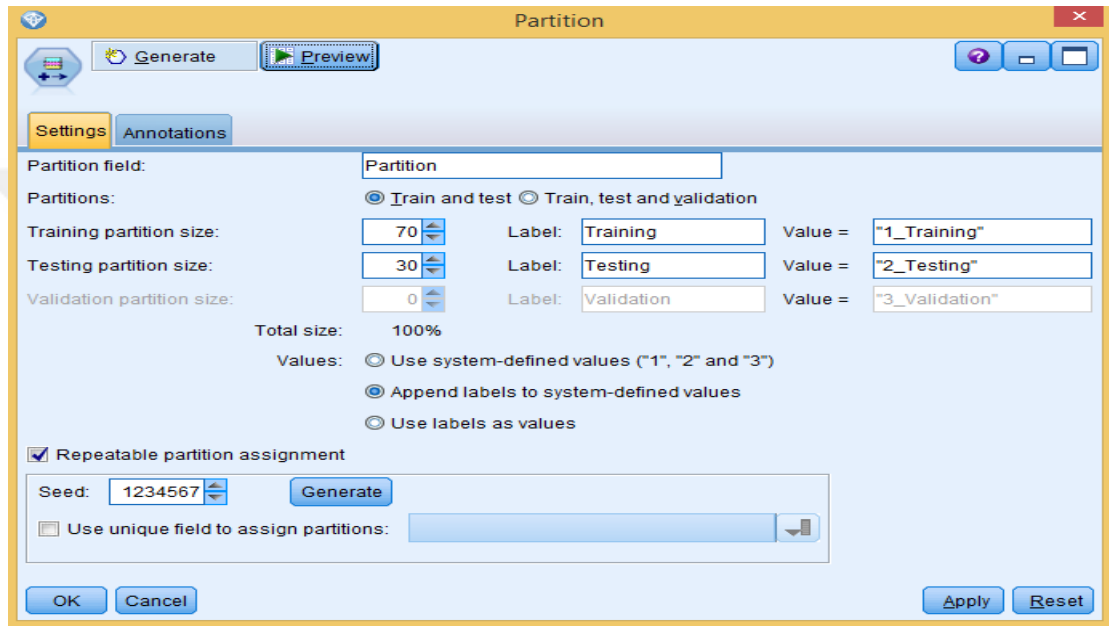
Tablo 4.1: Veri Dönüştürme İşlemi

Dönüştürme İşlemi Uygulanan Değişken	Değişken Değerinin İlk Hali	Değişken Değerinin Son Hali
SON KULLANIM ETİCARET	Son üç ay içinde	1 yıldan az
	Üç ay ile bir yıl arasında	
	Bir yıldan çok	1 yıldan fazla
	Hiç kullanmadı	Hiç kullanmadı

4.3. Veri Setinin Bölünmesi (training ve testing)

Veri setinin doğruluğunu test etmek için modelleme işleminden önce veriler eğitim ve test olmak üzere iki parçaya bölünür, bölünme “Field.Ops” paleti altında yer alan “partition” işlemcisi ile gerçekleştirilir. Burada %70 eğitim (training) ve %30 test (testing) seçerek analiz işlemine devam edilmiştir (Şekil 4.2).

Şekil 4.2: Verilerin Partiton İşlemcisi İle Bölünmesi



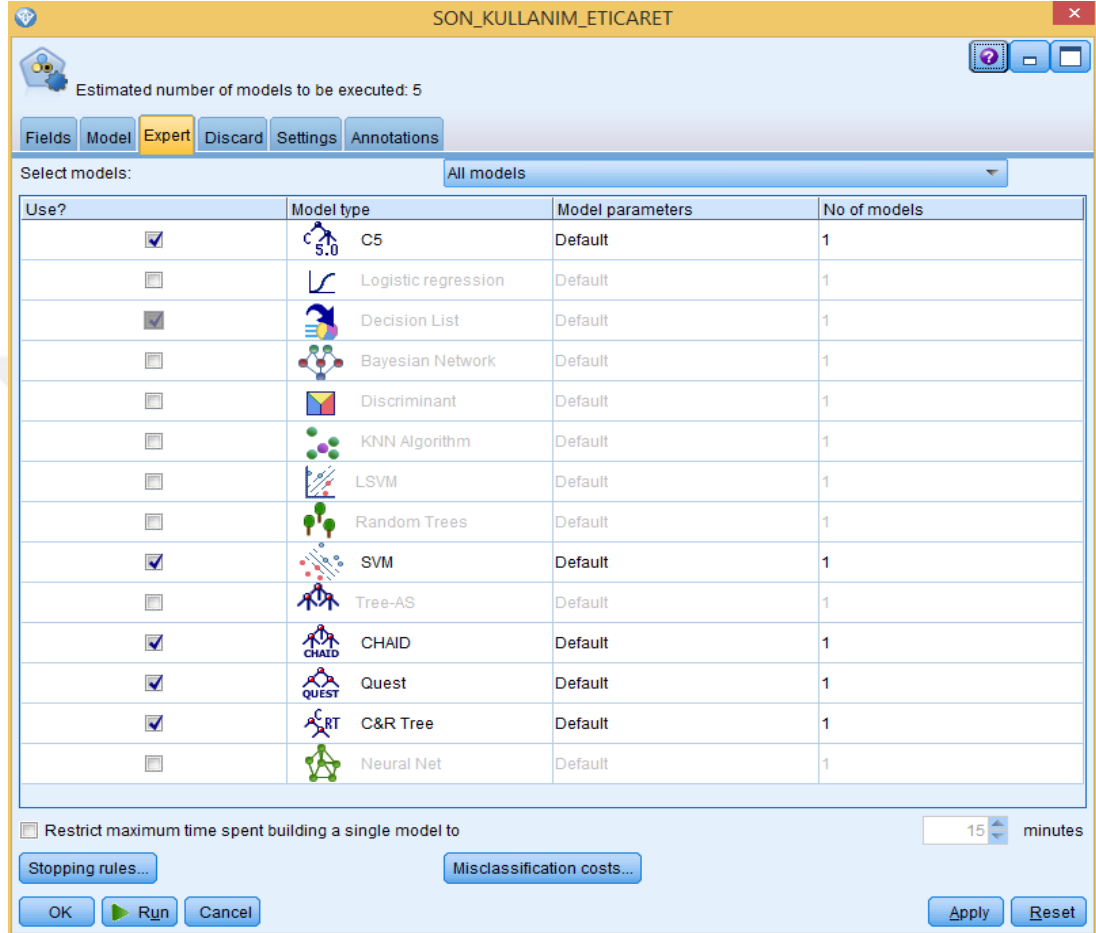
IBM SPSS Modeler programında bölme işlemi ikiye bölünüp analiz işlemi devam etmiştir.

4.4. Otomatik sınıflama işlemcisi ile kullanılacak modelin belirlenmesi

Otomatik modelleme işlemcisi farklı algoritmaları karşılaştırır ve modelleme için elimizdeki veri yapısına uygun olan en iyi algoritmanın belirlenmesini sağlar. Bağımlı değişkenin sürekli ya da kategorik olmasına göre kullanılacak otomatik modelleme işlemcileri bulunmaktadır. Elimizdeki veri kategorik ve sınıflama işlemi yapılacağı için otomatik sınıflama işlemcisi (Auto Classifier) kullanılacaktır (Şekil 4.3). Bu işlemci ile sınıflama ölçme düzeyinde bağımlı değişken için birden çok bağımlı değişken test eder ve karşılaştırır. Modelleme işleminde kullanılacak değişkenler kategorik ya da sürekli olabilir ancak bağımlı değişken kategorik olmak zorundadır.

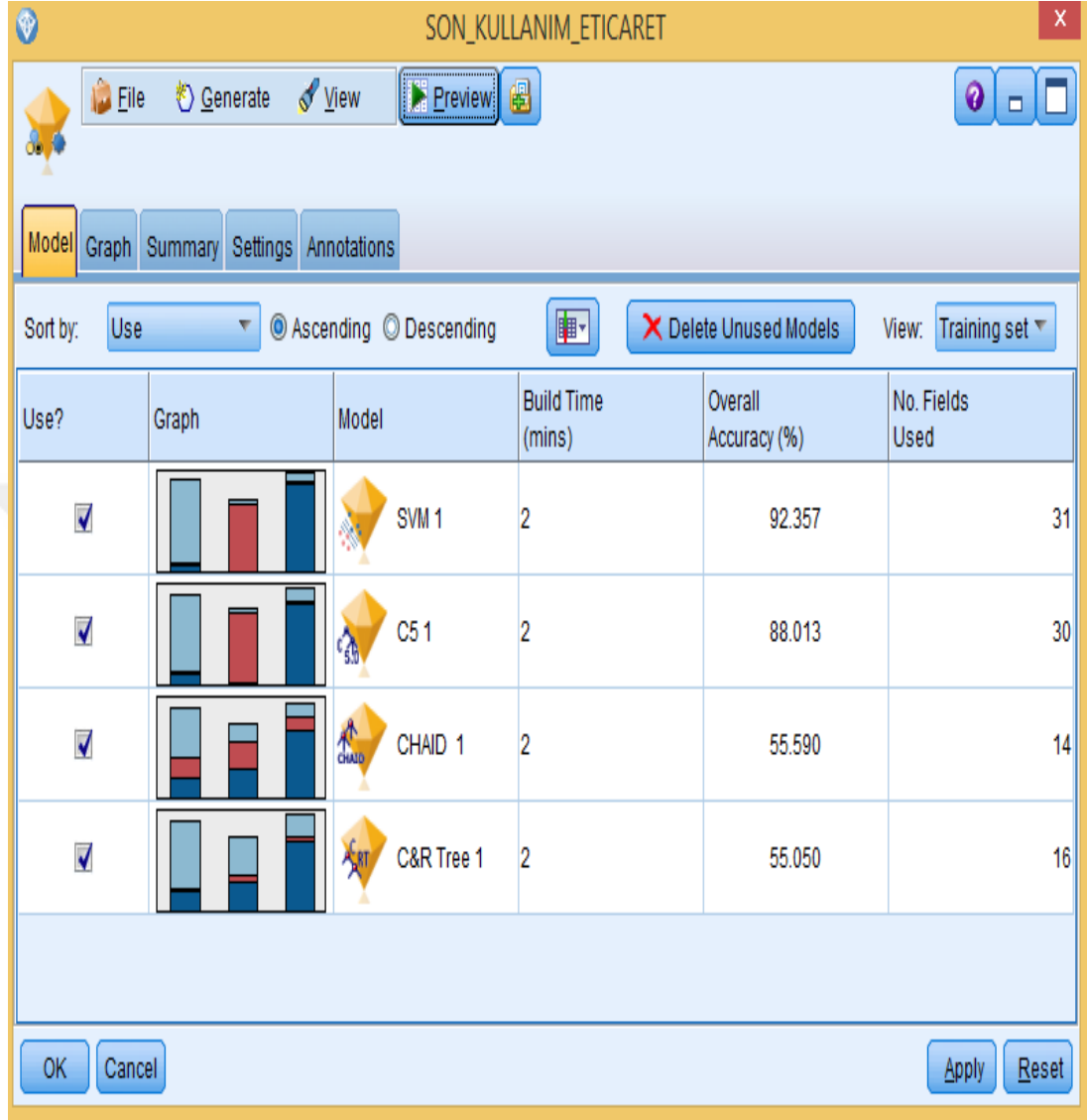
Otomatik sınıflama işlemcisi tezde bahsettiğimiz C5.0, CHAİD, C&RT algoritmaları ve Destek Vektör Makinalarını desteklemektedir.

Şekil 4.3: Auto Classifier İşlemcisi


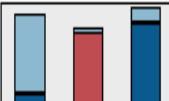
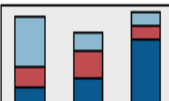
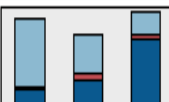


Auto Classifier işlemcisinde karar ağacı algoritmaları ve destek vektör makinaları seçildikten sonra nodül çalıştırılır. Çıkan sonuca göre elimizdeki veri yapısına uygun olarak DVM, C5.0, CHAİD, QUEST ve C&RT algoritmaları seçilmiş ve Auto Classifier işlemcisi çalıştırılmıştır.

Şekil 4.4: Auto Classifier İşlemcisi Sonuçları



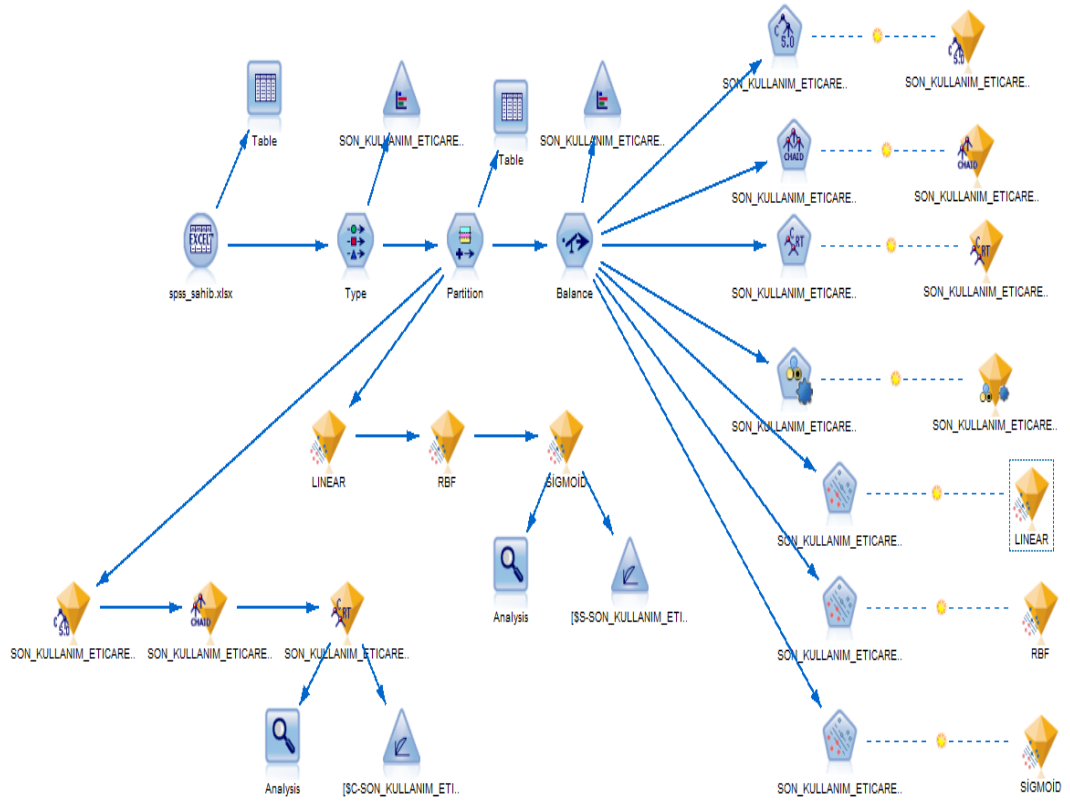
The screenshot shows the SON_KULLANIM_ETICARET Auto Classifier interface. The window title is "SON_KULLANIM_ETICARET". The menu bar includes "File", "Generate", "View", "Preview", and a help icon. The main area has tabs for "Model", "Graph", "Summary", "Settings", and "Annotations". Below the tabs, there are controls for "Sort by: Use", "Ascending", "Descending", "Delete Unused Models", and "View: Training set". The main table displays the results of the classification process.

Use?	Graph	Model	Build Time (mins)	Overall Accuracy (%)	No. Fields Used
<input checked="" type="checkbox"/>		SVM 1	2	92.357	31
<input checked="" type="checkbox"/>		C5 1	2	88.013	30
<input checked="" type="checkbox"/>		CHAID 1	2	55.590	14
<input checked="" type="checkbox"/>		C&R Tree 1	2	55.050	16

At the bottom of the window, there are buttons for "OK", "Cancel", "Apply", and "Reset".

Veri kümesini kullanılması ile kıyaslanan modellerin genel doğruluk oranlarına bakıldığında DVM (Support Vector Machine) %92,36 ile en yüksek genel doğruluk oranına sahiptir. Sonraki algoritmalar karar ağaçları algoritmaları olan C5.0 %88,01, Chaid %55,59 ve C&R Tree %55,05 algoritmaları olduğu gözlenmiştir. Karar ağaçları algoritmaları genel doğruluk oranı açısından en iyi gösterge C5.0 dir, DVM algoritması kadar iyi sonuç vermemektedir.

Şekil 4.5: Kurulan Modelin Ekran Görüntüsü

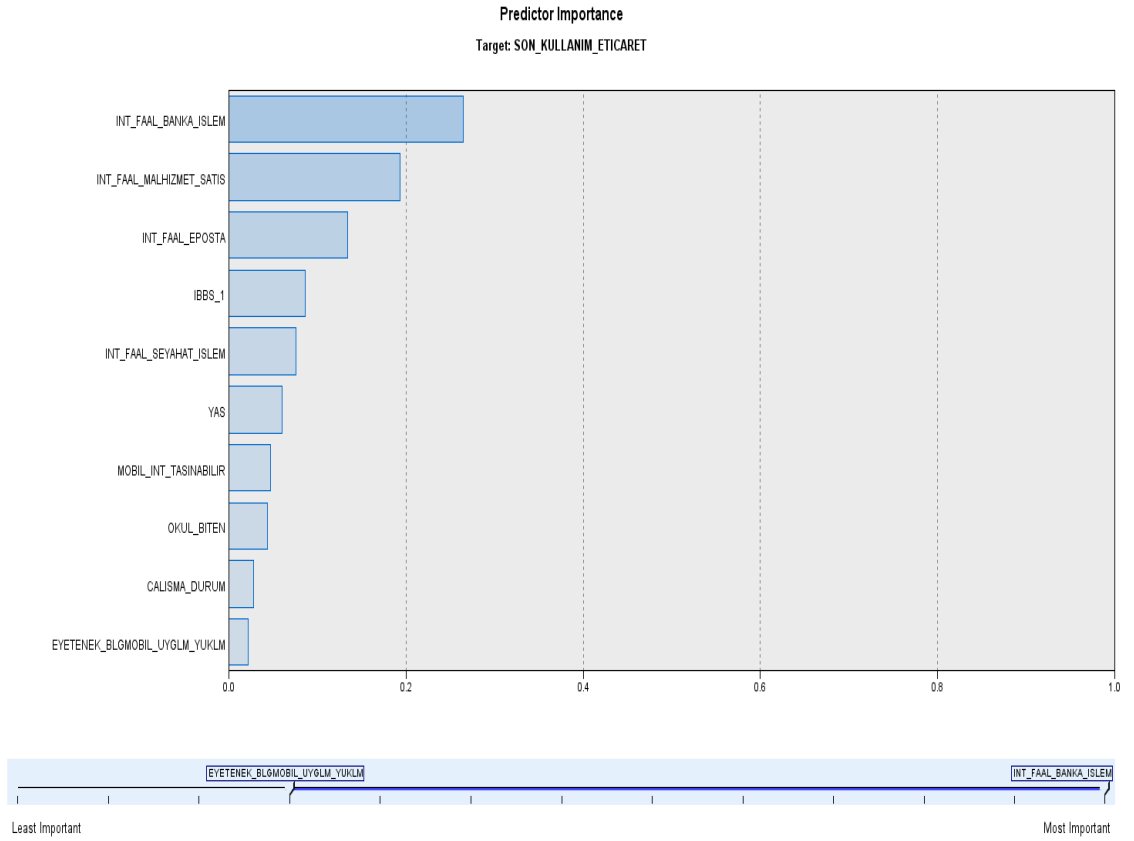


Auto Classifier işlemcisinin sonuçlarına göre, IBM SPSS Modeler programında altı farklı model kurulmuştur. Bu modellerin ekran görüntüsü Şekil 4.5’deki gibidir.

4.4.1. CHAID Algoritması

CHAID algoritmasının en önemli gördüğü değişken banka işlem olmuştur, malhizmet satış değişkeni takip etmiş ve diğer önemli gördüğü değişkenler Şekil 4.6’da yer almaktadır.

Şekil 4.6: CHAID Algoritmasına Göre Önemli Değişkenler

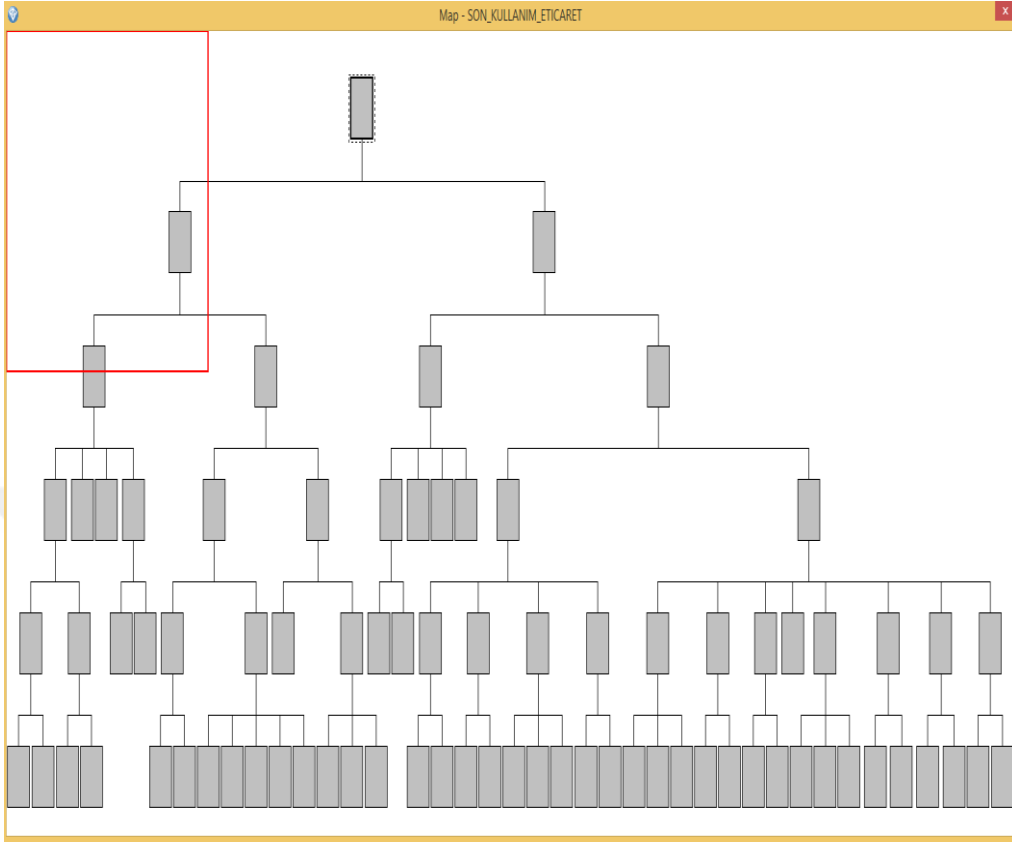


Tablo 4.2: CHAID Algoritmasında Kullanılan Değişkenlerin Önem Değerleri

Değişkenler	Değişken Önemi
Değişken_1 (int_faal_banka_islem)	0,26
Değişken_2 (int_faal_malhizmet_satis)	0,19
Değişken_3 (int_faal_eposta)	0,13
Değişken_4 (ibbs_1)	0,09
Değişken_5 (int_faal_seyahat_islem)	0,08
Değişken_6 (yas)	0,06
Değişken_7 (mobil_int_tasinabilir)	0,05
Değişken_8 (okul_biten)	0,04
Değişken_9 (calisma_durum)	0,03
Değişken_10 (eyetenek_blgmobil_uyglm_yuklm)	0,02

CHAID algoritması ile kurulan modelde en önemli değişken, 0,26 önem değeri ile “BANKA_ISLEM” değişkeni olmuştur. Bu değişkeni 0,19 önem değeri ile “MALHIZMET_SATIS” değişkeni, 0,13 önem değeri ile “EPOSTA” değişkeni takip etmiştir.

Şekil 4.7: CHAID Algoritması ile Oluşturulan Karar Ağacının Genel Görünümü



Karar ağacının ilk düğümünde bağımlı değişken olan son kullanım eticaret değişkeni yer almaktadır. Ağaç derinliği 5 olan karar ağacı, toplam 80 düğümden oluşmuştur. Karar ağacının başlangıç düğümü Şekil 4.8’de gösterilmiştir.

Şekil 4.8: CHAID Algoritması ile Oluşan Karar Ağacının Başlangıç Düğümü

Node 0		
Category	%	n
1 yıldan az	33.538	4421
1 yıldan fazla	33.880	4466
Hiç kullanmadı	32.582	4295
Total	100.000	13182

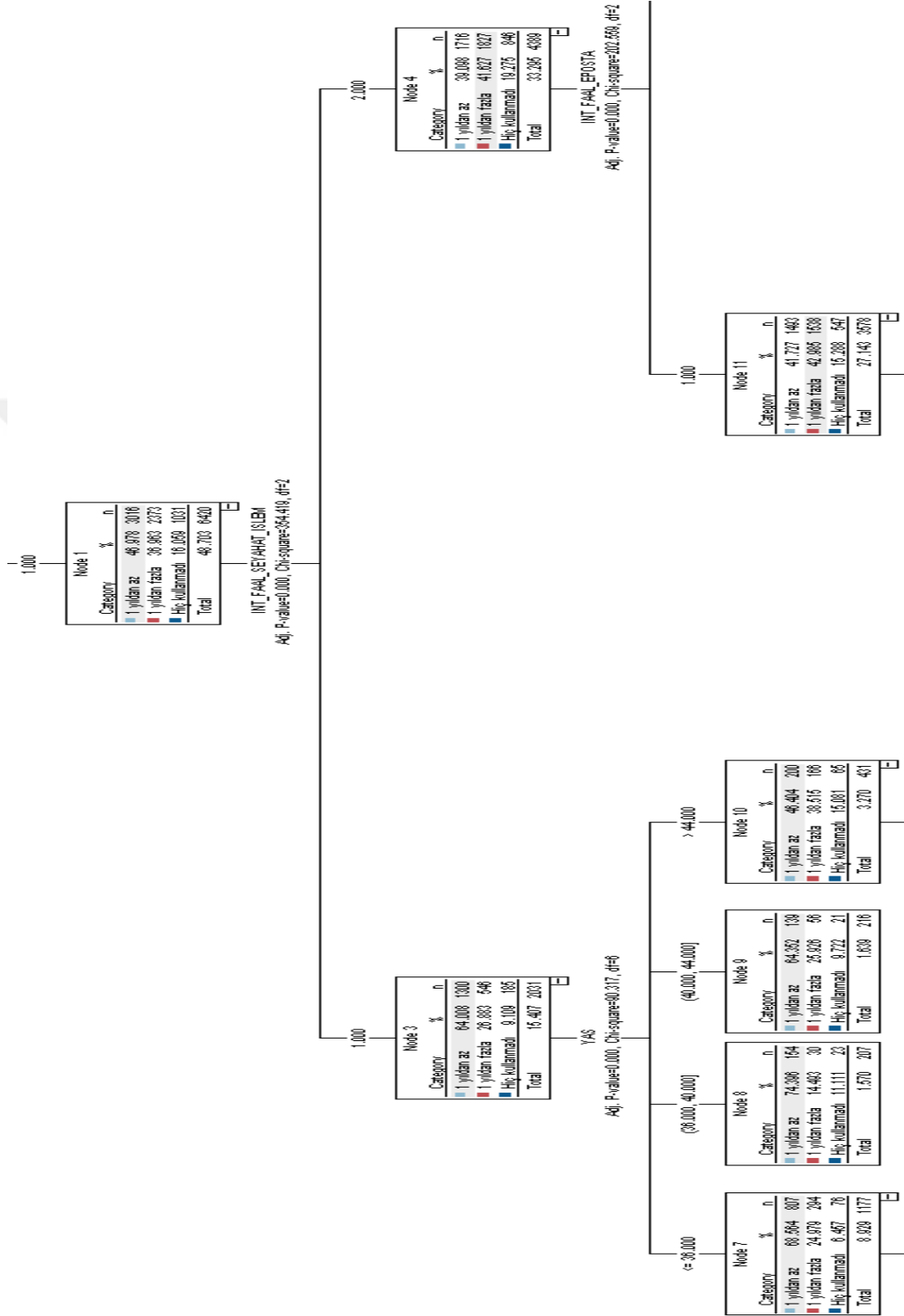
INT_FAAL_BANKA_ISLEM
Adj. P-value=0.000, Chi-square=1757.861, df=2

Başlangıç düğümünde yer alan 13182 ferdin % 33,538'i internet üzerinden mal veya hizmet satın alma ya da sipariş verme işlemini son 1 yılda, % 33,880'i 1 yıldan fazla süre içerisinde ve % 32,582'si ise hiç kullanmadı olarak nitelendirmiştir.

CHAID algoritmasını ile oluşturulan karar ağacı dallanmaya "INT_FAAL_BANKA_ISLEM" değişkeni ile 2 alt sınıf oluşturarak başlamıştır. Son 3 ay içerisinde internet bankacılığı kullanıp kullanmadığı sorusuna verilen "Evet" cevabı ile 1. düğüm, "Hayır" cevabı ile 2. düğüm oluşmuştur. İlk düğüme göre oluşan karar ağacı Şekil 4,9'da gösterilmiştir.



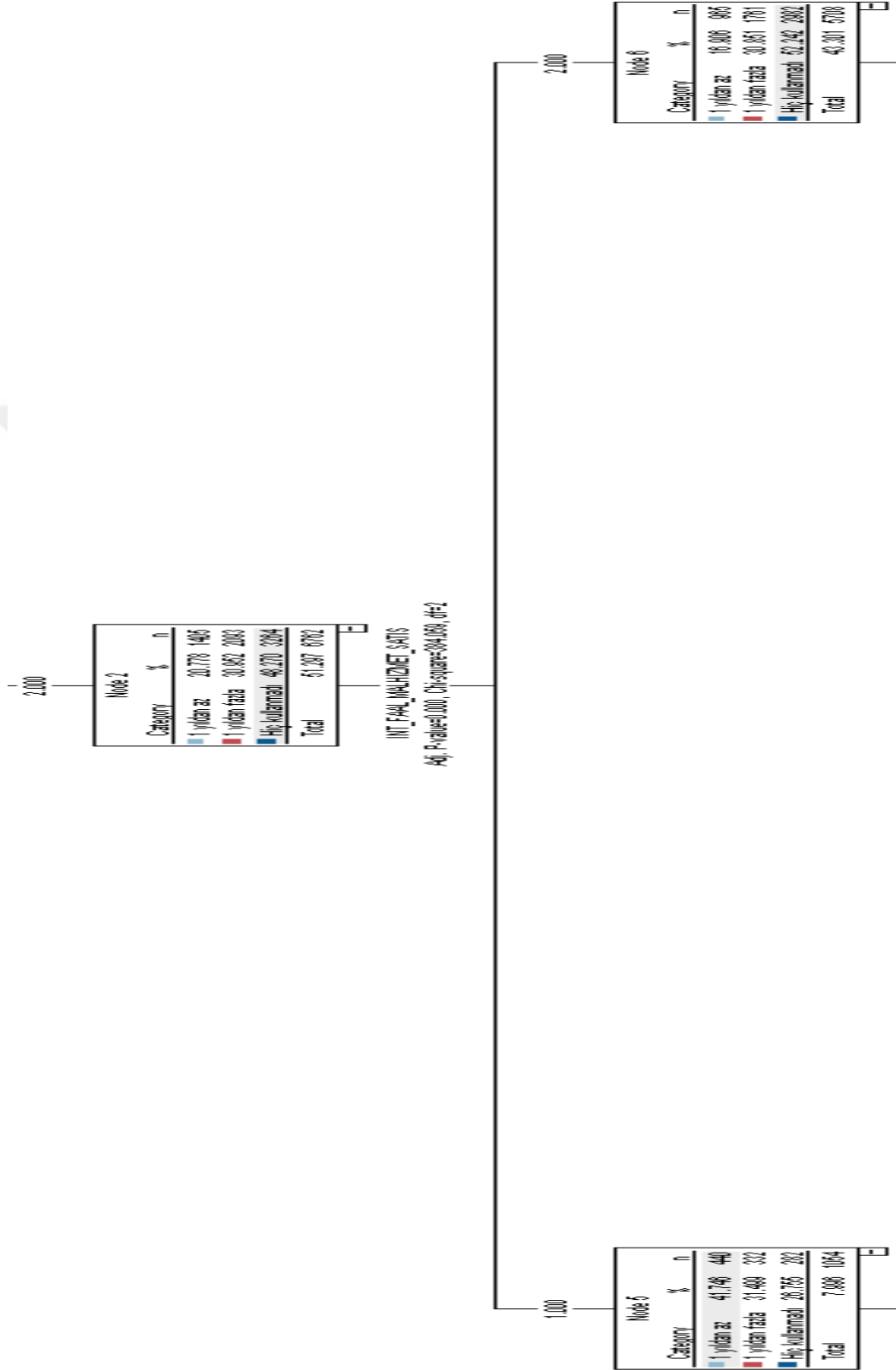
Şekil 4.9: CHAID Algoritmasının 1. Düzüme Göre Dallanması



Karar ağacının 1. düğümü incelendiğinde e-ticaret kullanımı 1 yıldan az olarak nitelendiren fertlerin oranının %46,978, 1 yıldan fazla nitelendirenlerin oranının %36,963, hiç kullanmayan ise %16,059 olduğu gözlenmektedir.

Düğüm 1 ‘SEYAHAT_İŞLEM’ değişkeni ile dallanmıştır. Seyahat ve konaklama ile ilgili hizmetleri kullanma (Otel rezervasyonu, bilet satış vb.) sorusuna verilen ‘Evet’ cevabı ile düğüm 3, ‘Hayır’ cevabı ile de düğüm 4 oluşmuştur. Düğüm 3’de e-ticaret kullanımını 1 yıldan az söyleyen fertlerin oranı %64,008, 1 yıldan fazla olduğunu söyleyen fertlerin %26,883, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %9,109’dur. Düğüm 4’de e-ticaret kullanımını 1 yıldan az olarak nitelendiren fertlerin oranı %39,098’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %41,627’ye yükselmiş, hiç kullanmadı nitelendiren oranı ise %19,275’e yükselmiştir.

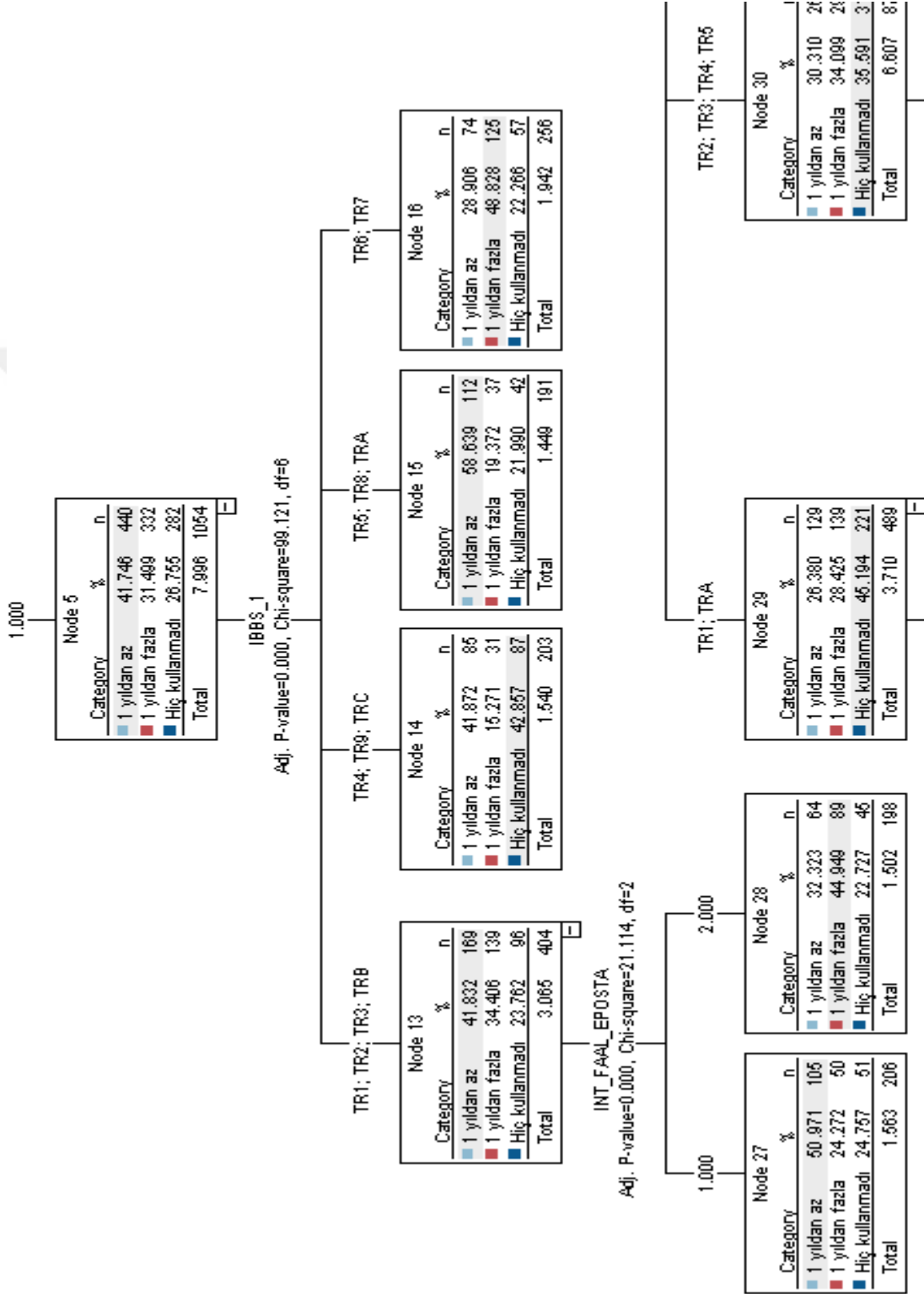
Şekil 4.10: CHAID Algoritmasının 2. Düzüme Göre Dallanması



Karar ağacının 2. düğümü incelendiğinde e-ticaret kullanımı 1 yıldan az olarak nitelendiren fertlerin oranının %20,778, 1 yıldan fazla nitelendirenlerin oranının %30,952, hiç kullanmayan ise %48,270 olduğu gözlenmektedir.

Düğüm 2 “MALHIZMET_SATIS” değişkeni ile dallanmıştır. Mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.) sorusuna verilen “Evet” cevabı ile düğüm 5, “Hayır” cevabı ile de düğüm 6 oluşmuştur. Düğüm 5’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %41,746, 1 yıldan fazla olduğunu söyleyen fertlerin %31,499, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %26,755’dir. Düğüm 6’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %16,906’a azalmış, 1 yıldan fazla nitelendirenlerin oranı %30,851’e azalmış, hiç kullanmadı nitelendiren oranı ise %52,242’e yükselmiştir.

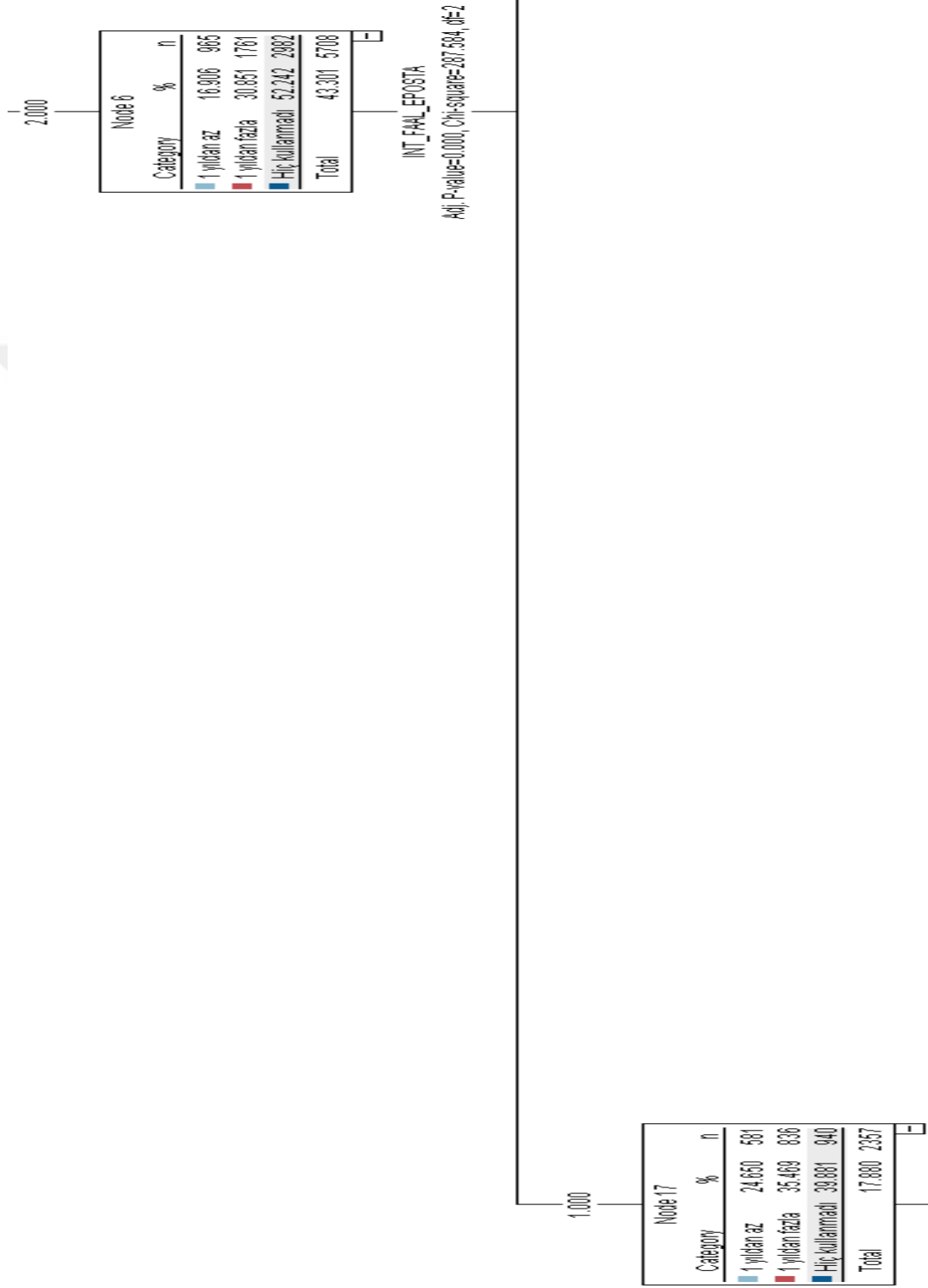
Şekil 4.11: CHAID Algoritmasının 5. Düzüme Göre Dallanması



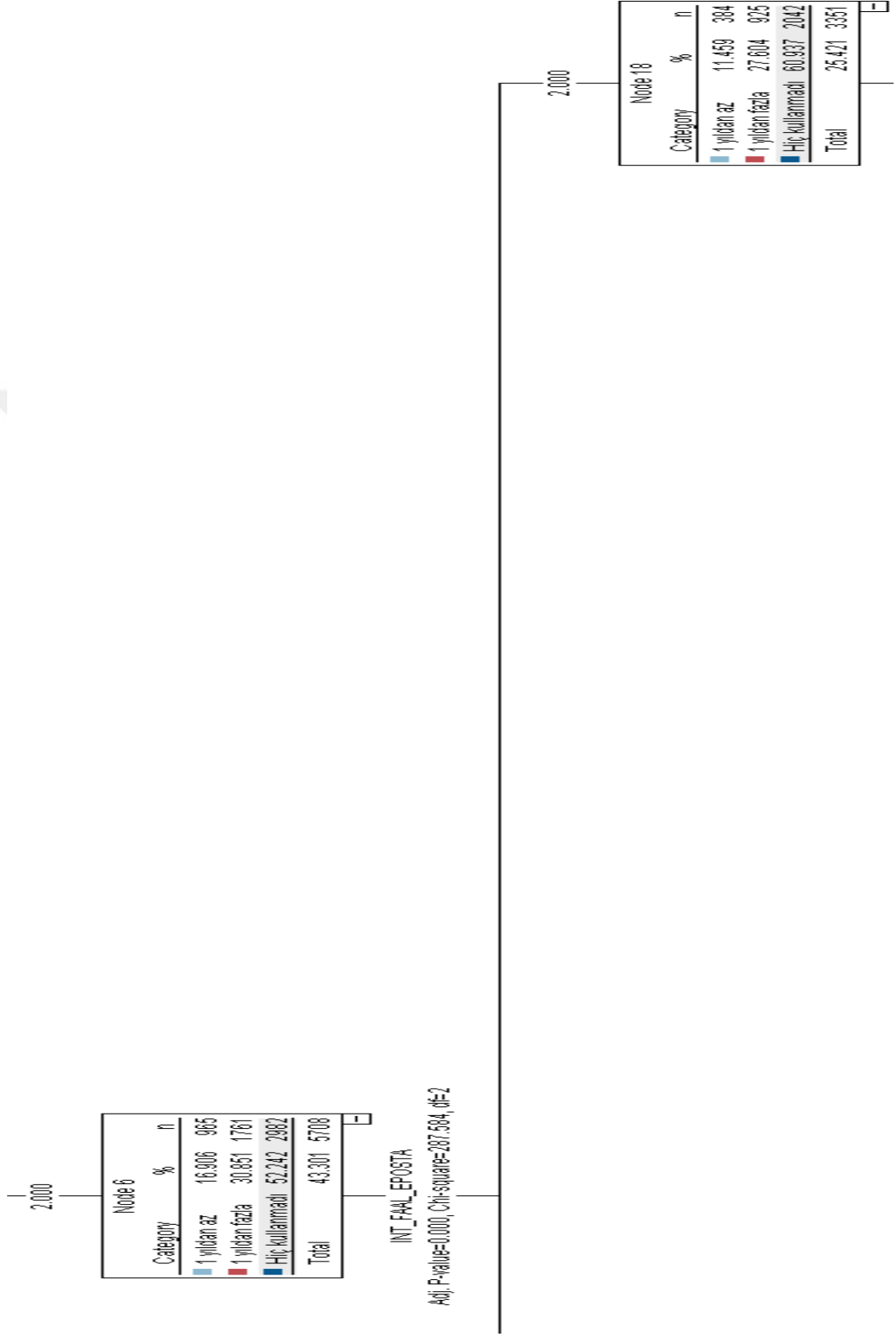
Düğüm 5 “IBBS_1” deęişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen “TR1; TR2; TR3; TRB” cevabı ile düğüm 13, “TR4; TR9; TRC” cevabı ile düğüm 14, “TR5; TR8; TRA” cevabı ile düğüm 15 ve , “TR6; TR7” cevabı ile düğüm 16 oluşmuştur. Düğüm 13’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %41,832, 1 yıldan fazla olduğunu söyleyen fertlerin %34,406, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise % 23,762’dir. Düğüm 14’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı önceki düğümdeki fertlere göre %41,872’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %15,271’e azalmış, hiç kullanmadı nitelendiren oranı ise %42,857’e yükselmiştir. Düğüm 15’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı önceki düğümdeki fertlere göre %58,639’e yükselmiş, 1 yıldan fazla olduğunu söyleyen fertlerin %19,372’e yükselmiş, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise % 21,990’a azalmıştır. Düğüm 16’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı önceki düğümdeki fertlere göre %28,906’a azalmış, 1 yıldan fazla nitelendirenlerin oranı %48,828’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %22,266’a yükselmiştir.

Düğüm 14 de dallanma meydana gelmemiştir fakat düğüm 13 “FAAL_EPOSTA” deęişkeni ile dallanmıştır. E-posta gönderme/alma sorusuna verilen “Evet” cevabı ile düğüm 27, “Hayır” cevabı ile de düğüm 28 oluşmuştur. Düğüm 27’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %50,971, 1 yıldan fazla olduğunu söyleyen fertlerin %24,272, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %24,757’dir. Düğüm 28’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %32,323’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %44,949’a yükselmiş, hiç kullanmadı nitelendiren oranı ise %22,727’e azalmıştır.

Şekil 4.12: CHAID Algoritmasının 6. Düzüme Göre Dalkanması



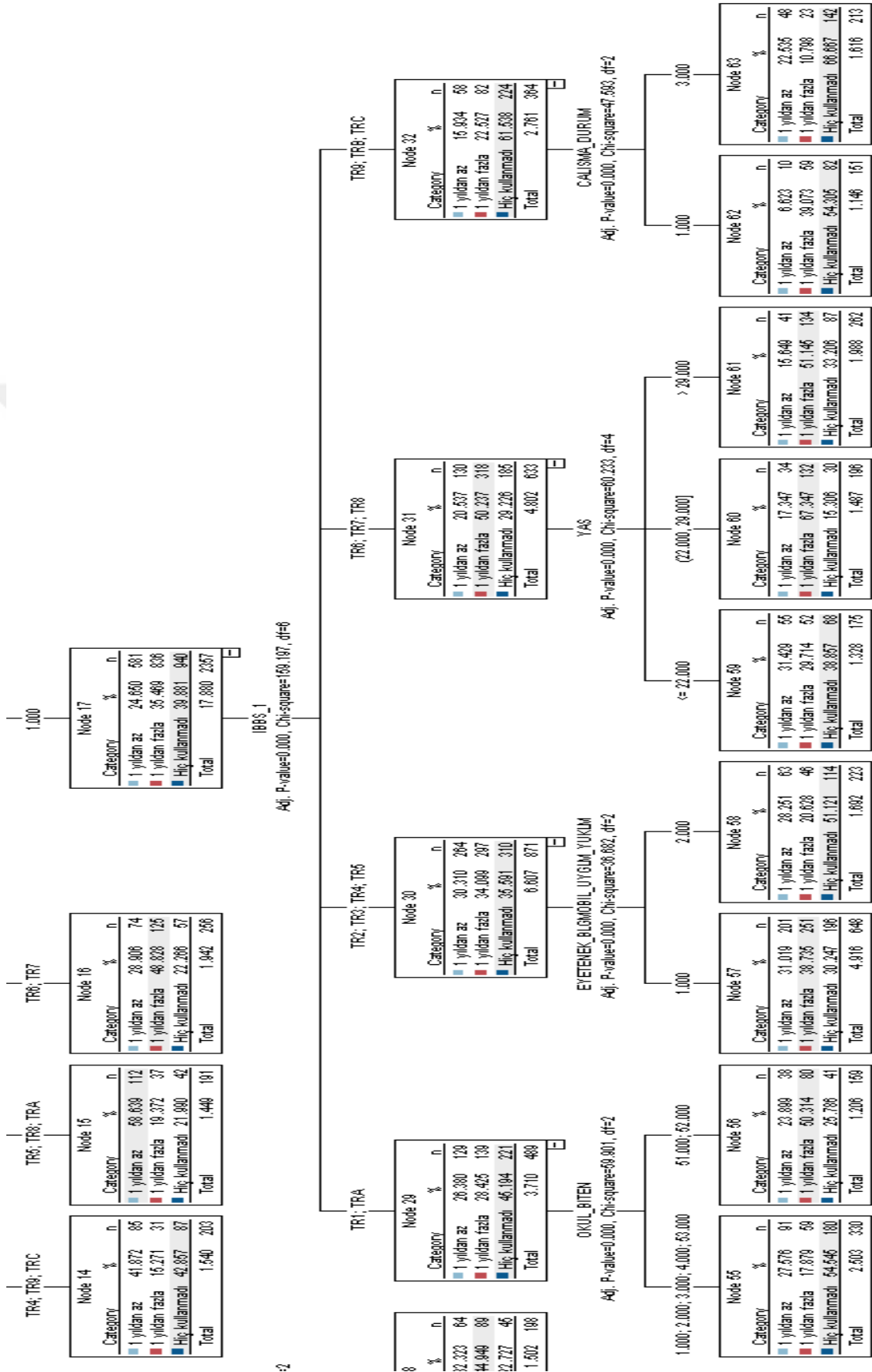
Şekil 4.13: Şekil 4.12'nin Devamı



Düğüm 6 “FAAL_EPOSTA” değişkeni ile dallanmıştır. E-posta gönderme/alma sorusuna verilen “Evet” cevabı ile düğüm 17, “Hayır” cevabı ile de düğüm 18 oluşmuştur. Düğüm 17’de e-posta kullanımı 1 yıldan az söyleyen fertlerin oranı %24,650, 1 yıldan fazla olduğunu söyleyen fertlerin %35,469, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %39,881’dir. Düğüm 18’de e-posta kullanımı 1 yıldan az olarak nitelendiren fertlerin oranı %11,459’a azalmış, 1 yıldan fazla nitelendirenlerin oranı %27,604’e azalmış, hiç kullanmadı nitelendiren oranı ise %60,937’e yükselmiştir.



Şekil 4.14: CHAID Algoritmasının 17. Düzüme Göre Dallanması



Düğüm 17 “IBBS_1” değişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen “TR1; TRA” cevabı ile düğüm 29, “TR2; TR3; TR4; TR5” cevabı ile düğüm 30, “TR6; TR7; TR8” cevabı ile düğüm 31 ve “TR9; TRB; TRC” cevabı ile de düğüm 32 oluşmuştur. Düğüm 29’da e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %26,380, 1 yıldan fazla olduğunu söyleyen fertlerin %28,425, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise % 45,194’dür. Düğüm 30’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %30,310’a yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %34,099’a yükselmiş, hiç kullanmadı nitelendiren oranı ise %35,591’e azalmıştır. Düğüm 31’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı önceki düğümdeki fertlere göre %20,537 azalmış, 1 yıldan fazla olduğunu söyleyen fertlerin %50,237 yükselmiş, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise % 29,226 azalmıştır. Düğüm 32’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %15,934’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %22,527’e azalmış, hiç kullanmadı nitelendiren oranı ise %61,538’e yükselmiştir.

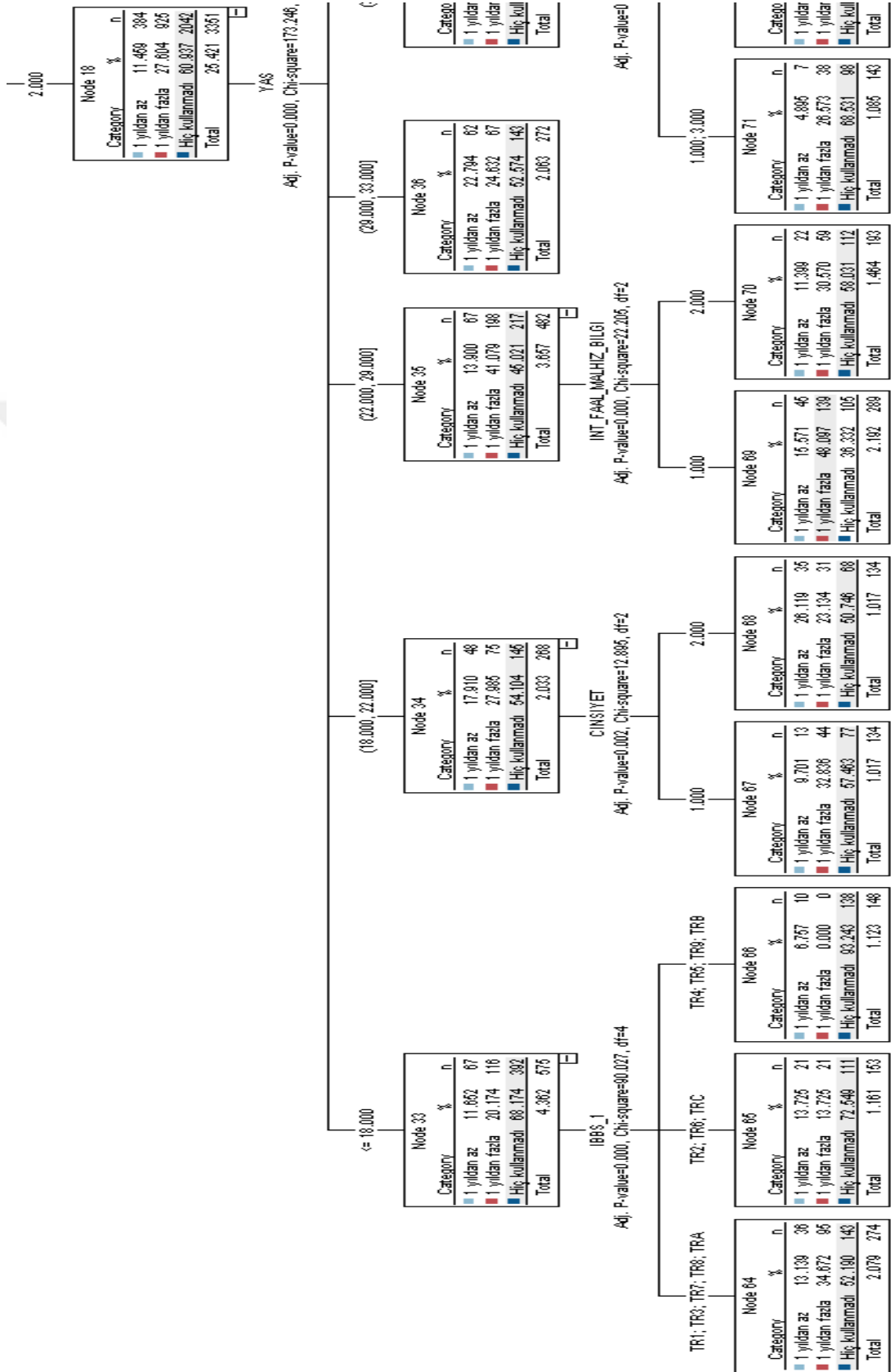
Düğüm 29 “OKUL_BITEN” değişkeni ile dallanmıştır. Eğitim durumu sorusuna verilen “Evet” cevabı ile düğüm 55, “Hayır” cevabı ile de düğüm 56 oluşmuştur. Düğüm 55’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %27,576, 1 yıldan fazla olduğunu söyleyen fertlerin %17,879, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %54,545’dir. Düğüm 56’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %23,899’a azalmış, 1 yıldan fazla nitelendirenlerin oranı %50,314’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %25,786’a azalmıştır.

Düğüm 30 “EYETENEK_BLGMOBIL_UYGLM_YUKLM” değişkeni ile dallanmıştır. Yazılım veya mobil uygulama (application) yükleme sorusuna verilen “Evet” cevabı ile düğüm 57, “Hayır” cevabı ile de düğüm 58 oluşmuştur. Düğüm 57’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %31,019, 1 yıldan fazla olduğunu söyleyen fertlerin %38,735, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %30,247’dir. Düğüm 58’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %28,251’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %20,628’e azalmış, hiç kullanmadı nitelendiren oranı ise %51,121’e yükselmiştir.

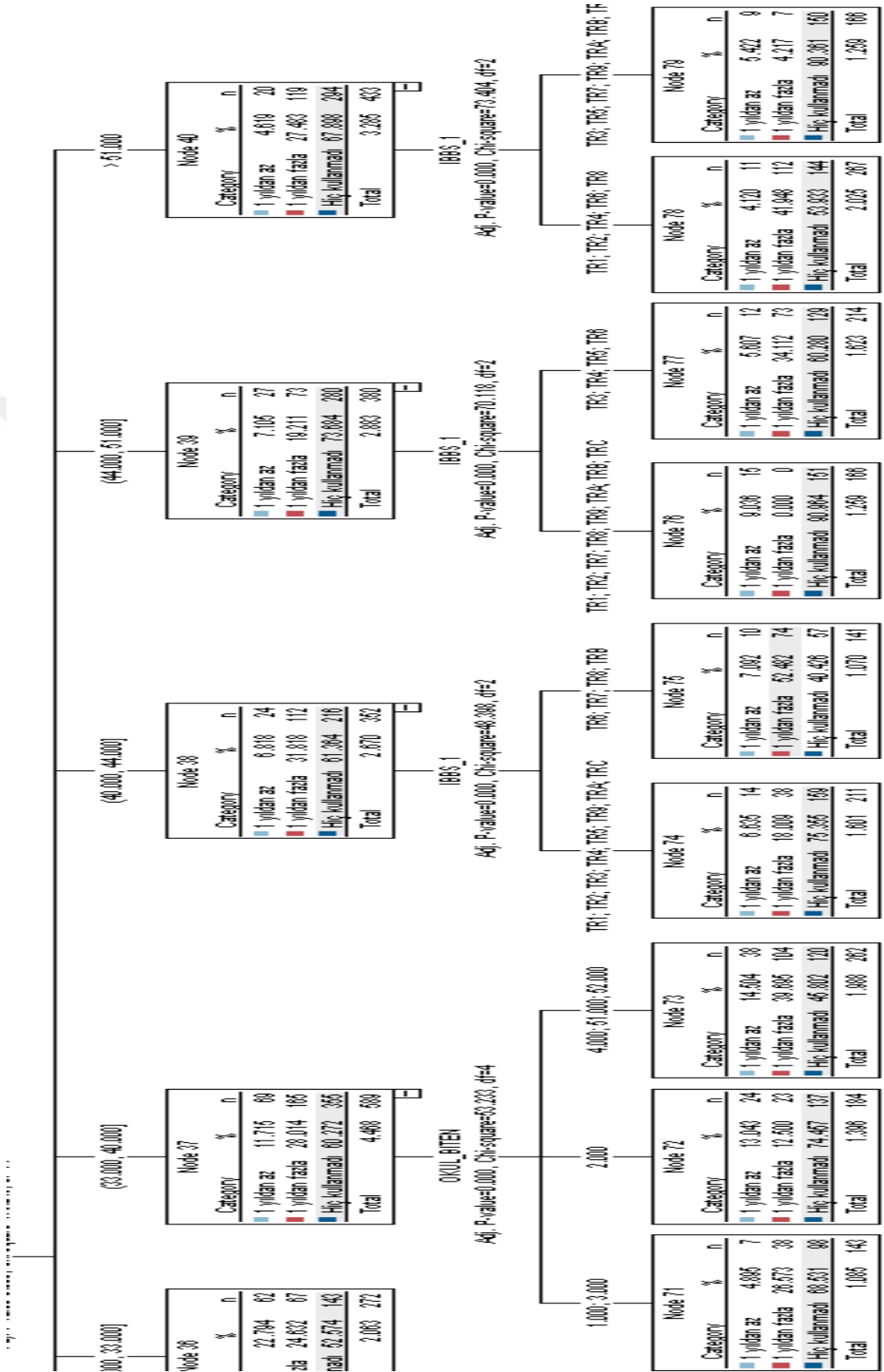
Düğüm 31 “YAS” değişkeni ile dallanmıştır. Ferdin yaşı sorusuna verilen “22 yaşdan az” düğüm 59, “22 ile 29 yaş” düğüm 60 ve “29 yaştan yukarı” cevabı ise düğüm 61 ile oluşmuştur. Düğüm 59’da e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %31,429, 1 yıldan fazla olduğunu söyleyen fertlerin %29,714, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %38,857’dir. Düğüm 60’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren %17,347’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %67,347’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %15,306’a azalmıştır. Düğüm 61’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı önceki düğümdeki fertlere göre %15,649’a azalmış, 1 yıldan fazla nitelendirenlerin oranı %51,145’e azalmış, hiç kullanmadı nitelendiren oranı ise %33,206’a yükselmiştir.

Düğüm 32 “CALISMA_DURUM” değişkeni ile dallanmıştır. Referans haftası çalışma durumu sorusuna verilen “Evet” cevabı ile düğüm 62, “Hayır” cevabı ile de düğüm 63 oluşmuştur. Düğüm 62’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %6,623, 1 yıldan fazla olduğunu söyleyen fertlerin %39,073, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %54,305’dir. Düğüm 63’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %22,535’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %10,798’e azalmış, hiç kullanmadı nitelendiren oranı ise %66,667’e yükselmiştir.

Şekil 4.15: CHAID Algoritmasının 18. Düzüme Göre Dallanması



Şekil 4.16: CHAID Algoritmasının 18. Düzüme Göre Dallanması



Düğüm 18 “YAS” değişkeni ile dallanmıştır. Ferdin yaşı sorusuna verilen “18 yaşdan az” cevabı ile düğüm 33 , “18 ile 22 yaş” cevabı ile düğüm 34, “22 ile 29 yaş” düğüm 35 ve “29 ile 33 yaş” cevabı ise düğüm 36, “33 ile 40 yaş” cevabı ile düğüm 37, “40 ile 44 yaş” cevabı ile de düğüm 38, “44 ile 51 yaş” cevabı ile düğüm 39 ve “51 yaş üzeri” cevabı ile de düğüm 40 oluşmuştur. Düğüm 33’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %11,652, 1 yıldan fazla olduğunu söyleyen fertlerin %20,174, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %68,174’dür. Düğüm 34’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %17,910’a yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %27,985’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %54,104’e azalmıştır. Düğüm 35’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı önceki düğümdeki fertlere göre %13,900 azalmış, 1 yıldan fazla olduğunu söyleyen fertlerin %41,079 yükselmiş, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %45,021 azalmıştır. Düğüm 36’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %22,794’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %24,632’e azalmış, hiç kullanmadı nitelendiren oranı ise %52,574’e yükselmiştir. Düğüm 37’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %11,715, 1 yıldan fazla olduğunu söyleyen fertlerin %28,014, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %60,272’dir. Düğüm 38’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %6,818’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %31,818’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %61,364’e yükselmiştir. Düğüm 39’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı önceki düğümdeki fertlere göre %7,105 yükselmiş, 1 yıldan fazla olduğunu söyleyen fertlerin %19,211 azalmış, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %73,684 yükselmiştir. Düğüm 40’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %4,619’a azalmış, 1 yıldan fazla nitelendirenlerin oranı %27,483’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %67,898’e azalmıştır.

Düğüm 33 “IBBS_1” değişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen “TR1; TR3; TR7; TR8; TRA” cevabı ile düğüm 64, “TR2; TR6; TRC” cevabı ile düğüm 65 ve “TR4; TR5; TR9; TRB” cevabı ile de düğüm 66 oluşmuştur. Düğüm 64’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %13,139, 1 yıldan fazla olduğunu söyleyen fertlerin oranı %34,672, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %52,190’dır.

Düğüm 65’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %13,725’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %13,725’e azalmış, hiç kullanmadı nitelendiren oranı ise %72,549’a yükselmiştir. Düğüm 66’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı önceki düğümdeki fertlere göre %6,757’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %0,000’a azalmış, hiç kullanmadı nitelendiren oranı ise %93,243’e yükselmiştir.

Düğüm 34 “CINSİYET” değişkeni ile dallanmıştır. Ferdin cinsiyeti sorusuna verilen “Evet” cevabı ile düğüm 67 , “Hayır” cevabı ile de düğüm 68 oluşmuştur. Düğüm 67’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %9,701, 1 yıldan fazla olduğunu söyleyen fertlerin %32,836, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %57,463’dür. Düğüm 68’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %26,119’a yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %23,134’e azalmış, hiç kullanmadı nitelendiren oranı ise %50,746’a azalmıştır.

Düğüm 36’da dallanma meydana gelmemiştir fakat düğüm 35 “MALHIZ_BILGI” değişkeni ile dallanmıştır. Mal ve hizmetler hakkında bilgi arama sorusuna verilen “Evet” cevabı ile düğüm 69, “Hayır” cevabı ile de düğüm 70 oluşmuştur. Düğüm 69’da e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %15,571, 1 yıldan fazla olduğunu söyleyen fertlerin %48,097, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %36,332’dir. Düğüm 70’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %11,399’a azalmış, 1 yıldan fazla nitelendirenlerin oranı %30,570’e azalmış, hiç kullanmadı nitelendiren oranı ise %58,031’e artmıştır.

Düğüm 37 “OKUL_BITEN” değişkeni ile dallanmıştır. Eğitim durumu sorusuna verilen cevap düğüm 71, düğüm 72 ve düğüm73 ile oluşmuştur. Düğüm 71’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %4,895, 1 yıldan fazla olduğunu söyleyen fertlerin %26,573, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %68,531’dir. Düğüm 72’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %13,043’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %12,500’e azalmış, hiç kullanmadı nitelendiren oranı ise %74,457’e yükselmiştir. Düğüm 73’te e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı önceki düğümdeki fertlere %14,504’e yükselmiş, 1 yıldan fazla

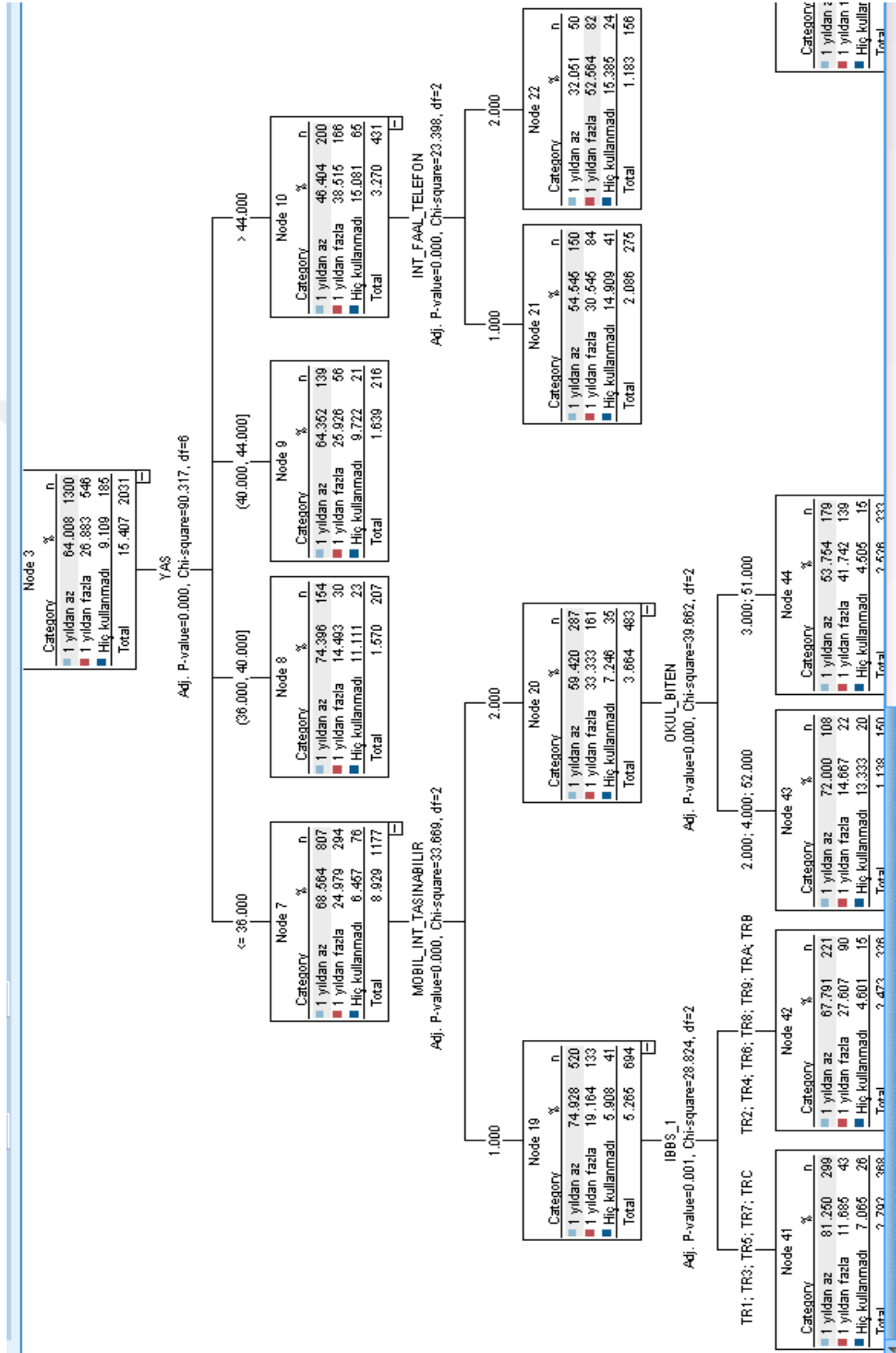
nitelendirenlerin oranı %39,695'e yükselmiş, hiç kullanmadı nitelendiren oranı ise %45,802'e azalmıştır.

Düğüm 38 "IBBS_1" değişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen "TR1; TR2; TR3; TR4; TR5; TR9; TRA; TRC" cevabı ile düğüm 74, "TR6; TR7; TR8 ;TRB" cevabı ile de düğüm 75 oluşmuştur. Düğüm 74'de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %6,635, 1 yıldan fazla olduğunu söyleyen fertlerin %18,009, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %75,355'dir. Düğüm 75'de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %7,092'e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %52,482'e yükselmiş, hiç kullanmadı nitelendiren oranı ise %40,426'a azalmıştır.

Düğüm 39 "IBBS_1" değişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen "TR1; TR2; TR7 ;TR8 ;TR9; TRA; TRB; TRC" cevabı ile düğüm 76, "TR3; TR4; TR5; TR6" cevabı ile de düğüm 77 oluşmuştur. Düğüm 76'da e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %9,036, 1 yıldan fazla olduğunu söyleyen fertlerin %0,000, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %90,964'dür. Düğüm 77'de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %5,607'e azalmış, 1 yıldan fazla nitelendirenlerin oranı %34,112'e yükselmiş, hiç kullanmadı nitelendiren oranı ise %60,280'e azalmıştır.

Düğüm 40 "IBBS_1" değişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen "TR1; TR2; TR4; TR6; TR8" cevabı ile düğüm 78, "TR3; TR5; TR7; TR9; TRA; TRB; TRC" cevabı ile de düğüm 79 oluşmuştur. Düğüm 78'de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %4,120, 1 yıldan fazla olduğunu söyleyen fertlerin %41,948, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %53,933'dür. Düğüm 79'de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %5,422'e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %4,217'e azalmış, hiç kullanmadı nitelendiren oranı ise %90,361'e yükselmiştir.

Şekil 4.17: CHAID Algoritmasının 3. Düzgüme Göre Dallanması



Düğüm 3 “YAS ” değişkeni ile dallanmıştır. Ferdin yaşı sorusuna verilen “36 yaştan az” cevabı ile düğüm 7, “36 ile 40 yaş” cevabı ile de düğüm 8, “40 ile 44 yaş” cevabı ile düğüm 9 ve “44 yaş üzeri” cevabı ile de düğüm 10 oluşmuştur. Düğüm 7’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %68,564, 1 yıldan fazla olduğunu söyleyen fertlerin %24,979, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %6,457’dir. Düğüm 8’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %74,396’a yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %14,493’e azalmış, hiç kullanmadı nitelendiren oranı ise %11,111’e yükselmiştir. Düğüm 9’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı önceki düğümdeki fertlere göre %64,352’e azalmış, 1 yıldan fazla olduğunu söyleyen fertlerin %25,926’a yükselmiş, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %9,722’e azalmıştır. Düğüm 10’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %46,404’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %38,515’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %15,081’e yükselmiştir.

Düğüm 10 “FAAL_TELEFON” değişkeni ile dallanmıştır. İnternet üzerinden telefonla görüşme/ video görüşmesi (Skype veya Facetime vb. kullanarak) sorusuna verilen “Evet” cevabı ile düğüm 21, “Hayır” cevabı ile de düğüm 22 oluşmuştur. Düğüm 21’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %54,545, 1 yıldan fazla olduğunu söyleyen fertlerin %30,545, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %14,909’dur. Düğüm 22’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %32,051’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %52,564’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %15,385’e yükselmiştir.

Node7 “MOBIL_INT_TASINABILIR” değişkeni ile dallanmıştır. Taşınabilir bilgisayar (dizüstü,netbook,tablet vb.) sorusuna verilen “1 yıldan az” cevabı ile Node19, “1 yıldan fazla” cevabı ile de Node20 oluşmuştur. Node19’da taşınabilir bilgisayar (dizüstü,netbook,tablet vb.) 1 yıldan az söyleyen fertlerin oranı %74,928, 1 yıldan fazla olduğunu söyleyen fertlerin %19,164, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %5,908’dir. Node20’de taşınabilir bilgisayar (dizüstü,netbook,tablet vb.) 1 yıldan az olarak nitelendiren fertlerin oranı %59,420’e

azalmış, 1 yıldan fazla nitelendirenlerin oranı %33,333'e yükselmiş, hiç kullanmadı nitelendiren oranı ise %7,246'a yükselmiştir.

Düğüm 19 "IBBS_1" değişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen "TR1; TR3; TR5; TR7; TRC" cevabı ile düğüm 41, "TR2; TR4; TR6; TR8; TR9; TRA; TRB" cevabı ile de düğüm 42 oluşmuştur. Düğüm 41'de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %81,250, 1 yıldan fazla olduğunu söyleyen fertlerin %11,685, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %7,065'dir. Düğüm 42'de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %67,791'e azalmış, 1 yıldan fazla nitelendirenlerin oranı %27,607'e yükselmiş, hiç kullanmadı nitelendiren oranı ise %4,601'e azalmıştır.

Düğüm 20 "OKUL_BITEN" değişkeni ile dallanmıştır. Eğitim durumu sorusuna verilen "Evet" cevabı ile düğüm 43, "Hayır" cevabı ile de düğüm 44 oluşmuştur. Düğüm 43'de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %72,000, 1 yıldan fazla olduğunu söyleyen fertlerin %14,667, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %13,333'dür. Düğüm 44'de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %53,754'e azalmış, 1 yıldan fazla nitelendirenlerin oranı %41,742'e yükselmiş, hiç kullanmadı nitelendiren oranı ise %4,505'e azalmıştır.

Eğitim kümesinde yer alan 8316 adet veri, test edilerek modelin doğruluk oranı belirlenmiştir. CHAID algoritmasının modeli doğru tahmin etme oranının %58,17 olduğu söylenebilir.

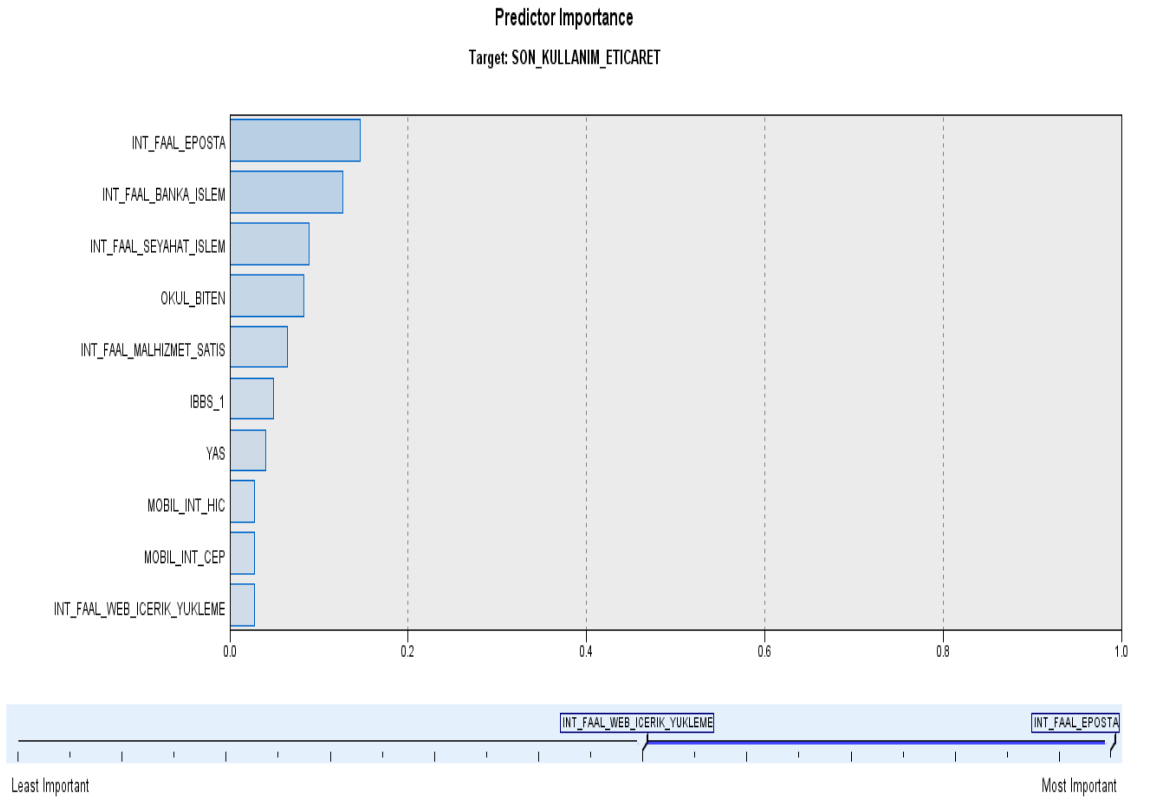
Tablo 4.3: CHAID Algoritmasının Performans Ölçüm Sonuçları

Sınıflandırma	Eğitim Kümesi		Test Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	4,965	%59,7	2,100	%58,17
Yanlış	3,351	%40,3	1,510	%41,83
Toplam	8,316	% 100	3,610	% 100

4.4.2. C&R Tree Algoritması

C&RT algoritmasının en önemli gördüğü değişken eposta olmuştur bunu banka işlem ve seyahat işlem değişkenleri takip etmiştir (Şekil 4.18). Şekil 4.19'da ise ağacın genel görünümü yer almaktadır.

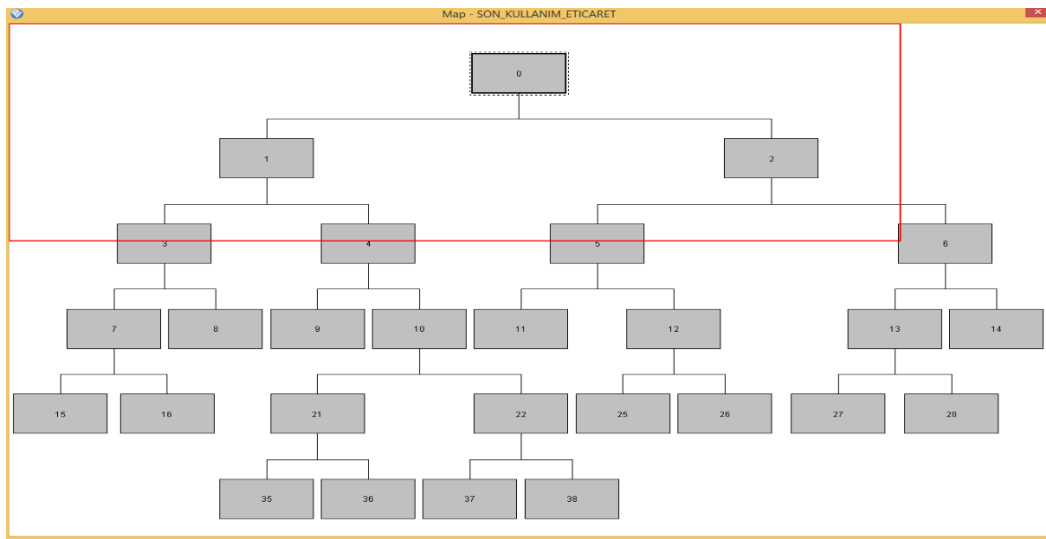
Şekil 4.18: C&RT Algoritmasına Göre Önemli Değişkenler



Tablo 4.4: C&R Tree Algoritmasında Kullanılan Değişkenlerin Önem Değerleri

Değişkenler	Değişken Önemi
Değişken_1 (int_faal_eposta)	0,15
Değişken_2 (int_faal_banka_islem)	0,13
Değişken_3 (int_faal_seyahat_islem)	0,09
Değişken_4 (okul_biten)	0,08
Değişken_5 (int_faal_malhizmet_satis)	0,06
Değişken_6 (ibbs_1)	0,05
Değişken_7 (yas)	0,04
Değişken_8 (mobil_int_tasinabilir)	0,03
Değişken_9 (mobil_int_cep)	0,03
Değişken_10 (int_faal_web_icerik_yukleme)	0,03

Şekil 4.19: C&RT Algoritması ile Oluşturulan Karar Ağacının Genel Görünümü



Şekil 4.20: C&R Tree Algoritması ile Oluşan Karar Ağacının Başlangıç Düzümü

SON_KULLANIM_ETICARET

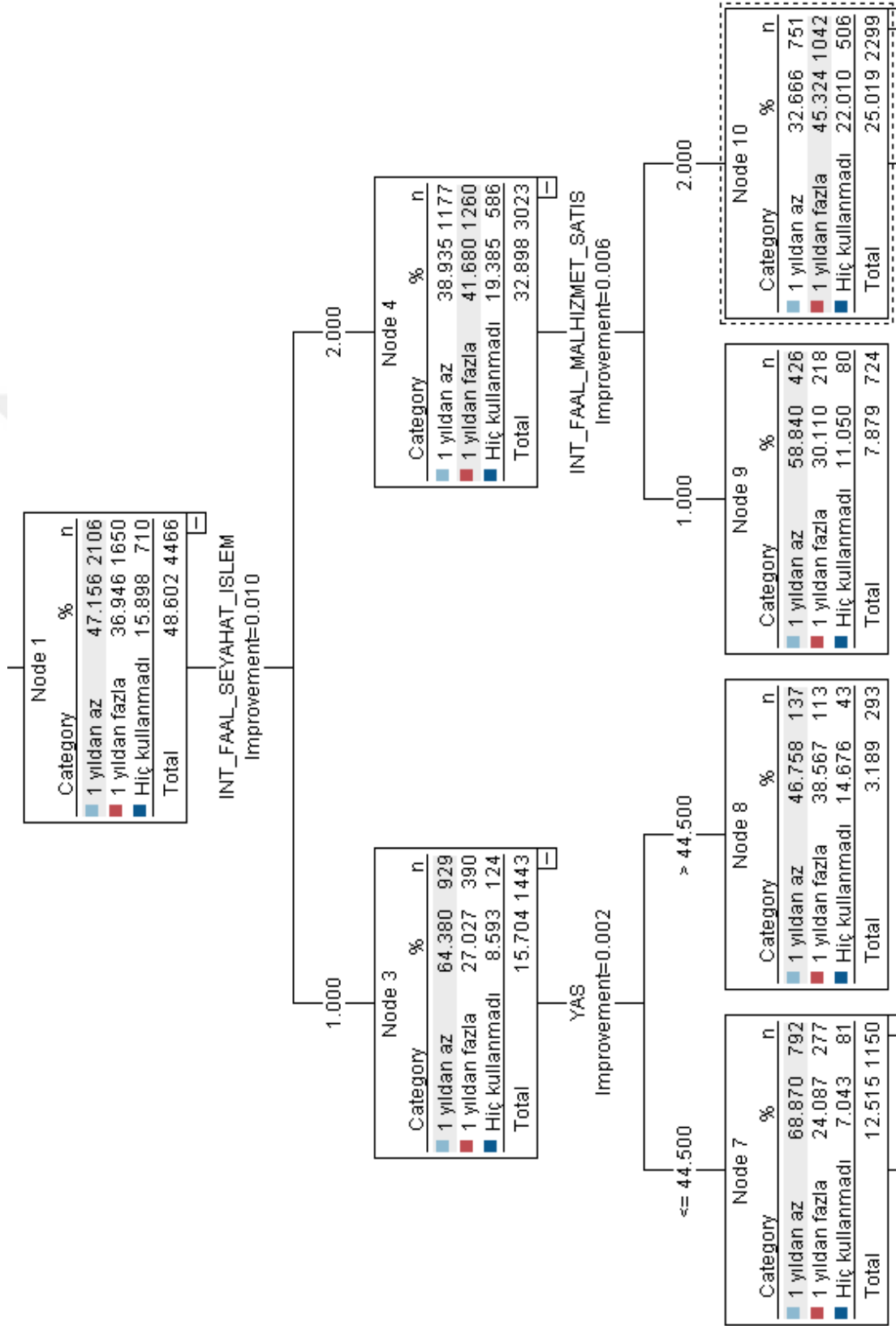
Node 0		
Category	%	n
1 yıldan az	33.638	3091
1 yıldan fazla	33.986	3123
Hiç kullanmadı	32.376	2975
Total	100.000	9189

INT_FAAL_BANKA_ISLEM
Improvement=0.044

Başlangıç düğümünde yer alan 9,189 ferдин %33,638’i son kullanım e-ticaret 1 yıldan az olarak nitelendirirken, %33,986’sı 1 yıldan fazla, %32,376’sı ise hiç kullanmadı olarak nitelendirmiştir.

C&R Tree algoritması ile oluşturulan karar ağacı dallanmaya “INT_FAAL_BANKA_ISLEM” sorusuna verilen “1 yıldan az, 1 yıldan çok, hiç kullanmadı” cevapları ile başlanmıştır. Buna göre 2 düzeydeki banka işlemlerine göre dallanma gerçekleşmiştir.

Şekil 4.21: C&R Tree Algoritmasının 1. Düzgüne Göre Dallanması



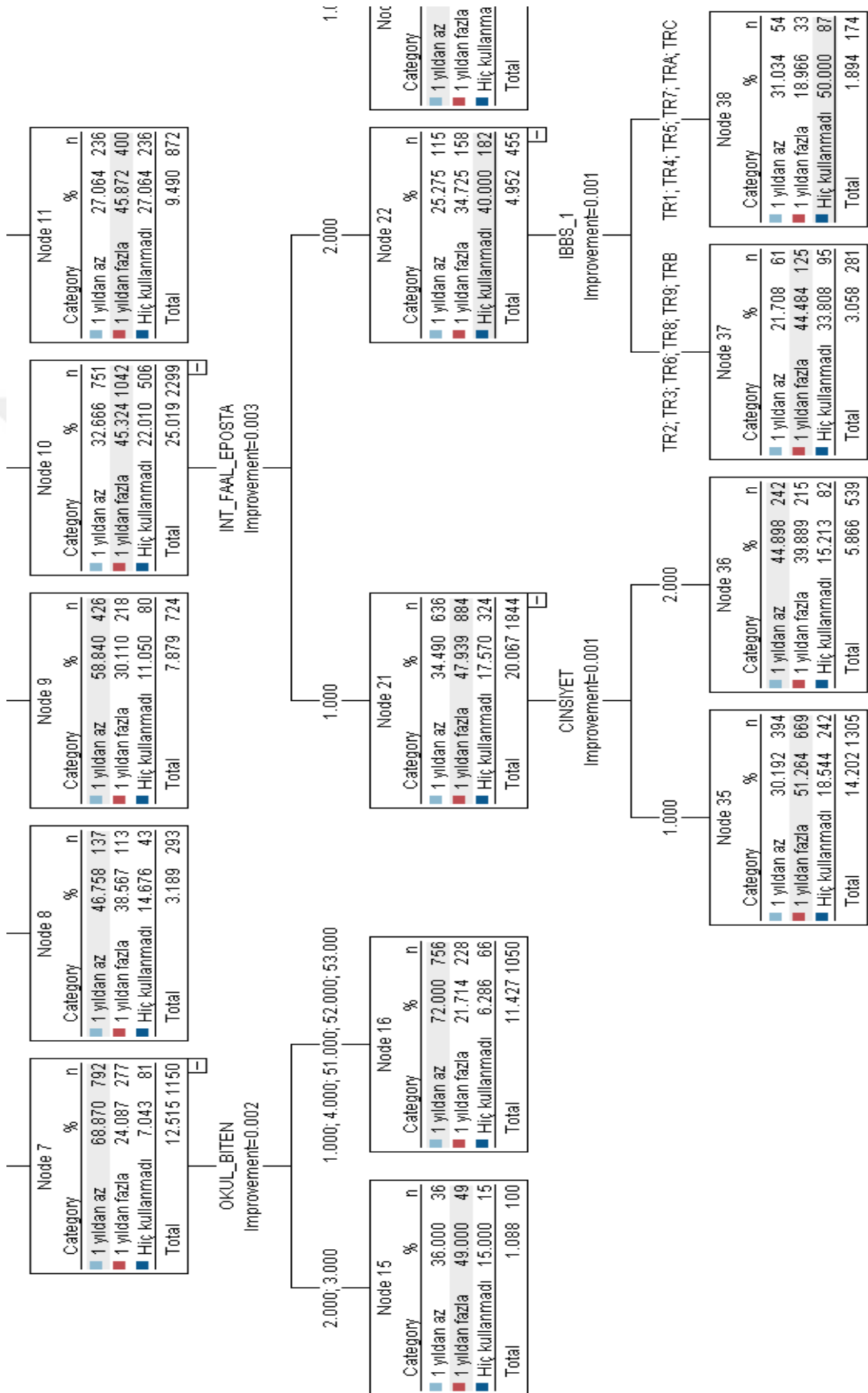
Karar ağacının 1. düğümü incelendiğinde e-ticaret kullanımı 1 yıldan az olarak nitelendiren fertlerin oranının %47,156, 1 yıldan fazla nitelendirenlerin oranının %36,946, hiç kullanmayan ise %15,898 olduğu gözlenmektedir.

Düğüm 1 “SEYAHAT_İŞLEM” değişkeni ile dallanmıştır. Seyahat ve konaklama ile ilgili hizmetleri kullanma (Otel rezervasyonu, bilet satış vb.) sorusuna verilen “Evet” cevabı ile düğüm 3, “Hayır” cevabı ile de düğüm 4 oluşmuştur. Düğüm 3’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %64,380, 1 yıldan fazla olduğunu söyleyen fertlerin %27,027, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise % 8,593’dür. Düğüm 4’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %38,935’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %41,680’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %19,385’e yükselmiştir.

Düğüm 3 “YAS” değişkeni ile dallanmıştır. Ferdin yaşı sorusuna verilen “45 yaşın altı” cevabı ile düğüm 7, “45 yaşın üzeri” cevabı ile de düğüm 8 oluşmuştur. Düğüm 7’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %68,870, 1 yıldan fazla olduğunu söyleyen fertlerin %24,087, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %7,043’dür. Düğüm 8’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %46,758’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %38,567’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %14,676’a yükselmiştir.

Düğüm 4 “MALHİZMET_SATIS” değişkeni ile dallanmıştır. Mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.) sorusuna verilen “Evet” cevabı ile düğüm 9, “Hayır” cevabı ile de düğüm 10 oluşmuştur. Düğüm 9’da e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %58,840, 1 yıldan fazla olduğunu söyleyen fertlerin %30,110, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise % 11,050’dir. Düğüm 10’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %32,666’a azalmış, 1 yıldan fazla nitelendirenlerin oranı %45,324’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %22,019’a yükselmiştir.

Şekil 4.22: C&R Tree Algoritmasının 7. ve 10. Düzüme Göre Dallanması



Düğüm 7 “OKUL_BITEN” değişkeni ile dallanmıştır. Eğitim durumu sorusuna verilen “Evet” cevabı ile düğüm 15, “Hayır” cevabı ile de düğüm 16 oluşmuştur. Düğüm 15’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %36,000, 1 yıldan fazla olduğunu söyleyen fertlerin %49,000, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %15,000’dir. Düğüm 16’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %72,000’a yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %21,714’e azalmış, hiç kullanmadı nitelendiren oranı ise %6,286’a azalmıştır.

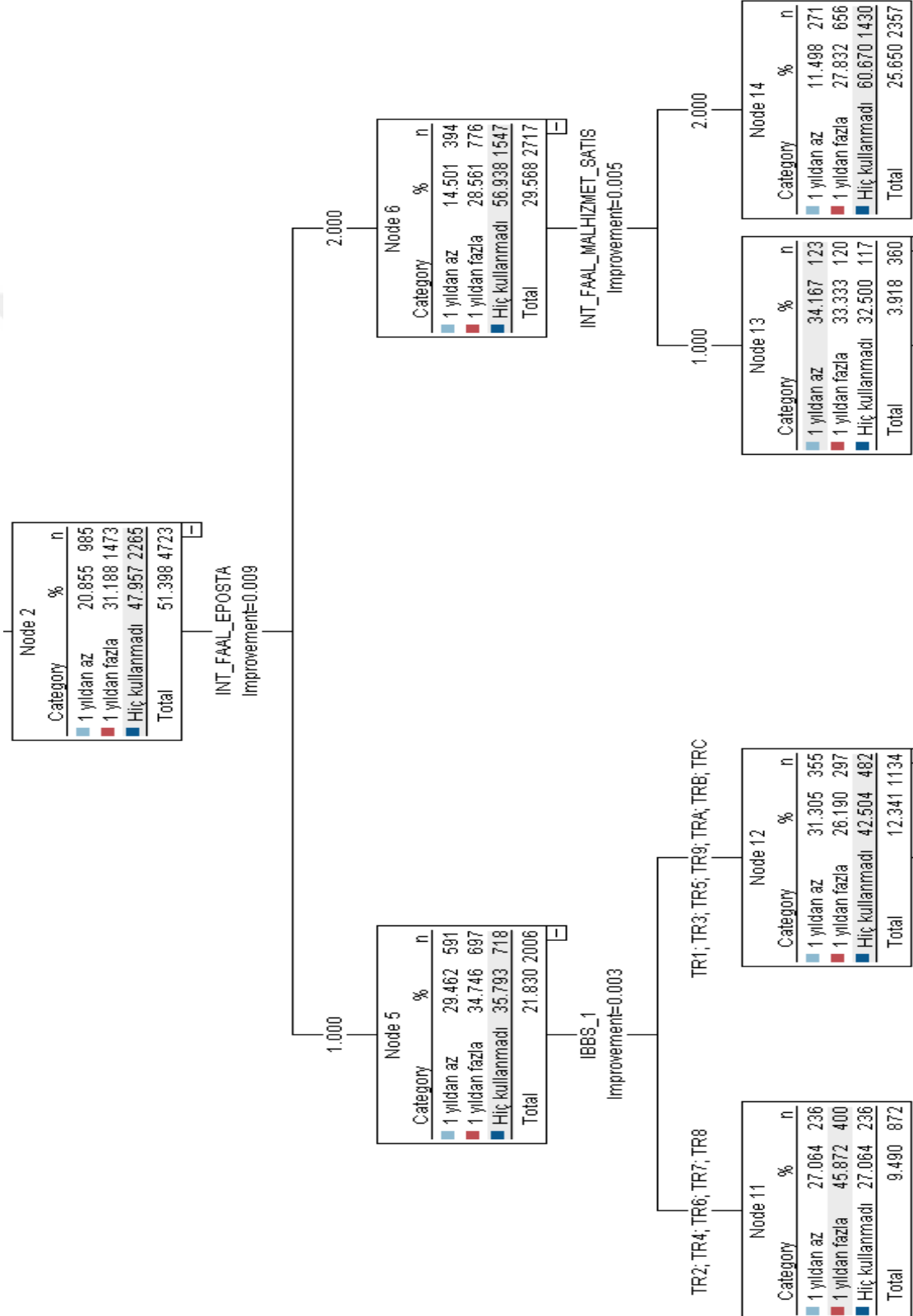
Düğüm 10 “FAAL_EPOSTA” değişkeni ile dallanmıştır. E-posta gönderme/alma sorusuna verilen “Evet” cevabı ile Düğüm 21, “Hayır” cevabı ile de düğüm 22 oluşmuştur. Düğüm 21’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %34,490, 1 yıldan fazla olduğunu söyleyen fertlerin %47,939, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %17,570’dir. Düğüm 22’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %25,275’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %34,725’e azalmış, hiç kullanmadı nitelendiren oranı ise %40,000’e yükselmiştir.

Düğüm 21 “CINSİYET” değişkeni ile dallanmıştır. Ferdin cinsiyeti sorusuna verilen “Evet” cevabı ile düğüm 35 , “Hayır” cevabı ile de düğüm 36 oluşmuştur. Düğüm 35’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %30,192, 1 yıldan fazla olduğunu söyleyen fertlerin %51,264, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %18,544’dür. Düğüm 36’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %44,898’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %39,889’a azalmış, hiç kullanmadı nitelendiren oranı ise %15,213’e azalmıştır.

Düğüm 22 “IBBS_1” değişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen “TR2; TR3; TR6; TR8; TR9; TRB” cevabı ile düğüm 37, “TR1; TR4; TR5; TR7; TRA; TRC” cevabı ile de düğüm 38 oluşmuştur. Düğüm 37’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %21,708, 1 yıldan fazla olduğunu söyleyen fertlerin %44,484, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %33,808’dir. Düğüm 38’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %31,034’e yükselmiş, 1

yıldan fazla nitelendirenlerin oranı %18,966'ya azalmış, hiç kullanmadı nitelendiren oranı ise %50,000'e yükselmiştir.

Şekil 4.23: C&R Tree Algoritmasının 2. Düzgüne Göre Dallanması



Karar ağacının 2.düğümü incelendiğinde e-ticaret kullanımı 1 yıldan az olarak nitelendiren fertlerin oranının %20,855, 1 yıldan fazla nitelendirenlerin oranının %31,188, hiç kullanmayan ise %47,957 olduğu gözlenmektedir.

Düğüm 2 “FAAL_EPOSTA” değişkeni ile dallanmıştır. E-posta gönderme/alma sorusuna verilen “Evet” cevabı ile düğüm 5, “Hayır” cevabı ile de düğüm 6 oluşmuştur. Düğüm 5’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %29,462, 1 yıldan fazla olduğunu söyleyen fertlerin %34,746, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %35,793’dür. Düğüm 6’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %14,501’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %28,561’e azalmış, hiç kullanmadı nitelendiren oranı ise %56,938’e yükselmiştir.

Düğüm 5 “IBBS_1” değişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen “TR2; TR4; TR6; TR7; TR8” cevabı ile düğüm 11, “TR1; TR3; TR5; TR9; TRA; TRB; TRC” cevabı ile de düğüm 12 oluşmuştur. Düğüm 11’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %27,064, 1 yıldan fazla olduğunu söyleyen fertlerin %45,872, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %27,064’dür. Düğüm 12’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %31,305’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %26,190’a azalmış, hiç kullanmadı nitelendiren oranı ise %42,504’e yükselmiştir.

Düğüm 12 “MALHIZMET_SATIS” değişkeni ile dallanmıştır. Mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.) sorusuna verilen “Evet” cevabı ile düğüm 25, “Hayır” cevabı ile de düğüm 26 oluşmuştur. Düğüm 25’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %56,000, 1 yıldan fazla olduğunu söyleyen fertlerin %16,000, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %28,000’dür. Düğüm 26’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %26,017’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %28,373’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %45,610’a yükselmiştir.

Düğüm 6 “MALHIZMET_SATIS” değişkeni ile dallanmıştır. Mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.) sorusuna verilen “Evet” cevabı ile düğüm 13, “Hayır” cevabı ile de düğüm 14 oluşmuştur. Düğüm 13’de e-ticaret

kullanımına 1 yıldan az söyleyen fertlerin oranı %34,167, 1 yıldan fazla olduğunu söyleyen fertlerin %33,333, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %32,500'dür. Düğüm 14'de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %11,498'e azalmış, 1 yıldan fazla nitelendirenlerin oranı %27,832'e azalmış, hiç kullanmadı nitelendiren oranı ise %60,670'e yükselmiştir.

Düğüm 14 dallanma meydana gelmemiştir fakat düğüm 13 "IBBS_1" değişkeni ile dallanmıştır. İstatistiki Bölge Birimleri Sınıflaması Düzey 1 (NUTS1) sorusuna verilen "TR1; TR2; TR3; TR6; TR7; TR9; TRA; TRB" cevabı ile düğüm 27, "TR4; TR5; TR8; TRC" cevabı ile de düğüm 28 oluşmuştur. Düğüm 27'de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %30,827, 1 yıldan fazla olduğunu söyleyen fertlerin %43,233, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %25,940'dır. Düğüm 28'de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %43,617'e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %5,319'a azalmış, hiç kullanmadı nitelendiren oranı ise %51,064'e yükselmiştir.

Eğitim kümesinde yer alan 8316 adet veri, test edilerek modelin doğruluk oranı belirlenmiştir. C&R Tree algoritmasının modeli doğru tahmin etme oranının % 60,19 olduğu söylenebilir.

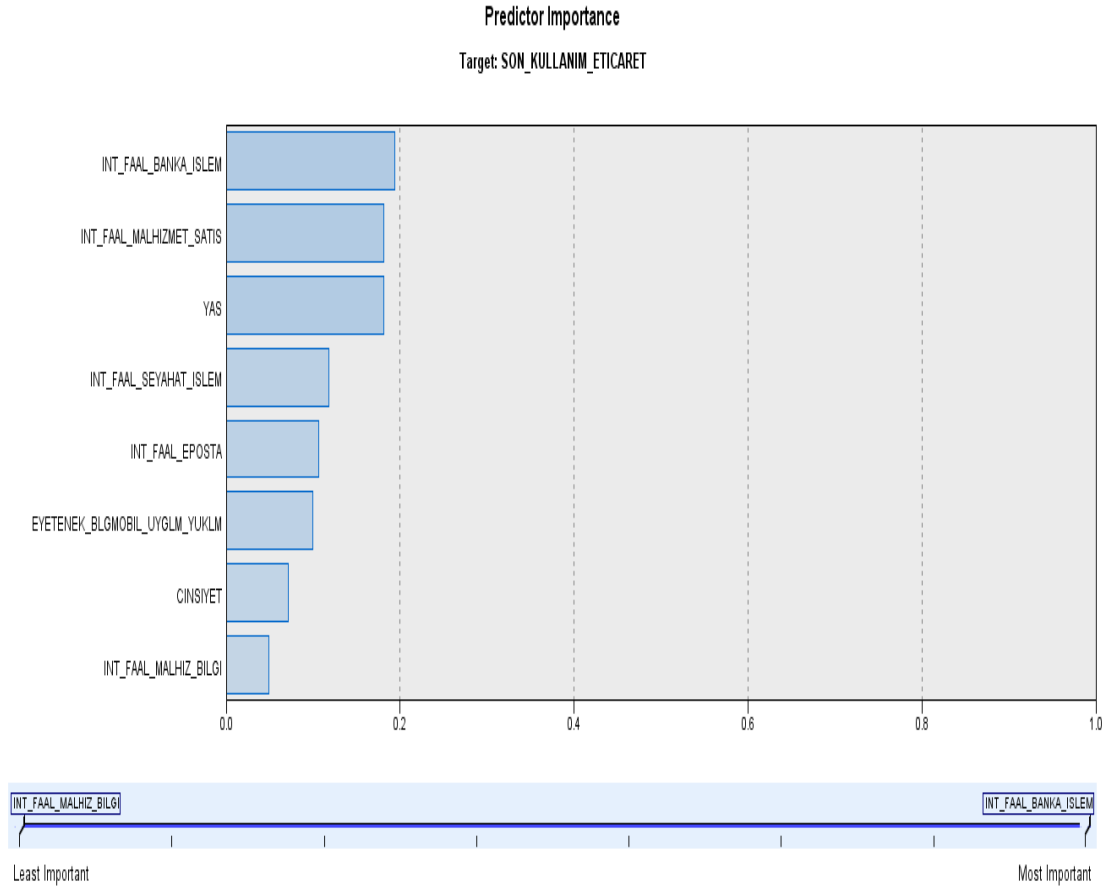
Tablo 4.5: C&R Tree Algoritmasının Performans Ölçüm Sonuçları

Sınıflandırma	Eğitim Kümesi		Test Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	4,979	%59,87	2,173	%60,19
Yanlış	3,337	%40,13	1,437	%39,81
Toplam	8,316	% 100	3,610	% 100

4.4.3. C5.0 Algoritması

C5.0 algoritmasının en önemli gördüğü değişken banka işlemi olmuştur bunu malhizmet satış ve yaş değişkenleri takip etmiştir (Şekil 4.24).

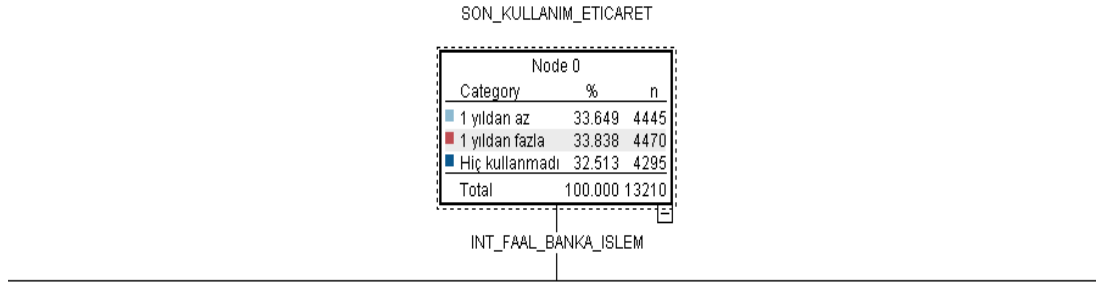
Şekil 4.24: C5.0 Algoritmasına Göre Önemli Değişkenler



Tablo 4.6: C5.0 Algoritmasında Kullanılan Değişkenlerin Önem Değerleri

Değişkenler	Değişken Önemi
Değişken _1(int_faal_banka_ıslem)	0,19
Değişken _2 (int_faal_malhızmet_satıs)	0,18
Değişken _3 (yas)	0,18
Değişken _4 (int_faal_seyahat_ıslem)	0,12
Değişken _5 (int_faal_eposta)	0,11
Değişken_6 (eyetenek_blgmobil_uyglm_yuklm)	0,10
Değişken _7 (cinsiyet)	0,07
Değişken _8 (int_faal_malhız_bılgı)	0,05

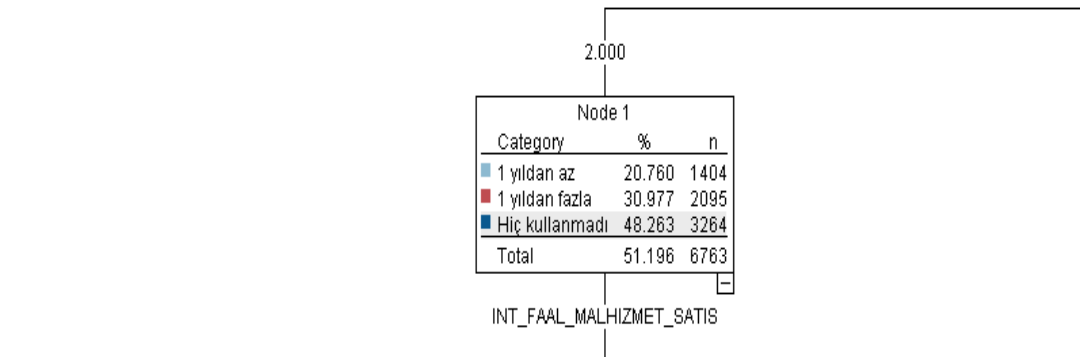
Şekil 4.25: C5.0 algoritması ile Oluşan Karar Ağacının Başlangıç Düğümü



Başlangıç düğümünde yer alan 13,210 ferdin %33,649’u son kullanım e-ticaret 1 yıldan az olarak nitelendirirken, %33,838’i 1 yıldan fazla, %32,513’ü ise hiç kullanmadı olarak nitelendirmiştir.

C5.0 algoritması ile oluşturulan karar ağacı dallanmaya “INT_FAAL_BANKA_ISLEM” sorusuna verilen “1 yıldan az, 1 yıldan çok, hiç kullanmadı” cevapları ile başlamıştır. Buna göre 2 düzeydeki banka işlemlerine göre dallanma gerçekleşmiştir.

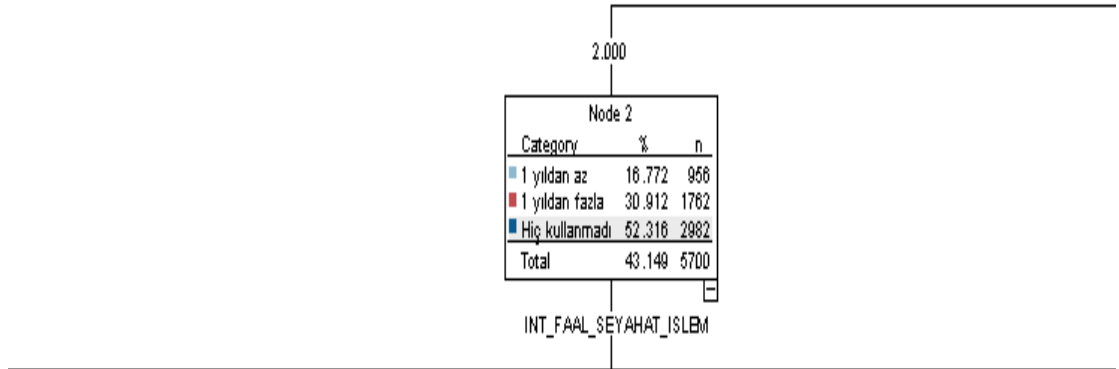
Şekil 4.26: C5.0 Algoritmasının 1. Düğüme Göre Dallanması



Karar ağacının 1. düğümü incelendiğinde e-ticaret kullanımı 1 yıldan az olarak nitelendiren fertlerin oranının %20,760, 1 yıldan fazla nitelendirenlerin oranının %30,977, hiç kullanmayan ise %48,263 olduğu gözlenmektedir.

Düğüm 1 “MALHIZMET_SATIS” değişkeni ile dallanmıştır. Mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.) sorusuna verilen “Evet” cevabı ile düğüm 2, “Hayır” cevabı ile de düğüm 393 oluşmuştur. Düğüm 2’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %16,772, 1 yıldan fazla olduğunu söyleyen fertlerin %30,912, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %52,316’dır. Düğüm 393’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %42,145’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %31,326’a yükselmiş, hiç kullanmadı nitelendiren oranı ise %26,539’a azalmıştır.

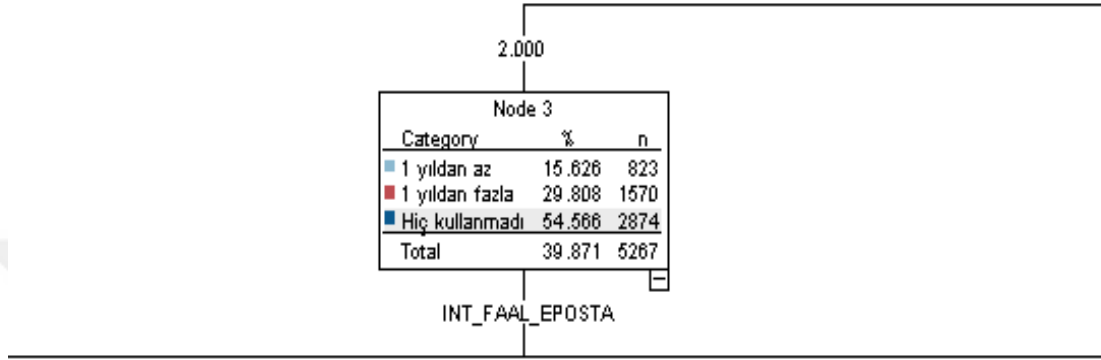
Şekil 4.27: C5.0 Algoritmasının 2. Düğüme Göre Dallanması



Düğüm 2 “SEYAHAT_IŞLEM” değişkeni ile dallanmıştır. Seyahat ve konaklama ile ilgili hizmetleri kullanma (Otel rezervasyonu, bilet satış vb.) sorusuna verilen “Evet” cevabı ile düğüm 3, “Hayır” cevabı ile de düğüm 330 oluşmuştur. Düğüm 3’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %15,626, 1 yıldan fazla olduğunu söyleyen fertlerin %29,808, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %54,566’dır. Düğüm 330’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %30,716’a yükselmiş, 1 yıldan fazla

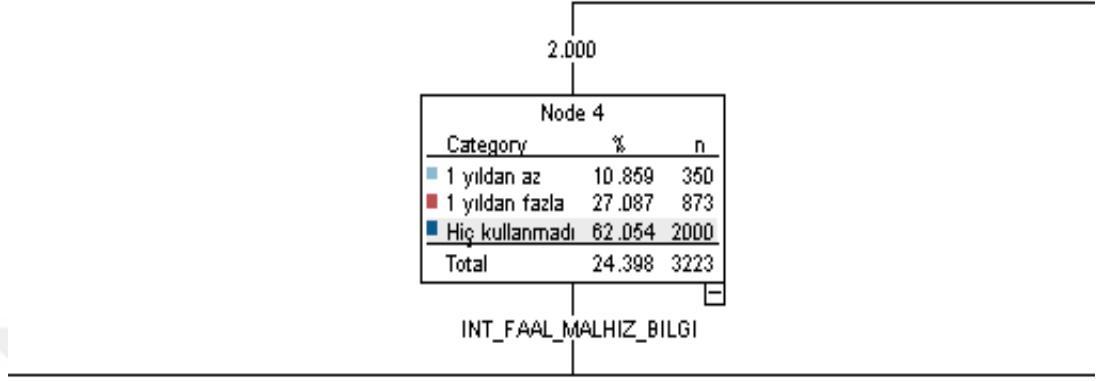
nitelendirenlerin oranı %44,342'e yükselmiş, hiç kullanmadı nitelendiren oranı ise %24,942'e azalmıştır.

Şekil 4.28: C5.0 Algoritmasının 3. Düzüme Göre Dallanması



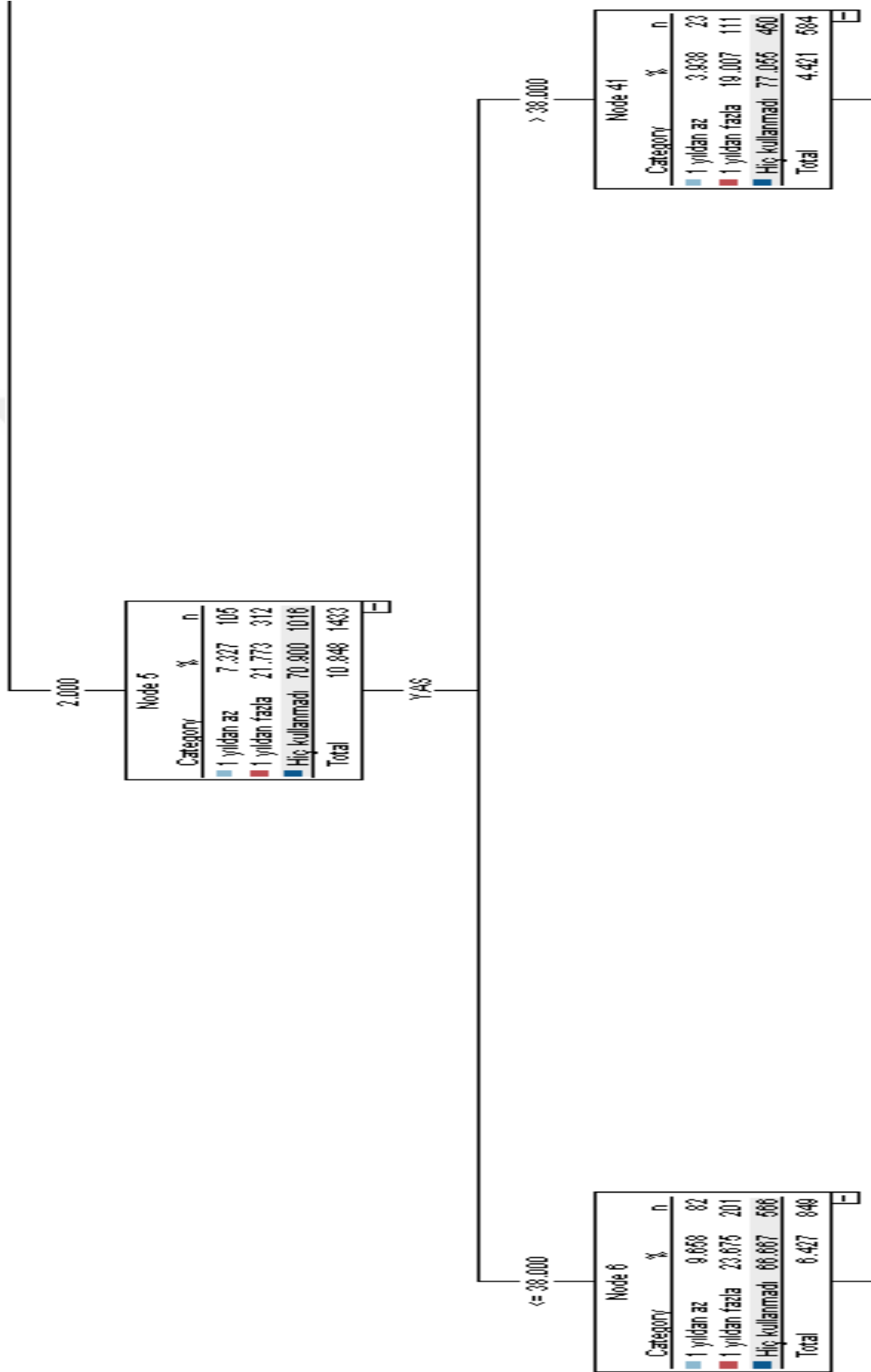
Düzüm 3 “FAAL_EPOSTA” deęişkeni ile dallanmıştır. E-posta gönderme/alma sorusuna verilen “Evet” cevabı ile düğüm 4, “Hayır” cevabı ile de düğüm 135 oluşmuştur. Düzüm 4’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %10,859, 1 yıldan fazla olduğunu söyleyen fertlerin %27,087, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %62,054’dür. Düzüm 135’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %23,141’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %34,100’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %42,759’a azalmıştır.

Şekil 4.29: C5.0 Algoritmasının 4. Düzümüne Göre Dallanması



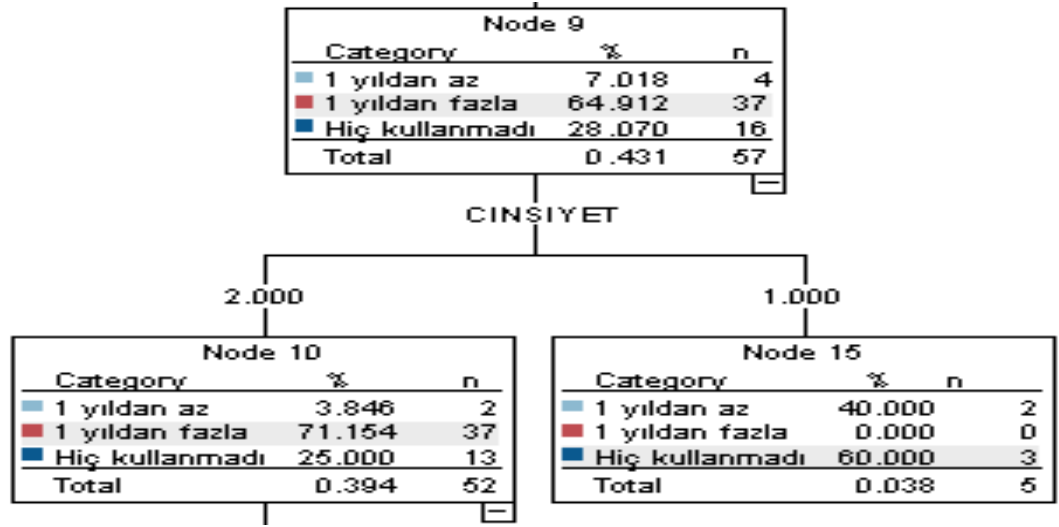
Düzüm 4 “FAAL_MALHIZ_BILGI” değişkeni ile dallanmıştır. Mal ve hizmetler hakkında bilgi arama sorusuna verilen “Evet” cevabı ile düğüm 5, “Hayır” cevabı ile de düğüm 60 oluşmuştur. Düzüm 5’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %7,327, 1 yıldan fazla olduğunu söyleyen fertlerin %21,773, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %70,900’dür. Düzüm 60’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %13,687’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %31,341’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %54,972’e azalmıştır.

Şekil 4.30: C5.0 Algoritmasının 5. Düzüme Göre Dallanması



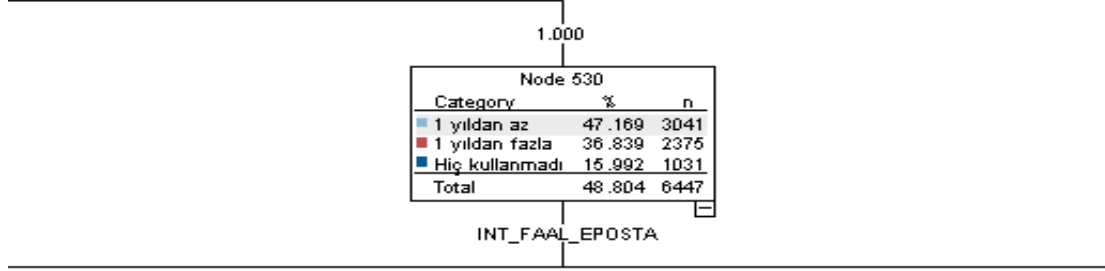
Düğüm 5 “YAS” değişkeni ile dallanmıştır. Ferdin yaşı sorusuna verilen “38 yaşdan az” cevabı ile düğüm 6, “38 yaş üzeri” cevabı ile de düğüm 41 oluşmuştur. Düğüm 6’da e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %9,658, 1 yıldan fazla olduğunu söyleyen fertlerin %23,675, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %66,667’dir. Düğüm 41’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %3,938’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %19,007’e azalmış, hiç kullanmadı nitelendiren oranı ise %77,055’e yükselmiştir.

Şekil 4.31: C5.0 Algoritmasının 9. Düğüme Göre Dallanması



Düğüm 9 “CINSIYET ” değişkeni ile dallanmıştır. Ferdin cinsiyeti sorusuna verilen “Evet” cevabı ile düğüm 10 , “Hayır” cevabı ile de düğüm 15 oluşmuştur. Düğüm 10’da e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %3,846, 1 yıldan fazla olduğunu söyleyen fertlerin %71,154, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %25,000’dir. Düğüm 15’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %40,000’a yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %0,000’a azalmış, hiç kullanmadı nitelendiren oranı ise %60,000’a yükselmiştir.

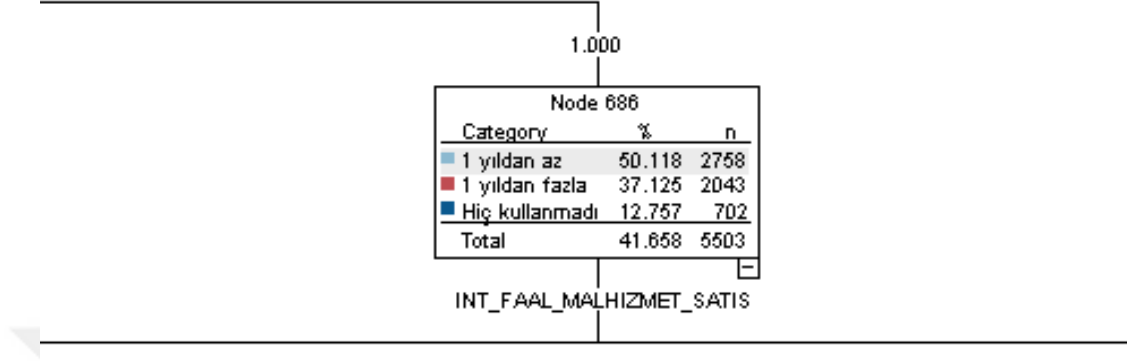
Şekil 4.32: C5.0 Algoritmasının 530. Düzüme Göre Dallanması



Karar ağacının 530. düğümü incelendiğinde eticaret kullanımı 1 yıldan az olarak nitelendiren fertlerin oranının %47,169, 1 yıldan fazla nitelendirenlerin oranının %36,839, hiç kullanmayan ise %15,992 olduğu gözlenmektedir.

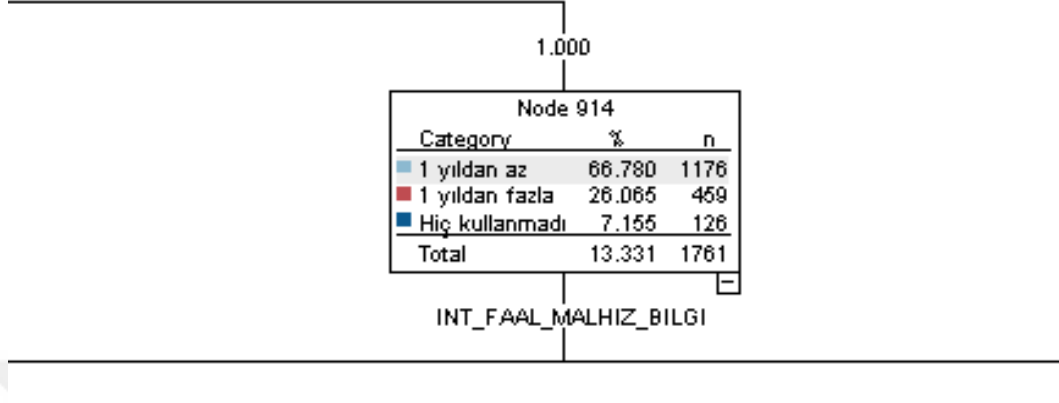
Düğüm 530 “FAAL_EPOSTA” değişkeni ile dallanmıştır. E-posta gönderme/alma sorusuna verilen “Evet” cevabı ile düğüm 531, “Hayır” cevabı ile de düğüm 686 oluşmuştur. Düğüm 531’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %29,979, 1 yıldan fazla olduğunu söyleyen fertlerin %35,169, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %34,852’dir. Düğüm 686’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %50,118’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %37,125’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %12,757’e azalmıştır.

Şekil 4.33: C5.0 Algoritmasının 686. Düzümüne Göre Dallanması



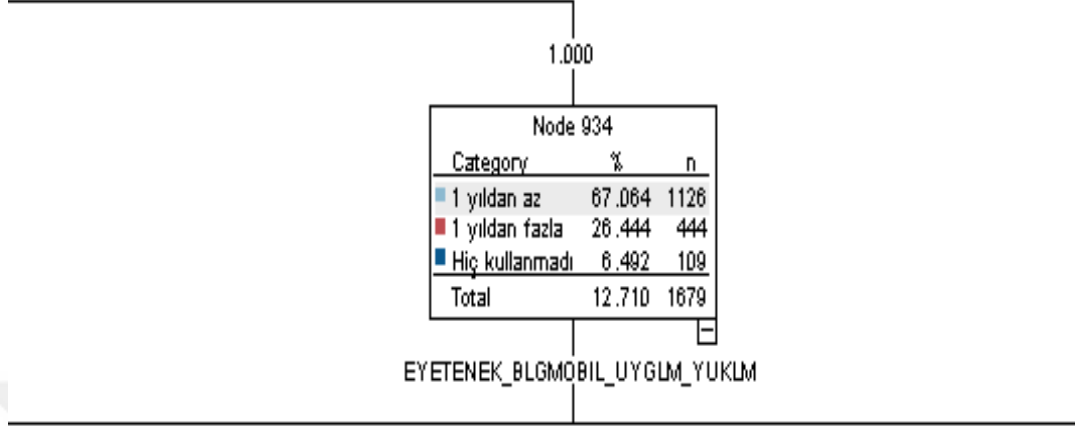
Düzüm 686 “MALHIZMET_SATIS” değişkeni ile dallanmıştır. Mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.) sorusuna verilen “Evet” cevabı ile düğüm 687, “Hayır” cevabı ile de düğüm 914 oluşmuştur. Düzüm 687’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %42,277, 1 yıldan fazla olduğunu söyleyen fertlerin %42,330, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %15,393’dür. Düzüm 914’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %66,780’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %26,065’e azalmış, hiç kullanmadı nitelendiren oranı ise %7,155’e azalmıştır.

Şekil 4.34: C5.0 Algoritmasının 914. Düzümüne Göre Dallanması



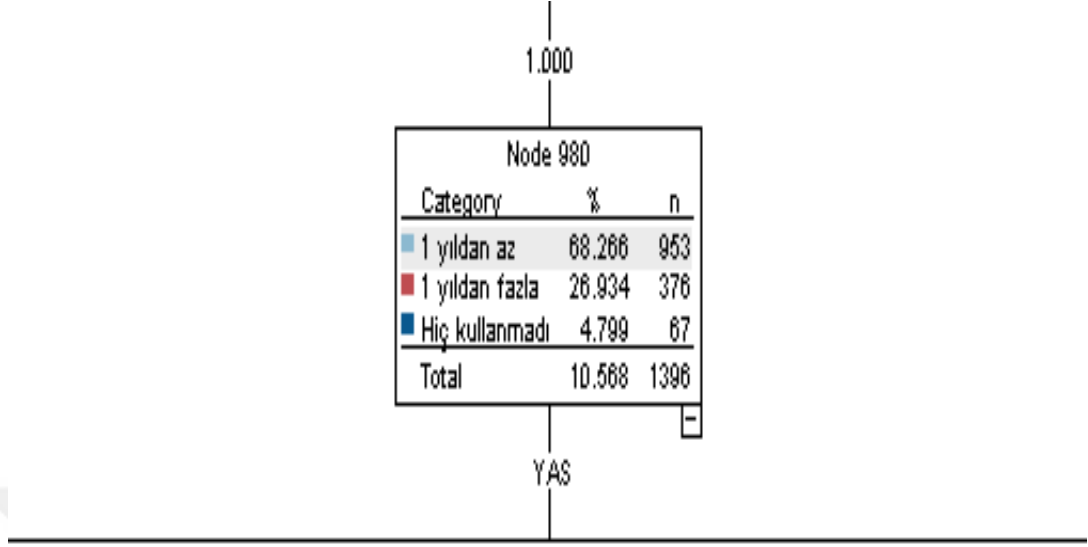
Düzüm 914 “FAAL_MALHIZ_BILGI” deęişkeni ile dallanmıştır. Mal ve hizmetler hakkında bilgi arama sorusuna verilen “Evet” cevabı ile düğüm 915, “Hayır” cevabı ile de düğüm 934 oluşmuştur. Düzüm 915’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %60,976, 1 yıldan fazla olduğunu söyleyen fertlerin %18,293, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %20,732’dir. Düzüm 934’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %67,064’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %26,444’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %6,492’e azalmıştır.

Şekil 4.35: C5.0 Algoritmasının 934. Düzgüne G6re Dalkanması



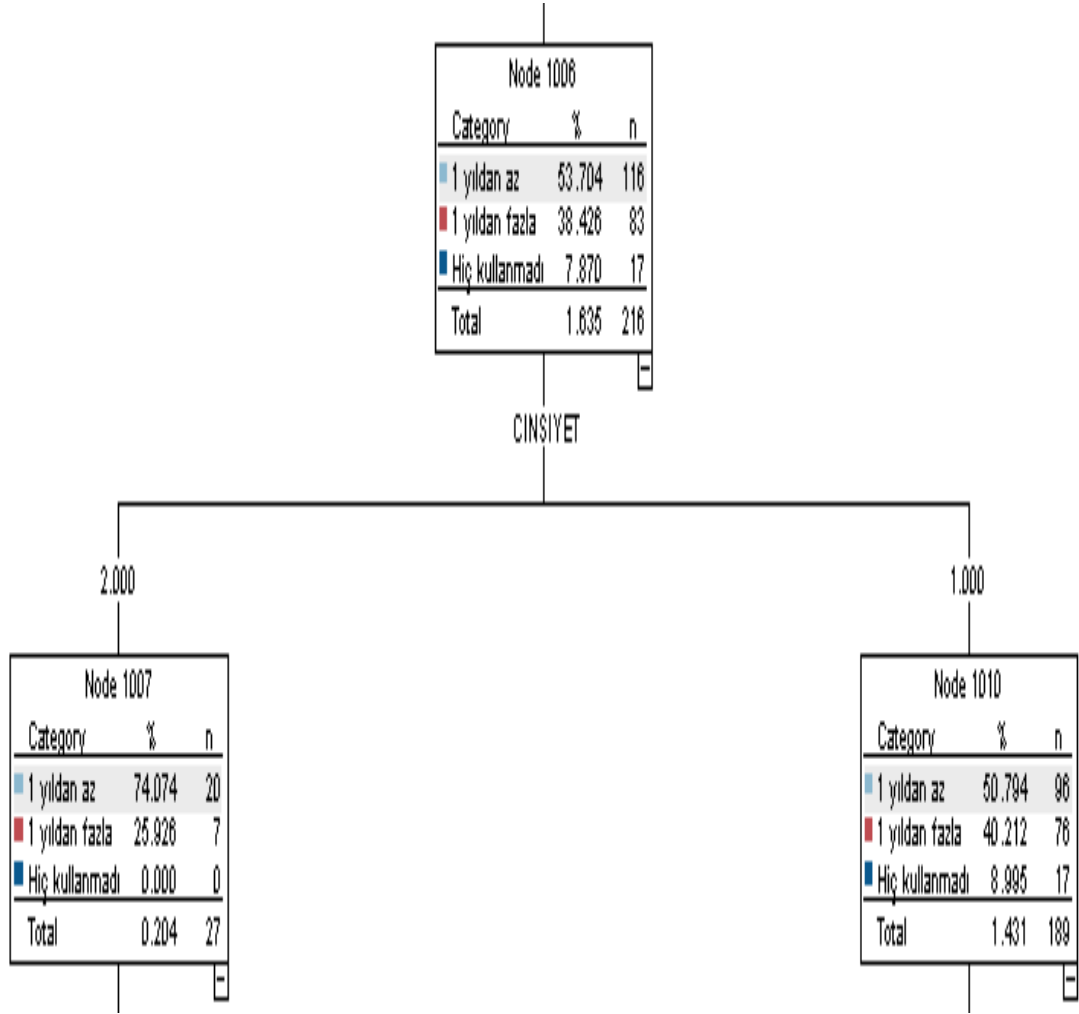
Düzgüm 934 “EYETENEK_BLGMOBIL_UYGLM_YUKLM” deęişkeni ile dallanmıştır. Yazılım veya mobil uygulama (application) yükleme sorusuna verilen “Evet” cevabı ile düzgüm 935, “Hayır” cevabı ile de düzgüm 980 oluşmuştur. Düzgüm 935’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %61,131, 1 yıldan fazla olduğunu söyleyen fertlerin %24,028, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %14,841’dir. Düzgüm 980’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %68,266’a yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %26,934’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %4,799’a azalmıştır.

Şekil 4.36: C5.0 Algoritmasının 980. Düzümüne Göre Dallanması



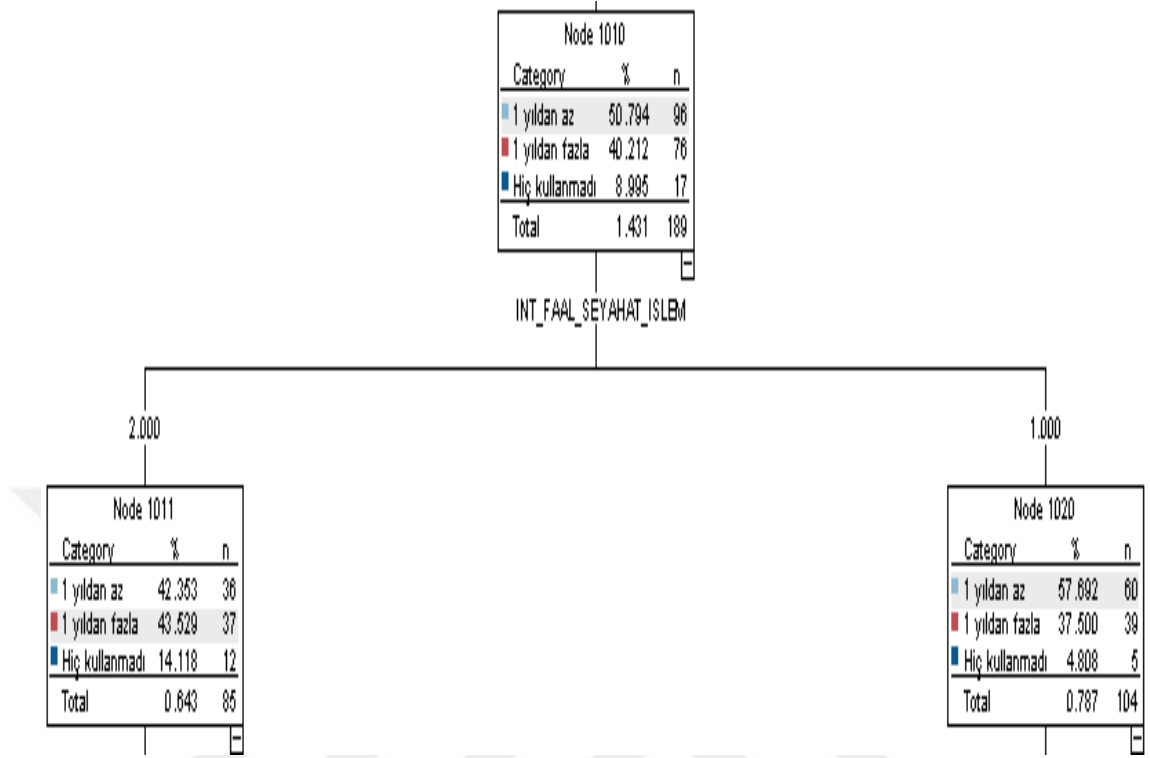
Düzüm 980 “YAS” değişkeni ile dallanmıştır. Ferdin yaşı sorusuna verilen “44 yaşdan az” cevabı ile düğüm 981, “44 yaş üzeri” cevabı ile de düğüm 1006 oluşmuştur. Düzüm 981’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %70,932, 1 yıldan fazla olduğunu söyleyen fertlerin %24,831, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %4,237’dir. Düzüm 1006’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %53,704’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %38,426’a yükselmiş, hiç kullanmadı nitelendiren oranı ise %7,870’e yükselmiştir.

Şekil 4.37: C5.0 Algoritmasının 1006. Düzüme Göre Dallanması



Düzüm 1006 “CINSIYET” deęişkeni ile dallanmıştır. Ferdin cinsiyeti sorusuna verilen “Evet” cevabı ile düğüm 1007, “Hayır” cevabı ile de düğüm 1010 oluşmuştur. Düzüm 1007’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %74,074, 1 yıldan fazla olduğunu söyleyen fertlerin %25,926, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %0,000’dır. Düzüm 1010’da e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %50,794’e azalmış, 1 yıldan fazla nitelendirenlerin oranı %40,212’e yükselmiş, hiç kullanmadı nitelendiren oranı ise %8,995’e yükselmiştir.

Şekil 4.38: C5.0 Algoritmasının 1010. Düzgüme Göre Dallanması



Düzgüm 1010 “SEYAHAT_İŞLEM” deęişkeni ile dallanmıştır. Seyahat ve konaklama ile ilgili hizmetleri kullanma (Otel rezervasyonu, bilet satış vb.) sorusuna verilen “Evet” cevabı ile düęüm 1011, “Hayır” cevabı ile de düęüm 1020 oluşmuştur. Düzgüm 1011’de e-ticaret kullanımına 1 yıldan az söyleyen fertlerin oranı %42,353, 1 yıldan fazla olduğunu söyleyen fertlerin %43,529, hiç kullanmadı olduğunu söyleyen fertlerin oranı ise %14,118’dir. Düzgüm 1020’de e-ticaret kullanımına 1 yıldan az olarak nitelendiren fertlerin oranı %57,692’e yükselmiş, 1 yıldan fazla nitelendirenlerin oranı %37,500’e azalmış, hiç kullanmadı nitelendiren oranı ise %4,808’e azalmıştır.

Eđitim kümesinde yer alan 8316 adet veri, test edilerek modelin doęruluk oranı belirlenmiştir. C5.0 algoritmasının modeli doęru tahmin etme oranının % 56,81 olduğu söylenebilir.

Tablo 4.7: C5.0 Algoritmasının Performans Ölçüm Sonuçları

Sınıflandırma	Eğitim Kümesi		Test Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	5,539	%66,61	2,051	%56,81
Yanlış	2,777	%33,39	1,559	%43,19
Toplam	8,316	% 100	3,610	% 100

4.5. RBF Kernel Fonksiyonu Kullanılarak Destek Vektör Makineleri Yardımı İle Değişkenlerin İncelenmesi

Hanehalkı bilişim teknolojileri kullanımı araştırmasında e-ticaret kullanımı incelendiğinde DVM modelinin sonuçları detayları ile aşağıda verilmiştir. Bağımlı değişkenimiz e-ticaret kullanımı (başarılı/başarısız) ve bağımsız değişkenler ise e-ticaret kullanımını etkileyen kalan diğer değişkenlerdir. İlk önce DVM modelimizin radyal tabanlı (RBF) kernel fonksiyonu kullanılarak ilgili sonuçlar elde edilmiştir. Clementine V.18 üzerinde DVM (Support Vector Machine) algoritmasında RBF fonksiyonu seçilmiştir.

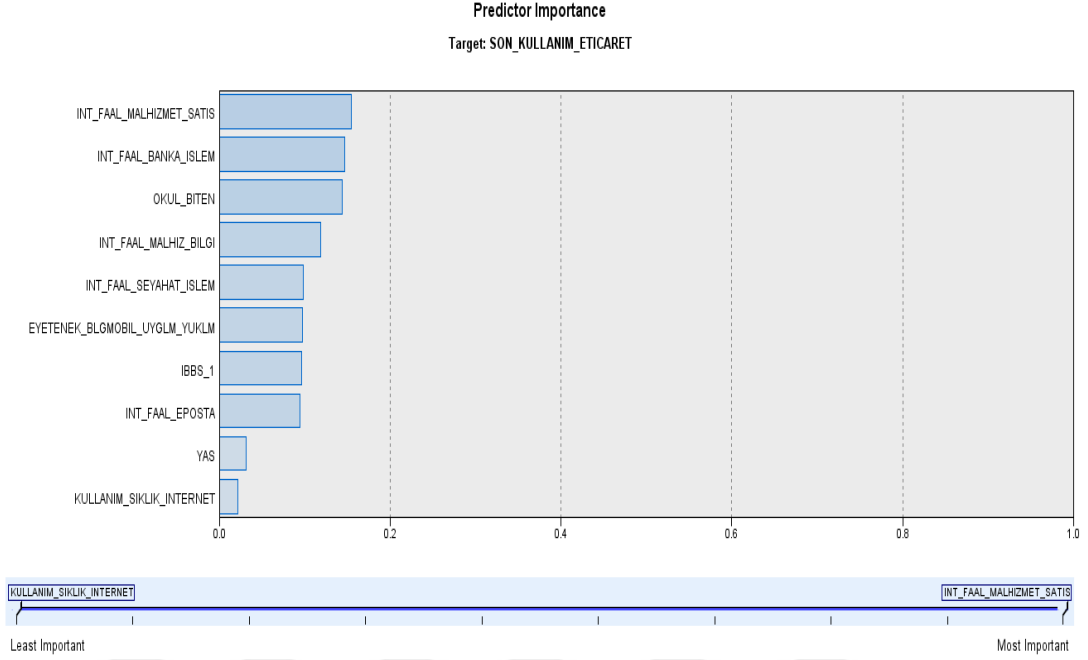
Hanehalkı bilişim teknolojileri kullanımı araştırmasında e-ticaret kullanımı incelendiğinde elde edilen bağımsız değişkenlerin göreceli etkileri kolaylıkla görülmektedir. En yüksek etkiye sahip ilk 10 değişkenin etki değerleri incelendiğinde, değişken 1 (mal veya hizmet satışı) %0,15 ile en yüksek etkiye sahip iken, değişken 2 (internet bankacılığı) %0,15 ile oldukça anlamlı bir etkiye sahip olduğu görülmektedir. Değişken 3 (eğitim durumu) ve değişken 4(mal ve hizmetler hakkında bilgi arama) sırası ile %0,14 ve %0,12 ile sonraki en önemli göreceli etkiye sahip değişkenlerdendir. Değişken 5 (Seyahat ve konaklama ile ilgili hizmetleri kullanma (otel rezervasyonu, bilet satış vb.)) %0,10 etkiye sahiptir. Değişken 6 (yazılım veya mobil uygulama (application) yükleme) ve değişken 7 (istatistiki bölge birimleri sınıflaması) sıra ile %0,10 ve %0,10 önemli göreceli etkiye sahip değişkenlerdendir. Değişken 8 (e-posta gönderme/alma) %0,09 anlamlı bir etkiye sahip olduğu görülmektedir. Değişken 9 (yaş) %0,03 etkiye sahiptir. Değişken 10

(son üç ay internet kullananların internet kullanım sıklığı) ise en az göreceli %0,02 etkiye sahiptir.

Tablo 4.8: RBF(Radyal Tabanlı)-DVM Değişkenlerin Göreceli Önem Oranları

Değişkenler	Değişken Önemi
Değişken_1(mal veya hizmet satışı)	0,15
Değişken_2 (internet bankacılığı)	0,15
Değişken_3 (eğitim durumu)	0,14
Değişken_4 (mal ve hizmetler hakkında bilgi arama)	0,12
Değişken_5 (Seyahat ve konaklama ile ilgili hizmetleri kullanma (Otel rezervasyonu, bilet satış vb.))	0,10
Değişken_6 (yazılım veya mobil uygulama (application) yükleme)	0,10
Değişken_7 (istatistiki bölge birimleri sınıflaması)	0,10
Değişken_8 (e-posta gönderme/alma)	0,09
Değişken_9 (yaş)fa	0,03
Değişken_10 (son üç ay internet kullananların internet kullanım sıklığı)	0,02

Şekil 4.39: RBF(Radyal Tabanlı)-DVM Değişkenlerin Göreceli Önem Oranları



RBF Doğruluk Oranı: DVM modeli RBF kernel fonksiyonu kullanılarak eğitim veri kümesinde eğitildikten sonra test veri kümesinde test edilerek %59,83 doğruluk oranına sahiptir.

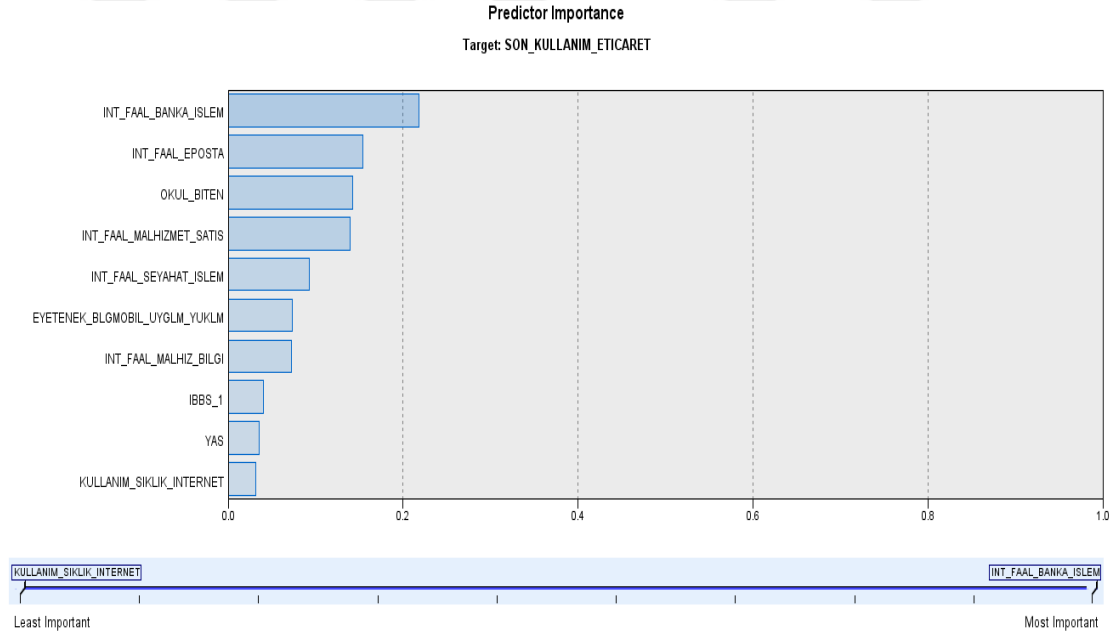
Tablo 4.9: RBF(Radyal Tabanlı)-DVM Değişkenlerin Performans Ölçüm Sonuçları

Sınıflandırma	Eğitim Kümesi		Test Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	5,571	%66,99	2,160	%59,83
Yanlış	2,745	%33,01	1,450	%40,17
Toplam	8,316	% 100	3,610	% 100

4.6. Doğrusal Kernel Fonksiyonu Kullanılarak Destek Vektör Makineleri Yardımı İle Değişkenlerin İncelenmesi

Doğrusal kernel fonksiyonu kullanılarak elde edilen DVM modelinde göreceli önem oranı en yüksek olan ve en etkili değişken %0,22 oranı ile İnternet bankacılığı değişkenidir. RBF DVM modelindeki gibi ilk iki en etkin değişkenler internet bankacılığı ile e-posta gönderme/alma %0,22 ve %0,15 oldukça anlamlı bir etkiye sahip olduğu görülmektedir. Yine aynı şekilde, sırasıyla eğitim durumu %0,14, mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.) %0,14, seyahat ve konaklama ile ilgili hizmetleri kullanma (Otel rezervasyonu, bilet satış vb.) %0,09, yazılım veya mobil uygulama (application) yükleme %0,07, mal ve hizmetler hakkında bilgi arama %0,07, istatistiki bölge birimleri sınıflaması %0,04, yaş %0,04, son üç ay internet kullananların internet kullanım sıklığı %0,03 göreceli önem oranları ile üçüncü, dördüncü,beşinci,altıncı, yedinci,sekiz ve dokuz önemli değişkenler olduğu doğrusal DVM modeli yardımı ile elde edilmiştir.

Şekil 4.40: Doğrusal-DVM Değişkenlerin Göreceli Önem Oranları



Tablo 4.10: Doğrusal-DVM Değişkenlerin Göreceli Önem Oranları

Değişkenler	Değişken Önemi
Değişken _1 (internet bankacılığı)	0,22
Değişken _2 (e-posta gönderme/alma)	0,15
Değişken _3 (eğitim durumu)	0,14
Değişken _4 (mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.))	0,14
Değişken _5 (seyahat ve konaklama ile ilgili hizmetleri kullanma (Otel rezervasyonu, bilet satış vb.))	0,09
Değişken _6 (yazılım veya mobil uygulama (application) yükleme)	0,07
Değişken _7 (mal ve hizmetler hakkında bilgi arama)	0,07
Değişken _8 (istatistiki bölge birimleri sınıflaması)	0,04
Değişken _9 (yaş)	0,04
Değişken _10 (son üç ay internet kullananların internet kullanım sıklığı)	0,03

Doğrusal Doğruluk Oranı: DVM modeli Doğrusal kernel fonksiyonu kullanılarak eğitim veri kümesinde eğitildikten sonra test veri kümesinde test edilerek %61,63 doğruluk oranına sahiptir.

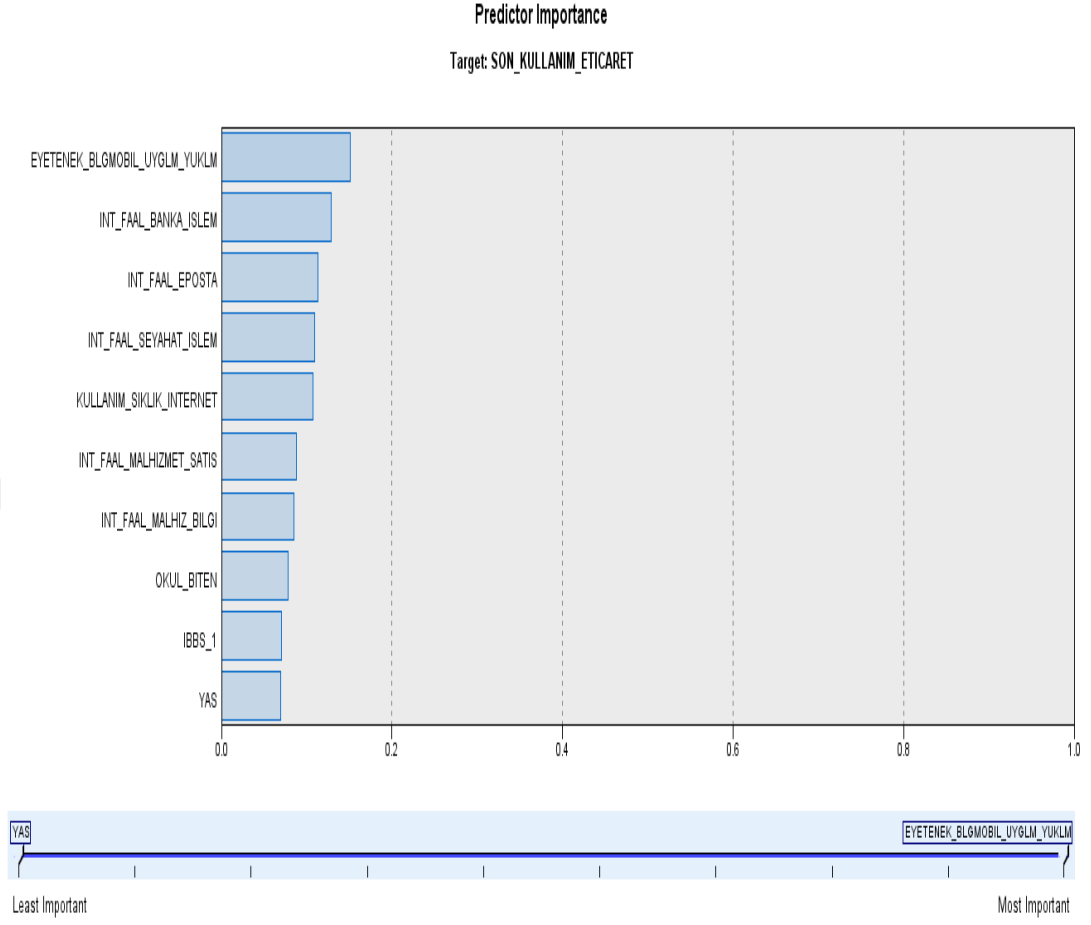
Tablo 4.11: Doğrusal-DVM Değişkenlerin Performans Ölçüm Sonuçları

Sınıflandırma	Eğitim Kümesi		Test Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	5,124	%61,62	2,225	%61,63
Yanlış	3,192	%38,38	1,385	%38,37
Toplam	8,316	% 100	3,610	% 100

4.7. Sigmoid Kernel Fonksiyonu Kullanarak DVM Modelinin İncelenmesi

Doğruluk oranı düşük olan DVM algoritmasıdır. Diğer algoritmalara göre farklı değişken önem sıraları vermiştir. Değişken 1 (yazılım veya mobil uygulama (application) yükleme) %0,15 ve değişken 2 (internet bankacılığı) %0,13 en etken iki değişken sigmoid DVM modeli tarafından belirlenmiştir. Diğer taraftan, %0,11 önem oranı ile değişken 3 (e-posta gönderme/alma), %0,11 önem oranı ile değişken 4 (seyahat ve konaklama ile ilgili hizmetleri kullanma (Otel rezervasyonu, bilet satış vb.)), %0,11 önem oranı ile değişken 5 (son üç ay internet kullananların internet kullanım sıklığı), %0,09 önem oranı ile değişken 6 (mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.)), %0,09 önem oranı ile değişken 7 (mal ve hizmetler hakkında bilgi arama), %0,08 önem oranı ile değişken 8 (eğitim durumu), %0,07 önem oranı ile değişken 9 (istatistiki bölge birimleri sınıflaması) ve son olarak ta %0,07 ile değişken 10 (yaş) göreceli olarak en az etken değişken olarak model tarafından belirtilmiştir.

Şekil 4.41: Sigmoid -DVM Değişkenlerin Göreceli Önem Oranları



Tablo 4.12: Sigmoid DVM Değişkenlerin Göreceli Önem Oranları

Değişkenler	Değişken Önemi
Değişken _1 (yazılım veya mobil uygulama (application) yükleme)	0,15
Değişken _2 (internet bankacılığı)	0,13
Değişken _3 (e-posta gönderme/alma)	0,11
Değişken _4 (seyahat ve konaklama ile ilgili hizmetleri kullanma (Otel rezervasyonu, bilet satış vb.))	0,11
Değişken _5 (son üç ay internet kullananların internet kullanım sıklığı)	0,11
Değişken _6 (mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.))	0,09
Değişken _7 (mal ve hizmetler hakkında bilgi arama)	0,09
Değişken _8 (eğitim durumu)	0,08
Değişken _9 (istatistiki bölge birimleri sınıflaması)	0,07
Değişken _10 (yaş)	0,07

Sigmoid Doğruluk Oranı: DVM modeli Sigmoid kernel fonksiyonu kullanılarak eğitim veri kümesinde eğitildikten sonra test veri kümesinde test edilerek %60,8 doğruluk oranına sahiptir.

Tablo 4.13: Sigmoid DVM Değişkenlerin Performans Ölçüm Sonuçları

Sınıflandırma	Eğitim Kümesi		Test Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	4,987	%59,97	2,195	%60,8
Yanlış	3,329	%40,03	1,415	%39,2
Toplam	8,316	% 100	3,610	% 100

5. SONUÇ

Bu tez çalışmasında amaç, TÜİK tarafından yapılan hanehalkı bilişim teknolojileri kullanım araştırmasına ait B grubu mikro veri seti kullanılarak e-ticaret kullanımı üzerinde etkili olabilecek faktörlerin karar ağacı algoritmaları ve destek vektör makineleri (DVM) ile belirlenmesidir.

TÜİK'ten temin edilen B grubu mikro veri seti; veri temizleme ve veri dönüştürme 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 e-ticaret kullanımı değişkeni, bağımlı değişken olarak belirlenmiştir. Hanehalkı bilişim teknolojileri kullanım araştırmasının kapsamı, Türkiye Cumhuriyeti sınırları içinde ikamet eden hanelerde bulunan tüm fertlerdir. Ancak “İnternet üzerinden mal veya hizmet satın alma ya da sipariş verme (e-posta, SMS ve MMS ile verilen siparişler hariç) işlemini en son ne zaman gerçekleştirdiniz?” sorusu, anketin uygulanması esnasında 16 yaş ve üzeri fertlere sorulmuştur. Bu nedenle analize 16 yaş ve üzeri fertler dahil edilmiştir.

Veri setinin ham halinde 274 değişken bulunmaktadır. Analiz, 108 bağımsız ve 1 bağımlı değişken olmak üzere 109 değişken ile gerçekleştirilmiştir. Veri ön işleme işleminden sonra değişken sayısı 56'ya düşmüştür.

Anketin uygulanması esnasında e-ticaret kullanımı değişkeni, 4 kategoride alınan cevaplar ile oluşturulmuştur. Sınıflandırma performansını arttırabilmek amacıyla 4 kategoriden oluşan bağımlı değişken 3 kategoriye dönüştürülmüştür. Bağımlı değişken üzerinde uygulanan dönüşüm sayesinde;

- CHAID algoritmasının modeli doğru tahmin etme oranı, % 58,17'e,
- C&R Tree algoritmasının modeli doğru tahmin etme oranı, % 60,19'a,
- C5.0 algoritmasının modeli doğru tahmin etme oranı, % 56,81'e,
- RBF(Radyal Tabanlı)-DVM Değişkenlerin doğru tahmin etme oranı, % 59,83'e,
- Doğrusal-DVM Değişkenlerin doğru tahmin etme oranı, % 61,63'e,

- Sigmoid DVM Değişkenlerin doğru tahmin etme oranı, % 60,8 olarak çıkmıştır.

Diler (2016), çalışmasında TÜİK aracılığıyla elde edilen Yaşam Memnuniyeti Araştırması B grubu mikro verileri üzerinde karar ağacı algoritmaları yöntemleri ile analiz yapılmıştır. Burada CHAID yöntemi eğitim verileri üzerinde %63.97 test verileri üzerinde %63.52, C&R Tree yöntemi eğitim verileri üzerinde %63.5 test verileri üzerinde %63.15 olarak bulmuştur.

Can (2017), çalışmasında TÜİK aracılığıyla elde edilen Türkiye Sağlık Araştırması (TSA)'ya ait B grubu mikro verileri üzerinde karar ağacı algoritmaları yöntemleri ile analiz yapılmıştır. Burada CHAID yöntemi eğitim verileri üzerinde %73.25 test verileri üzerinde %72.75, QUEST yöntemi eğitim verileri üzerinde %72.13 test verileri üzerinde %72.54, C&R Tree yöntemi eğitim verileri üzerinde %72.26 test verileri üzerinde %71.91 olarak bulmuştur.

Çalış, Kayapınar, ve Çetinyokuş (2014), çalışmasında bilgisayar ve internet güvenliği ile ilgili 10 soruluk anket, farklı demografik özelliklere sahip 300 kişiye uygulanmış ve SPSS Clementine programında veri madenciliğinde sınıflandırma yöntemlerinden biri olan karar ağaçları kullanılarak kişilerin demografik özelliklerine göre sorulara verilen cevapların doğruluk oranları, 4 farklı karar ağacında test edilmiştir. Burada doğruluk oranları C5.0 yönteminde %81.67, C&R Tree yönteminde %76.67, CHAID yönteminde %73 ve QUEST yönteminde %62,33 olarak bulunmuştur.

Kuzey ve Öner (2012), çalışmasında 2011 yılında bilgi çalışanları üzerinde yapılan anket yolu ile elde edilen 564 adet bilgi çalışanına ait veriler DVM ve Karar ağaçları modelleri uygulanarak analiz yapılmış ve kıyaslanmıştır. Burada karar ağaçları yöntemlerinden CHAID yöntemi eğitim verileri üzerinde %76.22 test verileri üzerinde %79.82, C5.0 yöntemi eğitim verileri üzerinde %74.67 test verileri üzerinde %78.07, C&R Tree yöntemi eğitim verileri üzerinde %78.00 test verileri üzerinde %75.44, Quest yöntemi eğitim verileri üzerinde %74.89 test verileri üzerinde %78.07 olarak bulmuştur. Diğer yandan DVM yöntemlerinden RBF Kernel Fonksiyonu ile eğitim verileri üzerinde %74.22 test verileri üzerinde %78.95, Doğrusal Kernel Fonksiyonu ile eğitim verileri üzerinde %74.89 test verileri üzerinde %78.95, Çok Terimli Kernel Fonksiyonu ile eğitim verileri üzerinde %74.89 test verileri üzerinde

%78.07 ve Sigmoid Kernel Fonksiyonu ile eğitim verileri üzerinde %69.78 test verileri üzerinde %74.56 olarak bulunmuştur.

Önceki çalışmalardan da görüldüğü üzere bu tez çalışmasında yapılan analizler sonucunda bulunmuş olan doğruluk oranları literatürdeki diğer çalışmalarla paralellik göstermektedir. Daha iyi sonuçların çıkmamasının sebebi olarak çalışmanın sosyal bilimler alanında yapılmış olması ve çalışmadaki E-ticaret kullanım sıklığı verilerinin dengeli dağılmamış olmasından dolayı verilerin simüle edilerek çoklanarak yapay dengeleme oluşturulması gösterilebilir.

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 ve DVM arasında otomatik karşılaştırma yapabilen bu işlemci, veri setine uygun ve en yüksek modeli tahmin etme oranına sahip algoritmaların DVM, C5.0, CHAID ve C&R Tree algoritmaları olduğu sonucuna varmıştır.

C5.0 algoritması, 8316 fert ve 8 değişken ile oluşturmuştur. E-ticaret kullanımını etkileyen en önemli 3 değişkeni;

- 0,19 önem değeri ile INT_FAAL_BANKA_ISLEM,
- 0,18 önem değeri ile INT_FAAL_MALHIZMET_SATIS,
- 0,18 önem değeri ile YAS olarak tespit etmiştir.

CHAID algoritması, 8316 fert ve 10 değişken ile oluşturmuştur. E-ticaret kullanımını etkileyen en önemli 3 değişkeni;

- 0,26 önem değeri ile INT_FAAL_BANKA_ISLEM,
- 0,19 önem değeri ile INT_FAAL_MALHIZMET_SATIS,
- 0,13 önem değeri ile INT_FAAL_EPOSTA olarak tespit etmiştir.

C&RT algoritması, 8316 fert ve 10 değişken ile oluşturmuştur. E-ticaret kullanımını etkileyen en önemli 3 değişkeni;

- 0,15 önem değeri ile INT_FAAL_EPOSTA,
- 0,13 önem değeri ile INT_FAAL_BANKA_ISLEM,
- 0,09 önem değeri ile INT_FAAL_SEYAHAT_ISLEM olarak tespit etmiştir.

Doğrusal kernel fonksiyonu, 8316 fert ve 10 değişken ile oluşturmuştur. E-ticaret kullanımını etkileyen en önemli 3 değişkeni;

- 0,22 önem değeri ile INT_FAAL_BANKA_ISLEM,
- 0,15 önem değeri ile INT_FAAL_EPOSTA,
- 0,14 önem değeri ile OKUL_BITEN olarak tespit etmiştir.

Sigmoid kernel fonksiyonu, 8316 fert ve 10 değişken ile oluşturmuştur. E-ticaret kullanımını etkileyen en önemli 3 değişkeni;

- 0,15 önem değeri ile EYETENEK_BLGMOBIL_UYGLM_YUKLM,
- 0,13 önem değeri ile INT_FAAL_BANKA_ISLEM,
- 0,11 önem değeri ile INT_FAAL_EPOSTA olarak tespit etmiştir.

RBF kernel fonksiyonu, 8316 fert ve 10 değişken ile oluşturmuştur. E-ticaret kullanımını etkileyen en önemli 3 değişkeni;

- 0,15 önem değeri ile INT_FAAL_MALHIZMET_SATIS,
- 0,15 önem değeri ile INT_FAAL_BANKA_ISLEM,
- 0,14 önem değeri ile OKUL_BITEN olarak tespit etmiştir.

Daha açık bir ifade ile analiz için kullanılan 6 algoritma incelendiğinde C5.0 algoritması için internet bankacılığı kullanma, CHAID algoritması için internet bankacılığı kullanma, C&RT algoritması için e-posta gönderme/alma, Doğrusal kernel fonksiyonu için internet bankacılığı kullanma, Sigmoid kernel fonksiyonu için yazılım veya mobil uygulama (application) yükleme ve RBF kernel fonksiyonu için mal veya hizmet satışı (gittigidiyor.com, sahibinden.com vb.) faktörleri e-ticaret kullanımını üzerinde etkili en önemli faktörler olduğu belirlenmiştir.

Modeli doğru tahmin etme oranları incelendiğinde kullanılan karar ağacı algoritmalarının birbirlerine çok yakın ve DVM (Support Vector Machine) ise yüksek performans gösterdiği söylenebilir. Analiz sonuçlarına göre en yüksek modeli doğru tahmin etme oranına sahip olan DVM (Support Vector Machine)'ni sırasıyla C5.0 algoritması, CHAID ve C&R Tree algoritmaları takip etmiştir. Ancak DVM (Support Vector Machine), karar ağacı algoritmalarına göre daha fazla veri ve değişken ile oluşmuştur. Algoritmalarda veri yapısına göre üçlü, dörtlü hatta beşli

dallanmalar yapan C5.0 algoritması, daha fazla alt gruplar oluşturarak ayrıntılı deęerlendirmeler yapmaya imkan tanımıştır.

Veri kümesinin kullanılması ile kıyaslanan modellerin genel doęruluk oranlarına bakıldığında DVM (Support Vector Machine) %92,36 ile en yüksek genel doęruluk oranına sahiptir. Sonraki algoritmalar karar aęaçları algoritmaları olan C5.0 %88,01, CHAID %55,59 ve C&R Tree %55,05 algoritmaları olduęu gözlenmiştir. Karar aęaçları algoritmaları genel doęruluk oranı açısından en iyi gösterge C5.0 dir, DVM algoritması kadar iyi sonuç vermemektedir.



KAYNAKLAR

- Akbal, E., Dogan, Ş., & Varol, N. (2017). Karar Ağaçları ile Telefon Dolandırıcılığı Verilerinin Analizi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 29(1), 171-177.
- 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.
- Akpınar, H. (2004). *Business Intelligence & Data Mining*.
- Aksu, M. Ç., & Karaman, E. (2017). Karar Ağaçları ile Bir Web Sitesinde Link Analizi ve Tespiti. *Acta INFOLOGICA*, 1(2), 84-91.
- Aktaş, R., Koçak, A. ve Acar, V. (2010). Kurumsal Kaynak Planlaması Teori ve Bilgisayar Destekli Uygulama Senaryoları, Gazi Kitabevi, Beşevler, Ankara.
- Aktaş, R., Koçak, A. ve Acar, V. (2010). Kurumsal Kaynak Planlaması Teori ve Bilgisayar Destekli Uygulama Senaryoları, Gazi Kitabevi, Beşevler, Ankara.
- Aktepe, C., Baş, M. ve Tolon, M. (2009). *Müşteri İlişkileri Yönetimi*, Detay Yayıncılık, Ankara.
- 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.
- Ay, D. ve Çil, İ. (2010). “Migros Türk A.Ş.De Birliktelik Kurallarının Yerleşim Düzeni Planlamada Kullanılması”, *Sakarya Üniversitesi, Endüstri Mühendisliği Bölümü, Adapazarı*.
- Ayhan, S., & Erdoğan, Ş. (2014). Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 9(1), 175-201.
- Babadağ, K. K. (2003). *Veri Madenciliği Yaklaşımı ve Veri Kalitesinin Artması İçin Kullanılması (uzmanlık tezi, basılmamış)*. Türkiye İstatistik Kurumu, Ankara.
- Bagad, V. (2011). *S., Management Information Systems, Technical Publications*.

- Bagad, V.S. (2011). Management Information Systems, Technical Publications, 2nd Edition, India.
- Bayoglu, L. (2010). Management Information Systems and Data Warehousing A Real Data Warehouse Retail Sample, Lap Lambert Academic Publishing, U.S.A.
- Bensghir, T. (1996). Bilgi Teknolojileri ve Örgütsel Değişim, TODAİE Yayın No:274.
- Berry, A. J. M. ve Linoff, S. G. (2004). Data Mining Techniques. Wiley Publication.
- Bertsekas, D.P. (1999). Nonlinear Programming, Second ed., MA.: Athena Scientific, Cambridge.
- Bilişik, M. T. (2011). Destek Vektör Makinesi, Çoklu Regresyon Ve Doğrusal Olmayan Programlama İle Perakendecilik Sektöründe Gelir Yönetimi İçin Dinamik Fiyatlandırma.
- Boswell, D. (2002). Introduction to Support Vector Machines, Caltech University, USA.
- Breiman, L., Friedman, J. ve Olshen, R. (1984). Classification and Regression Trees, Wadsworth Int.
- Burges, C.J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery 2, 121-167.
- Campbell, C. ve Ying, Y. (2011). Learning with Support Vector Machines, Morgan & Claypool Publishers, California, USA.
- Can, Onur (2017). Türkiye Sağlık Araştırmasının veri madenciliği teknikleri ile incelenmesi, Kafkas Üniversitesi; Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Kars.
- Chapelle, O. ve Haffner, P. (1999). "Support Vector machines for histogram-based image classification", IEEE Trans. on Neural Networks, Vol.10(3.5), pp.1055-1064.
- Cherkassky, V. ve Mulier, F. (1998). Learning From Data: Concepts, Theory and Methods, Wiley.

- Cherkassky, V. ve Shao, X. (1999). "Model Complexity control for regression using VC generalization bounds", IEEE Transactions on Neural Networks, vol.10, pp. 1075-1089.
- Christianini, N. ve Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel -Based Learning Methods, Cambridge University Press.
- Cortes, C. ve Vapnik, V. (1995). "Support Vector Networks", Machine Learning, vol.20, pp.273-297
- Coşkun, C. (2010). Veri Madenciliği Algoritmaları Karşılaştırılması (yüksek lisans tezi, basılmamış). DÜ, Fen Bilimleri Enstitüsü, Diyarbakır.
- Çalış, A., Kayapınar, S., & Çetinyokuş, T. (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama. Endüstri Mühendisliği Dergisi, 25(3-4), 2-19.
- Çelik, E., Atalay, M., & Bayer, H. (2014). Yapay Sinir Ağları ve Destek Vektör Makineleri ile Deprem Tahmininde Sismik Darbelerin Kullanılması.
- Dalkılıç, H. ve Dalkılıç, F. (2015). Karar ağaçları destekli vadeli mevduat hesabı. <http://ab.org.tr/ab15/kabul.html> 17. Akademik Bilişim Konferansı, 4-5 Şubat 2015, Eskişehir. Erişim Tarihi: 21.12.2015
- Demir, B., & Ertürk, S. (2007). Hiperspektral Görüntülerde Destek Vektör Makinaları ile Sınıflandırma Öncesi Dalgacık Uzayında Gürültü Giderimi Wavelet Denoising Before Support Vector Classification of Hyperspectral Images. 15. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı, 1-4.
- Diler, S. (2016). Veri madenciliği süreçleri ve karar ağaçları algoritmaları ile bir uygulama. Ulusal Tez Merkezi, 433080, 10-14.
- Dinçer, E. (2006). Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması. Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kocaeli.
- Dondurmacı, G. (2011). Veri Madenciliğinde Regresyon ağaçları ile Sınıflandırma ve Bir Uygulama (doktora tezi, basılmamış). MSGSU, Fen Bilimleri Enstitüsü, İstanbul.

- Dondurmaci, G. (2002). Bilgisayar, Ahmet Yesevi Üniversitesi Yayınları.
- Dondurmaci, G. (2011). Acar, Veri Madenciliğinde Regresyon Ağaçları ile Sınıflandırma ve Bir Uygulama, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, İstanbul.
- Elibol, H. ve Kesici, B. (2004). "Çağdaş İşletmecilik Açısından Elektronik Ticaret", Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi.
- Emel, G. G., & Taşkın, Ç. (2005). Veri Madenciliğinde Karar Ağaçları Ve Bir Satış Analizi Uygulaması. Sosyal Bilimler Dergisi, 6(2).
- Erişti, H., Uçar, A., & Demir, Y. (2007). Destek Vektör Makineler ve Dalgacık Dönüşümü Kullanarak Güç Sistemlerinde Geçici Rejim Bozulmalarının Sınıflandırılması. ile, 1, 4.
- Ersöz, T., Özseven, T., & Ersöz, F. (2017). Tüketicilerin Cep Telefonu Tercihlerinin Karar Ağacı İle Modellenmesi. Gümüşhane University Electronic Journal of the Institute of Social Science/Gümüşhane Üniversitesi Sosyal Bilimler Enstitüsü Elektronik Dergisi, 8(19).
- Gillo, M.W. (1972). "A Honeywell 600 program for an automatised survey analysis", Behavioral Science, vol.17, pp. 251-251.
- Golemanov, T. ve Golemanova, E. (2004). "ERP-System:Marketing Analysis Module", International Conference on Computer Systems and Technologies-CompSysTech'.
- Govindarajan, M. (2007). "Text Mining Technique for Data Mining Application". Engineering and Technology International Journal of Computer. 1(11): 3460-3465.
- Güner, N., & Çomak, E. (2011). Mühendislik öğrencilerinin matematik i derslerindeki başarısının destek vektör makineleri kullanılarak tahmin edilmesi. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 17(2), 87-96.
- Hamel, L. (2009). Knowledge and discovery with Support Vector Machines, Wiley, USA.

- Hastie, T. ve Tibshirani, R. (2008). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*, Springer, California.
- Haykin, S. (2001). *Neural Networks: A Comprehensive Foundation*, Tsinghua University Press.
- Hoşcan, Y. (2003). ‘Yönetim Bilişim Sistemi’, TC. Anadolu Üniversitesi Yayımı.
- Irmak, S., & Ercan, U. (2017). Karar Ağaçları Kullanılarak Türkiye Hanehalkı Zeytinyağı Tüketimi Görünümünün Belirlenmesi. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 13(3), 553-564.
- Ivanciuc, O. (2007). Applications of Support Vector Machines in Chemistry. In: *Reviews in Computational Chemistry, Volume 23*, Eds.: K. B.
- James, G., Witten, D., Hastie, T. ve Tibshirani, R. (2013). *An Introduction to Statistical Learning*, Springer, USA.
- Karahoca, A. ve Karahoca, D. (1998). *Yönetim Bilişim Sistemleri ve Uygulamaları*, Beta Yayınevi, İstanbul.
- Kass, G.V. (1980). “An exploratory technique for investigating large quantities of caregorical data”, *Applied Statistics*, vol. 29(2), pp.119-127.
- Kavzoğlu, T., & Çölkesen, İ. (2010). Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi. *Harita Dergisi*, 144(7), 73-82.
- Kaya, C., Erkamaz, O., Ayar, O., & Özer, M. (2017). C4. 5 Karar Ağacı Temelli Öznitelik Seçimi ile Video-Okülografi (VOG) Sinyallerinden Diyabetik Retinopati Hastalığının Sınıflandırılması Classification of Diabetic Retinopathy Disease. *Tıp Teknolojileri Ulusal Kongresi*, 75-78.
- Kayri, M. (2014). Karar ağaçları. *Karar Ağaçları Çalıştay*, Muş Alparslan Üniversitesi.
- Kecman, V. (2001). *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*, MIT Press, UK. Lipkowitz and T. R. Cundari. Wiley-VCH, Weinheim, 291–400.

- Köse, U., Güraksın, G. E., & Deperlioğlu, Ö. (2015, October). Diabetes determination via vortex optimization algorithm based support vector machines. In 2015 Medical Technologies National Conference (TIPTEKNO) (pp. 1-4). IEEE.
- Kuhn, H.W. ve Tucker, A.W. (1951). Nonlinear programming, Proceedings of 2nd Berkeley Symposium, Berkeley: University of California Press., 481–492.
- Kuzey, C., & Öner, E. S. E. N. (2012). Veri madenciliğinde destek vektör makinaları ve karar ağaçları yöntemlerini kullanarak bilgi çalışanlarının kurum performansı üzerine etkisinin ölçülmesi ve bir uygulama. Yayımlanmamış Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Laudon, K. C. ve Laudon, J. P. (2011). Yönetim Bilişim Sistemleri Dijital İşletmeyi Yönetme, Çeviri Editörü: Prof. Dr. Uğur Yozgat, 12. Basım, Nobel Akademik Yayıncılık Eğitim Danışmanlık Tic.Ltd. Şti.
- Loh, W.Y. ve Shih, X. (1997). “Split selection methods for classification trees”, *Statistica Sinica*, vol. 7, pp.815-840.
- Maimon, O. ve Rokach, L. (2005). “Introduction To Supervised Methods”, *Data Mining and Knowledge Discovery Handbook*, Ed. by Lior Rokach and Oded Maimon, Springer Science+Business Media Inc.
- Mercer, J. (1909). Functions of Positive and Negative Type and Their Connection with the Theory of Integral Equations, *Phil. Trans. Roy. Soc. London A*, 209, 415–446.
- Murthy, K. S. (1998). Automatic Construction of Decision Tree from Data: A Multi-Disciplinary Survey. Kluwer Academic Publishers. Siemens Corporate Research.
- Myatt, J. G. (2007). Making Sense of Data. Wiley Publication.
- Oğuzlar, A. (2003). Veri ön işleme, Erciyes Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi, 21: 67-76.
- Oğuzlar, A. (2004). CART analizi ile hanehalkı isgücü anketi sonuçlarının özetlenmesi. Atatürk Üniversitesi İİBF Dergisi 18(3-4):79-90.

- Omitaomu, O.A. (2006). Decision Trees. In Michael W. Berry and Murray Browne (Eds.), Lecture Notes in Data Mining, pp. 39-51, World Scientific Publishing of Hackensack, New Jersey, USA.
- Onan, A. (2015). “Şirket İflaslarının Tahmin Edilmesinde Karar Ağacı Algoritmalarının Karşılaştırmalı Başarım Analizi”. Bilişim Teknolojileri Dergisi, 8(1): 9-19.
- Öğüdücü, G. Ş. (2008). Veri madenciliği temel sınıflandırma yöntemleri. <http://ninova.itu.edu.tr/tr/dersler/bilisim-enstitusu/195/bbl-606/ekkaynaklar?g29172> İstanbul Teknik Üniversitesi, İstanbul.
- Özbudak, Ö. (2009). Yüz Resimlerinden Cinsiyet Tayini (yüksek lisans tezi, basılmamış). İTÜ, Fen bilimleri Enstitüsü, İstanbul.
- Özmen, Ş. (2013). Ağ Ekonomisindeki Yeni Ticaret Yolu E-Ticaret, İstanbul Bilgi Üniversitesi Yayınları, 5.Baskı, İstanbul.
- Pektaş, A. O. (2013). SPSS ile Veri Madenciliği. İstanbul: Ecem Basım Yayın.
- Steinwart, I. ve Chsirtmann, A. (2008). Support Vector Machines, Springer, USA.
- Quinlan, J.R. (1986). “Introduction of decison trees”, Machine Learning, vol.1, pp.81-106.
- Quinlan, J.R. (1989). “Inferring decission trees using minimum description length principle”, Information and Computations, vol.80, pp.227-248.
- Quinlan, J.R. (1993). C4.5: Programs for Machine Learning, Los Altos, Morgan Kaufmann.
- R.O., Hart, P.E. ve Stork, D.G. (2001). Pattern Classification, Wiley.
- Sezer, E. A., Bozkır, A. S., Yağız, S., & Gökçeoğlu, C. (2010). Karar Ağacı Derinliğinin CART Algoritmasında Kestirim Kapasitesine Etkisi: Bir Tünel Açma Makinesinin İlerleme Hızı Üzerinde Uygulama. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, Kayseri.
- Sezer, O. G., Erçil, A., & Keskinöz, M. (2005). Destek vektör makinesi kullanarak bağımsız bileşen tabanlı 3B nesne tanıma= Independent component based 3D object recognition using support vector machines.

- Silahtaroglu, G. (2013). Veri Madenciliği Kavram ve Algoritmaları, Papatya Yayınevi.
- Sonquist, J.A. (1971). Searching for Structure. Institute for Social Research, Ann Arbor, MI, Univ. of Michigan.
- Souza, J., Matwin, S. ve Japkowicz. (2002). Evaluation data mining Models: A Pattern Language. School of Information Technology and Engineering. University of Ottawa.
- SPSS. (1999). AnswerTree Algorithm Summary. U.S.A: SPSS Inc.
- SPSS. (2002). AnswerTree 3.1 User's Guide, U.S.A: SPSS Inc.
- SPSS. (2007). Clementine. Clementine 12.0 Algorithms Guide. Integral Solutions Limited.
- Şentürk, A. (2006). Veri Madenciliği Kavram ve Teknikler, Ekim Basım Yayın Dağıtım, Bursa.
- Şimşek, M. Ş. (2010). Yönetim Bilişim Sistemleri, Gazi Kitabevi
- Utku, A., Doğru, İ. A., & Akcayol, M. A. (2018, May). Decision tree based android malware detection system. In 2018 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.
- Uwizemungu, S. ve Raymond, L. (2004). "Essential Characteristics of An ERP System: Conceptualization and Operationalization", Sixth International Conference on Enterprise Information Systems, Porto, Portugal.
- Van Der Heijden, H. (2009). Designing Management Information Systems, Oxford University Press.
- Vercellis, C. (2009). Business intelligence : Data Mining and Optimization for Decision Making. Wiley Publication.
- Wang, L. (2005). Support Vector Machines: Theory and Applications, Springer, Netherlands.
- Witten, I.H., Frank, E. ve Hall, A. M. (2011). Data Mining : practical machine learning tools and techniques. Morgan Kaufmann Publishers.

- Yabanova, İ., & Yumurtacı, M. (2018). Destek vektör makineleri kullanarak dinamik yumurta ağırlıklarının sınıflandırılması. *Journal of the Faculty of Engineering & Architecture of Gazi University*, 33(2).
- Yakut, E. (2012). Veri Madenciliği 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.
- Yaşar, A. (2016). Olumlu Görüş Dışındaki Denetim Görüşlerinin Veri Madenciliği Yöntemleriyle Tahminine İlişkin Karar Ve Birliktelik Kuralları. *Mali Cozum Dergisi/Financial Analysis*, 26(133).
- Yenipinar, S. ve Gölcü, H. (2003) Elektronik Ticaret ve İnternette Pazarlama, Kahramanmaraş Sütçü İmam Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Lisans Tezi, Kahramanmaraş.
- Yıldıztepe, E., & Kocataş, A. (2018). Türkiye İşgücü Verilerinin Karar Ağacı Yöntemleriyle Analizi. *Çankırı Karatekin Üniversitesi İİBF Dergisi*, 8(2), 91-114.
- Yılmaz, A. K., Durak, M. Ş., Özdemir, E., Kiracı, K., & Tanrıverdi, G. (2017). Eskişehir İlinde Havayolu Yolcu Potansiyelinin Belirlenmesi Üzerine Bir Karar Ağacı Uygulaması. *Anadolu University of Sciences & Technology-B: Theoretical Sciences*, 5(1).
- Yılmaz, E. (2016). Kardiyotokogram verisinden fetal iyilik halinin belirlenmesi için bir karar destek sistemi. *Uludağ University Journal of The Faculty of Engineering*, 21(2), 331-340.
- Yozgat, U. (1998). Yönetim Bilişim Sistemleri, Beta Yayınevi.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	Sahib RAMAZANOV
Doğum Yeri	Bakü/Azərbaycan
Doğum Tarihi	01 Şubat 1994

LİSANS EĞİTİM BİLGİLERİ

Üniversite	Kafkas Üniversitesi
Fakülte	İktisadi İdari Bilimler Fakültesi
Bölüm	İşletme

İŞ DENEYİMİ

Çalıştığı Kurum	TURAN-SAM Uluslararası Bilimsel Hakemli Dergi
Görevi	Yazı İşleri Müdürü
Tecrübe Süresi	1 YIL

İLETİŞİM

Adres	Cumhuriyet Mahallesi. Şehit Yusuf Bey Caddesi. No:18 D:2
E-mail	sahibramazanli94@gmail.com