

**T.C.
ZONGULDAK KARAELMAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI**

Yüksek Lisans Tezi

**VERİ MADENCİLİĞİ: İSTANBUL MENKUL
KIYMETLER BORSASI ÖRNEĞİ**

Şebnem Koltan Yılmaz

Zonguldak 2008

**T.C.
ZONGULDAK KARAELMAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI**

Yüksek Lisans Tezi

**VERİ MADENCİLİĞİ: İSTANBUL MENKUL
KIYMETLER BORSASI ÖRNEĞİ**

**Hazırlayan
Şebnem Koltan Yılmaz**

**Tez Danışmanı
Yrd. Doç. Dr. Ali Sait Albayrak**

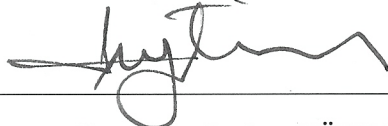
Zonguldak 2008

T.C.
ZONGULDAK KARAEMLAS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

TEZ ONAYI

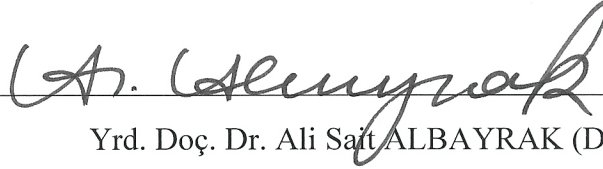
Enstitümüzün İşletme Anabilim Dalında 2004528201009 numaralı Şebnem Koltan Yılmaz'ın hazırladığı “**Veri Madenciliği: İstanbul Menkul Kıymetler Borsası Örneği**” konulu YÜKSEK LİSANS tezi ile ilgili TEZ SAVUNMA SINAVI, Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği uyarınca 20/06/2008 Pazartesi günü saat 14:00'te yapılmış, sorulan sorulara alınan cevaplar sonunda tezinin onayına OYBİRLİĞİYLE karar verilmiştir.

Başkan



Doç. Dr. Şevket TÜYLÜOĞLU

Üye



Yrd. Doç. Dr. Ali Saif ALBAYRAK (Danışman)

Üye



Yrd. Doç. Dr. Coşkun HAMZAÇEBİ

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

18.07.2008



Doç. Dr. Turhan KORKMAZ
Enstitü Müdürü

ÖZET

Kurum : ZKÜ Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı
Tez Başlığı : Veri Madenciliği: İstanbul Menkul Kıymetler Borsası Örneği
Tez Yazarı : Şebnem Koltan Yılmaz
Tez Danışmanı : Yrd. Doç. Dr. Ali Sait Albayrak
Tez Türü, Yılı : Yüksek Lisans Tezi, 2008
Sayfa Adedi : 80

Organizasyonlarda karar vermeye yönelik bilgiye erişim ihtiyacı arttıkça, bu amaca yönelik bilgi sistemleri kurulmaktadır. Bu sistemler karar almada kolaylık sağlarken bunu, önceden belirlenmiş ve sistem içerisinde yapılandırılmış konularda gerçekleştirebilmektedir. Veri boyutlarının artmasıyla birlikte önceden tahmin edilemeyen sorulara yönelik cevap bulan, ileriye dönük tahmin sistemlerine ihtiyaç duyulmuştur. Veri madenciliği, verilerden önceden bilinmeyen anlamlı bilgileri tanımlama ya da tahmin etme tekniklerini içermektedir. Bu bağlamda veri madenciliği teknikleri diğer birçok alanda olduğu gibi işletmecilik alanında da yaygın bir şekilde kullanılmaktadır.

Bu çalışmada amaç, yatırımcıların veya karar vericilerin işletmelerin performanslarını değerlendirmede dikkate alması gereken en önemli ölçütlerin saptanmasında veri madenciliği tekniklerinin uygun bir yöntem olarak kullanılabileceğini göstermektir. Çalışmada, veri madenciliği kavramları ve teknikleri açıklanmıştır. İstanbul Menkul Kıymetler Borsası Ulusal 100 endeksinde sanayi ve hizmet sektörlerinde faaliyet gösteren 173 işletmenin 2004-2006 yıllarına ait yıllık finansal göstergelerinden yararlanarak veri madenciliği tekniklerinden birisi olan karar ağaçları tekniği uygulanmıştır. Bu amaçla, ikili ağaç türetimi yerine çoklu ağaç türetimine olanak sağlayan CHAID algoritması kullanılmıştır. Seçilen finansal göstergelere göre sanayi ve hizmet sektörlerinde faaliyet gösteren firmaları ayıran en önemli özelliklerin saptanmasının yanında sermaye ve karlılık yapılarını belirleyen en önemli değişkenler saptanmıştır.

Anahtar Kelimeler: Veri Madenciliği, Karar Ağacı, CHAID Analizi, İMKB, Sermaye Yapısı

ABSTRACT

Institution : ZKÜ Institute of Social Sciences, Department of Management
Title : Data Mining: A Case Study of Istanbul Stock Exchange
Author : Şebnem Koltan Yılmaz
Adviser : Asst. Prof. Ali Sait Albayrak
Type of Thesis, Year : MSc. Thesis, 2008
Total Number of Pages : 80

It is getting more needs of knowledge access for decision making in organizations, management knowledge systems have been established for this aim. While these systems providing easiness to decision making, they realize that subjects predefined and costructed in system. With increasing data dimensions, to be needed prediction systems which answer for nonpredefined questions. Data mining include decriptive and predictive techniques for meaningfull knowledges which is unknown early from datas. For this context, data mining techniques are commonly used in business area as in the other areas.

The aim of the study indicate data mining techniques are available method that determining the most important factors which businesses must taken into consideration for evaluating their performances. In this study, data mining concepts and techniques are explained. Decision trees technique which is one of the data mining techniques is applied on 173 firms that operate in industry and service sectors in İstanbul Stock Exchange National 100 Index using their annual financial indicators from 2004 to 2006. For this aim, CHAID algorithm that derive multiple tree instead of binary tree is used. According to selected financial indicators, the most important factors of firms that operate in industry and service sectors are determined besides the most important variables for their profitabilities and capital structures.

Anahtar Kelimeler: Data Mining, Decision Tree, CHAID Analysis, ISE, Capital Stucture.

ÖNSÖZ

Kişisel ve kurumsal alanda bilgisayar kullanımının artması dijital veri yığınlarının ve veri tabanlarının boyutlarının artmasına sebep olmuştur. Dünyadaki veri miktarının olağanüstü artışı düşünüldüğünde, problemlere çözüm olacak bilgi bankalarının veri madenciliği alanındaki gelişmelerle sağlanacağı öngörülmektedir.

Karar ağaçları tekniği, kural çıkarmanın basitliği, sonuçlarının görselliği ve yorumlanmasının kolay olması nedenleriyle son yıllarda yaygın olarak uygulanan veri madenciliği tekniklerinden birisidir. Bu yaklaşımla bu çalışmada İMKB Ulusal 100 endeksinde sanayi ve hizmet sektöründe faaliyet gösteren işletmelerin finansal göstergelerinden yararlanılarak işletmelerin sermaye yapılarını, karlılık düzeylerini ve sektör ayrımını tanımlayan en önemli faktörler araştırılmaktadır. Ayrıca bu faktörlerden yararlanarak işletmelerin profilleri oluşturulmaktadır.

Tez çalışmamın her aşamasında beni yönlendiren, öneriler sunan ve hedefe ulaşmamda yol göstericim olan değerli hocam Yrd. Doç. Dr. Ali Sait ALBAYRAK'a teşekkür ederim.

Çalışmalarım sürecinde göstermiş oldukları sabır, destek ve yardımları için eşim Bülent YILMAZ'a ve aileme teşekkürlerimi sunarım.

İÇİNDEKİLER

Sayfa

ÖZET	iii
ABSTRACT.....	iv
ÖNSÖZ	v
İÇİNDEKİLER	vi
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ.....	ix
KISALTMALAR LİSTESİ.....	x
GİRİŞ	1
1. VERİ TABANLARINDA BİLGİ KEŞFİ	6
1.1. Veri Tabanlarında Bilgi Keşfi Aşamaları	7
1.2. Veri Ambarları (Data Warehouses)	11
1.3. Veri Madenciliği	12
1.3.1. Veri Madenciliği Tanımı ve Amaçları	12
1.3.2. Veri Madenciliği Süreci	15
1.3.3. Veri Madenciliği İşlevleri	17
1.4. Veri Madenciliği Yöntem ve Teknikleri	18
1.4.1. Sınıflama ve Regresyon Modelleri.....	18
1.4.2. Kümeleme	19
1.4.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler	19
1.5. Veri Madenciliği Uygulama Alanları.....	21
2. SINIFLANDIRMA VE KARAR AĞAÇLARI.....	25
2.1. Yapay Sinir Ağları (Artificial Neural Networks).....	26
2.2. Genetik Algoritmalar (Genetic Algorithms)	26
2.3. K- En Yakın Komşu (K-Nearest Neighbour).....	26
2.4. Bellek Tabanlı Nedenleme (Memory Based Reasoning).....	27
2.5. Naive – Bayes	27
2.6. Lojistik Regresyon (Logistic Regression).....	27
2.7. Karar Ağaçları (Decision Trees)	28

2.7.1. CRT Algoritması	31
2.7.2. CHAID Algoritması	31
2.7.3. QUEST Algoritması	33
2.7.4. MARS Algoritması	34
2.7.5. ID3 Algoritması	34
2.7.6. C4.5 Algoritması	34
3. VERİ MADENCİLİĞİ UYGULAMASI: İSTANBUL MENKUL KIYMETLER BORSASI (İMKB) ÖRNEĞİ	36
3.1. Çalışmanın Tanıtılması	36
3.1.1. Çalışmanın Amacı ve Önemi	36
3.1.2. Çalışmanın Kapsamı	37
3.1.3. Çalışmaya İlişkin Literatür Özeti	37
3.2. Çalışmanın Yöntemi	38
3.2.1. Veri Anlama ve Veri Seçimi Aşaması	39
3.2.2. Veri Önileme Aşaması	39
3.2.3. Veri Dönüştürme Aşaması	40
3.2.4. Modelleme Aşaması: Karar Ağacı Tekniği CHAID Algoritmasının Uygulanması	42
3.2.5. Değerlendirme Aşaması	43
3.2.5.1. CHAID Modeli Sonuçlarının Değerlendirilmesi	43
3.2.5.2. CHAID Karar Ağacı ile Elde Edilen İMKB100 Sektör Sonuçları	46
3.2.5.3. CHAID Karar Ağacı ile Elde Edilen İMKB100 KVB/TA Sermaye Yapısı Sonuçları	50
3.2.5.4. CHAID Karar Ağacı ile Elde Edilen İMKB100 TB/TA Sermaye Yapısı Sonuçları	54
3.2.5.5. CHAID Karar Ağacı ile Elde Edilen İMKB100 NK/TA Karlılık Sonuçları	58
3.2.5.6. CHAID Karar Ağacı ile Elde Edilen İMKB100 NK/ÖS Karlılık Sonuçları	63
SONUÇ	68
KAYNAKÇA	74

TABLÖLAR LİSTESİ

Sayfa

Tablo 3.1: Çalışmada Kullanılan Finansal Oranlar Tablosu	40
Tablo 3.2: Sektör Sınıfı CHAID Algoritma Sonuçları.....	44
Tablo 3.3: Kısa Vadeli Borç Oranı CHAID Algoritma Sonuçları	44
Tablo 3.4: Toplam Borç Oranı CHAID Algoritma Sonuçları.....	45
Tablo 3.5: Aktiflerin Karlılığı CHAID Algoritma Sonuçları.....	45
Tablo 3.6: Özsermayenin Karlılığı CHAID Algoritma Sonuçları	45
Tablo 3.7: Sektör Sınıfı Profilleri ve Profilleri Oluşturan Değişkenler	47
Tablo 3.8: Kısa Vadeli Borç Oranı Profilleri ve Profilleri Oluşturan Değişkenler....	50
Tablo 3.9: Toplam Borç Oranı Profilleri ve Profilleri Oluşturan Değişkenler	55
Tablo 3.10: Aktiflerin Karlılığı Profilleri ve Profilleri Oluşturan Değişkenler	59
Tablo 3.11: Özsermayenin Karlılığı Profilleri ve Profilleri Oluşturan Değişkenler..	64

ŞEKİLLER LİSTESİ

Sayfa

Şekil 1.1: Veri Tabanlı Bilgi Keşfi Süreci	8
Şekil 1.2: Veri Madenciliği İlgili Alanları	14
Şekil 2.1: Karar Ağacı Örneği.....	29
Şekil 2.2: Karar Ağacı Değişken İlişkisi.....	32
Şekil 3.1: İMKB Sektör Sınıfı Karar Ağacı	48
Şekil 3.2: Kısa Vadeli Borç Oranı Karar Ağacı.....	52
Şekil 3.3: Toplam Borç Oranı Karar Ağacı	57
Şekil 3.4: Aktiflerin Karlılığı Karar Ağacı	60
Şekil 3.5: Özsermayenin Karlılığı Karar Ağacı	65

KISALTMALAR LİSTESİ

BTY	: Bellek Tabanlı Yönetim
CBS	: Coğrafi Bilgi Sistemleri
CHAID	: Chi-squared Automatic Interaction Detection
CRT	: Classification and Regression Tree
İMKB	: İstanbul Menkul Kıymetler Borsası
İMKB100	İstanbul Menkul Kıymetler Borsası Ulusal 100 Endeksi
İSDH	: İşletme Sermayesi Devir Hızı
KOBİ	: Küçük ve Orta Büyüklükteki İşletmeler
MARS	: Multivariate Adaptive Regression Splines
QUEST	: Quick Unbiased Efficient Statical Tree
RFM	: Recency, Frequency and Monetary Value
SLIQ	: Supervised Learning in Quest
SPRINT	: Scalable Paralleizable Induction of Decision Trees
VM	: Veri Madenciliği
VTBK	: Veri Tabanlarında Bilgi Keşfi
VTYS	: Veri Tabanı Yönetim Sistemleri
YSA	: Yapay Sinir Ağları

GİRİŞ

Bilgi teknolojilerinde dikkati çekecek ölçülerde yaşanan gelişmeler ile depolanan veriler çok daha büyük boyutlara ulaşmaktadır. Bilgisayarlar daha büyük miktardaki veriyi saklayabilmekte ve daha kısa sürede işleyebilmektedir. Bilgisayar ağlarının gelişimi ile açık ağ sistemlerinin yanı sıra kapalı ağ sistemlerinin de kullanılması sonucunda veriye başka bilgisayarlardan da hızla ulaşılabilir. Verinin sayısal olarak kullanılabilmesi sonucu daha güvenilir, doğru, hızlı ve ayrıntılı bilgiye ulaşılabilir.

Ham veri kendi başına değersizdir. Veri, bilgisayar sistemleriyle belirli bir amaca yönelik işlenerek bilgiye dönüşmektedir. Birbiriyle ilişkili verilerin bir arada toplanmasıyla bilgiye erişim ihtiyacı doğmuştur. Organizasyonlarda bu amaca yönelik bilgi sistemleri kurulmaktadır. Bu bilgi sistemleri, satışlar, faturalar, ödemeler gibi işlemlerin takip edilmesini sağlarken, karar almaya yönelik ihtiyaçlara da girdi oluşturmaktadır. Ancak bu sistemler karar almada kolaylık sağlarken bunu, önceden belirlenmiş ve sistem içerisinde yapılandırılmış konularda gerçekleştirebilmektedir. Kurum içi ve dışı bilgilerin yanı sıra önceden tahmin edilemeyen sorulara yönelik cevap bulan, ileriye dönük tahmin sistemlerine ihtiyaç duyulmuştur. Bu bağlamda veri madenciliği (VM) teknikleri diğer birçok alanda olduğu gibi işletmecilik alanında da yaygın bir şekilde kullanılmaktadır. VM, bilgi teknolojilerinin doğal gelişim sürecinin sonucu olarak da değerlendirilebilir. Çok büyük ölçekli veriler, farklı alanlardaki büyük ölçekli veri tabanları içinde değerli verileri bulunduran bir veri madeni gibi düşünülebilir. VM ise bu verilerden önceden bilinmeyen anlamlı bilgileri üretme süreci olarak tanımlanmaktadır. Diğer bir anlatımla VM, veri tabanlarında yer alan büyük hacimli verilerde bulunan gizli bilgilerin, örüntülerin, kuralların ve ilişkilerin gelişmiş çok boyutlu tekniklerle keşfedilmesi süreci olarak tanımlanabilir. VM, bu süreci bilgisayarı, makine öğrenmesini (machine learning), veri tabanı veya veri ambarı yönetimini, matematiksel algoritmaları ve istatistik teknikleri kullanarak gerçekleştirmektedir.

VM yerine veri tabanlarından bilgi keşfi kavramı (VTBK) da kullanılmaktadır. Veri tabanlarından bilgi keşifleri ile yüksek düzeyli bilgiler veri tabanlarındaki verilerden çekilmeye ve farklı açılardan değerlendirilmeye başlanmıştır. Büyük ölçekli veri tabanlarından bilgi çıkarmak, birçok araştırmacı için veri tabanı sistemlerindeki bir araştırma konusu olmuştur. VM ile keşfedilen bilgi, çevrimiçi (online) veri takibi, iş yönetimi, ürün kontrol sistemleri, pazarlama, tıp, finans, kimya, coğrafi bilgi sistemleri (CBS), mühendislik, görüntü tanıma ve robot görüş sistemleri, uzay bilimleri, sosyal bilimler, davranış bilimleri, meteoroloji ve atmosfer bilimleri gibi alanlarda yaygın bir şekilde kullanılmaya başlanmıştır.

VTBK ifadesi ilk defa Piatetsky-Shapiro (1991) tarafından 1989 yılında gerçekleştirilen ilk VTBK çalışma grubu toplantısında kullanılmış, konuyla ilgili kavram ve tanımlar ortaya konulmuştur. VM terimi de VTBK'nin bir bileşeni olarak tanımlanmıştır. Daha sonra VM yöntemleri geliştirmeye yönelik çalışmalar devam etmiş, Agrawal vd. (1995) nicel birliktelik kurallarının madenciliği için hızlı bir algoritma olan Apriori'yi önermişlerdir. VTBK'nin temel süreçleri üzerinde hiyerarşi arayışının sonucunda, Fayyad vd. (1996) veri tabanlarında bilgi keşfinin süreçlerine ve VM'nin bu süreçteki yerine yönelik bir akış sunmuşlardır. Ayrıca, VM'nin temel özelliklerini irdemişlerdir. VM'nin uygulama yaygınlığı kazanmasıyla birlikte uygulamaya yönelik çalışmalar ağırlık kazanmaya başlamış ve Berson vd. (1999) VM'nin en yaygın kullanım alanı olan müşteri ilişkileri yönetimi kapsamında VM yöntemleri ve uygulamalarına yer vermişlerdir. Ayrıca, yeni nesil yöntemlerden olan karar ağacı yöntemlerinin başlıcalarından CRT¹ (Classification and Regression Tree) ve CHAID (Chi-squared Automatic Interaction Detection) yöntemlerini incelemişlerdir. VM bilişim ve istatistik olmak üzere iki bakış altında incelenmektedir. Ziarko (1991), Elder and Pregipon (1995), VM'yi, istatistik alanındaki birçok yöntemi kullanmasına rağmen, nesnelerin nitelik değerlerine bağlı çıkarım yapmada bilinen istatistik metotlardan ayırmaktadır. Hastie vd. (2001), VM'ye istatistik perspektiften yaklaşmış ve VM ile istatistik arasındaki ilişkiyi 'İstatistik Öğrenme' kavramı ile kurmuşlardır. Moss and Atre (2003) VM ile

¹ Literatürde C&RT ya da CART kısaltmaları da kullanılmaktadır.

istatistik analiz yöntemlerini karşılaştırmış ve aralarındaki farklılıkları ortaya koymuşlardır.

VM’de kullanılan modeller, tahmin edici (predictive) ve tanımlayıcı (descriptive) olmak üzere iki ana başlık altında incelenmektedir. VM modellerini gördükleri işlevlere göre sınıflama (classification) ve regresyon (regression) modelleri, kümeleme (clustering) modelleri ve birliktelik kuralları (association rules) ve ardışık zamanlı örüntüler (sequential patterns) olmak üzere üç ana başlık altında incelemek de mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir.

Bu çalışmada, VM modellerinden biri olan sınıflama modelinin karar ağacı tekniği kullanılmıştır. Sınıflama modeli, sınıfı tanımlanmış mevcut verilerden yararlanarak, sınıfı belli olmayan verilerin sınıfını tahmin etmek için kullanılan VM modelidir. Karar ağacı, çok sayıda kayıt içeren bir veri kümesini, bir dizi karar kuralları uygulayarak daha küçük kümelere bölmek için kullanılan bir yapıdır.

VM’de karar ağacı oluşturmak için çeşitli algoritmalar geliştirilmiştir. Bu algoritmalarından biri olan CHAID algoritması Kaas (1980) tarafından en iyi bölmeyi hesaplamak için istatistik olarak anlamlı bir farklılığın olmadığı, hedef değişkene uyan çiftlerde tahmin değişkeninin olası kategori çiftini birleştirmesiyle oluşturulmuştur (SPSS, 1999:2).

Haughton ve Oulabi (1999), çalışmasında bir doğrudan pazarlama modelini CHAID algoritması ile gerçekleştirmiş, CRT ve CHAID algoritmalarının sonuçlarını karşılaştırmıştır. Akpınar (2000), uygulamasında kişilerin geri ödemelerini düzenli veya düzensiz yapmalarına bağlı olarak, kredi değerlendirmesini iyi ve kötü şeklinde sınıflandırmak üzere CHAID algoritmasını kullanmıştır. Grobler vd. (2002), okul etkinliği ve çeşitli bağımsız değişkenler arasındaki ilişkiyi tahmin etmede CHAID yönteminden yararlanmıştır. Doğan ve Özdamar (2003), CHAID analizi yardımıyla ailelerin çocuk isteğine etki eden faktörlere ulaşmada bağımsız değişkenlerin birleşmiş kategorilerini ve alt gruplarını tahmin eden bir çalışma yapmıştır. Van

Diepen ve Franses (2006), CHAID algoritmasından tahmin edilen sonuçların güven sınırlarını hesaplamak için önerdiği yöntemle, benzetim tekniğiyle elde edilmiş veriler kullanmıştır. Laura vd. (2006), market üyelik kartlarının sınıflamasını CHAID yöntemi kullanarak yapmıştır. Koyuncugil ve Özgülbaş (2006), CHAID algoritmasını kullanarak İstanbul Menkul Kıymetler Borsası (İMKB)'nda işlem gören Küçük ve Orta Ölçekli İşletmeler (KOBİ)'in finansal profilleri ile güçlü ve zayıf yönlerinin belirlenmesini amaçlamıştır. McCarty vd. (2007), iki farklı veri seti kullanarak doğrudan pazarlama bölümlemesi için RFM (Recency, Frequency, and Monetary Value), CHAID ve lojistik regresyon modellerinin karşılaştırmasını yapmıştır. Koyuncugil (2007), CHAID algoritması ile İMKB şirketleri özelinde risk ölçümlemesindeki değişimleri otomatik olarak değerlendirmeyi ve öncü risk göstergelerini sektörel bazda belirlemeyi sağlayacak VM'ye dayalı bir sektörel risk modeli oluşturmuştur. Türe vd. (2008), meme kanserli hastalarda yinelemesiz sağkalım süresini etkileyen risk faktörlerinin belirlenmesinde karar ağacı yöntemlerinden CRT, CHAID, QUEST (Quick Unbiased Efficient Statical Tree), C4.5 ve ID3 ile Kaplan-Meier analizini birlikte kullanmıştır. Dolgun (2008), Türkiye'deki hastaneleri karar ağaçları tekniğini kullanarak durumlarına ait sonuçlar çıkarmış ve bu sonuçlardan elde ettiği verilerle kümeleme analizi yaparak gruplandırmıştır.

Bu tezde, karar ağaçları kullanılarak İstanbul Menkul Kıymetler Borsası (İMKB) örneği incelenmiştir. Çalışma giriş ve sonuç hariç üç bölümden oluşmaktadır.

Giriş bölümünde, veri madenciliği ve karar ağaçları hakkında genel bilgi verilmiş, konuya ilişkin yerli ve yabancı literatür taraması yapılmış ve tez çalışmasının amacı hakkında genel bir tanımlama yapılmıştır.

Birinci bölümde, veri tabanlarında bilgi keşfi, veri ambarı ve veri madenciliğine ilişkin kavramsal bir çerçeve sunulmaktadır. Veri madenciliği yöntem ve teknikleri gördükleri işleve göre incelenmiş ve veri madenciliği uygulama alanları açıklanmıştır.

İkinci bölümde, sınıflandırma ve karar ağaçları teknikleri ayrıntılı olarak incelenmiştir. Ayrıca karar ağaçları tekniğinde kullanılan algoritmalar açıklanmış ve bunların birbirlerine göre avantaj ve dezavantajları tartışılmıştır.

Üçüncü bölümde, karar ağaçları tekniği ile CHAID algoritması kullanılarak İMKB Ulusal 100 endeksindeki firmaların sektör ve finansal göstergelerinin birbirleriyle olan ilişkileri irdelenmiştir. Çalışmada öncelikle İMKB100 sanayi ve hizmet sektörlerinde faaliyet gösteren firmaların karlılık, likitide, varlık kullanım etkinliği, sermaye yapısı ve işletme büyüklüğü göstergelerinden hangisine göre anlamsal farklılık gösterdiği araştırılmaktadır. Daha sonra sırasıyla sermaye yapısı göstergelerinden kısa vadeli borç oranı, toplam borç oranı; karlılık göstergelerinden aktiflerin karlılığı, özsermayenin karlılığı değişkenlerinin en önemli belirleyicileri belirlenmektedir.

Sonuç bölümünde ise elde edilen bulgular ışığında araştırma sonuçları değerlendirilip bazı önerilerde bulunmaktadır.

1. VERİ TABANLARINDA BİLGİ KEŞFİ

Belirlenen amaçlar doğrultusunda ölçülebilen ve sayılabilen her şey veri olarak tanımlanmaktadır. Sayılar, harfler veri olduğu gibi bunlardan elde edilen bilgi de yine veri olarak kullanılmaktadır. Günümüzde her alanda artan hızla veri yığınları meydana gelmektedir.

Veriler hacim olarak sayfalarca yer kaplarlar ancak kullanım değerleri azdır. Oysa, sayılar düzenlenip özetlendiğinde, harfler düzenlenerek anlamlı cümleler haline dönüştürüldüğünde, notalar sıralanıp bir melodi oluşturulduğunda ve bilgisayar ekranındaki noktalar (verileri) bir araya getirilerek bir ağaç resmi veya bir grafik oluşturulduğunda veriler bilgiye dönüştürülmüş olur. Bilgi verilere göre hacim olarak daha az yer tutar ancak kullanım değeri olarak daha güçlüdür (Gürsakal, 2001:48).

Veri saklama birimlerinde depolanan veri topluluklarına “dosya” adı verilmektedir. Karmaşık dosya yapıları, çok sayıda dosya arası ilişki ve kullanıcıların dosyaya erişimi söz konusu olduğunda geleneksel dosya sisteminin yetersiz olduğu görülmüştür. Bu sorunu çözmek üzere veriyi saklama ve erişim konusunda yeni yazılım teknolojilerine yönelme başlamış ve Veri Tabanı Yönetim Sistemleri (VTYS) yaklaşımı ortaya çıkmıştır. VTYS yaklaşımında veri girişi ve depolanması, veriye erişen uygulama programlarından bağımsızdır. Klasik dosya sisteminde ise, kayıt desenleri ve dosya yapılarında ortaya çıkabilecek en ufak değişiklik bile uygulama programlarının değişmesine ve yeniden derlenmesine neden olmaktadır. Veri tabanı sistemleri, birbirleriyle ilişkili veri ve programlar topluluğundan oluşmaktadır. Veri topluluğu bir “veri tabanı” olarak değerlendirilir (Özkan, 2006:4).

Veri tabanı, sistematik erişim imkanı olan, yönetilebilir, güncellenebilir, taşınabilir, birbirleri arasında tanımlı ilişkiler bulunabilen bilgiler kümesidir (Usgurlu, 2007:2). Veri tabanında bir kuruluşa ait bilgiler düzenli bir şekilde tutulmakta ve çeşitli yazılımlar aracılığıyla yönetilmektedir.

Verinin bilgiye dönüşümünün geleneksel yöntemi, klasik çözümleme ve yorumlamaya dayanmaktadır. Örneğin, sağlık sektöründe uzmanlar sağlık verilerindeki değişiklikleri ve güncel eğilimleri düzenli olarak analiz etmekte ve rapor ayrıntılarını saklamaktadırlar. Çünkü bu raporlar gelecekteki karar verme ve planlama süreçlerinin temelini oluşturmaktadır. Farklı uygulamalarda da gök bilimciler, gezegenler ve gök cisimlerinin uzaktan algılama görüntüleri yoluyla gök cisimlerini ve gezegenleri kategorilere ayırmaktadırlar. Bilimsel olarak, pazarlama, finans, sağlık, perakende ya da diğer alanlarda da veri analizi için klasik yaklaşımlar ayrıntılı bir ya da daha çok analize dayanmaktadır. Bu uygulamalar için veri setlerinin klasik araştırma şekli yavaş, pahalı ve subjektif olmaktadır. Veri tabanı sistemlerinin artan kullanımı ve hacimlerindeki bu olağanüstü artış, organizasyonların elde toplanan bu verilerinin pek çok alanda uygulanamaz olmasına sebep olmaktadır. Geleneksel sorgu veya raporlama araçlarının veri yığınları karşısında yetersiz kalması, Veri Tabanlarında Bilgi Keşfi – VTBK (Knowledge Discovery In Databases) adı altında sürekli ve yeni arayışlara neden olmaktadır (Fayyad vd., 1996:37).

VM ve bilgi keşfi alanındaki dikkate değer faktörler, veri toplamadaki artış patlaması, veri ambarlarındaki verinin depolanması, ağ ve intranetlerden veriye erişimin artması, küresel ekonomideki pazar payını arttıran rekabet baskısı, elektronik ticaret, VM yazılımlarının gelişimi, bilgisayar gücü ve depolama kapasitesindeki muazzam artışlardır (ITRG, 2003:2).

1.1. Veri Tabanlarında Bilgi Keşfi Aşamaları

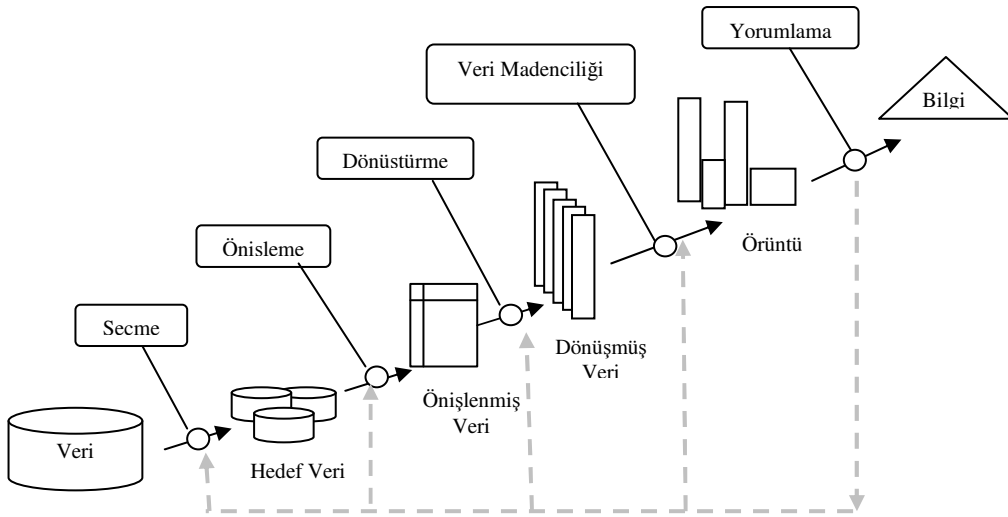
Küresel iş süreçlerinin bir parçası olan VM genellikle bilgi keşfi işlemi olarak da bilinmektedir.

Fayyad vd. (1996), VTBK aşamalarını, veriyi anlama, hedef veri seti oluşturma, veri temizleme ve önileme, veri indirgeme, amaçları seçme, model analizi ve hipotez seçme, VM, örüntüleri yorumlama, bilgi üzerinde harekete geçme olarak 9 aşamada incelemektedir. Veri madenciliği bu aşamaların odak noktasını oluşturmaktadır ve model gösterimi, model değerlendirme ve araştırma olarak 3 aşamada incelenmiştir. Fu (1997)'ya göre VTBK aşamaları, veri temizleme, veri dönüştürme,

VM, modelleme, yorumlama ve değerlendirme aşamalarından oluşmaktadır. VM ise VTBK'nın anahtar konusudur. Fayyad (1997), VTBK'nın veri hazırlama, örüntü arama, veri analizi, bilgi değerlendirme, yenileme gibi yinelemeli adımlardan oluştuğunu belirtmiştir. VM'yi ise bilgi ya da örüntülerin algoritmik anlamıyla ilgilenen VTBK aşaması olarak değerlendirmiştir. Hegland (2001), VTBK'yı veri sorgulama, veri temizleme, veri analizi, bilginin sunumunu içeren prosedürler olarak ele almıştır.

Bu tezde, VTBK aşamaları için, Şekil 1.1: Veri Tabanlarında Bilgi Keşfi Süreci'nde gösterildiği gibi veri seçimi, veri önileme, veri dönüştürme, VM, yorumlama/değerlendirme olarak altı aşama benimsenmiştir.

Şekil 1.1: Veri Tabanlarında Bilgi Keşfi Süreci



Kaynak: Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Symth, (1996); "From Data Mining to Knowledge Discovery in Databases," *AI Magazine*, Cilt 17, Sayı 3, s. 41.

(1) Veri Seçimi

Veri seçme işlemi, üzerinde çalışılacak veritabanından veya diğer veri kaynaklarından verilerin seçilerek veri dosyası oluşturulması işlemidir.

Veri seçimi, VM için kullanılabilecek verinin değerini belirleme işi ile ilgilenir. Pek çok durumda işlemsel veri tabanları, veri depolama ya da sorgulama işlemi için yeterli olmamıştır. Ancak gün geçtikçe iş uygulamalarının hepsini desteklemiştir (Kleissner, 1998:297).

Bu nedenle, bilgi teknolojisi organizasyonları stratejik karar destek aktivitelerini desteklemek için veri ambarları uygulamalarına yönelmişlerdir.

(2) Veri Önışleme

Veri kalitesi, VM’de anahtar bir konudur. VM’de güvenilirliğin artırılması için, veri ön işleme yapılmalıdır. Aksi halde hatalı kullanılan veriler hatalı sonuçlara sebep olmaktadır. Veri ön işleme, çoğu durumda yarı otomatik olan ve zaman isteyen bir VM aşamasıdır. Veri sayısındaki artış ve buna bağlı olarak çok sayıda verinin ön işlemeden geçirilmesi, veri ön işleme için gerekli etkin teknikleri önemli hale getirmiştir (Oğuzlar, 2003:70).

Veri ön işleme aşağıdaki sebeplerden dolayı verilere uygulanmaktadır (Oğuzlar, 2003:70):

1. Verilere uygulanacak tekniğin varsayımlarını sağlamak.
2. Verilerin doğasının anlaşılması ve anlamlı veri analizinin başarılması.
3. Verilen bir veri kümesinden daha sağlıklı bilginin üretilmesi.

Önişleme, uç (extreme), sapan (outlier), eksik (missing) veya hatalı birim değerlerinin düzeltilmesi gibi işlemleri içerir.

Bazı veri temizleme işlemleri, VM süreci başlamadan ya işlemsel veri tabanlarındaki kaynak veri üzerinde ya da veri ambarı yönetiminin bir kısmında süren uygulamalarda yapılmaktadır. Diğer veri temizleme uygulamaları ise veri zenginleştirme ve kodlama aşamasının bir parçası olarak veri fazlalığı fark edildikten sonra başlatılabilir (Kleissner, 1998:297).

Çok sayıda veri ön işleme tekniği mevcuttur. Bunlardan biri olan veri temizleme (data cleaning) verilerdeki gürültünün giderilmesi ve tutarsızlıkların düzeltilmesi için uygulanır. Bir diğer teknik olan veri birleştirme (data integration) ise farklı kaynaklı verileri uygun bir veri tabanında birleştirir. Normalleştirme gibi veri dönüştürme yöntemleri (data transformations) uygulanabilir Veri indirgemedede (data reduction) ise fazla olan bazı değişkenlerin atılması ve birleştirilmesi veya

kümeleme yolu ile veri büyüklüğünün azaltılması amaçlanır. Bu sözü edilen veri ön işleme teknikleri, VM'den önce uygulanarak elde edilen sonuçların kalitesi ve/veya VM için harcanacak zaman azaltılmış olur (Han and Kamber, 2001:105).

(3) Veri Dönüştürme

Bilgi keşfi sürecinin bu aşaması veri zenginleştirme aşaması olarak da adlandırılmaktadır. Veri ambarlarında ya da işlemsel veri tabanlarında mevcut olmayan ek veri, anlamlı VM çalışmalarını yönetmek için önemlidir.

Bir veri ambarındaki veri, kodlama aşamasındaki birtakım dönüşümlerden yararlanmaktadır. Bu aşama, yeterince anlamlı olmayan bazı kayıtların silinmesi, bölgesel kodların adreslere dönüştürülmesi, numaraların satın alma tarihlerine dönüştürülmesi, yaş sınıflarının doğum tarihlerine dönüştürülmesi gibi bilgileri dönüştürmeyi ve kaydetmeyi içermektedir. Bu dönüşümler, VM ya da VM'nin bölümleri olarak bilinen diğer aşamalara hazırlık için en iyi şekilde yapılmalıdır. Veri zenginleştirme ve kodlama aşamasının sonunda veri, VM uygulaması için hazır hale gelir (Kleissner, 1998:297).

Veri dönüştürme; düzeltme, birleştirme, genelleştirme ve normalleştirme gibi değişik işlemlerden birini veya bir kaçını içerebilir. Veri normalleştirme en sık kullanılan veri dönüştürme işlemlerinden birisidir (Roiger and Geatz, 2003:156).

(4) Veri Madenciliği

Bilgi keşfi sürecinin son aşaması olan VM, veriyi özetlemek ve gözlemlenemeyen ilişkileri bulmak için incelenen veri gruplarının analizidir (Hand vd., 2001:3). Veri yığınları içerisinde, gelecek ile ilgili tahminde bulunulmasını sağlayabilecek örüntülerin bilgisayar programı kullanılarak aranmasıdır (İlhan, 2005:2). Madencilik teriminin kullanılma sebebi, büyük veri yığını arasından uygun olanı arama ve seçme işleminin maden arama işine benzetilmesindendir (İnan, 2003:5).

VM araçları, veri tabanları ya da düz dosyalarda depolanan verileri desteklemesine karşın veri birleştirildikten, temizlendikten ve VM için hazırlandıktan

sonra veri pazarları ve veri ambarlarında toplanan veri üzerinde en etkili şekilde kullanılmaktadır (Kleissner, 1998:297).

1.2. Veri Ambarları (Data Warehouses)

Veri ambarı, işletmelerin işlemsel verilerinin belirli bir sistem ve metodoloji altında toplanarak konularına göre, sorgulama, analiz, tahmin gibi karar destek sistemleri özelliklerinin uygulandığı bir süreçtir.

Veri ambarı, iş dünyasında bilgiye hemen ulaşmak amacıyla karar vericiler için tasarlanmış bir bilgisayar sistemidir. Burada karar vericiden kasıt, ihtiyaç duyulan bilgiyi kullanabilen yöneticilerdir. Ambar, verilerini var olan sistemlerden (genel hesap defterleri, insan kaynakları, sipariş girdileri) alarak bu verileri yöneticilerin kullanması için saklamaktadır. Depolanmış bilgileri programcılardan daha çok yöneticiler kullanılmaktadır. Veri ambarı kullanıcıları, özel yazılımlar kullanmaktadırlar. Bu yazılımlar kullanıcının bilgiye ihtiyacı olduğu zaman, bu bilgileri meydana getirmesini ve ulaşmasını sağlamaktadır (Kuzey, 2000:3).

Veri ambarcılığı, organizasyonların teknolojilerini geliştirmek ve düşük maliyet sağlamak amacıyla karar destek amaçlı nitelikli verileri yaratmak, kullanmak ve sürdürmek için yapılan işlem sürecidir (Haley, 1999:2).

İş organizasyonlarında bilgi akış mimarisinde veri ambarları iki amaçla oluşturulmaktadır (Usgurlu, 2007:4):

1. Hareketsel ve organizasyonel görevler arasındaki depo ve analitik stratejik verilerin birikimini sağlar. Bu veriler daha sonra yeniden kullanılmak üzere arşivlenir. Veri ambarları verilerin sorgulanabildiği ve analiz yapılabilirdiği bir depodur.
2. Veri ambarlarının pazarda yeni fırsatlar bulmaya, rekabete katkı, yoğun proje çevirimi, iş, envanter, ürün maliyetlerinin azalmasının yanında farklı işlere ait verilerin ilişkilendirilmesi, karar destek ve alınan bilgiye hızlı cevap verebilme gibi katkıları vardır.

VM'yi istatistik yöntemlere üstün kılan özelliği, çok fazla miktarda veriyle çalışılabilir olmasıdır. İstatistikte, anakütleden seçilen bir örneklem üzerinde çalışarak genelleştirme yapılmaya çalışılır. Fakat bu durumun gelecekteki işletme ihtiyaçlarını tam olarak karşılayamama, iş çevresindeki gelişmelere ve değişimlere cevap verememe gibi eksik yönleri vardır. Bazen VM teknikleriyle daha basit ve ayrıntılı ve uygulanabilir kararlar alınabilmektedir. Tüm veriyle çalışan VM'nde, bütün veriyi sağlayan organ veri ambarıdır. Bu amaçla VM'nin, veri ambarına görüldüğünden daha fazla ihtiyacı vardır. Ayrıca veri ambarı çok basit ve cevabı kullanışlı olacak soruların cevabını hızlı bir şekilde alarak VM'nin işini kolaylaştırmaktadır (Biçen, 2002:24).

1.3. Veri Madenciliği

1.3.1. Veri Madenciliği Tanımı ve Amaçları

Dijital veri elde etme ve depolama teknolojisindeki gelişmeler, çok büyük miktarlardaki veri tabanlarının artmasıyla sonuçlanmıştır. Bu artış süpermarkete ait veriler, kredi kartı kullanma kayıtları, telefon arama detayları ve devlet istatistikleri gibi günlük, olağan olaylardan astronomik miktarlardaki görüntüler, moleküler veri tabanları ve tıbbi kayıtlar gibi farklı olaylara kadar insan çabalarının her alanında meydana gelmiştir (Hand vd., 2001:3).

VM genel anlamda; büyük miktarda veri içerisinde, gizli kalmış, değerli, kullanılabilir bilgilerin açığa çıkarılması biçiminde ifade edilmektedir. Bu ifade irdelendiğinde yakın zamana kadar pek duyulmamış bir terim olan VM'nin, aslında uzun zamandır gerçekleştirilmesine karşın evrim geçirerek 1990'lardan itibaren isimlendirildiği görülmektedir. Öyle ki, bu konu en iyi biçimde 'istatistikçiler, VM'ni elle yaparlar,' ifadesiyle özetlenebilir. Bu ifade her ne kadar VM'nin temel mantığının klasik istatistik teknikler olduğunun altını çizse de, herhangi bir istatistik yazılımı veya yöntemini VM olarak ifade etmek doğru olmamaktadır. Bununla beraber VM'yi, istatistik tekniklerin evrim geçirmiş hali olarak düşünmenin de mümkün olmasının yanı sıra; VM'nin, kendisine özgü tanımlayıcı özellikleri vardır (Koyuncugil ve Özgülbaş, 2006:4).

VM, bir veri kümesi içerisinde keşfedilmemiş örüntüleri bulmayı hedefleyen teknikler bütünüdür. VM'nin amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar verme modelleri yaratmaktır (Koyuncugil, 2007:7).

VM'ndeki amaç, toplanmış olan bilgilerin bir takım istatistik yöntemlerle incelenip ilgili kurum ve yönetim destek sistemlerinde kullanılmak üzere değerlendirilmesidir. Veri madencisinin geleneksel yöntemlerde olduğunun aksine başlangıçta herhangi bir amacı ya da varmak istediği bir kavram yoktur. Yapılacak analizlerden sonra elde edilen verilerin bir istatistikçi gözü ile incelenip daha önceden düşünülmemiş kavramların ortaya çıkarılması, başarılı bir VM süreci olarak kabul edilmektedir (Vahaplar, 2003:4).

VM, birleşik verilerdeki gizli bilgileri bulmak ve iş uzmanlığını arttırmak amacıyla yapılan yeni bir karar destek analiz işlemidir. Bazı anahtar kelimeler kullanılarak 4 aşamalı ayrıntılı VM tanımı şöyledir (Kleissner, 1998:296):

1. VM, bir süreçtir.

VM, veri çözümleme için yapılan tek zamanlı aktivite değildir. Çözümlemenin devamlılığı ve yeni aşamalardaki uygulamalar üzerinde sürekli ve tekrarlı olarak yapılan bir süreçtir.

2. VM, karar destek araçlarının niteliğini yükseltir.

Günümüz karar destek araçları, veri hakkındaki hipotezleri doğrulamak için “varsayım-yol gösterme” araçlarıdır. VM araçları “keşif-yol gösterme” araçlarıdır ve “varsayım-yol gösterme” araçlarının niteliğini yükseltir. VM araçları hipotez üreticilerdir. Bu araçlar, iş deneyimlerinde yeni anlayışları keşfetmek için veri yığınlarını çözümlerler.

3. VM, gizlenmiş bilgileri bulur.

VM, yeni algoritmalar ve metodolojiler kullanarak gizli bilgilerden örüntüler bulur. Bu süreç içerisinde zaman zaman aşırı yüklemeler de olur. Deneyimler

göstermektedir ki VM, iş uygulamaları içine ilave edilen anlayışları sağlayan gizli bilgileri bulmada çok etkili olmuştur.

4. VM, iş uzmanları için kavrayış dağıtıcı bir sistemdir.

VM araçlarının yönü, zamanla, istatistikçiler ve yapay zeka uzmanları yerine pazarlama uzmanlarına, iş analistlerine ve iş uzmanlarına doğru değişmiştir. VM araçları istatistik tekniklerin yerine geçen bir araç değildir. Ancak yapay zekanın karmaşıklıklarını ya da istatistik analiz detaylarını içermeyen iş verilerini daha iyi anlamak için iş uzmanlarını güçlendirmeye yöneliktir.

VM, disiplinler arası bir uygulamadır. İstatistik, veri tabanı teknolojisi, makine öğrenmesi, örüntü tanımlama, yapay zeka gibi alanların hepsinde bir etkisi vardır (Bkz: Şekil 1.2: Veri Madenciliği İlgili Alanları). Bu disiplinler arası kesin sınırları tanımlamak zordur, bu nedenle VM ve bu disiplinlerin her biri arasındaki kesin sınırları tanımlamak da zordur. Bir insana göre VM olan tanımlar bir diğeri için istatistik, veri tabanı ya da makine öğrenmesi problemi olabilmektedir (Hand vd., 2001:3).

Şekil 1.2: Veri Madenciliği İlgili Alanları



Kaynak: Adil Baykasoğlu (2005); “Veri Madenciliği ve Çimento Sektöründe Bir Uygulama,” Akademik Bilişim 2005, 02-04 Şubat 2005, Gaziantep, s. 2.

VM çeşitli endüstrilerdeki organizasyonlar tarafından kullanılmaktadır. Bu tip analizler için temel ihtiyaç müşteri verisi, uygun hesaplama gücü ve uygun yazılım algoritmasıdır. Bir organizasyon, müşteri veritabanlarıyla, karar vermede önemli müşteri bilgileri için VM'den bazı faydalar türetebilmektedir. VM'nin kullanımı bir örnekle şöyle açıklanabilir. Bir finansal kuruluş müşterileri üzerinde hesap numarası, hesap tipi, hesap bakiyesi, hesabını tuttuğu zaman uzunluğu, coğrafi konumu gibi pek çok değişken söz konusudur. Müşterilerin hesap kapatma oranlarının tahmin edilmeyecek oranda fazla olduğu düşünülürse VM, bu problemi çözmek için faydalı olabilmektedir. Proje, benzer mülkiyet, benzer hesap türleri, benzer bakiyelere sahip olanlar arasındaki ortak bilgilere karar verecektir. Böyle örüntüler ortaya çıkarılırsa adımlar hesap kapatma oranını azaltmak için kullanılacaktır (ITRG, 2003:2).

VM uygulamalarının etkili olabilmesi için gerçekleştirmek zorunda olduğu 3 koşul vardır. Birincisi, VM bölüme özgü veri yerine organizasyon çapında veriye ulaşmalıdır. Organizasyonun verisi, sık sık açık kaynaklı ya da maliyetli verilere eklenmektedir. Bu şekilde oluşturulan veri tabanı veri ambarı olarak adlandırılmaktadır. Veri entegrasyonu süresince uygulama, verileri, yeni oturmuş, türetilmiş özellikleri (çift tanımlamaları kaldırarak, eksik değerleri doldurarak) temizlemektedir. İkinci olarak, bir VM uygulaması veri ambarlarındaki bilgiyi işlemek zorundadır. Son aşamada ise VM işlenmiş veriyi karar vermeye imkan verecek şekilde düzenlemeli ve sunmalıdır (Simoudis, 1996:26).

1.3.2. Veri Madenciliği Süreci

VM altı adımlı bir süreç olarak incelenebilir (Larose, 2005:7).

(1) Araştırma Probleminin Tanımlanması (Business Understanding)

Bu aşama veri madenciliği sürecinin en önemli aşamasıdır. Araştırma probleminin (konusunun) tanımlanması aşaması araştırmanın amacını, mevcut durumun değerlendirilmesini, veri madenciliğinin amaçlarını ve proje planlama sürecinin belirlenmesini kapsamaktadır.

(2) Verileri Tanıma Aşaması (Data Understanding)

Veri anlama aşaması veri toplamakla başlamaktadır. Daha sonra benzer verileri bir araya getirme, veri niteliklerini tanımlama, verileri keşfetme, gizli bilgileri sınıflandırma ile sürece devam etmektedir. Diğer bir anlatımla bu aşama ilk verilerin toplanması, verinin tanımlanması, verilerin keşfedilmesi ve verilerin kalitesinin incelenmesi faaliyetlerini içerir.

(3) Veri Hazırlama Aşaması (Data Preperation)

Veri hazırlama aşaması, ham veriden başlayarak son veriye kadar yapılması gereken bütün düzenlemeleri içermektedir. Veri hazırlama, görevleri çok zamanlı olarak yerine getirmekle gerçekleştirebilmektedir. Görevler tablo, kayıt, veri dönüşümü ve modelleme araçları için veri temizleme gibi özellikleri içermektedir.

(4) Modelleme Aşaması (Modelling)

Bu aşamada, verilerden bilgi çekmek için ileri çözümleme yöntemleri kullanıldığından VM sürecinin en gösterişli aşamasıdır. Bu aşama uygun modelleme tekniğinin seçimi, test tasarımının üretimi, model geliştirme ve tahmin işlemlerini içermektedir. Uygun modellerin seçilip uygulanmasıyla birlikte parametreler en uygun değişkenlere dönüştürülmektedir. VM, her problem tipi için farklı yöntemler içermektedir. Bazı yöntemler, veri tipi için uygun değildir ya da özel tanımlamalar gerektirmektedir. Bu nedenle gerekli olduğunda 3. aşama olan veri hazırlama aşamasına geri dönülür.

(5) Değerlendirme Aşaması (Evaluation)

Değerlendirme aşamasında, uygun model ya da modeller kurulduktan sonra, VM sonuçlarının araştırma probleminin amaçlarını gerçekleştirip gerçekleştirmediği değerlendirilir. Bu aşama sonuçların değerlendirilmesi, veri madenciliği sürecinin gözden geçirilmesi ve sonraki adımların ne olacağı hususlarını içermektedir. Bu aşamanın sonunda VM sonuçlarının kullanımı üzerindeki karara varılmaktadır.

(6) Uygulama Aşaması (Deployment)

Son aşama olan uygulama aşaması, araştırmacının tüm emeklerinin karşılığını aldığı bir aşamadır. Bu aşamada VM süreciyle üretilen bilgiler, pratik işletme problemlerinin çözümünde kullanılmaktadır. Bu aşamada elde edilen bilgilerin uygulanabilmesine yönelik bir plan hazırlama, gözden geçirme ve bakım faaliyetlerini içerir. Ayrıca bu aşamada nihai araştırma raporunun yazılması ve projenin gözden geçirilmesi işlemleri yer almaktadır.

VM’de veri kümesinin büyüklüğünden kaynaklanan en fazla zaman alıcı aşama, verilerin ön işlemde geçirilmesi aşamasıdır. VM uygulamalarında kaynakların %80’i verilerin ön işlemde geçirilmesi ve temizlenmesi süreçleri için harcanmaktadır (Piramuthu, 2004:483).

1.3.3. Veri Madenciliği İşlevleri

Genel olarak VM işlevleri ikiye ayrılabilir:

(1) Tahmin Edici Veri Madenciliği

Tahmin edici model (predictive model), açık bir şekilde önceden bilinen değerlerin kullanıldığı bir model geliştirmek için bilinen sonuçları kullanmaktadır (Edelstein, 2000:2).

Tahmin edici VM bir model ya da modeller kümesi oluşturup mevcut veri seti üzerinde sonuçlar sergiler ve yeni veri kümelerinin davranışlarının nasıl olacağını tahmin etmeye çalışır (Vahaplar, 2003:11). Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır (Özekes, 2003:67). Örneğin, bir banka vermiş olduğu kredilere ilişkin verilere sahip olabilir. Bu veri setinde bağımsız değişkenler kredi alan müşterinin özellikleri, bağımlı değişken ise kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceği tahmininde kullanılmaktadır (Akpınar, 2000:3).

(2) Tanımlayıcı Veri Madenciliği

Tanımlayıcı VM modeli (descriptive) tahmin edici modelin tersine karar vericilere yol göstermek için kullanılan mevcut verideki örüntüleri tanımlamaktadır (Edelstein, 2000:2).

Tanımlayıcı VM, veri setini kısa ve özet biçimde tanımlamakta ve verinin çarpıcı özelliklerini göstermektedir (Vahaplar, 2003:11). Tanımlayıcı modellerde, politika geliştirme ve karar verme aşamalarında kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır (Özekes, 2003:67). X-Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu aileler ile çocuğu olmayan ve geliri X-Y aralığından az olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere örnektir (Akpınar, 2000:3).

1.4. Veri Madenciliği Yöntem ve Teknikleri

Bir VM modeliyle aşağıdaki işlemlerden bir veya birkaçı gerçekleştirilebilir:

Sınıflama (Classification) ve Regresyon (Regression) Modelleri,

Kümeleme (Clustering) Modelleri ve

Birliktelik Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns).

Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir.

1.4.1. Sınıflama ve Regresyon Modelleri

En yaygın uygulanan VM tekniklerinden biri olan sınıflama, büyük veri kümelerini sınıflandıran model oluşturmak için önceden sınıflandırılmış örnekleri kullanır. Özellikle hile tespiti ve kredi riski uygulamaları bu tip analizlere en uygun uygulamalardır. Bu yöntem yaygın olarak karar ağaçları ve nöron ağlarını esas alan algoritmaları kullanır.

Başlıca sınıflandırma teknikleri aşağıdaki gibidir:

- Yapay Sinir Ağları (Artificial Neural Networks)
- Genetik Algoritmalar (Genetic Algorithms)
- K- En Yakın Komşu (K-Nearest Neighbour)
- Bellek Temelli Nedenleme (Memory Based Reasoning)
- Naive - Bayes
- Lojistik Regresyon (Logistic Regression)
- Karar Ağaçları (Decision Trees)

1.4.2. Kümeleme

Kümeleme analizi, nesnelerin alt dizinlere gruplanmasını yapan bir işlemdir. Böylece nesneler, örneklenen kitle özelliklerini iyi yansıtan etkili bir temsil gücüne sahip olmaktadır. Sınıflamanın aksine, yeniden tanımlanmış sınıflara dayalı değildir. Kümeleme, bir denetimsiz öğrenme (unsupervised learning) yöntemidir (Aydoğan, 2003:88) .

Amaç, elemanların birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere (gruplara) bölünmesidir (Durmuş ve İplikçi, 2007).

Kümeleme algoritması genelde astronomi, nüfus bilimi, bankacılık vb. uygulamalarda kullanılmaktadır (Aydoğan, 2003:18).

1.4.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler

Birliktelik kuralları ile bir ilişkide yer alan niteliklerin değerleri arasındaki bağımlılıklar, anahtarda yer almayan diğer niteliklerin gruplandırılması ile bulunur. Bu kurallar ilk olarak Agrawal tarafından 1994'te geliştirilmiştir (Altınışik, 2006:17).

Çok sayıda verinin depolandığı bir veri tabanı içinde çeşitli nitelikler arasında hemen fark edilmeyen birtakım ilişkiler mevcut olabilir. Bu tip ilişkilerin ortaya çıkartılması stratejik kararların alınmasına yardımcı olabilir. Ancak, bu ilişkilerin çok sayıda verinin içinden elde edilmesi basit bir süreç değildir. Bu süreç birliktelik kuralı madenciliği (association rule mining) olarak adlandırılmaktadır. Veriler arasındaki ilişkiler, eğer-sonra ifadeleri ile aşağıdaki gibi gösterilmektedir (Ghosh ve Nath, 204:124):

Eğer <bazı şartlar sağlanırsa> sonra <bazı niteliklerin değerlerini tahmin et>

Birliktelik kurallarının analizi süreci market sepeti analizi olarak da adlandırılır. Market sepeti analizinde müşteri ile ilgili veri hareketlerinden gelecekte müşterinin nasıl bir tercih yapacağına dair sonuçlar tahmin edilmektedir (Altınışik, 2006:17).

“Çocuk maması alanların %40’ı makarna da satın alır” örneğinde bir bağıntı bulunmaktadır. Çocuk maması alanların çocuk bezi, makarna alanların ketçap alacağını tahmin etmek kolay olmaktadır. Ancak örnek incelendiğinde, sonucu çıkarmak için bütün olasılıkları göz önüne alarak kolayca aklımıza gelmeyen ürün birliktelikleri ortaya çıkartılmaktadır (Altınışik, 2006:12).

Zaman içinde sıralı örüntüler örneği “ilk üç taksitinden iki veya daha fazlasını geç ödemiş olan müşteriler %60 olasılıkla kanuni takibe gidiyor” cümlesiyle verilebilir. Davranış skoru (behavioral score), başvuru skorundan farklı olarak kredi almış ve taksitleri ödeyen bir kişinin sonraki taksitlerini ödeme/geciktirme davranışını notlamayı amaçlar (Alpaydın, 2007:3).

“X şirketinin hisse fiyatları ile Y şirketinin hisse fiyatları benzer hareket ediyor” örneğinde ise benzer zaman sıraları görülmektedir. Amaç zaman içindeki iki hareket serisi arasında bağıntı kurmaktır. Bunlar iki malın zaman içindeki satış miktarları olabilir. Örneğin dondurma satışları ile kola satışları arasında pozitif, dondurma satışları ile sahlep satışları arasında negatif bir bağıntı beklenebilir (Alpaydın, 2007:3).

En yaygın birliktelik kuralı algoritmaları arasında GRI (The Generalized Rule Induction), Apriori sayılabilir.

1.5. Veri Madenciliği Uygulama Alanları

VM her geçen gün yeni ve farklı alanlarda kullanılmaya başlamakla birlikte günümüzde daha yaygın olarak kullanıldığı alanlar şöyledir:

Pazarlama: VM pazarlama alanında, müşterilerin satın alma örüntülerinin belirlenmesi, müşterilerin demografik özellikleri arasındaki ilişkilerin bulunması, posta kampanyalarında cevap verme oranının artırılması, mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması konularında kullanılmaktadır. Pazar sepeti analizi (market basket analysis), müşteri ilişkileri yönetimi (customer relationship management), müşteri değerlendirme (customer value analysis), satış tahmin (sales forecasting) konularında da kullanılmaktadır (Akpınar, 2000:3).

Banka ve Sigortacılık: Bankacılık sektöründe farklı finansal göstergeler arasında gizli korelasyonların bulunmasında kullanılmaktadır (Akpınar, 2000:3). Kredi ve kredi kartı sahtekarlığı tahminlerinde, risk değerlendirmede, müşteri eğilim analizlerinde, kar analizi gibi alanlarda VM kullanılmaktadır. Finans ve borsa kuruluşları ise stok fiyat tahminlerinde, gümrük ölçümleme, portföy yönetimi gibi alanlarda VM yöntemlerini kullanabilirler (Baykasoğlu, 2004:5). Ekonomik ve finansal yatırımları kararlaştırma, iflas ya da başarısızlık tahmini gibi alanlarda VM yaygın olarak uygulanmaktadır (Toktaş ve Demirhan, 2004:3).

Biyoloji, Tıp ve Genetik: Bitki türlerinin ıslahı, gen haritasının analizi, genetik hastalıkların tespiti, fizyolojik parametrelerin analizi ve değerlendirilmesinde kullanılmaktadır (İlhan, 2005:5).

Vysis firması ilaç sanayisi için protein analizini yapay sinir ağı (YSA) algoritmasını kullanarak yapmaktadır. Rochester Kanseri Merkezi bölümü araştırmalarında bir karar tekniğini kullanmaktadır. California Hastanesi veriden bilgi üretmek için “Information Discovery” adlı ürünü kullanmaktadır (DEU, 2004).

Sağlık alanında yapılan birçok VM araştırmalarında hastaların elektronik tıbbi kayıtları ve idari işleri belgeleyen veriler kullanılmaktadır. Bu verilerden yararlanılarak farklı tahminler yapılabilir. Örneğin bunlardan bazıları, belirli bir hastalığa sahip kişilerin ortak özelliklerinin tahmin edilmesi, tıbbi tedaviden sonra hastaların durumlarının tahmin edilmesi, hastane maliyetlerinin tahmin edilmesi, ölüm oranları ve salgın hastalıkların tahmin edilmesi gibi uygulamalardır (Yıldırım vd., 2007:3).

Yüzey Analizi ve Coğrafi Bilgi Sistemleri: VM, coğrafi bilgi sistemleri, belediyeçilik, mühendislik, ziraat-ormancılık, çevre uygulamalarında kullanılmaktadır.

CBS'nin işletme uygulamalarında yardımcı olabileceği konular, yer seçimi, alan planlaması, kaynak dağıtımı ve planlaması, proje kontrolü ve benzeri konulardır. CBS sadece haritalar ve kayıtlardan ibaret değildir. Aynı zamanda CBS yatırım kararlarına, müşteri hizmetlerini iyileştirilmesi çalışmalarına, satış artırma ve pazarlama çalışmalarına, müşteri yönelimlerine, rakipler hakkında bilgi edinme çalışmalarına ve diğer stratejik konulara yardımcı olan güçlü bir araçtır (Işık, 2003:36).

Bilim, Mühendislik Verileri: Günümüzde laboratuvar veya bilgisayar ortamında sistemlerin benzetimi ve analizi sürecinde yüksek miktarda bilimsel veri üretilmektedir. Elde edilen bu verilerin anlamlandırılması için VM çok uygun bir ortam sağlamaktadır. Çimento deneylerinden elde edilen verilerden mukavemet analizi, üretim sistemlerinin benzetiminden elde edilen verilerden sistem performansını etkileyen faktörlerin ve kuralların çıkarılması, deprem verilerinin analizi ile deprem ve etkilerinin tahmini, kalite kontrol uygulamaları gibi pek çok uygulama örnek olarak verilebilir (Baykasoğlu, 2005:5).

Eğitim Sektörü Verileri: Öğrenci işlerinde veriler analiz edilerek öğrencilerin başarı ve başarısızlık nedenleri, başarının artırılması için hangi konulara ağırlık verilmesi gerektiği, üniversite giriş puanları ile okul başarısı arasında bir ilişkinin var

olup olmadığı gibi soruların cevabı bulunarak eğitim kalitesi ve performansı artırılabilir (Baykasoğlu, 2005:5).

İnternet ve Ağ Verileri: İnternet ve ağ üzerindeki veriler hem hacim hem de karmaşıklık olarak hızla artmaktadır. Ağ madenciliği özetle internetten faydalı bilginin keşfi olarak tanımlanabilir. Ağ VM birçok web sunucusu veya çevrimiçi (online) servisten kullanıcı erişim desenlerinin analiz ve keşfi için kullanılır (Cooley vd., 2004:6).

Metin Madenciliği: Çok büyük ve anlamsız metin yığınları arasından anlamlı ilişkiler elde etmede kullanılmaktadır (İlhan, 2005:5). Ana amaç, dokümanlar arasında ayrıca elle bir tasnif gerekmeden benzerlik hesaplayabilmektir. Bu genelde otomatik olarak çıkarılan anahtar sözcüklerin tekrar sayısı sayesinde yapılmaktadır. Polis kayıtlarında mevcut rapora benzer kaç adet ve hangi raporların olduğu sorusuna yanıt bulunabilmektedir. Ürün tasarım dokümanları arasında mevcut tasarım için kullanılacak ne tür dosyaların olduğu sorusuna yanıt bulunabilmektedir (Baykasoğlu, 2005:5).

İş ve Elektronik Ticaret Verileri: İş süreçleri sırasında geniş çaplarda veri üretilmektedir. Bu veri, karar verme mekanizmalarında etkin olarak kullanılmaktadır. Personel veri tabanlarında yapılabilecek analizler ile performansa etki eden faktörlerin saptanması ve yeni personel seçiminde kullanılacak değişik kuralları türetebilmektedir. Müşteri veri tabanlarının analizi ile reklam ve promosyon ile ilgili pek çok faydalı bilgiye ulaşmak mümkündür (Baykasoğlu, 2005:5).

Kimya: Yeni kimyasal moleküllerin keşfi ve sınıflandırılması, yem ilaç türlerinin keşfinde kullanılmaktadır (İlhan, 2005:13).

Meteoroloji ve Atmosfer Bilimleri: Bölgesel iklim, yağış haritaları oluşturma, hava tahminleri, ozon tabakası deliklerinin tespiti, çeşitli okyanus hareketlerinin belirlenmesinde kullanılmaktadır (İlhan, 2005:13).

Uzay Bilimleri ve Teknolojisi: Gezegen yüzey şekilleri, gezegen yerleşimleri, yeni galaksilerin keşfi, yıldızların konumları gibi konuların gruplandırılmasında kullanılmaktadır (İlhan, 2005:13).

Görüntü Tanıma ve Robot Görüş Sistemleri: Çeşitli sensörler aracılığı ile tespit edilen görüntülerden yola çıkarak engel tanıma, yol tanıma, yüz tanıma, parmak izi tanıma gibi tekniklerde kullanılmaktadır (İlhan, 2005:13).

Sosyal Bilimler ve Davranış Bilimleri: Anketler düzenlenerek toplumun belirli bir konu hakkındaki eğilimleri belirlenerek anlamlı sonuçlar çıkarılmaktadır. Örnek olarak seçim anketleri verilebilir (Altınışik, 2006:6).

Bu bölümde veri tabanlarında bilgi keşfiyle (VTBK) ilgili temel kavramlar, VTBK'nın aşamaları, veri ambarları, veri madenciliğinin aşamaları, VM yöntem ve teknikleri ile VM'nin uygulama alanları hususlarına değinilmiştir. Bir sonraki bölümde genel anlamda veri madenciliğinde kullanılan yöntem ve teknikler kısaca tanıtıldıktan sonra, bu çalışmada kullanılan karar ağacı algoritmaları ayrıntılı bir biçimde açıklanmaktadır.

2. SINIFLANDIRMA VE KARAR AĞAÇLARI

Sınıflama modeli, sınıfı tanımlanmış mevcut verilerden yararlanarak sınıfı belli olmayan verilerin sınıfını tahmin etmek için kullanılan VM modelidir. Sınıflama iki adım içeren bir işlemdir. Birinci adımda tahmin için kullanılacak bir model oluşturulmaktadır. İkinci adımda, oluşturulan bu model sınıfı belli olmayan veriler üzerinde uygulanarak sınıflar tahmin edilmektedir (Han ve Kamber 2001:35).

Sınıflandırma ve Regresyon modellerine şu örnekler verilebilir. “Genç kadınlar küçük araba satın alır, yaşlı, zengin erkekler büyük, lüks araba satın alır” cümlesi bir sınıflandırma modelini açıklamaktadır. Amaç bir malın özellikleri ile müşteri özelliklerini eşleştirmektir. Böylece bir müşteri için ideal ürün veya bir ürün için ideal müşteri profili çıkarılabilir. Örneğin bir otomobil satıcısı şirketin geçmişteki müşteri hareketlerinin analizi ile yukarıdaki gibi iki kural gerçekleştirebilirse genç kadınların okuduğu bir dergiye reklam verirken küçük otomobil modelinin reklamını verebilmektedir (Alpaydın, 2007:4).

“Ev sahibi olan, evli, aynı iş yerinde beş yıldan fazladır çalışan, geçmiş kredilerinde geç ödemesi bir ayı geçmemiş bir erkeğin kredi skoru 825’dir” cümlesi bir regresyon örneğidir. Başvuru değerlemede bir finans kurumu kredi için başvuran kişi ile ilgili finansal güvenilirliğini 0 ile 1000 arasında bir değerle ölçebilir. Bu değer, kişinin özellikleri ve geçmiş kredi hareketlerine dayanılarak hesaplanır (Alpaydın, 2007:5).

Başlıca sınıflandırma teknikleri aşağıdaki gibidir:

- Yapay Sinir Ağları (Artificial Neural Networks)
- Genetik Algoritmalar (Genetic Algorithms)
- K- En Yakın Komşu (K – Nearest Neighbour)
- Bellek Temelli Nedenleme (Memory Based Reasoning)
- Naive - Bayes

- Lojistik Regresyon (Logistic Regression)
- Karar Ağaçları (Decision Trees)

2.1. Yapay Sinir Ağları (Artificial Neural Networks)

YSA, insan beyninin çalışma ilkelerinden ilham alınarak geliştirilmiş, ağırlıklı bağlantılar denilen tek yönlü iletişim kanalları vasıtası ile birbirleriyle haberleşen, her biri kendi hafızasına sahip birçok işlem elemanından (nöronlardan) oluşan paralel ve dağıtık bilgi işleme yapılarıdır (Erdem ve Uzun, 2005:13).

YSA, çok değişkenli ve değişkenler arasında karmaşık, karşılıklı etkileşimin bulunduğu veya tek bir çözüm kümesinin bulunmadığı durumlarda başarılı sonuçlar üreten bir yapay zeka teknolojisidir. Bu özellikleri nedeniyle yapay sinir ağı teknolojisi mali başarısızlık alanında kullanıma uygun bir araç olarak görülmektedir (Benli, 2005:34).

2.2. Genetik Algoritmalar (Genetic Algorithms)

Genetik algoritmalar, karmaşık matematiksel problemlerin çözümünde en sık kullanılan evrimsel algoritmalar olup, doğal evrim felsefesi üzerine kurulmuştur. Genetik algoritmalarındaki temel fikir, rekabet ve kontrollü değişim süreci içinde bilgi yapıları popülasyonunu muhafaza etmektir. Popülasyondaki her bir yapı somut problem için aday bir çözüm temsil eder ve uygunluk değerlerine göre rekabet sürecinde hangi yapıların yenilerini oluşturmak için yetkinliğinin olduğu belirlenir. Yeni yapılar, çaprazlama ve mutasyon gibi genetik operatörler kullanılarak oluşturulur. Yeni popülasyon oluşturma süreci, popülasyon evrim geçirene kadar devam eder. Genetik algoritmalar, optimizasyon problemlerinin yanı sıra sınıflama amacıyla kullanılmaktadır. Ayrıca VM’de, diğer algoritmaların uygunluğunu değerlendirmek amacıyla da kullanılabilirler (Toktaş ve Demirhan, 2004:5).

2.3. K- En Yakın Komşu (K-Nearest Neighbour)

Bu teknikte tüm örneklemeler bir örüntü uzayında saklanır. Algoritma, bilinmeyen bir örneklemin hangi sınıfa dahil olduğunu belirlemek için örüntü uzayını araştırarak bilinmeyen örnekleme en yakın olan k örneklemini bulur. Daha sonra,

bilinmeyen örneklem, k en yakın komşu içinden en çok benzediği sınıfa atanır. K -en yakın komşu algoritması, aynı zamanda, bilinmeyen örneklem için bir gerçek değerin tahmininde de kullanılabilir (Toktaş ve Demirhan, 2004:5).

2.4. Bellek Tabanlı Nedenleme (Memory Based Reasoning)

Bellek tabanlı yöntemler, denetimli öğrenmenin kullanıldığı VM tekniklerindendir. Bu tekniğin temel özelliği, daha önceki deneyimlerden faydalanarak mevcut problemlere benzer durumları tanımlayıp geçmiş benzer problemlere getirilen uygun çözümleri mevcut problemlere uygulamaya çalışmaktır.

Bellek tabanlı yönetim tekniğinin (BTY) performansını belirleyen iki fonksiyon vardır. Bunlar, uzaklık ve kombinasyon fonksiyonlarıdır. Uzaklık fonksiyonu iki kayıt arasındaki uzaklığın bulunmasına, kombinasyon fonksiyonu ise sonuçların anlamlı sonuç sunacak şekilde birleştirilmesine olanak sağlar. BTY tekniğinin sözü edilen fonksiyonları kullanmasının bir faydası her türlü veri tipi için geçerliliğinin olmasıdır. Bu tekniğin uygulamada çok avantajlı yanları olmasının (her türlü veriye uygulanabilirliği, adaptasyon kolaylığı) dışında, geçmiş tarihi verileri saklama maliyeti bu yöntemi oldukça pahalı bir teknik haline getirmektedir. Yeni kayıtların sınıflandırılması bu kayıtlara en yakın komşu kayıtların sistemde taranacağı anlamına geldiğinden, bu teknik YSA ya da karar ağaçları tekniklerinden çok daha fazla zaman alıcı olmaktadır (Biçen, 2002:30-31).

2.5. Naive – Bayes

Bir sınıflandırma sorununun olasılık terimleriyle açıklanabileceği varsayımına dayanır. Bayes kuralı, bir veri grubunda bir özelliğin olasılığını tahmin etme yöntemidir. Belirli bir veri değerinde çeşitli varsayımların olasılığını araştırır (Baykal, 2003).

2.6. Lojistik Regresyon (Logistic Regression)

Lojistik Regresyon Analizinin kullanım amacı, istatistikte kullanılan diğer model yapılandırma teknikleri ile aynıdır. En az değişkeni kullanarak en iyi uyuma sahip olacak şekilde bağımlı ile bağımsız değişkenler arasındaki ilişkiyi uygun olarak tanımlayabilen bir model kurmaktır. Lojistik regresyon modelleri, son yıllarda

biyoloji, tıp, ekonomi, tarım ve veterinerlik ve taşıma sahalarında yaygın olarak kullanılmaktadır (Bircan, 2004:186).

2.7. Karar Ağaçları (Decision Trees)

Karar ağaçları, sınıfları bilinen örnek veriden tümevarım yöntemiyle öğrenilen ağaç şekilli bir karar yapısı çeşididir (Sun ve Li, 2008:2).

Bir karar ağacı, basit karar verme adımları uygulanarak, büyük miktarlardaki kayıtları, çok küçük kayıt gruplarına bölerek kullanılan bir yapıdır. Her başarılı bölme işlemiyle, sonuç gruplarının üyeleri bir diğeriyle çok daha benzer hale gelmektedir (Berry, 2004:194).

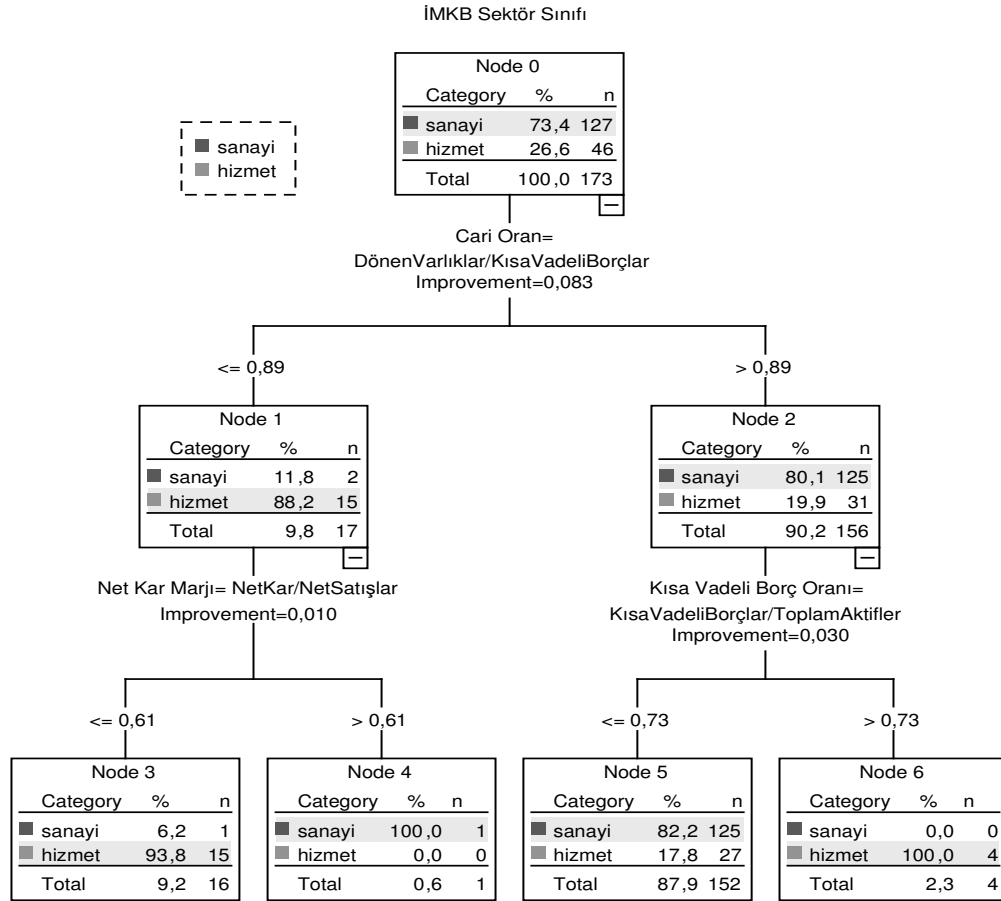
Karar ağaçları Bierman ve Friedman tarafından 1973 yılında önerilmiş olup değişkenleri parçalayarak bir ağaç oluşturmaya dayanmaktadır. Karar ağacında, tanımlanan sorunun cevabı gruplara ayrılmaktadır. Soruya verilecek bir ölçüt belirlendikten sonra setler arasındaki riski maksimize edecek şekilde cevaplar bölünmektedir. En iyi bölünmeyi bulmak için her soruda bu işlem tekrar edilmektedir. Bir soru için grup oluşturulduktan ve gruplar arasındaki risk maksimize edildikten sonra oluşan iki grup için aynı işlemler devam ettirilmektedir. Bu işlemler istatistik olarak anlamlı bir fark bulunana kadar devam ettirilip istatistik olarak anlamlı bir fark bulunmadığı durumlarda ise son verilmektedir. Ayırıştırma işlemi tamamlandıktan sonra ise o grup içerisinde yer alan gözlemlerin oranına göre grup değerlendirilmektedir (Thomas, 2000:159).

Büyük veri tabanlarının kullanıldığı pek çok sınıflama probleminde ve karmaşık ya da hata içeren bilgilerde karar ağaçları yararlı bir çözüm olmaktadır (Aitkenhead, 2008:18).

Karar ağacı, ağaç yapısında olan bir akış şeması şeklindedir. Düğümler üzerinde niteliklerin test işlemi yapılmaktadır. Test işleminin sonucuna göre ise dallanma meydana gelmektedir. Sonuç olarak ağaç sınıflar ile son bulmaktadır. Karar ağaçları genellikle banka kredisinin onaylanması, kredi kartı ve sigorta işlemlerinde

risk analizi yapmak gibi işlemler için kullanılmaktadır (Altınışik, 2006:13). Şekil 2.1’de bir karar ağacı örneği verilmektedir.

Şekil 2.1: Karar Ağacı Örneği



Bir karar ağacı aşağıdaki özelliklerle tanımlanabilir (Aitkenhead, 2008:19):

- Bir durumun spesifik nitelikleriyle ilgili Boolean (doğru/yanlış) sorularının yinelemeli olarak kullanımıdır.
- Nitelikler değerler (tamsayı, ondalık sayı vb.) ya da özellikler (açık/kapalı, karakteristik özellikler vb.) kullanılabilir.
- Her soru ya da kök düğüm, tahminlere dayanan çıkış düğümleri ya da kök düğümleri içeren iki adet düğümden oluşmaktadır.

- Çıktı tahminleri biri ya da diğerrinin konumlandığı tartışmaların konusunu içeren sınıflarla ifade edilmektedir.

Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, VM’de kuruluşlarının kolay olması, yorumlanmalarının kolay olması, veri tabanı sistemlerine kolayca entegre edilebilmeleri, güvenilirliklerinin daha iyi olması nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahip olan bir tekniktir (Olafsson, 2008:1436).

Karar ağacı temelli analizlerin yaygın olarak kullanıldığı alanlar şunlardır:

- Belirli bir sınıfın olası üyesi olacak elemanların belirlenmesi,
- Çeşitli vakaların yüksek, orta, düşük risk grupları gibi çeşitli kategorilere ayrılması,
- Parametrik modellerin kurulmasında kullanılmak üzere çok sayıdaki değişkenden en önemlilerinin seçilmesi,
- Gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulması,
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması,
- Kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikli değişkenlere dönüştürülmesidir (Akpınar, 2000:17).

Karar Ağaçları yöntemleri, temel anlamda hedef (bağımlı) değişkeni, tahmin edici değişkenlere göre ayırma mantığına dayansa da; bünyesinde değişik amaçlara hizmet eden birbirinden farklı algoritmalara sahiptir (Koyuncugil, 2006:72).

Karar ağacı oluşturmak için geliştirilen bu algoritmalar arasında CHAID (Chi-Squared Automatic Interaction Detector), Exhaustive CHAID, CRT (Classification and Regression Trees), ID3, C4.5, MARS (Multivariate Adaptive Regression Splines), QUEST (Quick, Unbiased, Efficient Statistical Tree), C5.0, SLIQ (Supervised Learning in Quest), SPRINT (Scalable Parallelizable Induction of Decision Trees) yer almaktadır.

2.7.1. CRT Algoritması

CRT (CART yada C&RT), sınıflama ve regresyon için kullanılan özyinelemeli bölme metodu Breiman tarafından 1984'te geliştirilmiştir (SPSS,1999:3). CRT, öncelikle bütün veri seti olmak üzere iki alt düğümü tekrarlı bir şekilde yaratmak için bütün tahmin değişkenlerinin hepsini kullanan veri setinin alt kümelerine bölünmesiyle kurulmuştur (Türe vd., 2008:2). Bu algoritmada, her aşamada ilgili küme kendinden daha homojen olan iki alt kümeye ayrılmaktadır. Ayrım işlemi kategorik bağımlı değişkenler için gini, twoing; sürekli değişkenler için en küçük kareler sapması (Least-Squared Deviation) indeks hesaplamalarına göre yapılmaktadır. Bu hesaplamalarda kar, maliyet değerleri ve değişken kategorileri arasındaki önceliklerin tanımlanabilmesi gibi sağlanan çeşitli esneklikler, CRT algoritmasının günümüzde yaygın olarak tercih edilmesine neden olmaktadır (İmamoğlu, 2005:51).

2.7.2. CHAID Algoritması

CHAID metodu 1980'de Kaas tarafından en iyi bölmeyi hesaplamak için istatistik olarak anlamlı bir farklılığın olmadığı, hedef değişkene uyan çiftlerde tahmin değişkeninin olası kategori çiftini birleştirmesiyle oluşturulmuştur (SPSS, 1999:2). CRT algoritmasına benzemektedir; fakat veriyi bölümlere ayırırken farklı bir yol kullanmaktadır. En uygun bölümleri seçmek için kullanılan entropy veya gini metrikleri yerine chi-square testi kullanılmaktadır (İmamoğlu, 2005:50). Herhangi bir düğümdeki en iyi bölmeyi hesaplamak için tahmin değişkenleri kategorisinin herhangi kullanılabilir parçası hedef değişkene uyan bir çiftin içinde istatistik olarak anlamlı bir fark kalmayınca kadar birleştirilmektedir (Türe vd., 2008:3).

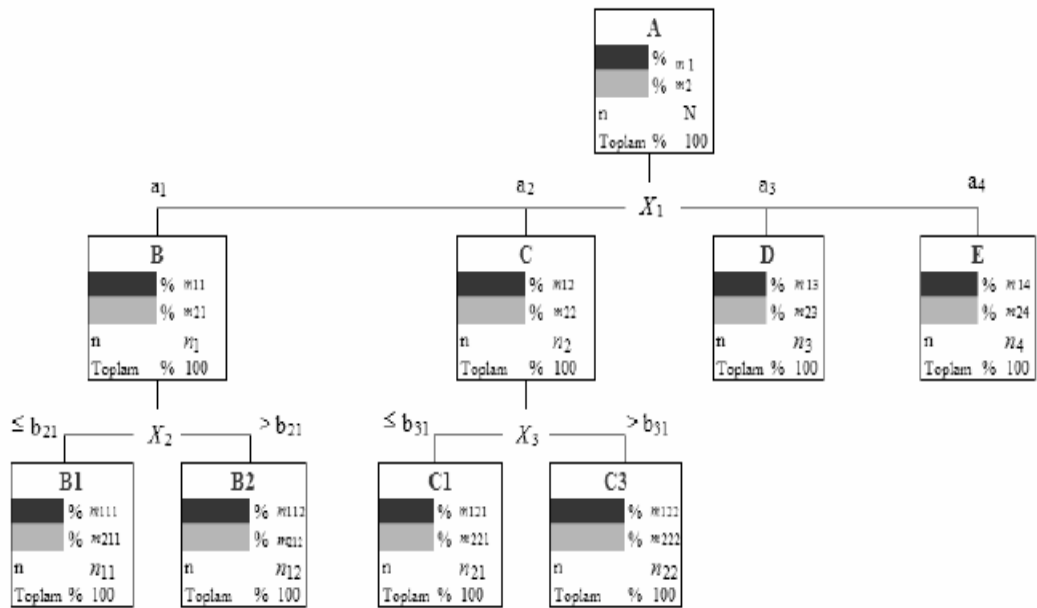
CHAID ile diğer yöntemler arasındaki en önemli farklılıklardan birisi, ağaç türetimidir. ID3, C4.5, ve CRT ikili ağaçlar türetirken, CHAID çoklu ağaçlar türetmektedir (Türe vd., 2008:4).

Y'nin hedef değişken olduğu varsayılarak, Şekil 2.2: Karar Ağacı Değişken İlişkisi dikkate alındığında (Koyuncugil, 2007:17):

- X1, X2 ve X3 olmak üzere sadece 3 değişken hedef Y değişkeniyle istatistik olarak önemli ilişkiye sahiptir.

- X1 değişkeni Y hedef değişkeniyle istatistik olarak en önemli ilişkiye sahiptir.
- X2 değişkeni X1 değişkeniyle $X_1=a_1$ olması koşuluyla istatistik olarak önemli ilişkiye sahiptir.
- X3 değişkeni X1 değişkeniyle $X_1=a_2$ olması koşuluyla istatistiksel açıdan önemli ilişkiye sahiptir.

Şekil 2.2: Karar Ağacı Değişken İlişkisi



Kaynak: Ali Serhan Koyuncu (2007); *Borsa Şirketlerinin Sektörel Risk Profillerinin Veri Madenciliğiyle Belirlenmesi*, Sermaye Piyasası Kurulu Araştırma Raporu, Araştırma Dairesi, Ankara.

CHAID algoritması aşağıda verilmektedir (SPSS, 2002:205):

1. Her bir tahmin edici değişken için X'in, Y hedef değişkeninin dikkate alınan en az öneme sahip (p değeri en büyük) kategori çiftini bulmaktadır. Yöntem, Y'nin ölçüm düzeyine bağlı olarak p değerlerini hesaplamaktadır.

a. Bağımlı değişken Y sürekli ise F testi kullanılır.

b. Y nominal ölçekli bir değişken ise X'in kategorileri satırlarda ve Y'nin kategorileri sütunlarda olacak biçimde çift yönlü bir sınıflandırma tablosu düzenlenir. Pearson ki-kare testi veya Pearson olabilirlik testi kullanılır.

c. Y ordinal ölçekli bir değişken ise bir Y birliktelik modeli uygulanır. Olabilirlik testi kullanılır.

2. En büyük p değerine sahip X'in kategori çifti için, p değeri önceden belirlenmiş $\alpha_{\text{birleştirme}}$ ölçütüyle kıyaslanır.

a. p değeri $\alpha_{\text{birleştirme}}$ 'den büyük ise bu çifti tek kategori altında birleştir. X'in yeni kategori kümesi için algoritma Adım1'den başlatılır.

b. p değeri $\alpha_{\text{birleştirme}}$ 'den küçük ise Adım3'e geçilir.

3. X'in ve Y'nin kategori kümesi için uygun Bonferroni düzeltmesi kullanılarak düzeltilmiş p değeri hesaplanır.

4. En küçük düzeltilmiş p değerine sahip (en önemli) X tahmin edici değişkeni seçilerek, p değeri önceden tanımlanmış alfa düzeyi $\alpha_{\text{böl}}$ ile kıyaslanır.

a. p değeri, $\alpha_{\text{böl}}$ değerine eşit veya küçük ise düğüm X'in kategori kümesini temel alınarak bölünür.

b. p değeri, $\alpha_{\text{böl}}$ değerinden büyük ise bu düğüm uç düğüm olduğu için bölünmez.

5. Ağaç büyüme süreci, durma kuralları sağlanana kadar tekrarlanır.

2.7.3. QUEST Algoritması

QUEST algoritması hızlı, tarafsız, etkili, istatistik ağacı anlamına gelmektedir. Yöntem, 1997'de Loh ve Shih tarafından tanıtılmıştır. CRT gibi ikili karar ağacının yerine getirdiği işlevleri yapan ağaç yapısında bir sınıflama algoritmasıdır. İkili ağaç üretme nedeni bölme, durdurma kuralları ve budama gibi tekniklere izin veren ikili ağaçların kullanılmasıdır. CHAID ve CRT'den farklı olarak, ağaç dallarını genişletme işlemi süresince eşzamanlı olarak bölme noktası seçimini ve değişken seçimini QUEST'in ayrı ayrı ele almasıdır (SPSS, 1999:4).

2.7.4. MARS Algoritması

MARS algoritması çok değişkenli uyumlu regresyon uzanımları tekniği anlamına gelmektedir. Yöntem, 2001'de Steinberg tarafından tanıtılmıştır. Bu yöntem, büyük, karmaşık veri tabanlarına uygulanabilen regresyon analizi temelli bir tekniktir (Francis, 2006:2). MARS geleneksel yöntemlere göre son derece zor olan, çok boyutlu verilerin içinde gizlenmiş karmaşık veri yapısını, en uygun veri dönüşümlerini ve verilerin karşılıklı etkileşimlerini belirleyebilme avantajı ile regresyon modellemesinde yeni bir yaklaşımdır. Küçük veri tabanları söz konusu olduğunda bile iyi bir regresyon modelinin geliştirilmesi çok zaman ve emek istemektedir. Bununla birlikte, MARS yöntemiyle geniş veri tabanları ve çok karmaşık veri yapıları için regresyon modelleri kolayca geliştirilebilmektedir (Tunay, 2001:432-433).

2.7.5. ID3 Algoritması

Veri tabanında çok sayıda değişken varsa ve eğitim kümesi çok fazla kayıt içeriyorsa, fakat fazla hesaplama yapmadan uygun bir karar ağacı oluşturulmak isteniyorsa ID3 algoritması kullanılabilir. ID3 yinelemeli yapıya sahip bir algoritmadır. (İmamoğlu, 2005:60).

2.7.6. C4.5 Algoritması

Veriyi özyinelemeli olarak alt kümelere ayırarak bir sınıflama karar ağacı oluşturur. C4.5 algoritması ID3 algoritmasının geliştirilmiş bir versiyonudur. Algoritma sürekli değerlere sahip nitelikler için kullanılabilir, budama işlemi yapılabilmektedir ve karar üretilmesi gerçekleştirilebilir (İmamoğlu, 2005:55).

Bu bölümde veri madenciliğinde kullanılan yöntem ve teknikler kısaca tanıtıldıktan sonra, bu çalışmada kullanılan karar ağacı algoritmaları ayrıntılı bir biçimde açıklanmıştır. Bir sonraki bölümde, tezin amacı, önemi, kapsamı açıklanmış ve çalışmanın yöntemi veri madenciliği sürecine uygun olarak açıklanmıştır. Uygulamada, karar ağaçları tekniklerinden CHAID kullanılarak İMKB Ulusal 100 endeksindeki firmaların sektör ve finansal göstergelerinin birbirleriyle olan ilişkileri irdelenmiştir. Çalışmada öncelikle İMKB100 sanayi ve hizmet sektörlerinde faaliyet

gösteren firmaların karlılık, likitide, varlık kullanım etkinliği, sermaye yapısı ve işletme büyüklüğü göstergelerinden hangisine göre anlamsal farklılık gösterdiği araştırılmaktadır. Daha sonra sırasıyla sermaye yapısı göstergelerinden kısa vadeli borç oranı, toplam borç oranı; karlılık göstergelerinden aktiflerin karlılığı, özsermayenin karlılığı değişkenlerinin en önemli belirleyicileri belirlenmektedir.

3. VERİ MADENCİLİĞİ UYGULAMASI: İSTANBUL MENKUL KIYMETLER BORSASI (İMKB) ÖRNEĞİ

3.1. Çalışmanın Tanıtılması

Bu çalışmada, İMKB100 sanayi ve hizmet sektöründe faaliyette bulunan 173 işletmenin 2004-2006 yıllarına ait yıllık finansal göstergelerinden yararlanarak VM tekniklerinden birisi olan karar ağaçları tekniği, CHAID algoritması uygulanarak seçilen finansal göstergelere göre firmaları ayıran en önemli özelliklerin saptanmasının yanında sermaye ve karlılık yapılarını belirleyen en önemli değişkenler araştırılmaktadır.

3.1.1. Çalışmanın Amacı ve Önemi

Bu çalışmanın temel amacı, yatırımcıların veya karar verme konumunda olanların, işletmeleri belirli yönlerden performanslarını değerlendirmede dikkate alması gereken en önemli ölçütlerin (değişkenlerin) saptanmasında veri madenciliği tekniklerinden karar ağacı algoritmalarının uygun bir yöntem olarak kullanılabileceğini göstermektir. Bu amaçla bugüne kadar daha çok regresyon analizi, diskriminant analizi ve lojistik regresyon analizi gibi çok değişkenli istatistik tekniklerden daha çok yararlanıldığı görülmektedir. Ülkemizde karar ağacı algoritmalarının kullanımı ise oldukça yenidir. Karar ağacı algoritmalarının en önemli avantajları, parametrik olmayan yöntemler arasında olması nedeniyle diğer çok değişkenli tekniklerde sağlanması gereken istatistik varsayımların olmamasıdır. Ayrıca karar ağacı algoritmalarının bağımlı ve bağımsız değişkenler arasındaki ilişkilerin yönünü, önem sırasını görselleştirmesi diğer avantajları arasındadır. Bu özelliği özellikle elde edilen sonuçların yorumunu oldukça basitleştirerek daha somut ve kullanışlı hale getirmektedir.

Bu bağlamda bu çalışmada öncelikle İMKB100 sanayi ve hizmet sektöründe faaliyet gösteren firmaları ayırmada etkili olan en önemli karlılık, likitide, varlık kullanım etkinliği, sermaye yapısı ve işletme büyüklüğü değişkenleri araştırılmaktadır. Daha sonra sermaye yapısı göstergelerinden kısa vadeli borç oranını, toplam borç oranını ve karlılık göstergelerinden özsermayenin karlılığı ve aktiflerin karlılığını belirleyen en önemli finansal göstergeler araştırılmaktadır.

3.1.2. Çalışmanın Kapsamı

Çalışmanın kapsamına 2004-2006 yılları arasında İMKB100 sanayi ve hizmet sektörlerinde faaliyet gösteren 173 şirket alınmıştır. Bu şirketlerin, 2004-2006 yıllarına ait bilanço ve gelir tablolarına ilişkin veriler İMKB'nin ağ sayfasından elde edilmiştir.

3.1.3. Çalışmaya İlişkin Literatür Özeti

Bir firmanın sermaye yapısı, kararlılığı üzerinde önemli rol oynamaktadır. Yüksek oranda borç, bir şirkete büyümekteyken karlılık getirebilmektedir. Ancak, özellikle firmanın büyümesi yavaşlamış ya da geçici olarak azalmışsa, finansal sıkıntı ve iflas olasılığını artırabilmektedir (Liu, 1999:6). Sermaye yapısı (borç/özsermaye veya aktifler) konusunda yapılan çalışmaların temelini Modigliani ve Miler'in ilintisizlik kuramı oluşturmaktadır (Korkmaz vd., 2007:81). Modigliani ve Miler'den sonra pek çok teorik çalışma yapılmış olmasına rağmen, karlılık ve kaldıraç oranları arasında tutarlı tahminlere ulaşılamamıştır (Huang ve Song, 2006:17).

Literatürde çeşitli sektörlerdeki firmaların sermaye yapısı ve karlılığını belirlemeyi amaçlayan çalışmalar yapılmıştır. Kale vd. (1991), makalesinde optimum sermaye yapısını, kişisel ve anonim şirket vergilendirme çerçevesinde optimum borç oranı ve işletme riski arasındaki ilişkiyi regresyon analiziyle araştırmış ve açıklamıştır. Haris ve Raviv (1991), firma maliyetleri, asimetrik bilgi, ürün/girdi pazar etkileşimleri üzerine teorileri incelemiş ve karşılaştırmıştır. Durukan (1997), çalışmasında Türkiye ekonomisinde faaliyet gösteren ve hisse senetleri İstanbul Menkul Kıymetler Borsası'nda işlem gören 68 firmanın sermaye yapılarını etkileyen faktörleri regresyon analizi ile saptamayı amaçlamıştır. Pastor ve Veronesi (2003), ortalama karlılığı içinde hisse senedi değerlemesi için regresyon analizi tekniği ile bir yaklaşım geliştirmiştir. Sogorb (2005), çalışmasında Küçük ve Orta Büyüklükteki İşletmelerin (KOBİ) sermaye yapısı üzerinde firma karakteristiklerinin etkisini regresyon analizi ile test etmeyi amaçlamıştır. Istaitieh ve Fernandez (2006), ürün faktör pazarları ve firmaların sermaye yapısına ilişkin yapılan deneysel modellere ilişkin fikirler ve önerilerde bulunmuştur. Berger ve Di Patti (2006), banka

endüstrisindeki bir uygulamasında yeni bir yaklaşım denemesiyle sermaye yapısı ve firma performansını denklemler sistemi yaklaşımıyla incelemiştir. Akyüz vd. (2006), çalışmasında mikro, küçük ve orta ölçekteki firma sahiplerinin finanslama tercihleri ve sermaye yapılarını araştırmayı amaçlamıştır. Tang ve Jang (2007), regresyon analizi kullanarak yaptığı çalışmada sermaye yapısı teorileri ve daha önceki ampirik çalışmalar arasındaki çelişkileri onaylamaktadır. Margaritis ve Psillaki (2007), çalışmasında parametrik olmayan veri zarflama analizi ve regresyon analizini kullanarak, firma karlılığı ve kaldıraç oranları arasındaki ilişkiyi incelemiştir. Ters nedensellik ilişkisinde olduğu kadar firma performansı üzerindeki kaldıraç oranlarının etkisini de göz önünde bulundurmuştur. Korkmaz vd. (2007), çalışmasında hisse senetleri İMKB’de işlem gören ve imalat sektöründe yer alan KOBİ niteliğindeki firmaların sermaye yapıları ile sermaye yapılarını etkileyen faktörleri regresyon analiziyle araştırmıştır.

Bu çalışmalar, sermaye yapısını ve karlılığı simgeleyen ve tanımlayan değişkenlerin saptanmasında ve elde edilen sonuçların yorumlanmasında yararlanılacak ölçütlerin saptanması hususlarında yol gösterici olmuşlardır.

3.2. Çalışmanın Yöntemi

Profillerin oluşturulması için görsel, kolay anlaşılır, kolay yorumlanabilir ve kural çıkarımına izin vermesi nedeniyle VM yöntemlerinden karar ağaçları kullanılmıştır. Çalışmada İMKB100 endeksindeki şirketlerin profilleri, karar ağacı yöntemlerinden CHAID ile belirlenmiştir. CHAID’in analiz aracı olarak seçilme nedenlerinin başında sürekli ve kategorik tüm değişken tipleriyle çalışabilmesi gelmektedir.

Bununla beraber, sürekli bağımlı değişkenler otomatik olarak analizin amacına uygun olarak kategorize edilmektedir. CHAID, Ki-Kare metriği vasıtasıyla, ilişki düzeyine göre farklılık rastlanan grupları ayrı ayrı sınıflamakta ve ağacın yaprakları, ikili değil, verideki farklı yapı sayısı kadar dallanmaktadır. Bu algoritmaya ilişkin ayrıntılı açıklamalar ikinci bölümde verilmiştir.

3.2.1. Veri Anlama ve Veri Seçimi Aşaması

Tez çalışması için en uygun veri tabanı oluşturulurken hisse senedi ulusal pazarda işlem görmekte olan firmaların adları ve faaliyet konuları ile ilgili bilgilere ihtiyaç duyulmuştur. Bu bilgiler İMKB tarafından yayınlanmış olan İMKB100 endeksindeki firmalardan sağlanmıştır.

Değerlendirmeye alınacak firmalarda bulunması gereken temel nitelikler belirlendikten sonra bunlara sahip olan firmaları kapsayan düzenli bir finansal veri tabanına ulaşılmaya çalışılmıştır. Bu araştırma adımı sonucunda, İMKB’de hazırlanmış ve yayınlanmış olan Ocak 2004-Aralık 2006 dönemini içeren finansal tablo verileri veri tabanına aktarılmıştır.

Kararlaştırılan özellikteki örneğin oluşturulabilmesi için kapsamdaki firmaların hesap dönemleri de incelenmiştir. Oranların hesaplanıp değerlendirilmesinde daha sağlıklı bir sonuca ulaşılabilmesi için İMKB’de kaydı bulunan, Ulusal Pazarda işlem gören ve finansal hesap dönemi 1 Ocak- 31 Aralık olan firmalar değerlendirmeye alınmıştır.

3.2.2. Veri Önışleme Aşaması

Bir ekonomideki finansal firmalar, birikim fazlası olan kişi ve kuruluşlarla belli amaçlar -mevcut firmaların finansal durumlarını iyileştirmek, gerekli yatırımları gerçekleştirebilmek gibi- için bu birikimleri kullanmak isteyenlerin karşılaşmalarını sağlarlar. Finansal firma faaliyetleri, finansal olmayan firmalardan farklı olarak belli bir üretim süreci içermemektedir. Çeşitli araçlar kullanarak, fon aktarımları konusunda hizmet veren bu firmaların finansal yapıları da diğer üretim ve hizmet firmalarından son derece farklıdır (Erken, 1998:38).

Finansal firmaların müşterilerinden topladıkları mevduatların finansal tablolarındaki gösterim şekli, bu firmalarda yüksek kaldıraç oranlarına sebep olmaktadır. Bu firmaların finansal analizlerinde kullanılan birçok oranın, kendilerine özgü bir şekilde yorumlanması gerekmektedir ve tüm sektörlerde geçerli olabilecek değerlendirme ölçütlerinin oluşturulması mümkün değildir. Çalışmada daha tutarlı ve

anlamli bulgular elde edebilmek ve saglikli sonuclara ulasmak icin finansal firmalar degerlendirme disIi bırakılmıştır.

Çalışma örneğinin belirlenmesi aşamasında ilk olarak ulusal pazarda işlem görmekte olan İMKB100 endeksindeki firmaların ad ve faaliyet konuları veri tabanına alınmıştır. Daha sonra çalışmada temel alınan finansal firma tanımının kapsamında bulunduğu saptanan firmalar listeden elenerek temizlenmiştir. Söz konusu 173 işletmeye ait finansal göstergeler veri tabanına aktarılmıştır.

3.2.3. Veri Dönüştürme Aşaması

Verilerin analizinde yıllar arası enflasyon etkisini yok etmek için finansal tablolardan alınan finansal göstergelere enflasyon düzeltmesi yapılmıştır. Daha sonra çalışmada kullanılan finansal oranlar hesaplanarak veriler düzenlenmiştir. Tablo 3.1’de çalışmada kullanılan finansal oranlar verilmiştir.

Tablo 3.1: Çalışmada Kullanılan Finansal Oranlar Tablosu

Değişken Sınıfı	Simge	Değişkenin Tanımı
Karlılık Oranları	NK/TA	Aktiflerin Karlılığı=Net Kar/Toplam Aktifler
	$NK/ÖS$	Özsermayenin Karlılığı=Net Kar/Özsermaye
	NK/NS	Net Kar Marjı=Net Kar/Net Satışlar
	ERO	Ekonomik Rantabilite Oranı=(VÖK+Finansman Giderleri)/Toplam Aktifler
Likitide Oranları	$DÖV/KVB$	Cari Oran=Döner Varlıklar/Kısa Vadeli Borçlar
	ATO	Likitide Oranı=(Döner Varlıklar-Stoklar)/Kısa Vadeli Borçlar
	$NİS/NS$	İşleme Sermayesinin Net Satışlara Oranı=(DÖV-KVB)/Net Satışlar
	$KVAL/DÖV$	Kısa Vadeli Alacaklar/Döner Varlıklar=KVAL/DÖV
Varlık Kullanım Etkinliği	NS/TA	Aktif Devir Hızı=Net Satışlar/Toplam Aktifler
	SDH	Stok Devir Hızı=Satışların Maliyeti/Stoklar
	$NS/DÖV$	İşletme Sermayesi Devir Hızı (İSDH)=Net Satışlar/Döner Varlıklar
	$NS/ÖS$	Özsermaye Devir Hızı=Net Satışlar/Özsermaye
Sermaye Yapısı Oranları	KVB/TA	Kısa Vadeli Borç Oranı=Kısa Vadeli Borçlar/Toplam Aktifler
	TB/TA	Toplam Borç Oranı=Toplam Borçlar/Toplam Aktifler
	$DÖV/TA$	Döner Varlıklar/Toplam Aktifler
	$DV/ÖS$	Duran Varlıklar/Özsermaye
İşletme Büyüklüğü	$İB1$	Toplam Aktifler
	$İB2$	Net Satışlar

Çalışmada bağımlı değişken olarak sektör kukla değişkeni (sanayi=1 ve hizmet=2); sermaye yapısı göstergelerinden toplam borç oranı (TB/TA), kısa vadeli borç oranı (KVB/TA); karlılık göstergelerinden ise aktiflerin karlılığı (NK/TA) ve özsermayenin karlılığı ($NK/ÖS$) değişkenleri alınmıştır.

Literatürde sermaye yapısını tanımlayan değişkenler olarak kaldıraç oranları yaygın olarak kullanılmaktadır. Çözümlemede sadece TB/TA ve KVB/TA kullanılmıştır. Karlılığı etkilen faktörler ise aktiflerin karlılığı ve özsermayenin karlılığı olacağı düşüncesiyle, özsermaye ve aktiflerin karlılığına göre kaldıraç oranları kullanılmaktadır. Firmaların faaliyet gösterdikleri sektöre bağlı olarak ise diğer finansal oranların birbirleriyle ilişkisi incelenmektedir.

Araştırmada kullanılan bağımlı değişkenlerin tanımları şöyledir:

KVB/TA : Kısa Vadeli Borç Oranı = Kısa Vadeli Borçlar/Toplam Aktifler

TB/TA : Toplam Borç Oranı = Toplam Borçlar/Toplam Aktifler

NK/TA : Aktiflerin Karlılığı = Net Kar/Toplam Aktifler

$NK/ÖS$: Özsermayenin Karlılığı = Net Kar/Özsermaye

Kaldıraç oranları, işletmenin borç ödeyebilme yeteneğini ölçmektedir. Diğer veriler sabit iken, yüksek bir kaldıraç oranı, daha riskli bir işletmeyi ifade etmektedir. Nakit akımı aşırı bir şekilde azalırsa, işletme borçlarını ödeyememekte ve teknik anlamda güçlüğü düşmektedir.

Toplam Borçlar/Toplam Aktifler oranı, işletmenin yatırımlarının finansmanında ne ölçüde borç kullandığını göstermektedir. Söz konusu oranın yüksek olması, işletmenin spekülative (vurgunsal) tarzda finanse edildiğini, kreditorler açısından emniyet marjının dar olduğunu, işletmenin faiz ve anapara taksitlerini ödeyememe nedeniyle mali yönden güç duruma düşme olasılığının yüksek olduğunu göstermektedir.

Kısa Vadeli Borçlar/Toplam Aktifler oranı, bir işletmenin toplam varlıkları içinde kısa vadeli borçlarının payını göstermektedir. Bu oranın $\frac{1}{3}$ olması idealdir (Aktan ve Bodur, 2006:9).

Karlılık ya da rantabilite oranları, üretim faaliyetinde bulunan bir işletmenin kar durumunu açıklayan göstergelerdir. Yatırımların karlılığını gösteren ana oranlar ise özsermaye karlılığı ve toplam aktif karlılığı oranlarıdır.

Özsermaye Karlılığı = Net Kar/Özsermaye oranı, işletmelerin özsermayelerinin performansını ölçmektedir. Özellikle aynı sanayi dalında faaliyet gösteren işletmeleri kıyaslarken bu orandan yararlanılmaktadır. Eğer oran büyükse işletmenin iyi bir yatırım yaptığını ve giderlerini sıkı bir şekilde kontrol altında tuttuğunu söylenebilir.

Toplam Aktif Karlılığı = Net Kar/Toplam Aktifler oranı, işletmenin yaptığı tüm yatırımlara karşılık vergiden sonra ne kazandığını göstermektedir. Yani işletmenin yatırımlarından ne oranda getiri sağladığını ölçen bir orandır (Aktan ve Bodur, 2006:8).

3.2.4. Modelleme Aşaması: Karar Ağacı Tekniği CHAID Algoritmasının Uygulanması

Karar ağaçları algoritmaları uygulama bakımından bir hedef değişken (bağımlı değişken) ve hedef değişkeni açıklamaya yönelik kullanılacak açıklayıcı değişkenler (bağımsız değişkenler) olmak üzere iki grup değişken ile gerçekleştirilmektedir. Bu çalışmada, sektör, sermaye yapısı ve karlılık bağımlı değişkenleri, Tablo 3.1’de tanımlarıyla birlikte yer verilen finansal değişkenler ile açıklanmıştır. Dolayısıyla, sektör, TB/TA , KVB/TA , NK/TA ve $NK/ÖS$ göstergesi hedef değişken; diğer finansal değişkenler de açıklayıcı olarak yer almıştır.

SPSS 15 yazılımı ile CHAID algoritması kullanılarak $\alpha_{\text{birleştir}} = \alpha_{\text{böl}}=0,10$ olarak belirlenmiş ve karar ağaçları elde edilmiştir.

3.2.5. Değerlendirme Aşaması

Çalışmada karar ağaçları algoritması bakımından hedef değişken olan sektör, sermaye yapısı (toplam borç oranı ve kısa vadeli borç oranı), karlılık (aktiflerin karlılığı ve özsermayenin karlılığı) bulgularına göre 173 işletmenin CHAID model sonuçları ve profilleri verilmiştir.

3.2.5.1. CHAID Modeli Sonuçlarının Değerlendirilmesi

Risk değeri, sonuç modeli ve modelleri oluşturmak için bilgi sağlayan CHAID model tablosunun en uygun ağaç yapısını belirlemede bir göstergedir. Terminal düğüm sayısı da en uygun ağaç yapısı için bir gösterge olarak kullanılmaktadır. Aşağıda verilen tablolar, hedef değişkenlere ait 12, 10 ve 8 farklı ağaç yapısından oluşmaktadır. Ağaç yapılarının derinliği ve düğüm sayıları, ağaç 1'den ağaç 12'ye kadar ve ağaç 1'den ağaç 8'e kadar artmaktadır. Risk değerleri ise standart hata, yerine koyma (resubstitution) ve çapraz geçerlilik (cross validation) yüzdelerinden oluşmaktadır. En uygun modellerin seçiminde toplam ve nihai düğüm sayıları, ağaç derinliği, yerine koyma, çapraz geçerlilik ile açıklanan varyans yüzdeleri ölçütlerinden yararlanılmıştır. Hedef değişken olarak sektör sınıfını tanımlayan modelde riskin ölçüsü yanlış sınıflandırma oranını göstermektedir. Bu nedenle risk yüzdesi, en uygun ağaç yapısının seçiminde dengeli ve minimum olmalıdır. Ayrıca, kategorik olmayan bağımlı değişkenlerin değerlendirilmesinde riskin bir ölçüsü olarak varyans kullanılmaktadır. Bu nedenle açıklanan varyans yüzdesi de en uygun ağaç seçiminde bir ölçüt olarak kullanılmaktadır. Tablo 3.3 – Tablo 3.6 tablolarındaki açıklanan varyans yüzdeleri, 1'den açıklanmayan varyans oranı (risk/ilgili değişkenin varyansı) çıkartılarak hesaplanmıştır. Tablo 3.3'te *KVB/TA* standart sapması 0,199 ve varyansı 0,040'tır. En uygun ağaç yapısı olarak seçilen Ağaç 5'in risk değeri 0,011 (%1,1)'dir. Açıklanan varyans değeri $(1 - (0,011/0,040))$ 0,72'dir.

Modelin uygulanması sırasında çapraz geçerlilik kuralı seçilmiş ve farklı ağaç yapılarını izlemek için algoritmanın durdurma kurallarına (stopping rules) uygun farklı düğüm sayıları denenmiştir. Tablolarda sektör bağımlı değişkeni için yerine koyma ve çapraz geçerlilik riskleri birbirine en yakın olan ağaç yapısı seçilmiştir.

Bununla birlikte *KVB/TA*, *TB/TA*, *NK/TA* ve *NK/ÖS* değişkenleri için açıklanan varyans yüzdesi hesaplanmış %60'dan büyük olan ağaç yapıları en uygun ağaç yapısı olarak belirlenmiştir. En uygun ağaç yapıları tablolar üzerinde gösterilmiştir. Bu sonuçla sektör, sermaye yapısı ve karlılık hedef değişkenlerine göre elde edilen tablolar ve kullanılan ağaç yapıları aşağıdaki gibidir.

Tablo 3.2: Sektör Sınıfı CHAID Algoritma Sonuçları

Tüm Olası Ağaçlar	Karar Düğüm Sayıları			Yerine Koyma (%)		Çapraz Geçerlilik (%)	
	Toplam	Terminal	Ağaç Derinliği	Risk	Standart Hata	Risk	Standart Hata
Ağaç 1	1	1	0	26,6	3,4	26,6	3,4
Ağaç 2	3	2	1	26,6	3,4	26,6	3,4
Ağaç 3	4	3	1	25,4	3,3	30,1	3,5
Ağaç 4	7	5	2	17,3	2,9	18,5	3,0
Ağaç 5	8	5	2	19,7	3,0	30,6	3,5
Ağaç 6	13	9	2	12,1	2,5	24,3	3,3
Ağaç 7	10	7	3	17,3	2,9	17,9	2,9
Ağaç 8	12	8	3	16,2	2,8	24,3	3,3
Ağaç 9	16	10	3	13,3	2,6	21,4	3,1
Ağaç 10	17	11	3	11,0	2,4	21,4	3,1
Ağaç 11	23	14	4	8,7	2,1	22,5	3,2
Ağaç 12	25	15	5	8,7	2,1	20,8	3,1

Tablo 3.3: Kısa Vadeli Borç Oranı CHAID Algoritma Sonuçları

Tüm Olası Ağaçlar	Karar Düğüm Sayıları			Yerine Koyma (%)		Çapraz Geçerlilik (%)		Açıklanan Varyans (%)
	Toplam	Terminal	Ağaç Derinliği	Risk	Standart Hata	Risk	Standart Hata	
Ağaç 1	1	1	0	3,9	0,6	4,0	0,6	2
Ağaç 2	7	6	1	1,9	0,3	2,2	0,6	52
Ağaç 3	6	4	2	1,9	0,3	2,1	0,3	52
Ağaç 4	8	5	3	1,9	0,3	1,9	0,3	52
Ağaç 5	13	9	3	1,1	0,2	2,5	0,7	72 *
Ağaç 6	19	14	3	0,7	0,1	2,2	0,4	82
Ağaç 7	27	17	3	0,7	0,1	2,5	0,5	82
Ağaç 8	36	24	3	0,4	0,1	1,4	0,2	90
Ağaç 9	25	18	4	0,7	0,1	2,3	0,5	82
Ağaç 10	42	27	5	0,4	0,1	2,3	0,6	90

* Kısa vadeli borç oranı standart sapması (σ) 0,199 ve varyansı (σ^2) 0,040'tır. Açıklanan varyans oranı ($1 - (0,011 / 0,040)$) 0,72'dir.

Tablo 3.4: Toplam Borç Oranı CHAID Algoritma Sonuçları

Tüm Olası Ağaçlar	Karar Düzüm Sayıları			Yerine Koyma (%)		Çapraz Geçerlilik (%)		Açıklanan Varyans (%)
	Toplam	Terminal	Ağaç Derinliği	Risk	Standart Hata	Risk	Standart Hata	
Ağaç 1	4	3	1	2,6	0,3	3,8	0,3	44
Ağaç 2	7	6	1	2,1	0,3	2,4	0,3	55
Ağaç 3	7	5	2	2,4	0,3	3,0	0,3	49
Ağaç 4	13	9	2	1,7	0,3	2,1	0,3	64 *
Ağaç 5	15	10	2	1,4	0,2	2,5	0,4	70
Ağaç 6	19	13	3	1,2	0,2	3,1	0,4	74
Ağaç 7	24	16	3	0,9	0,1	2,4	0,3	81
Ağaç 8	30	20	3	0,8	0,1	2,0	0,2	83
Ağaç 9	36	23	4	0,7	0,1	2,2	0,3	85
Ağaç 10	75	52	5	0,5	0,1	3,2	0,7	89

* Toplam borç oranı standart sapması (σ) 0,216 ve varyansı (σ^2) 0,047'dir. Açıklanan varyans oranı (1- (0,017/ 0,047)) 0,64'tür.

Tablo 3.5: Aktiflerin Karlılığı CHAID Algoritma Sonuçları

Tüm Olası Ağaçlar	Karar Düzüm Sayıları			Yerine Koyma (%)		Çapraz Geçerlilik (%)		Açıklanan Varyans (%)
	Toplam	Terminal	Ağaç Derinliği	Risk	Standart Hata	Risk	Standart Hata	
Ağaç 1	3	2	1	1,5	0,6	1,7	0,7	10
Ağaç 2	5	3	2	1,5	0,6	1,9	0,7	10
Ağaç 3	7	4	2	1,3	0,6	1,9	0,8	22
Ağaç 4	7	4	3	1,3	0,4	1,7	0,7	22
Ağaç 5	15	9	3	1,1	0,4	2,0	0,7	34
Ağaç 6	11	6	3	1,2	0,4	1,8	0,7	28
Ağaç 7	12	7	3	1,2	0,4	2,1	0,7	28
Ağaç 8	19	12	4	0,6	0,1	2,5	0,9	64 *

* Aktiflerin karlılığı standart sapması (σ) 0,129 ve varyansı (σ^2) 0,017'dir. Açıklanan varyans oranı (1- (0,006/ 0,017)) 0,64'tür.

Tablo 3.6: Özsermayenin Karlılığı CHAID Algoritma Sonuçları

Tüm Olası Ağaçlar	Karar Düzüm Sayıları			Yerine Koyma (%)		Çapraz Geçerlilik (%)		Açıklanan Varyans (%)
	Toplam	Terminal	Ağaç Derinliği	Risk	Standart Hata	Risk	Standart Hata	
Ağaç 1	1	1	0	25,2	10,8	25,5	10,9	00
Ağaç 2	3	2	1	22,4	9,9	26,8	10,4	11
Ağaç 3	5	3	2	16,4	5,0	23,8	7,2	35
Ağaç 4	7	4	3	20,3	7,6	23,8	9,3	20
Ağaç 5	9	5	3	8,6	2,9	23,6	7,0	66 *
Ağaç 6	13	9	3	7,2	2,7	32,6	11,9	72
Ağaç 7	19	13	4	6,9	2,7	24,4	6,6	73
Ağaç 8	24	16	5	6,9	2,7	69,6	26,7	73

* Özsermayenin karlılığı standart sapması (σ) 0,503 ve varyansı (σ^2) 0,253'tür. Açıklanan varyans oranı (1- (0,086/ 0,253)) 0,64'tür.

3.2.5.2. CHAID Karar Ağacı ile Elde Edilen İMKB100 Sektör Sonuçları

Çalışmada, karar ağaçları algoritması bakımından hedef değişken seçilen sektör bulgularına göre araştırma kapsamında İMKB100 endeksinde yer alan 173 işletmeden %73,4'ünün (127 işletme) sanayi sektöründe, %26,6'sının (46 işletme) hizmet sektöründe olduğu anlaşılmaktadır. Şekil 3.1'de verilen CHAID karar ağacı ile işletmelerin sektör profilleri oluşturulmuş ve işletmeler 7 farklı profilde gruplanmıştır. İşletmeleri klasik yöntemle “sanayi” ya da “hizmet” sektörü olarak iki grupta toplamak mümkünken, bu çalışma ile sektör durumu ve özelliklerine göre 7 farklı profilde gruplandırmak söz konusu olmuştur.

Tablo 3.7'de CHAID karar ağacı ile elde edilen İMKB100 endeksindeki 7 işletme profiline, profili oluşturan düğümlere ve sektör değişkenine etki eden değişkenlere yer verilmiştir. Profilleri ve sektör durumunu değerlendirmek için Tablo 3.7 ile Şekil 3.1 birlikte incelenmektedir. Şekil 3.1, 4 ana daldan oluşan CHAID karar ağacını ve profillerin tamamını tek bir şekilde göstermektedir.

İMKB100 endeksindeki şirketlerin sektörler itibariyle profillerinin oluşturulmasında, en önemli değişkenin işletme sermayesinin net satışlara oranı ($p=0,000$) olduğu anlaşılmaktadır. İşletme sermayesinin net satışlara oranı değerlerine göre ağacın dalları dört ana dalda (Node 1-Node 4) toplanmıştır. Bu dört düğümü oluşturan alt düğümler de profilleri oluşturmaktadır. İMKB100 endeksindeki firmaların sektör profillerinin oluşturulmasında etkili olan en önemli değişkenler sırasıyla işletme sermayesinin net satışlara oranı, stok devir hızı ve ekonomik rantabilite oranı değişkenleridir.

Profil 1 (Node 1) ve 2 (Node 2): Tablo 3.7 ve Şekil 3.1'de görüldüğü gibi 1. ve 2. profil, işletme sermayesinin net satışlara oranı değişkeninden oluşmaktadır. İşletme sermayesinin net satışlara oranı $-0,04$ 'e eşit ya da küçük olan işletmeler birinci profili, $-0,04$ ile $0,09$ arasında olan işletmeler ikinci profili oluşturmaktadır. Birinci profildeki işletmelerin %11,8'i (2 işletme) sanayi sektöründe, %88,2'si (15 işletme) hizmet sektöründe yer alırken; ikinci profildeki işletmelerin %71,4'ü (25 işletme) sanayi sektöründe; %28,6'sı (10 işletme) hizmet sektöründe yer almaktadır. İşletme sermayesinin net satışlara oranı negatif olan birinci profil, işletmelerin kısa

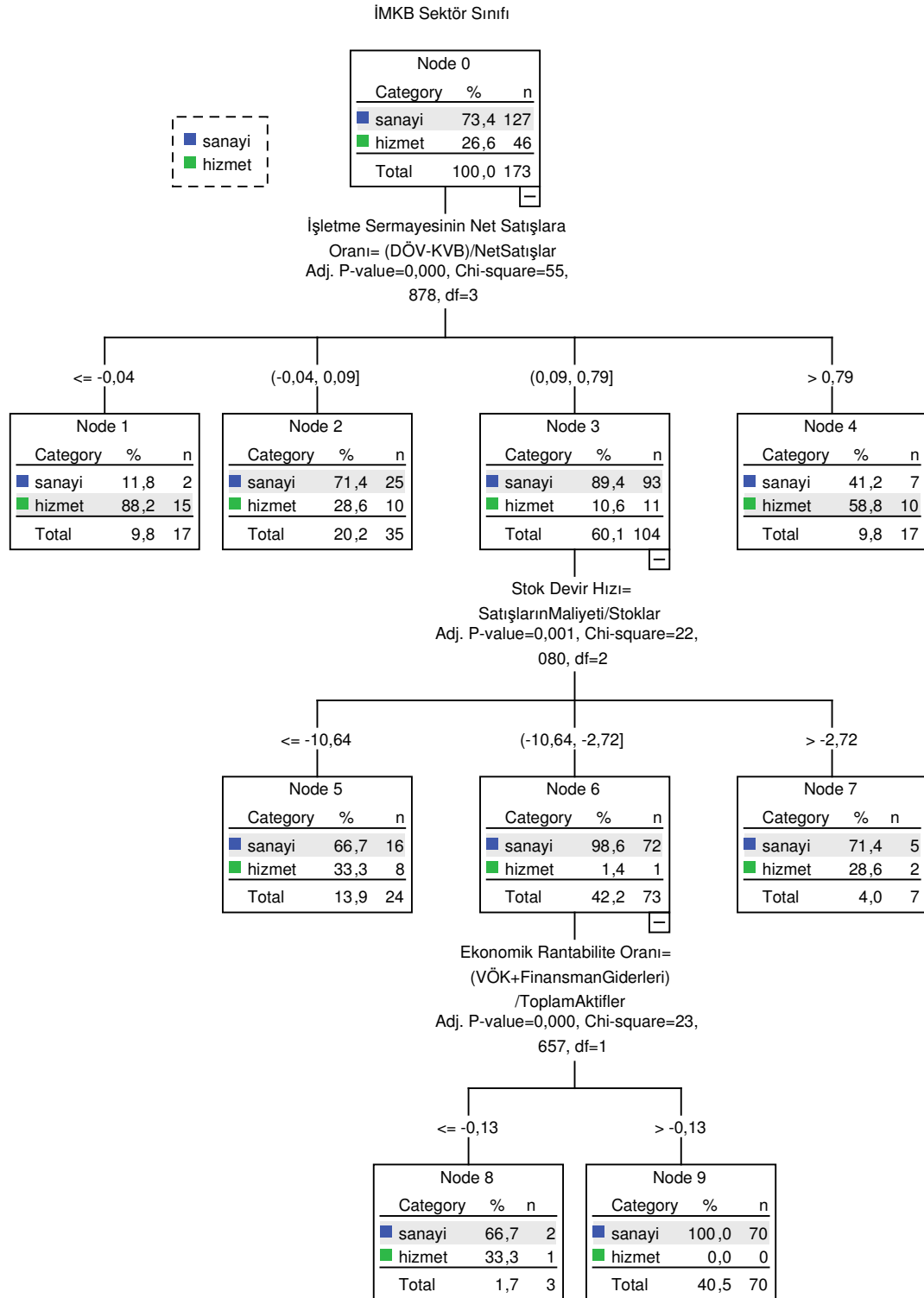
vadeli borçlarının yüksek olduğunu göstermekte ve faaliyetleri süresince tam kapasitede karlı ve verimli çalışabilmesini güçleştirmektedir. Net işletme sermayesinin büyüklüğü, işletmelerin üretimini devam ettirebilmesi ve yükümlülüklerini karşılayamama riskini azaltması açısından önem taşımaktadır. İkinci profilde yer alan işletmelerin işletme satış oranlarının sermayeye bağımlı olmadan, yeterli bir satış düzeyine ulaşabildiği ve diğer finansal oranlarının da yeterli olması durumunda satışların, işletme faaliyetleriyle gerçekleşebileceği görülmektedir. Buna karşın işletmeler, işletme sermayesine bağlı olmadan satışlarını yeterli düzeye getirmek için önlemler almalı ve bu düzeyde kendi faaliyetlerinin dönüş hızını koruyabilmelidir.

Tablo 3.7: Sektör Sınıfı Profilleri ve Profilleri Oluşturan Değişkenler

Profiller	Düğüm	İşletme Sermayesi/ Net Satış Oranı	Stok Devir Hızı	Ekonomik Rantabilite Oranı
Profil 1	1	$\leq -0,04$		
Profil 2	2	$-0,04 - 0,09$		
Profil 3	3, 5	$0,09 - 0,79$	$\leq -10,64$	
Profil 4	3, 6, 8	$0,09 - 0,79$	$(-10,64) - (-2,72)$	$\leq -0,13$
Profil 5	3, 6, 9	$0,09 - 0,79$	$(-10,64) - (-2,72)$	$> -0,13$
Profil 6	3, 7	$0,09 - 0,79$	$> (-2,72)$	
Profil 7	4	$> 0,79$		

Profil 3 (Node 3): Bu profil, işletme sermayesinin net satışlara oranı ve stok devir hızı değişkenlerinden oluşmaktadır ve işletmelerin %66,7'si (16 işletme) sanayi sektöründe, %33,3'ü (8 işletme) hizmet sektöründe yer almaktadır. Bu profili oluşturan işletmelerin işletme sermayesinin net satışlara oranı 0,09 ile 0,79 arasında iken, stok devir hızı -10,64'e eşit ya da küçüktür. Stok devir hızının negatif olması işletmede fazla stok bulundurulduğunu, satışların net işletme sermayesine bağlı olduğunu ve bu nedenle işletme satışlarının azaldığını göstermektedir. Bu da stok tutma maliyetinin artmasına ve stoklarda bozulma riskine sebep olmaktadır. İşletme sermayesinin net satışlara oranının yüksek olması, satışların net işletme sermayesine bağlı olması stok devir hızının yavaş olmasında da önemli bir etkidir.

Şekil 3.1: İMKB Sektör Sınıfı Karar Ağacı *



* Şekilde geçen Node, Category, Total İngilizce sözcükleri sırasıyla düğüm, kategori ve toplam anlamına gelmektedir.

Profil 4 (Node 8) ve 5 (Node 9): Toplam 73 işletmenin olduğu dördüncü ve beşinci profiller, işletme sermayesinin net satışlara oranı, stok devir hızı ve ekonomik rantabilite oranı değişkenlerinden oluşmaktadır. Dördüncü profildeki işletmelerin %66,7'si (2 işletme) sanayi sektöründe, %33,3'ü (1 işletme) hizmet sektöründe yer alırken beşinci profildeki işletmelerin tamamı (70 işletme) sanayi sektöründe yer almaktadır. Bu profili oluşturan işletmelerin, işletmelerin işletme sermayesinin net satışlara oranı 0,09 ile 0,79 arasında, stok devir hızı -10,64 ile -2,72 arasında değişmektedir. Dördüncü profildeki işletmelerin ekonomik rantabilite oranları -0,13'ten eşit ya da küçük iken beşinci profilde ise -0,13'ten büyüktür. Ekonomik rantabilite oranının negatif olması firmaların faaliyetlerinin finansmanında ağırlıklı olarak yabancı kaynak kullandığını göstermektedir. Ayrıca söz konusu firmaların stok devir hızının negatif olması stok tutma maliyetini yükselterek, firmanın finansman gereksinimini arttırabilir ve stokta bulunan malların çeşitli nedenlerle satış kabiliyetini yitirme olasılığını yükselebilir. İşletme sermayesinin net satışlara oranı pozitif olan söz konusu firmaların satışlarının işletme sermayesini karşılayacak düzeyde olmadığını ya da ihtiyaçtan fazla net işletme sermayesi bulunduğunu, alacakların ve stokların dönüş hızının yeterli olmadığını göstermektedir.

Profil 6 (Node 7): Bu profil, işletme sermayesinin net satışlara oranı ve stok devir hızı değişkenlerinden oluşmaktadır. Altıncı profili oluşturan işletmelerin %71,4'ü (5 işletme) sanayi sektöründe, %28,6'sı (2 işletme) hizmet sektöründe yer almaktadır. Bu profilde yer alan işletmelerin işletme sermayesinin net satışlara oranı 0,09 ile 0,79 arasında iken, stok devir hızları -2,72'den büyüktür. İşletmelerin, satışlarından çok özsermayesine bağlı kalması ve stok devir hızının düşük olması kredi veren kurumlar açısından riskli bir durum oluşturmaktadır. İşletmelerin stok devir hızını arttırıcı önlemler alması gerekmektedir.

Profil 7 (Node 4): Toplam 17 işletmenin yer aldığı yedinci profili, işletme sermayesinin net satışlara oranı 0,79'dan büyük olan işletmeler oluşturmaktadır. Bu profildeki işletmelerin %41,2'si (7 işletme) sanayi sektöründe, %58,8'i (10 işletme) hizmet sektöründe yer almaktadır. İşletme sermayesinin net satışlara oranının çok yüksek olması işletmelerin satışlarının çok düşük olduğunu ya da sermayesinin

ihtiyacın çok üzerinde olduğunu göstermektedir. Karlılığını satışlarından çok sermayeye bağımlı olması işletmelerin faaliyetlerini devam ettirmesi ve kredibilitesini koruması bakımından yüksek risk oluşturmaktadır.

3.2.5.3. CHAID Karar Ağacı ile Elde Edilen İMKB100 KVB/TA Sermaye Yapısı Sonuçları

Şekil 3.2’de verilen CHAID karar ağacı ile işletmelerin kısa vadeli borç oranı profilleri oluşturulmuş ve işletmeler 9 farklı profilde gruplanmıştır.

Tablo 3.8’de CHAID karar ağacı ile elde edilen İMKB100 endeksindeki 9 işletme profiline, profili oluşturan düğümlere ve kısa vadeli borç oranı değişkenine etki eden değişkenlere yer verilmiştir. Profilleri ve kısa vadeli borç oranını değerlendirmek için Tablo 3.8 ile Şekil 3.2 birlikte incelenmektedir. Şekil 3.2, 6 ana daldan oluşan CHAID karar ağacını ve profillerin tamamını göstermektedir.

İMKB100 sanayi ve hizmet sektörlerinde faaliyet gösteren işletmelerin *KVB/TA* sermaye yapısını tanımlamada istatistik olarak en önemli değişken cari oran değişkenidir. Tüm işletmeler için *KVB/TA* ortalaması 0,301 iken standart sapması 0,199’dur. Cari oran değerlerine göre ağacın dalları altı ana dalda (Node 1-Node 6) toplanmıştır. Bu altı düğümü oluşturan alt düğümler de profilleri oluşturmuştur. *KVB/TA* sermaye yapısına ilişkin profillerinin oluşturulmasında etkili olan en önemli değişkenler sırasıyla cari oran, aktif devir hızı, özsermaye devir hızı ve işletme sermayesinin net satışlara oranı değişkenleridir.

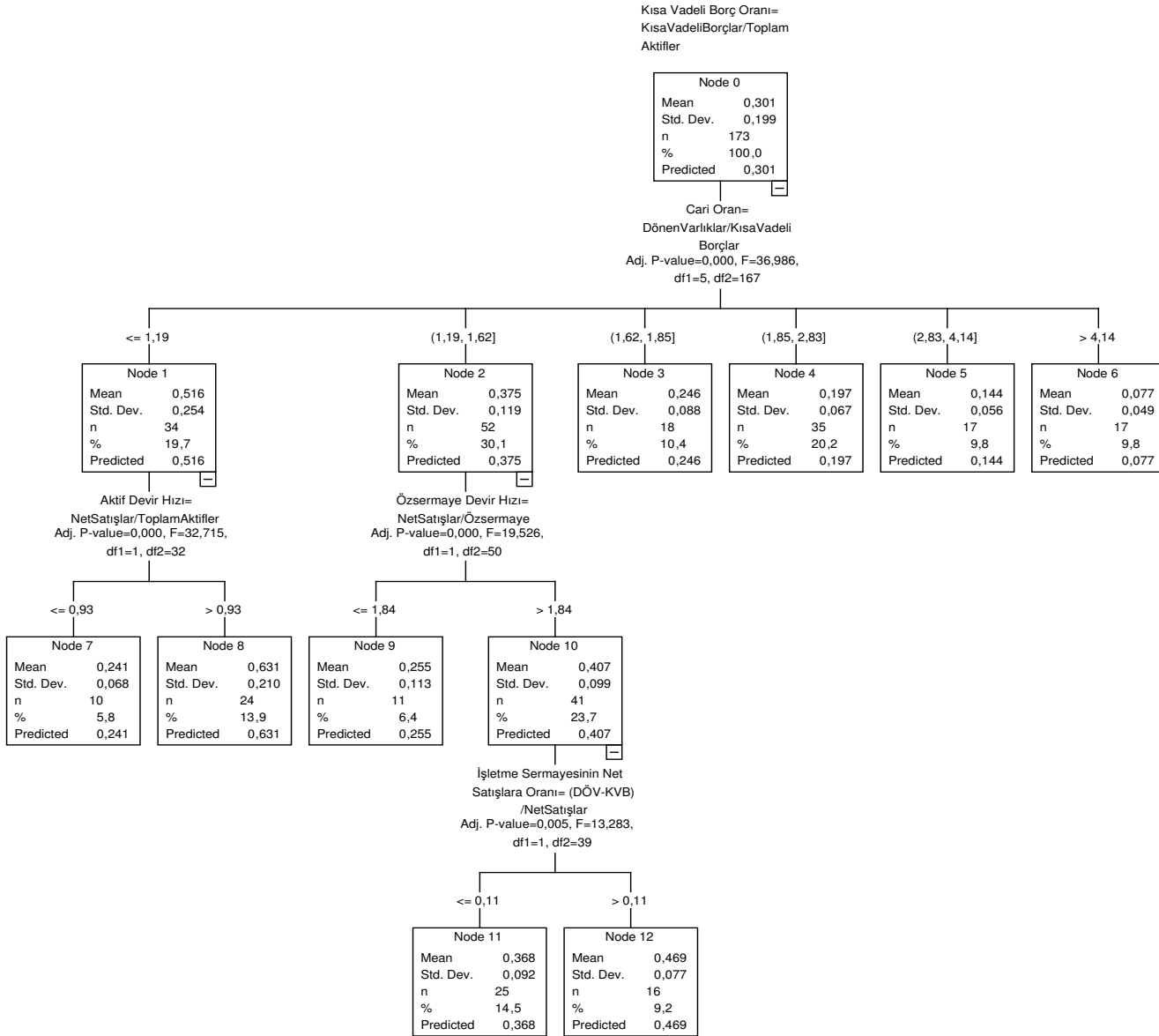
Profil 1 (Node 7) ve 2 (Node 8): 34 işletmenin (%19,7) yer aldığı 1. ve 2. profillerdeki işletmelerin *KVB/TA* sermaye yapısını tanımlamada en önemli değişkenler cari oran ve aktif devir hızıdır. Bu profili oluşturan işletmelerin, cari oranı 1,19’a eşit ve küçük iken birinci profildeki işletmelerin aktif devir hızı 0,93’e eşit ve küçük, ikinci profildeki işletmelerin aktif devir hızı 0,93’ten büyüktür. Toplam kısa vadeli borç oranı ortalaması 0,516 ve standart sapması 0,254 olan birinci ve ikinci profildeki işletmelerin kısa vadeli borç oranı ortalamaları (ve standart sapmaları) sırasıyla 0,241 (0,068) ve 0,631 (0,210)’dur. Cari oran işletmelerin borçlarını ödeyebilme gücünü göstermektedir. Bu oranın batılı finans

kurumlarınca 2 olması genel kabul gördüğü halde, yüksek enflasyonlu ve kıt fon kaynaklarına sahip ülkelerde endüstri ve sektörlerin değişik özellikleri göz önünde bulundurulduğunda cari oranın 1,5 olması genellikle yeterli olarak kabul edilmektedir. Bu nedenle bu profillerdeki işletmelerin borç ödeyebilme gücü yetersizdir. Aktif devir hızı işletmelerin sahip oldukları varlıklar ile yarattıkları satış hacmi başarısını ölçmektedir. Birinci profilde yer alan işletmelerin aktif devir hızının düşük olması, işletmelerin varlıklarına aşırı yatırım yapıldığının göstergesidir. Ayrıca, işletme varlıkları içerisinde duran varlıklar önemli bir yer tutuyorsa, aktif devir hızının düşmesine sebep olmaktadır. Bu profildeki işletmelerin gelecekteki kar durumu, talebin gelecekteki gelişimine bağlı olduğundan belirsizlik ifade etmektedir. İkinci profilde yer alan işletmelerin aktif devir hızının yüksek olması, işletmelerin karlılık oranının yüksek olduğunu göstermektedir. Ancak cari oranın düşük olması, firmaların sabit yatırımlarının fazla olması nedeniyle aktif devir hızlarının yükselmiş olabileceğini de göstermektedir.

Tablo 3.8: Kısa Vadeli Borç Oranı Profilleri ve Profilleri Oluşturan Değişkenler

Profiller	Düğüm	Cari Oran	Aktif Devir Hızı	Özsermaye Devir Hızı	İşletme Sermayesinin Net Satışlara Oranı
Profil 1	1, 7	$\leq 1,19$	$\leq 0,93$		
Profil 2	1, 8	$\leq 1,19$	$> 0,93$		
Profil 3	2, 9	1,19 – 1,62		$\leq 1,84$	
Profil 4	2, 10, 11	1,19 – 1,62		$> 1,84$	$\leq 0,11$
Profil 5	2, 10, 12	1,19 – 1,62			$> 0,11$
Profil 6	3	1,62 – 1,85			
Profil 7	4	1,85 – 2,83			
Profil 8	5	2,83 – 4,14			
Profil 9	6	$> 4,14$			

Şekil 3.2: Kısa Vadeli Borç Oranı Karar Ağacı *



* Şekilde geçen Node, Mean, Std. Dev., Predicted İngilizce sözcükleri sırasıyla düğüm, ortalama, standart sapma ve tahmini değer anlamına gelmektedir.

Profil 3 (Node 9), 4 (Node 11) ve 5 (Node 12): Tablo 3.8 ve Şekil 3.2’de görüleceği gibi toplam 52 işletmeden oluşan profilleri oluşturan en önemli değişkenler cari oran, özsermaye devir hızı ve işletme sermayesinin net satışlara oranıdır. Bu profilde yer alan tüm işletmelerin cari oran değeri 1,19 ve 1,62 arasında iken, kısa vadeli borç oranı ortalaması 0,375 ve standart sapması 0,119’dur. Özsermaye devir hızı 1,84’e eşit ya da küçük olan 11 işletme (%6,4) üçüncü profili oluştururken, bu profilin cari oran ortalaması 0,255 ve standart sapması 0,113’tür. Dördüncü ve beşinci profillerdeki işletmelerin cari oran ortalaması 0,407 ve standart sapması 0,099’dur. Özsermaye devir hızı 1,84’ten büyük olan dördüncü profilde yer alan 25 işletmenin (%14,5) işletme sermayesinin net satışlara oranı 0,11’e eşit ya da küçük iken, beşinci profilde yer alan 16 işletmenin (%9,2) ise 0,11’den büyüktür. Dördüncü ve beşinci profilde yer alan işletmelerin kısa vadeli borç oranı ortalamaları (ve standart sapmaları) sırasıyla 0,368 (0,092) ve 0,469 (0,077)’dir. Üçüncü profildeki işletmelerin cari oranı yeterli olmakla birlikte özsermaye devir hızının düşük olması özsermayelerinin verimli kullanılmadığını göstermektedir. Cari oranın düşmesi beklentisi işletmelerin kısa vadeli borçlarını ödemediği zorluğa düşeceklerinin göstergesi olabilmektedir. Buna karşın özsermaye devir hızının yüksek olduğu görülen dördüncü ve beşinci profillerde, sermayenin verimli kullanıldığı görülmektedir. İşletme sermayesinin net satışlara oranı %20 düzeyinde olması kabul görmektedir. Bu değerin çok düşük olduğu dördüncü profildeki işletmeler, cari oran ve özsermaye devir hızları yüksek olmasına karşın işletme sermayesine bağlı olmadan satışlarını sürdürebilmek ve faaliyetlerinin dönüş hızını koruyabilmek için gerekli önlemler alınmalıdır. Beşinci profilde işletme sermayesinin net satışlarına oranının %11’den yüksek olması, değer artışı ve satış oranlarının sermayeye bağımlı olmadan, yeterli bir satış düzeyine ulaşabildiği ve satışların, işletme faaliyetleriyle gerçekleşebileceği görülmektedir.

Profil 6 (Node 3) ve 7 (Node 4) : Cari oran değeri 1,62 ve 1,85 arasında olan 18 işletme (%10,4) altıncı profili, cari oran değeri 1,85 ve 2,83 arasında olan 35 işletme (%20,2) ise yedinci profili oluşturmaktadır. Bu işletmelerin kısa vadeli borç oranı ortalamaları (ve standart sapmaları) sırasıyla 0,246 (0,088) ve 0,197 (0,067)’dir. Cari oran değerinin batılı finans kurumlarınca 2 olması genel kabul

gördüğü halde, yüksek enflasyonlu ve kıt fon kaynaklarına sahip ülkelerde endüstri ve sektörlerin değişik özellikleri göz önünde bulundurulduğunda cari oranın 1,5 olması genellikle yeterli olarak kabul edilmektedir. Bu nedenle cari oran değeri en iyi olan, yükümlülüklerini karşılamada en iyi durumda olan işletmeler altıncı ve yedinci profillerde yer almaktadır. Ancak, yedinci profildeki işletmelerin ihtiyaçtan fazla sermayeye sahip olma olasılığı artmaktadır.

Profil 8 (Node 5) ve 9 (Node 6): Sekizinci profil cari oran değeri 2,83 ve 4,14 arasında olan 17 işletmeden (%9,8) ve dokuzuncu profil cari oran değeri 4,14'ten büyük olan 17 işletmeden (%9,8) oluşmaktadır. Bu işletmelerin kısa vadeli borç oranı ortalamaları (ve standart sapmaları) sırasıyla 0,144 (0,056) ve 0,077 (0,049)'dur. Bu profilde yer alan işletmelerin kısa vadeli borçları ödemede likitide durumunun çok yüksek olduğu görülmektedir. Ayrıca, fazla değer artışı işletmelerin yatırım yapmakta başarısız olduğunu göstermektedir.

3.2.5.4. CHAID Karar Ağacı ile Elde Edilen İMKB100 TB/TA Sermaye Yapısı Sonuçları

Şekil 3.3'te verilen CHAID karar ağacı ile işletmelerin toplam borç oranı profilleri oluşturulmuş ve işletmeler 9 farklı profilde gruplanmıştır.

Tablo 3.9'da CHAID karar ağacı ile elde edilen İMKB100 endeksindeki 9 işletme profiline, profili oluşturan düğümlere ve toplam borç oranı değişkenine etki eden değişkenlere yer verilmiştir. Profilleri ve toplam borç oranını değerlendirmek için Tablo 3.9 ile Şekil 3.3 birlikte incelenmektedir. Şekil 3.3, 6 ana daldan oluşan CHAID karar ağacını ve profillerin tamamını tek bir şekilde göstermektedir.

İMKB100 endeksindeki şirketlerin profillendirilmesi, toplam borç oranı durumu ile istatistik olarak en önemli değişkenin cari oran ($p=0,000$) olduğu anlaşılmaktadır. Tüm işletmelerin *TB/TA* ortalaması 0,421 iken standart sapması 0,216'dır. Cari oran değerlerine göre ağacın dalları altı ana dalda (Node 1-Node 6) toplanmıştır. Bu dört düğümü oluşturan alt düğümler de profilleri oluşturmuştur. *TB/TA* sermaye yapısına ilişkin profillerinin oluşturulmasında etkili olan en önemli

değişkenler sırasıyla cari oran, işletme sermayesi devir hızı, işletme büyüklüğü (toplam aktifler) ve aktif devir hızı değişkenleridir.

Profil 1 (Node 1): Bu profilde yer alan 34 işletmenin (%19,7) ortalama toplam borç oranı 0,65 iken standart sapması ise 0,217'dir. Standart sapma istatistiği (0,217) en büyük olan işletmeler bu profilde yer almaktadır. Bu durum birinci profildeki işletmelerin cari oran açısından türdeş olmayan bir yapıya sahip olduğunu göstermektedir. Cari oran, işletmenin likitide durumunu yansıtmaktadır. İşletmenin çalışma sermayesinin yeterliliği konusunda bilgi vermektedir. Oranın 1'den büyük olması olumlu bir işarettir. Ancak zamanla azalması sonucunda nakit sıkıntısı yaşanabilmektedir.

Tablo 3.9: Toplam Borç Oranı Profilleri ve Profilleri Oluşturan Değişkenler

Profiller	Düğüm	Cari Oran	İşletme Sermayesi Devir Hızı	İşletme Büyüklüğü	Aktif Devir Hızı
Profil 1	1	$\leq 1,19$			
Profil 2	2, 7	1,19 – 1,52	$\leq 2,01$		
Profil 3	2, 8	1,19 – 1,52	$> 2,01$		
Profil 4	3, 9	1,52 – 1,85		$\leq 3,1 \text{ E8}$	
Profil 5	3, 10	1,52 – 1,85		$> 3,1 \text{ E8}$	
Profil 6	4, 11	1,85 – 2,83			$\leq 0,38$
Profil 7	4, 12	1,85 – 2,83			$> 0,38$
Profil 8	5	2,83 – 4,14			
Profil 9	6	$> 4,14$			

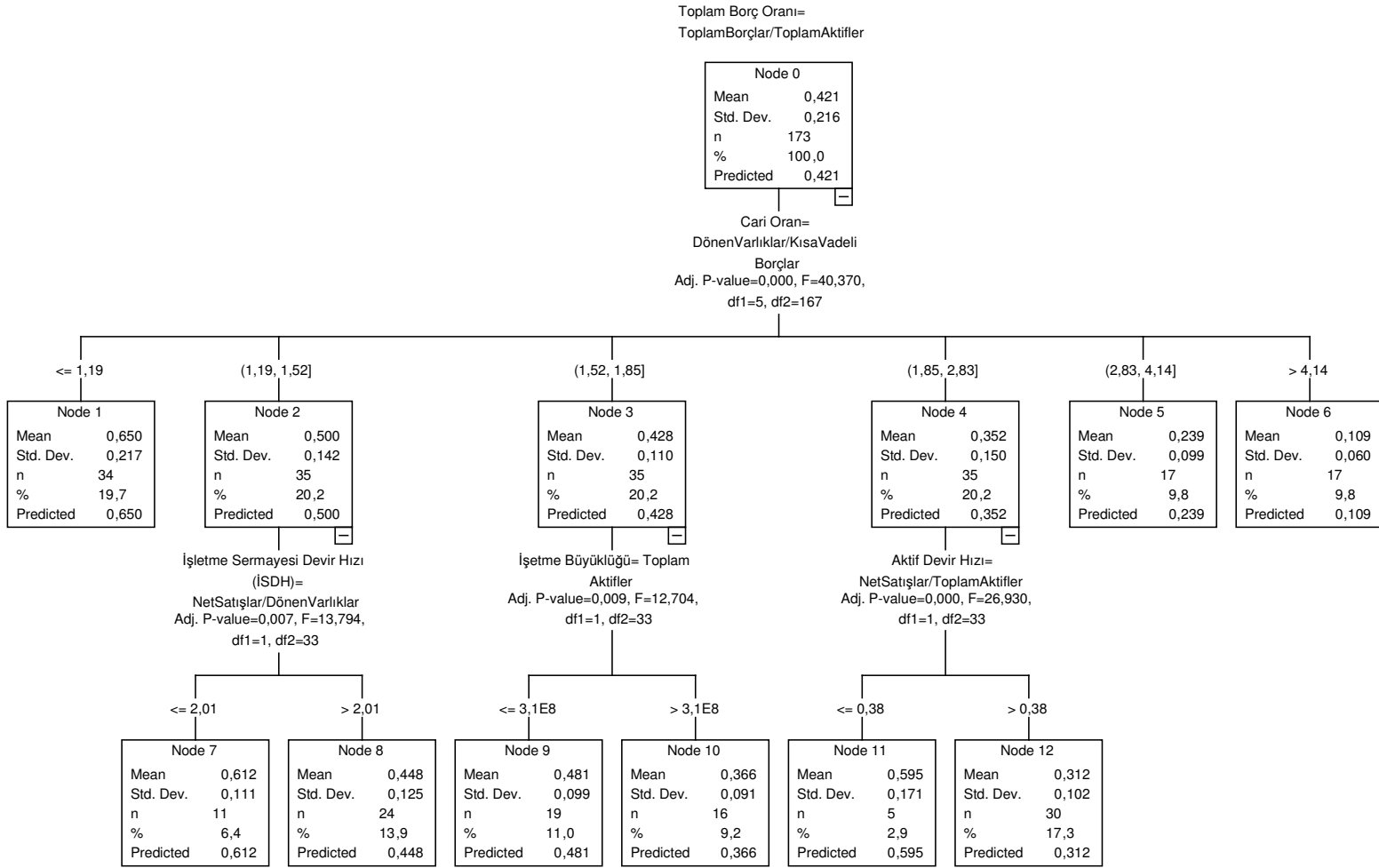
Profil 2 (Node 7) ve 3 (Node 8): İkinci ve üçüncü profilde yer alan toplam 35 işletmenin (%20,2) TB/TA oranını tanımlayan en önemli göstergeleri cari oran ve işletme sermayesi devir hızı değişkenleridir. Profilleri oluşturan işletmelerin, cari oranı 1,19 ile 1,52 arasındadır. İkinci profilde yer alan toplam 11 işletmenin İSDH oranı 2,01'e eşit ya da küçük iken, üçüncü profilde yer alan 24 işletmenin İSDH oranı 2,01'den büyüktür. İkinci ve üçüncü profillerde yer alan işletmelerin toplam borç oranı ortalamaları (ve standart sapmaları) sırasıyla 0,612 (0,111) ve 0,448 (0,125)'dir. Bu profillerde, alacak ve stoklara ayrılması gereken net işletme

sermayesinin çok olması, net işletme sermayesi devir hızının düşmesine sebep olmaktadır. Cari oranın da yeterli olması sonucunda bu işletmelerin faaliyetlerini rahatlıkla gerçekleştirebildiği görülmektedir.

Profil 4 (Node 9) ve 5 (Node 10): Toplam 35 işletmenin (%20,2) yer aldığı dördüncü ve beşinci profillerdeki işletmelerin *TB/TA* sermaye yapısını etkileyen en önemli değişkenler sırasıyla cari oran ve işletme büyüklüğü (toplam aktifler) değişkenleridir. Bu iki profili oluşturan işletmelerin, cari oranı 1,52 ile 1,85 arasında değişmekte ve dördüncü profilde işletmelerin aktifleri 310.000.000 YTL'ye eşit ya da küçük iken beşinci profilde işletmelerin aktifleri 310.000.000 YTL'den büyüktür. Dördüncü profildeki 19 işletmenin ve beşinci profildeki 16 işletmenin toplam borç oranı ortalamaları (ve standart sapmaları) sırasıyla 0,481 (0,099) ve 0,366 (0,091)'dir. Cari oranın 2'ye yakın olması işletmeler için olumlu bir gösterge olmasıyla birlikte işletmenin aktifleri küçülse de büyüse de işletmelerin borçlarını ödeyebilme yeteneğini sürdürebilmesi için korunmalıdır.

Profil 6 (Node 11) ve 7 (Node 12): Altıncı ve yedinci profillerde yer alan toplam 35 işletmenin (%20,2) *TB/TA* oranını tanımlayan en önemli göstergeler cari oran ve aktif devir hızı değişkenleridir. Bu profili oluşturan işletmelerin, cari oranı 1,85 ile 2,83 arasında değişirken, toplam borç oranı ortalaması 0,352 ve standart sapması 0,150'dir. Altıncı profilde yer alan toplam 5 işletmenin (%2,9) aktif devir hızı oranı 0,38'e eşit ya da küçük iken, yedinci profilde yer alan 30 işletmenin (%17,3) aktif devir hızı oranı 0,38'den büyüktür. Altıncı ve yedinci profillerde yer alan işletmelerin toplam borç oranı ortalamaları (ve standart sapmaları) sırasıyla 0,595 (0,171) ve 0,312 (0,102)'dir. Aktif devir hızı işletmelerin varlıklarının verimliliğini göstermektedir. Bu değer yüksek olması, verimliliğin iyi olduğunu göstermektedir. Cari oran değeri yüksek olmakla birlikte 2'den yüksek olan değerler gereksiz olarak değerlendirilmektedir. İşletmeler, fazla likit varlıklarını yatırım amaçlı kullanmalıdırlar.

Şekil 3.3: Toplam Borç Oranı Karar Ağacı *



* Şekilde geçen Node, Mean, Std. Dev., Predicted İngilizce sözcükleri sırasıyla düğüm, ortalama, standart sapma ve tahmini değer anlamına gelmektedir.

Profil 8 (Node 5) ve 9 (Node 6): Her ikisi de 17 işletmeden (%9,8) oluşan profillerden cari oranı 2,83 ile 4,14 arasında olan firmalar sekizinci profili, cari oranı 4,14'ten büyük olan firmalar dokuzuncu profili oluşturmaktadır. Sekizinci ve dokuzuncu profillerde yer alan işletmelerin toplam borç oranı ortalamaları (ve standart sapmaları) sırasıyla 0,239 (0,099) ve 0,109 (0,060)'tır. Bu profilde yer alan işletmelerin kısa vadeli borçları ödemede likitide durumunun çok yüksek olduğu görülmektedir. Ancak fazla değer artışı işletmelerin yatırım yapmakta başarısız olduğunu göstermektedir.

3.2.5.5. CHAID Karar Ağacı ile Elde Edilen İMKB100 NK/TA Karlılık Sonuçları

Şekil 3.4'te verilen CHAID karar ağacı ile işletmelerin aktiflerin karlılığı profilleri oluşturulmuş ve işletmeler 12 farklı profilde gruplanmıştır.

Tablo 3.10'da CHAID karar ağacı ile elde edilen İMKB100 endeksindeki 12 işletme profiline, profili oluşturan düğümlere, aktiflerin karlılığı değişkenine etki eden en önemli değişkenlere yer verilmiştir. Profilleri ve aktiflerin karlılığını değerlendirmek için Tablo 3.10 ile Şekil 3.4 birlikte incelenmektedir. Şekil 3.4, 2 ana daldan oluşan CHAID karar ağacını ve profillerin tamamını göstermektedir.

İMKB100 endeksindeki şirketlerin profillendirilmesi, aktiflerin karlılığı ile istatistik olarak en önemli değişkenin cari oran ($p=0,000$) olduğu anlaşılmaktadır. Tüm işletmeler için *NK/TA* ortalaması 0,061 ve standart sapması 0,129'dur. Cari oran değerlerine göre ağacın dalları iki ana dalda (Node 1-Node 2) toplanmıştır. Bu iki düğümü oluşturan alt düğümler de profilleri oluşturmaktadır. *NK/TA* karlılığına ilişkin profillerinin oluşturulmasında etkili olan en önemli değişkenler sırasıyla cari oran, toplam borç oranı, özsermaye devir hızı, kısa vadeli borç oranı, dönen varlıklar/toplam aktifler oranı, işletme sermayesinin net satışlara oranı ve işletme sermayesi devir hızı değişkenleridir.

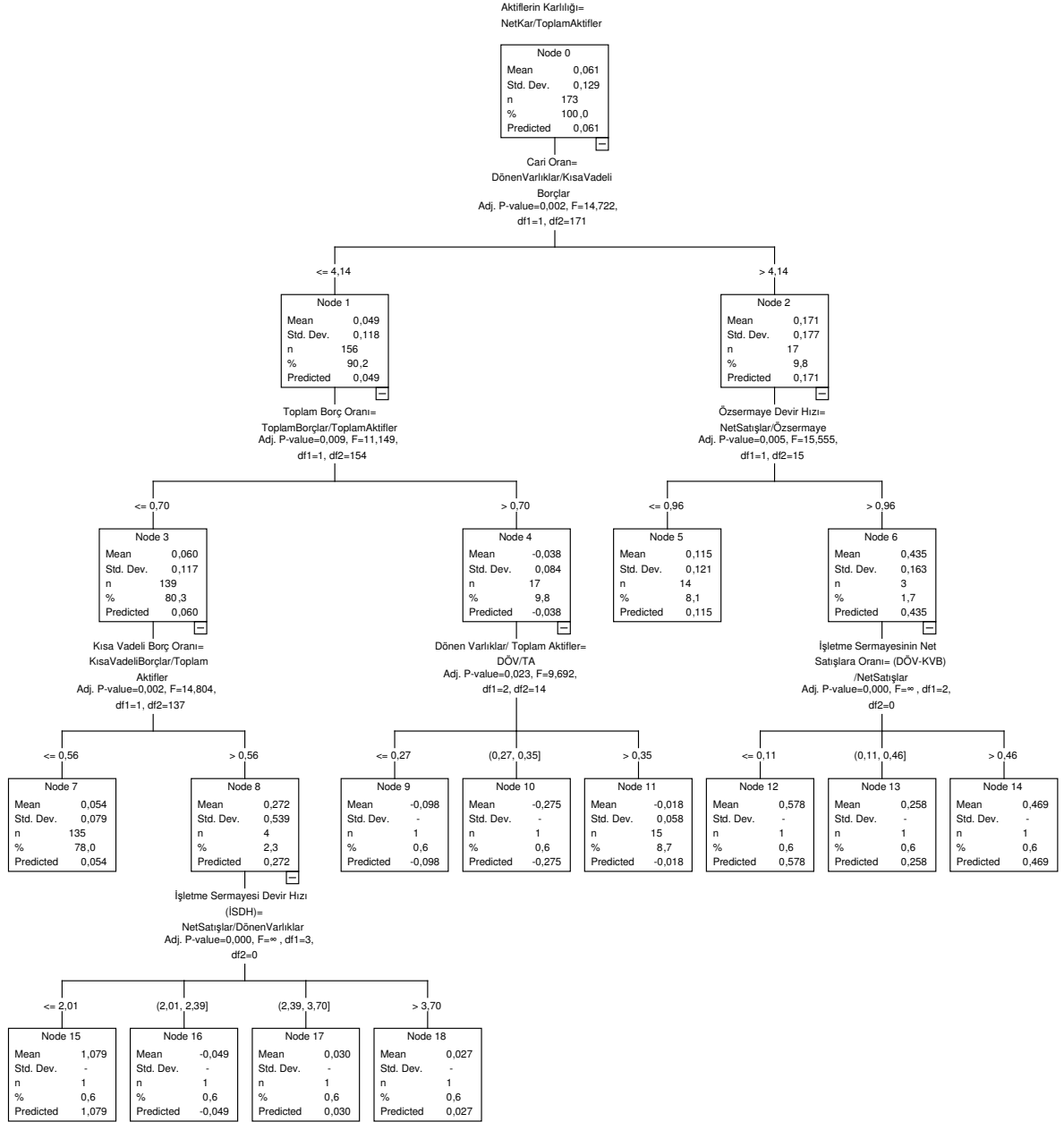
Profil 1 (Node 7): Tüm işletmelerin %78'ini oluşturan toplam 135 işletmenin yer aldığı birinci profildeki işletmelerin aktiflerin karlılığını tanımlamada istatistik olarak en önemli değişkenler sırasıyla cari oran, toplam borç oranı ve kısa vadeli

borç oranıdır. Bu profilde yer alan işletmelerin, cari oranı 4,14'e eşit ya da küçük, toplam borç oranı 0,70'e eşit ya da küçük ve kısa vadeli borç oranı 0,56'ya eşit ya da küçüktür. Ayrıca bu profilde yer alan toplam 135 işletmenin aktiflerin karlılığı ortalaması (ve standart sapması) sırasıyla 0,054 (0,079)'dur. Bu profilde yer alan işletmelerin cari oran değeri, bir başka deyişle, borçlarını ödeyebilme gücü oldukça yüksektir. Toplam borç oranı değerinin %70'den küçük olması olumlu değerlendirilirken enflasyonist ülkelerde bu oranın %60'tan yüksek olması normal karşılanmaktadır. Kısa vadeli borç oranının %33'ten küçük olması beklenirken bu oranın da %56'dan küçük olması normal görülebilir. Birinci profildeki işletmelerin toplam borç oranı olumlu düzeydeyken kısa vadeli borç oranı da yeterli seviyededir. Bu nedenle birinci profildeki işletmelerin mali yapısının oldukça iyi olduğu söylenebilir.

Tablo 3.10: Aktiflerin Karlılığı Profilleri ve Profilleri Oluşturan Değişkenler

Profiller	Düğüm	Cari Oran	Toplam Borç Oranı	Kısa Vadeli Borç Oranı	İşletme Sermayesi Devir Hızı	Dönen Varlıklar/ Toplam Aktifler	Özsermaye Devir Hızı	İşletme Sermayesinin Net Satışlara Oranı
Profil 1	1, 3, 7	≤ 4,14	≤ 0,70	≤ 0,56				
Profil 2	1, 3, 8, 15	≤ 4,14	≤ 0,70	> 0,56	≤ 2,01			
Profil 3	1, 3, 8, 16	≤ 4,14	≤ 0,70	> 0,56	2,01 – 2,39			
Profil 4	1, 3, 8, 17	≤ 4,14	≤ 0,70	> 0,56	2,39 – 3,70			
Profil 5	1, 3, 8, 18	≤ 4,14	≤ 0,70	> 0,56	> 3,70			
Profil 6	1, 4, 9	≤ 4,14	> 0,70			≤ 0,27		
Profil 7	1, 4, 10	≤ 4,14	> 0,70			0,27 – 0,35		
Profil 8	1, 4, 11	≤ 4,14	> 0,70			> 0,35		
Profil 9	2, 5	> 4,14					≤ 0,96	
Profil 10	2, 6, 12	> 4,14					> 0,96	≤ 0,11
Profil 11	2, 6, 13	> 4,14					> 0,96	0,11 – 0,46
Profil 12	2, 6, 14	> 4,14					> 0,96	> 0,46

Şekil 3.4: Aktiflerin Karlılığı Karar Ağacı *



* Şekilde geçen Node, Mean, Std. Dev., Predicted İngilizce sözcükleri sırasıyla düğüm, ortalama, standart sapma ve tahmini değer anlamına gelmektedir.

Profil 2 (Node 15), 3 (Node 16), 4 (Node 17) ve 5 (Node 18): Toplam 4 işletmenin (%2,3) yer aldığı profillerde *NK/TA* karlılığını tanımlayan değişkenler sırasıyla cari oran, toplam borç oranı, kısa vadeli borç oranı ve işletme sermayesi devir hızı (İSDH) değişkenleridir. Bu profildeki işletmelerin cari oranı 4,14'e eşit ya da küçük, toplam borç oranı 0,70'e eşit ve küçük, kısa vadeli borç oranı 0,56'dan büyük iken, işletme sermayesi devir hızları sırasıyla 2,01'e eşit veya daha küçük, 2,01-2,39, 2,39-3,70 ve 3,70'den büyüktür. Bu profillerdeki işletmelerin aktiflerin karlılığı ortalamaları (ve standart sapmaları) sırasıyla 0,049 (0,118), 0,060 (0,117) ve 0,272 (0,539)'dur. Kısa vadeli borç oranı ortalamaları sırasıyla 1,079, -0,049, 0,030 ve 0,027'dir. İkinci profilde İSDH 2,01'den küçük 1 işletme (%0,6), üçüncü profilde İSDH 2,01 ve 2,39 arasında olan 1 işletme (%0,6), dördüncü profilde İSDH 2,39 ve 3,70 arasında olan 1 işletme (%0,6), beşinci profilde İSDH 3,70'ten büyük 1 işletme (%0,6) yer almaktadır. Bu profillerdeki işletmelerin kısa vadeli borçlarını ödeyebilme gücü oldukça iyi ve toplam borç oranı da olumlu düzeydedir. Ancak işletmelerin kısa vadeli borç oranı yüksektir, bu nedenle bu profillerdeki işletmelerin kredi veren kurumlar açısından riskli durumda olduğu görülmektedir. İSDH değeri, şirketlerin faaliyetlerini yürütebilmeleri için gereken minimum dönen varlıkları (işletme sermayesi) ile hangi ölçüde satış hacmi başarısı gösterdiklerini ölçen bir orandır. Bu profillerdeki İSDH değerinin yüksek olması dönen varlıkların satışlara bağlı olarak arttığını göstermektedir. Cari oran değerinin olumlu olması etkin işletme sermayesi yönetimi sonucunu göstermektedir. İkinci, üçüncü, dördüncü ve beşinci profillerin hepsinde İSDH yüksek iken dönen varlıkların hızlı artması İSDH oranını düşürmektedir. Bu nedenle ek finansman desteğine ihtiyaç duyulmasına yol açmakta ve kar marjı düşmektedir. İkinci profildeki düşme eğilimi satışlar artarken dönen varlıkların daha hızlı arttığını ve stok devir hızının yavaşlamasını göstermektedir. Buna karşın üçüncü ve dördüncü profilde İSDH oranı artmakta ve beşinci profilde en yüksek düzeye çıkmaktadır. Bu nedenle birinci profildeki işletmelerin kar marjı en düşük iken beşinci profilde yer alan işletmelerin kar marjı en yüksektir.

Profil 6 (Node 9), 7 (Node 10) ve 8 (Node 11): *NK/TA* karlılığını tanımlayan cari oran, toplam borç oranı ve dönen varlıklar/toplam aktifler değişkenlerinden oluşan bu profillerde toplam 17 işletme (%9,8) yer almaktadır. Bu profillerdeki

işletmelerin cari oranı 4,14'ten küçük, toplam borç oranı 0,70'ten büyüktür. Dönen varlıklar/toplam aktifler oranı 0,27'den küçük olan 1 işletme (%0,6) altıncı profili, 0,27 ve 0,35 arasında olan 1 işletme (%0,6) yedinci profili, 0,35'ten büyük olan 15 işletme (%8,7) sekizinci profili oluşturmaktadır. Bu profillerdeki işletmelerin ortalama aktif karlılık oranları sırasıyla -0,098, -0,275 ve -0,018'dir. Altıncı, yedinci ve sekizinci profillerde yer alan işletmelerin borç ödeyebilme kapasitesi yüksek iken, toplam borç oranı değeri de yüksektir. Toplam borç oranının yüksek olması işletmeleri kredi verenler açısından riskli olduğunu gösterir. Dönen varlıklar/toplam aktifler oranı işletmelerin mali yapısı hakkında bilgi vermektedir. Altıncı ve yedinci profillerde bu oranın düşük olması dönen varlıkların toplam aktifler içerisinde düşük pay aldığını ve duran varlıklarının yüksek olduğunu göstermektedir. Sekizinci profilde ise bu pay artmakta ve işletmenin dönen varlıklarının toplam aktiflerin %35'inden fazlasını oluşturduğunu ve faaliyetlerini kendi kaynaklarıyla döndürebildiğini göstermektedir.

Profil 9 (Node 5): Tüm işletmelerin %8,1'ini oluşturan toplam 14 işletmenin yer aldığı dokuzuncu profilde, aktiflerin karlılığını tanımlamada istatistik olarak anlamlı değişkenler cari oran ve özsermaye devir hızıdır. Bu profillerde yer alan işletmelerin, cari oranı 4,14'ten büyük iken özsermaye devir hızı 0,96'dan küçüktür. Dokuzuncu profilde yer alan işletmelerin aktiflerin karlılığı ortalaması (ve standart sapması) sırasıyla 0,115 (0,121)'dir. Borçlarını ödemede likitide durumunun çok yüksek olduğu görülen dokuzuncu profildeki işletmelerin cari oran değerinin yüksek olması likit kaynakların atıl bir şekilde işletmede tutulduğunu göstermektedir. Özsermaye devir hızının düşük olması, işletmelerin özsermayelerini verimli kullanamadığının göstergesidir. Öz sermaye devir hızının düşüklüğü, atıl öz sermayenin olması sebebiyle gerçekleşmişse yöneticilerin öz sermayeyi azaltıcı (örneğin temettü dağıtarak) önlemler alması veya satış hacmini artırıcı önlemler alması gerekmektedir.

Profil 10 (Node 12), 11 (Node 13) ve 12 (Node 14): Cari oran, özsermaye devir hızı ve işletme sermayesinin net satışlara oranı değişkenlerinden oluşan bu profillerde toplam 3 işletme (%1,7) yer almaktadır. Bu profillerdeki işletmelerin cari

oranları 4,14'ten büyük, özsermaye devir hızları 0,96'dan büyüktür. Her birinde 1 işletme yer alan profillerde işletme sermayesinin net satışlara oranları sırasıyla 0,11'e eşit ve küçük, 0,11-0,46 ve 0,46'dan büyüktür. Bu profillerde yer alan işletmelerin aktiflerin karlılığı ortalamaları sırasıyla 0,578, 0,258 ve 0,469'dur. Onuncu, on birinci ve on ikinci profillerde yer alan işletmelerin cari oranlarının yeterli düzeyin çok üstünde olması gereğinden fazla likit varlıkların elde tutulduğunu ve bu nedenle olası gelirlerden işletmenin mahrum bırakıldığını göstermekle birlikte bu profillerde, özsermaye devir hızının yüksek olması özsermayenin etkin bir şekilde kullanıldığını ve özsermaye karlılığının yüksek olduğunu göstermektedir. Ayrıca işletme sermayesinin net satışlara oranı değişkeni incelendiğinde, onuncu profilde yer alan işletmelerin işletme sermayesinin az olduğu ve satışlarının yeterli düzeyde olmadığı, on birinci profildeki işletmelerin işletme sermayesinin yeterli düzeye geldiği ve işletme sermayesine bağlı olmadan satış hacminin iyi olduğu görülmektedir. Ancak bu profilde yer alan işletmelerin işletme sermayesinde görülen yüksek düzeydeki artış, satışların yetersizliğinden kaynaklanabilir. On ikinci profildeki işletmelerin işletme sermayesinin net satışlara oranının yüksek olması, işletme sermayesinin gereğinden fazla oluşunu ve satışların işletme sermayesine bağlı olarak gerçekleştiği görülmektedir.

3.2.5.6.CHAID Karar Ağacı ile Elde Edilen İMKB100 NK/ÖS Karlılık Sonuçları

Şekil 3.5'te verilen CHAID karar ağacı ile işletmelerin özsermayenin karlılığı profilleri oluşturulmuş ve işletmeler 5 farklı profilde gruplanmıştır.

Tablo 3.11'de CHAID karar ağacı ile elde edilen İMKB100 endeksindeki 5 işletme profiline, profili oluşturan düğümlere ve özsermayenin karlılığını tanımlayan en önemli değişkenlere yer verilmiştir. İMKB100 sanayi ve hizmet sektörlerinde faaliyet gösteren işletmelerin *NK/TA* karlılığını tanımlamada istatistik olarak en önemli değişken özsermaye devir hızı değişkenidir. Tüm işletmeler için *NK/ÖS* ortalaması 0,108 iken standart sapması 0,503'tür. Özsermaye devir hızı değerlerine göre ağacın dalları iki ana dalda (Node 1-Node 2) toplanmıştır. Bu iki düğümü oluşturan alt düğümler de profilleri oluşturmaktadır.

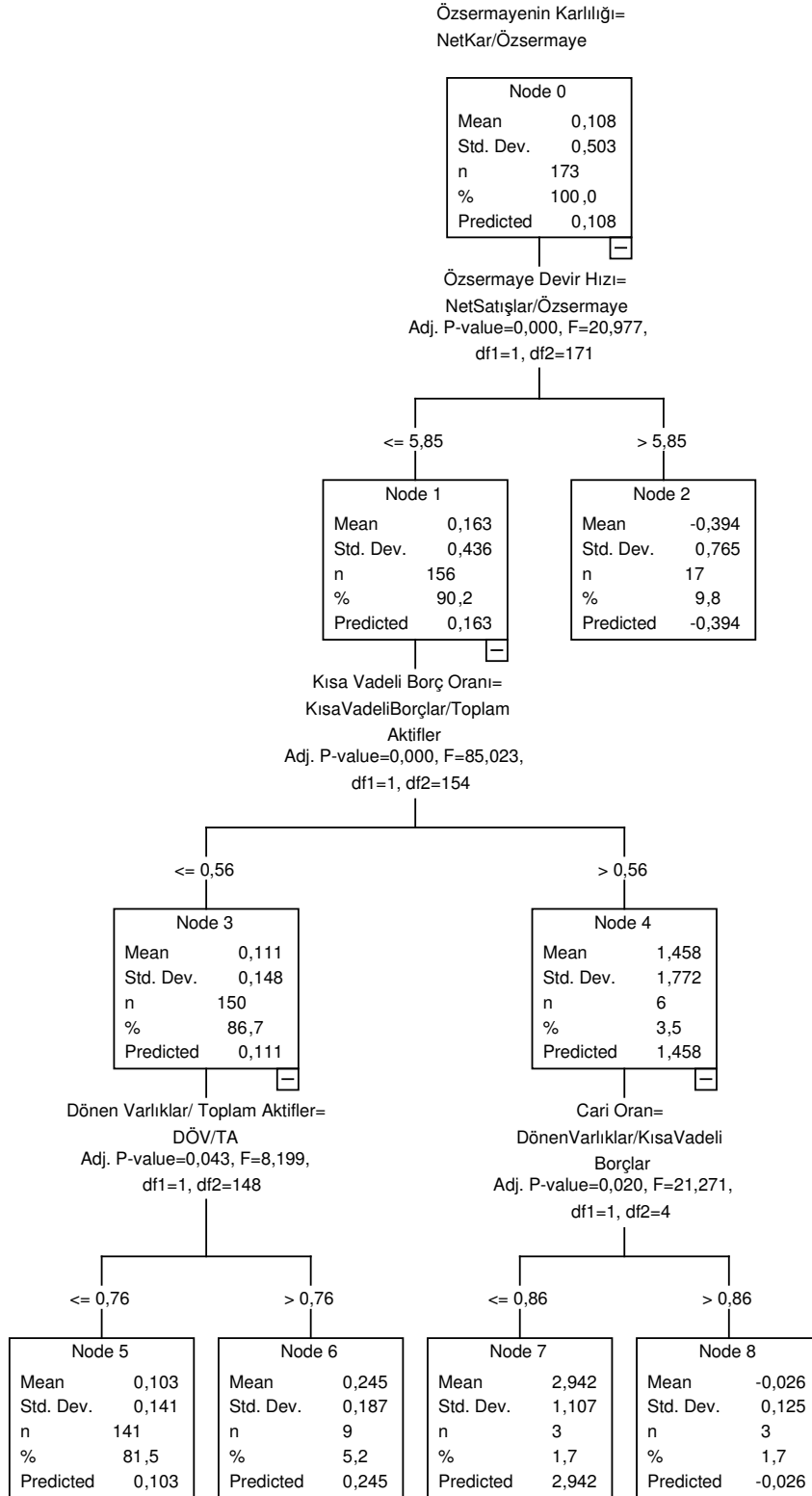
Tüm işletmelerin profilleri ve özsermayenin karlılığını değerlendirmek için Tablo 3.11 ile Şekil 3.5 birlikte incelenmektedir. Şekil 3.5, iki ana daldan oluşan CHAID karar ağacını ve profillerin tamamını göstermektedir. *NK/ÖS* karlılığına ilişkin profillerinin oluşturulmasında etkili olan en önemli değişkenler sırasıyla özsermaye devir hızı, kısa vadeli borç oranı, dönen varlıklar/toplam aktifler oranı ve cari oran değişkenleridir.

Profil 1 (Node 5) ve 2 (Node 6): Tüm işletmelerin %86,7'ini oluşturan toplam 150 işletmenin yer aldığı birinci ve ikinci profiller, özsermaye devir hızı, kısa vadeli borç oranı ve dönen varlıklar/toplam aktifler değişkenlerinden oluşmaktadır. Bu profilleri oluşturan işletmelerin, özsermaye devir hızları 5,85'e eşit ya da küçük, kısa vadeli borç oranı 0,56'ya eşit ya da küçük iken, birinci profilde dönen varlıklar/toplam aktifler oranı 0,76'ya eşit ya da küçük, ikinci profilde 0,76'dan büyüktür. Birinci ve ikinci profildeki işletmelerin özsermayenin karlılığı ortalamaları (ve standart sapmaları) sırasıyla 0,103 (0,141) ve 0,245 (0,187)'dir. Özsermaye devir hızı iyi olmasına karşın düşme eğilimi de gösteren profillerde bulunan bu işletmelerin kısa vadeli borç oranının iyi olması birinci ve ikinci profillerdeki işletmelerin sermayelerini verimli kullandığı görülmektedir. Dönen varlıklar/toplam aktifler değeri incelendiğinde birinci profildeki işletmelerin daha riskli durumda olduğu görülmektedir. İşletmelerin faaliyetlerini sürdürebilmek için sermayeye bağlı olmaksızın kendi faaliyet karlarıyla varlıklarını sürdürmeleri önemlidir. Bu nedenle, ikinci profildeki işletmelerin dönen varlıklarının toplam aktifleri içerisinde daha fazla yer alması mali yapılarının daha iyi olduğunun bir göstergesidir.

Tablo 3.11: Özsermayenin Karlılığı Profilleri ve Profilleri Oluşturan Değişkenler

Profiller	Düğümmler	Özsermaye Devir Hızı	Kısa Vadeli Borç Oranı	Dönen Varlıklar/ Toplam Aktifler	Cari Oran
Profil 1	1, 3, 5	$\leq 5,85$	$\leq 0,56$	$\leq 0,76$	
Profil 2	1, 3, 6	$\leq 5,85$	$\leq 0,56$	$> 0,76$	
Profil 3	1, 4, 7	$\leq 5,85$	$> 0,56$		$\leq 0,86$
Profil 4	1, 4, 8	$\leq 5,85$	$> 0,56$		$> 0,86$
Profil 5	2	$> 5,85$			

Şekil 3.5: Özsermayenin Karlılığı Karar Ağacı *



* Şekilde geçen Node, Mean, Std. Dev., Predicted İngilizce sözcükleri sırasıyla düğüm, ortalama, standart sapma ve tahmini değer anlamına gelmektedir.

Profil 3 (Node 7) ve 4 (Node 8): Tüm işletmelerin %3,5'ini oluşturan toplam 6 işletmenin yer aldığı üçüncü ve dördüncü profiller, özsermaye devir hızı, kısa vadeli borç oranı ve cari oran değişkenlerinden oluşmaktadır. Bu profilleri oluşturan işletmelerin, özsermaye devir hızları 5,85'e eşit ya da küçük, kısa vadeli borç oranı 0,56'dan büyük iken üçüncü profilde cari oran 0,86'ya eşit ya da küçük, dördüncü profilde 0,86'dan büyüktür. Bu iki profilde (Node 7 ve Node 8) yer alan 6 işletmenin özsermaye karlılığı ortalamaları (ve standart sapmaları) sırasıyla 2,942 (1,107) ve -0,026 (0,125)'dir. Özsermaye devir hızının düşük olması nedeniyle üçüncü ve dördüncü profillerdeki işletmelerin sermayelerini verimli kullanılmadığı görülmektedir. İşletmelerin toplam varlıkları içerisindeki kısa vadeli borç fonunu gösteren kısa vadeli borç oranı, riskin de göstergesi olduğu için bu profillerde yer alan 6 işletmenin riskli yapıda olduğu görülmektedir. Üçüncü profil cari oranla birlikte değerlendirildiğinde likitide durumunun oldukça riskli olduğu, dördüncü profildeki işletmelerin de cari oran değerinin yükselmesi üçüncü profile oranla daha iyi durumda olduğunu göstermektedir. Buna karşın dördüncü profilde yer alan işletmelerin de cari oran değerinin yeterli düzeyin altında olması bu profildeki işletmelerin riskli yapısını ortaya koymaktadır.

Profil 5 (Node 2): Özsermaye devir hızı 5,85'ten büyük olan 17 işletmenin (%9,8) yer aldığı beşinci profilde özsermayenin karlılığı ortalaması -0,394 ve standart sapması 0,765'tir. Özsermaye devir hızı bir işletmenin özsermayesini ne ölçüde verimli kullandığını göstermektedir. Özsermaye devir hızının yüksek olması işletmelerin özsermayelerinin etkin bir şekilde kullanıldığını göstermekle birlikte özsermayenin yetersiz olabileceğinin ve işletmelerin faaliyetlerinin finansmanında ağırlıklı olarak yabancı kaynakları kullanmasının bir sonucu da olabilmektedir. Büyük ölçüde yabancı kaynaklarla dönen bir işletmede ise özsermaye devir hızının yüksek olması, özsermayenin yetersiz olduğunu göstermektedir.

Bu bölümde İMKB sanayi ve hizmet sektörlerinde faaliyet gösteren 173 işletmeye ilişkin karlılık, likitide, varlık kullanım etkinliği, sermaye yapısı ve işletme büyüklüğü göstergelerinden yararlanarak işletmelerin sektör profillerinin, sermaye yapılarının (*KVB/TA* ve *TB/TA*) ve karlılık düzeylerinin (*NK/TA* ve *NK/ÖS*) en

önemli belirleyicileri veri madenciliği tekniklerinden CHAID karar ağacı algoritmasıyla araştırılmıştır. Araştırma sonuçlarına göre sektör profillerinin tanımlanmasında işletme sermayesinin net satışlara oranı, stok devir hızı ve ekonomik rantabilite göstergelerinin en önemli değişkenler olduğu görülmüştür.

Firmaların *KVB/TA* sermaye yapılarının ve profillerinin belirlenmesinde etkili olan en önemli değişkenlerin sırasıyla cari oran, aktif devir hızı, özsermaye devir hızı ve işletme sermayesinin net satışlara oranı olduğu görülmüştür. Firmaların *TB/TA* sermaye yapılarının ve profillerinin tanımlanmasında en önemli değişkenler ise sırasıyla şunlardır: Cari oran, işletme sermayesi devir hızı, işletme büyüklüğü (toplam aktifler) ve aktif devir hızı.

İMKB100 sanayi ve hizmet sektörlerinde faaliyette bulunan 173 işletmenin karlılık düzeylerinden aktiflerin karlılığını belirleyen en önemli göstergelerin sırasıyla cari oran, toplam borç oranı, özsermaye devir hızı, kısa vadeli borç oranı, dönen varlıkların toplam aktiflere oranı, işletme sermayesinin net satışlara oranı ve işletme sermayesi devir hızı değişkenleridir. Özsermaye karlılığını belirleyen en önemli değişkenler arasında ise sırasıyla özsermaye devir hızı, kısa vadeli borç oranı, dönen varlıkların toplam aktiflere oranı ve cari oran değişkenleri yer almaktadır.

SONUÇ

Verinin bilgiye dönüşümünün geleneksel yöntemi, klasik çözümleme ve yorumlamaya dayanmaktadır. Ancak günümüzde veri miktarındaki olağanüstü artış, sağlık, eğitim, ticaret, askeri alanlar, alışveriş, devlet sektörü, özel sektör ve pek çok alanda verilerin işlenmesi ve bu verilerin değerlendirilerek bilgiye dönüştürülmesi bir zorunluluk haline gelmiştir. Böylece bu alanlardaki gelişmeler daha hızlı ve verimli bir şekilde gerçekleşebilmektedir. Veri madenciliği bu noktada uygun teknikler kullanarak gizli kalmış bilgileri ortaya çıkarma özelliğiyle gereklidir. Veri madenciliğinde amaç önceden bilinmeyen bilgileri ortaya çıkarmak, yeni örüntülerden sonuçları tahmin etmek ya da tanımlamaktır. Veri madenciliği, bilgi teknolojilerinin doğal gelişim sürecinin sonucu olarak da değerlendirilebilir.

Veri madenciliği teknikleri diğer birçok alanda olduğu gibi işletmecilik alanında da yaygın bir şekilde kullanılmaktadır. Bu çalışmada, veri madenciliği tekniklerinden karar ağaçları tekniği kullanılarak İstanbul Menkul Kıymetler Borsası'ndan elde edilen veriler üzerinde yöntemin uygulanabilirliği gösterilmiştir. Karar ağacı algoritmalarının en önemli avantajları, parametrik olmayan yöntemler arasında olması nedeniyle diğer çok değişkenli tekniklerde sağlanması gereken istatistik varsayımların söz konusu olmamasıdır. Çünkü veri madenciliği sonucunda oluşturmuş olduğumuz tüm kurallar kesinlikle kullanılabilir veya kullanılamaz diye bir yargıya önceden varmamız mümkün değildir. Ayrıca karar ağacı algoritmalarının bağımlı ve bağımsız değişkenler arasındaki ilişkilerin yönünü, önem sırasını görselleştirmesi diğer avantajları arasındadır. Bu özelliği özellikle elde edilen sonuçların yorumunu oldukça basitleştirerek daha somut ve kullanışlı hale getirmektedir.

Çalışma kapsamına İstanbul Menkul Kıymetler Borsası Ulusal 100 endeksi sanayi ve hizmet sektörlerinde faaliyet gösteren 173 şirket alınmış ve şirketlere ait finansal bilgiler kullanılarak elde edilen verilere CHAID algoritması uygulanmış işletmelerin birbirlerine göre konumları ortaya konmuştur. CHAID ile diğer algoritmalar arasındaki en önemli farklılıklardan birisi, ağacın yapraklarının ikili değil verideki farklı yapı sayısı kadar dallanmasıdır. Bu özelliği nedeniyle, daha fazla

alt gruplarla değerlendirme yapmak ve daha homojen gruplardan sonuç çıkarmak mümkün olmaktadır. Bu bağlamda bu çalışmada öncelikle İMKB100 sanayi ve hizmet sektöründe faaliyet gösteren firmaları ayırmada etkili olan en önemli karlılık, likitide, varlık kullanım etkinliği, sermaye yapısı ve işletme büyüklüğü değişkenleri araştırılmıştır. Daha sonra sermaye yapısı göstergelerinden kısa vadeli borç oranını, toplam borç oranını ve karlılık göstergelerinden özsermayenin karlılığını ve aktiflerin karlılığını belirleyen en önemli finansal göstergeler araştırılmıştır.

CHAID algoritmasından elde edilen ağaç yapıları risk değerine göre belirlenmiştir. Risk değeri, yanlış sınıflandırma oranını göstermektedir ve en uygun ağaç yapısının seçiminde dengeli ve minimum olmalıdır. Bu nedenle farklı karar ağacı yapıları içerisinde en uygun olan ağaç yapısı, yerine koyma (resubstitution) ve çapraz geçerlilik (cross-validation) risklerinin birbirine en yakın olduğu ağaç yapısı olarak belirlenmiştir. Ayrıca bağımlı değişken sürekli olan modellerde riskin ölçümü varyans olduğundan, bu modellerde en uygun çözümün belirlenmesinde açıklanan varyans ölçüsünden de yararlanılmıştır. İMKB100 endeksinde sanayi ve hizmet sektöründe yer alan işletmeleri ayırmada seçilen en uygun ağaç yapısının yerine koyma risk değeri %17,3 ve çapraz geçerlilik risk değeri %17,9'dur. *KVB/TA*, *TB/TA*, *NK/TA* ve *NK/ÖS* bağımlı değişkenleri için risk değerlerinin yanı sıra, bu değişkenlere ait açıklanan varyans değeri %60'ın üzerinde olan ağaç yapıları seçilmiştir. Kısa vadeli borç oranının etkileyen en önemli değişkenleri belirleyen ağaç yapısının risk değerleri sırasıyla %1,1 ve %2,5 iken toplam borç oranını etkileyen en önemli değişkenleri belirleyen ağaç yapısının risk değerleri %1,7 ve %2,1'dir. Bu risk değerleri, aktiflerin karlılığı değişkeni için sırasıyla %0,6 ve %2,5 iken özsermayenin karlılığı değişkeni için %8,6 ve %23,6'dır. *KVB/TA*, *TB/TA*, *NK/TA* ve *NK/ÖS* bağımlı değişkenlerinin açıklanan varyans değerleri ise sırasıyla %72, %64, %64 ve %66'dır. Çalışmada incelenen ve yorumlanan en uygun ağaç yapılarına göre elde edilen bulguların değerlendirilmesi sonucunda şu hususlar ortaya konmuştur:

Çalışmada, sanayi ve hizmet sektörlerinde faaliyet gösteren firmaları ayırmada etkili olan değişkenler işletme sermayesinin net satışlara oranı, stok devir hızı ve

ekonomik rantabilite oranı değişkenlerinden oluşmaktadır. 1. ana dalda işletme sermayesinin net satışlara oranı negatif olması, 1. profildeki 17 işletmenin (sanayi: 2, hizmet:15) kısa vadeli borçların yüksek olduğunun göstergesidir ve yükümlülüklerini karşılayabilme riskini arttırmaktadır. 2. ana daldaki 35 işletmenin (sanayi: 25, hizmet: 10) oluşturduğu 2. profilde, işletme sermayesinin net satışlara oranı değerinin yükselmesiyle birlikte işletme sermayesi satışlara bağımlı olmaktan çıkmakta ve satışlar artmaktadır. 3. ana daldaki toplam 104 işletmenin (sanayi: 93, hizmet: 11) yer aldığı profillerde elde edilen bulgular değerlendirildiğinde, 4. ve 5. profillerde ekonomik rantabilite oranının negatif olması işletme faaliyetlerinin finansmanında ağırlıklı olarak yabancı kaynak kullanıldığını göstermektedir. 3., 4. ve 5. profillerde stok devir hızının negatif olması ise işletmelerde fazla stok bulundurulduğunu, satışların net işletme sermayesine bağlı olduğunu ve bu nedenle işletme satışlarının azaldığını göstermektedir. 4. ana dalda yer alan 7. profildeki 17 işletmenin (sanayi:7, hizmet:10), ihtiyaçlarının çok üzerinde işletme sermayesi bulunmakta ve kredi verecek kurumlar tarafından riskli bir yapıyı oluşturmaktadır.

Kısa vadeli borç oranını (*KVB/TA*) tanımlamada etkili değişkenler cari oran, aktif devir hızı, özsermaye devir hızı ve işletme sermayesinin net satışlara oranı değişkenlerinden oluşmaktadır. Bu bağımlı değişkeni oluşturan profillerde, cari oran değeri 1,19 ve küçük, 1,19 ve 4,14 aralığında değişen ve 4,14'ten büyük olmak üzere 6 ana dala ayrılmıştır. Cari oran değeri düşük olan 1. ana daldaki profillerde aktif devir hızı yüksek olan 24 işletme 2. profilde yer almaktadır. Bu profildeki işletmelerin karlılığının yüksek olmasına karşın 1. profildeki 10 işletmenin aktif devir hızının düşme eğiliminde olması bu işletmelerin satışlarını arttırıcı önlemler almasını gerektirmektedir. 2. ana dalda toplam 52 işletmenin yer aldığı 3., 4. ve 5. profillerde işletmelerin cari oran değerleri yeterli iken 3. profilde özsermayelerinin yeterince verimli kullanılmadığı ve atıl kaldığı görülmektedir. 4. ve 5. profillerde, işletmelerin özsermayelerini verimli kullandıkları sonucuna karşın 4. profildeki işletmelerin satışlarının yeterli düzey olmadığı, işletme sermayesinin verimli kullanılması gerektiği görülmektedir. Buna karşın 5. profildeki işletmelerin net satışlarının yeterli düzeye ulaşmaya başladığı ve satışların sermayeye bağımlı olmadan gerçekleştiği görülmektedir. 3. ve 4. ana dalda yer alan toplam 53 işletme,

cari oran değeri en iyi olan, yükümlülüklerini karşılamada en iyi durumda olan işletmelerdir. 5. ve 6. ana dalda yer alan 8. ve 9. profillerdeki toplam 34 işletmenin ise gereğinden fazla dönen varlıklara sahip oldukları ve yatırım yapmada yetersiz oldukları (likit varlıklarının atıl kaldığı) görülmektedir.

Toplam borç oranının (TB/TA) tanımlamada etkili değişkenler, cari oran, işletme sermayesi devir hızı, işletme büyüklüğü ve aktif devir hızı değişkenlerinden oluşmaktadır. Cari oranın yeterli olması, işletmelerin borçlarını ödeyebilme yeteneğinin güçlü olduğunu göstermektedir. 1. ve 2. ana dallardaki 2. ve 3. profillerde yer alan 69 işletmenin alacak ve stoklara ayırması gereken net işletme sermayesinin fazla olması net işletme sermayesi devir hızının düşmesine sebep olmaktadır. 3. ana daldaki 4. ve 5. profillerde, cari oran değeri işletme aktiflerinin büyük ya da küçüklüğüne göre alt dallara ayrılmaktayken cari oranın bu değişkenden bağımsız olarak uygun seviyeyi koruması gerektiği düşünülmektedir. 4. ana dalda, 35 işletmenin yer aldığı 6. ve 7. profillerde, aktif devir hızı işletmelerin verimli olduğunu göstermektedir. 5. ana dalda ve 6. ana dalda, 17 işletmenin yer aldığı 8. ve 9. profillerde ise cari oranın çok yüksek olması nedeniyle işletmelerin yatırım yapmada başarısız olduğunu göstermektedir.

Aktiflerin karlılığını (NK/TA) tanımlamada etkili olan değişkenler, cari oran, toplam borç oranı, kısa vadeli borç oranı, işletme sermayesi devir hızı, dönen varlıklar/toplam aktifler, özsermaye devir hızı ve işletme sermayesinin net satışlara oranı değişkenleridir. Bu bağımlı değişkeni oluşturan profiller incelendiğinde, 1. ana dalda yer alan toplam 156 işletme içerisinde 1. profildeki 135 işletmenin toplam borç oranını ve kısa vadeli borç oranı oldukça iyidir. Bu da bu işletmelerin mali yapılarının çok iyi olduğunun bir göstergesidir. 2., 3., 4. ve 5. profillerde yer alan toplam 4 işletmenin toplam borç oranı değeri olumlu olmakla birlikte kısa vadeli borç oranı değerlendirildiğinde, bu işletmelerin mali yapılarının kredi veren firmalar açısından riskli olduğu görülmektedir. İşletme sermayesi devir hızı, işletmelerin karlılıklarını göstermektedir. 2. profile bu oran düşük olmasına karşın diğer profillerde yer alan işletmelerin karlılığı yüksektir. 6., 7. ve 8. profillerdeki 17 işletmenin toplam borç oranının yüksek olması, işletmelerin riskli grupta olduğunu

göstermekteyken dönen varlıklar/toplam aktifler oranı da mali yapılarının riskli olduğunu ifade etmektedir. 8. profilde yer alan 15 işletme, diğerlerine oranla toplam aktifler içerisindeki dönen varlıklar oranı yüksek olduğundan kendi faaliyetleriyle varlıklarını sürdürebilme yeteneği daha yüksektir. 2. ana dalda bulunan toplam 17 işletmenin borçlarını ödeyebilme yeteneği, işletmelerin normalin çok üstünde dönen varlıklara sahip olduğunu göstermektedir. 9. profilde yer alan işletmelerin özsermaye devir hızının düşük olması, sermayelerini verimli bir şekilde kullanamadıklarını göstermektedir. Özsermaye devir hızı yüksek olan 10., 11. ve 12. profillerde verimlilik oldukça iyi iken, 10. profildeki işletmelerin sermayelerini arttırmaları gerekmektedir. 11. profildeki işletmelerde, sermaye ve satışlar yeterli düzeye ulaşmaya başlamaktadır. Ancak bu işletmeler, satışlarıyla faaliyetlerini sürdürme yeteneğini korumaya çalışmalıdırlar. 12. profilde yer alan işletmelerin satışlarının az olduğu ve bu nedenle faaliyetlerini işletme sermayesine bağlı olarak devam ettirdiği görülmektedir.

Özsermayenin karlılığını ($NK/ÖS$) tanımlamada etkili olan değişkenler, özsermaye devir hızı, kısa vadeli borç oranı, dönen varlıklar/ toplam aktifler ve cari oran değişkenleridir. $NK/ÖS$ bağımlı değişkenin profilleri incelendiğinde, 1. ana daldaki düğümlerden oluşan profillerin özsermaye devir hızı, işletmelerin verimli olduklarını göstermekle birlikte düşme eğiliminde olması verimliliğin azaldığına işaret eder. Kısa vadeli borç oranı çok iyi olan toplam 150 işletmeden 2. profilde yer alan 9 işletme, dönen varlıklar/toplam aktifler oranının da yüksek olması nedeniyle en iyi mali yapıya sahip olan işletmelerdir. Buna karşın 1. profildeki işletmeler, faaliyetlerini sermayeden karşılama ihtimali nedeniyle risk oluşturmaktadırlar. 1. ana dalda yer alan, kısa vadeli borç oranı yüksek 3. ve 4. profiller, kredi veren firmalar tarafından riskli durumda bulunmaktadır. Toplam 6 işletmenin bulunduğu bu profillerin cari oran değeri, işletmelerin borçlarını ödeyebilme yeteneğinin de düşük olduğunu göstermektedir. 2. ana daldaki tek profil olan 5. profilin özsermaye devir hızı oldukça yüksektir. Bu profilde yer alan 17 işletmenin özsermaye devir hızının yüksek olması, işletmelerin verimliliği ifade edebileceği gibi işletmelerin finansmanında ağırlıklı olarak yabancı kaynaklı kaynakları kullandıklarının da göstergesi olabilmektedir.

Çalışmanın sonuçlarıyla görüldüğü gibi karar ağaçları tekniğiyle işletmelerin birbirlerine göre konumları ortaya konmuş ve sektör, *TB/TA*, *KVB/TA*, *NK/TA* ve *NK/ÖS* değişkenlerini etkileyen en önemli değişkenler saptanmıştır. İşletmelerin finans piyasalarındaki durumlarının değerlendirilmesi, mevcut yatırımcılar, potansiyel yatırımcılar, mali analistler ve diğer kesimlerce önem taşımaktadır. İlgili kesimlerin özellikle portföy oluşturmaları, portföyde yer alan finansal varlıkları yenilemeleri, ne zaman alışı ve satışı geçeceğine karar vermeleri ve pazar fırsatlarından yararlanmaları şirketlerin finansal göstergelerinin analizinden elde edilecek doğru bilgiler ve bunların sağlıklı bilimsel yorumlarıyla mümkündür. Şirketlere ait finansal bilgiler yanında, ekonomik gidişatla ve sektörlerin ekonomik yapılarıyla ilgili bilgiler de yatırım kararlarının doğruluğunu artırmakta, pazar analizlerinin sağlıklı yapılmasına imkan vermektedir. Bu bağlamda veri madenciliği yol göstermede uygun bir teknik olmuştur. Sınıflandırmanın geniş olması ve sonuçların görselliği ve kolay yorumlanabilir olması nedeniyle bulgular karar ağacı tekniğinin ihtiyaca yönelik cevapları ortaya çıkarmakta etkili olduğunu göstermektedir. Bununla birlikte karar ağaçlarıyla ortaya çıkan sonuçlar farklı yöntemlerde veri ve değişken olarak da kullanılabilmekte ve bu yolla yeni bilgiler elde edilebilmektedir.

KAYNAKÇA

- Agrawal, R., H. Mannila, R. Srikant, H. Toivonen, ve A. I. Verkamo (1995); *Fast Discovery of Association Rules, Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Chapter 12, s. 307-328.
- Aitkenhead, M. J. (2008); "A Co-Evolving Decision Tree Classification," *Expert Systems with Applications*, Cilt 34, Sayı 1, s. 18-25.
- Akpınar, Haldun (2000); "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği," *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, Cilt 29, Sayı 1, s.1-22.
- Aktan Bora ve Bora Bodur (2006); "Oranlar Aracılığı ile Finansal Durumunuzu Nasıl Çözümlersiniz?," *Journal of Yaşar University*, Cilt 1, Sayı 1, s.1-13.
- Akyüz, Cemil, İlker Akyüz, Hasan Serin ve Hicabi Cindik (2006); "The Financing Preferences and Capital Structure of Micro, Small and Medium Sized Firm Owners in Forest Products Industry in Turkey," *Forest Policy and Economics*, Cilt 8, Sayı 3, s. 301–311.
- Alpaydın, Ethem (2000); "Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri," *Bilişim 2000 Eğitim Semineri*, www.Cmpe.Boun.Edu.Tr /~Ethem/Files/Papers/Veri-Maden_2k-Notlar.Doc, (Erişim Tarihi: 06.04.2007).
- Altınışık, Umut (2006); "Öğrenci Bilgi Sisteminde Veri Madenciliğinin Uygulanması," Yayınlanmamış Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli.
- Aydoğan, Fatih (2003); "E-Ticarette Veri Madenciliği Yaklaşımlarıyla Müşteriye Hizmet Sunan Akıllı Modüllerin Tasarımı ve Gerçekleştirimi," Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Baykal, Nazife (2003); "Veri Tabanı Ve Veri Madenciliği," *Tıp Bilişimi Güz Okulu*, http://www.Turkmia.Org/Eski/File/231verimadenciligi_Baykal.Ppt, (Erişim Tarihi: 07.03.2005).
- Baykasoğlu, Adil (2005); "Veri Madenciliği ve Çimento Sektöründe Bir Uygulama," *Akademik Bilişim 2005*, 02- 04 Şubat 2005, Gaziantep.
- Benli Keskin, Yasemin (2005); "Bankalarda Mali Başarısızlığın Öngörülmesi Lojistik Regresyon Ve Yapay Sinir Ağı Karşılaştırması," *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, Sayı 6, s. 31-46.
- Berger, Allen N. ve Emilia Bonaccorsi Di Patti (2006); "Capital Structure and Firm Performance: A New Approach to Testing Agency Theory and An Application to The Banking Industry," *Journal of Banking and Finance*, Cilt 30, Sayı 4, s. 1065–1102.

- Berry, Michael J. (2004); *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, John Wiley and Sons Incorporated, USA.
- Berson, A., S. Smith, ve K. Thearling (1999); *Building Data Mining Applications for CRM*, McGraw Hill, USA.
- Biçen, Pelin (2002); “Veri Madenciliği: Sınıflandırma ve Tahmin Yöntemlerini Kullanarak Bir Uygulama,” Yayınlanmamış Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Bircan, Hüdaverdi (2004); “Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama,” *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Sayı 2, s. 185-208.
- Cooley, Robert, Bamshad Mobasher ve Jaideep Srivastava (2004); “Web Mining: Information and Pattern Discovery on The World Wide Web,” [Http://www.users.cs.umn.edu/~mobasher/webminer/survey/survey.html](http://www.users.cs.umn.edu/~mobasher/webminer/survey/survey.html), (Erişim Tarihi: 20.04.2004)
- DEU (2004); “Veri Madencilik Uygulamaları ve Eğilimi,” [Http://courses.cs.deu.edu.tr/cse572/veri%20madencilikuygulamalar.doc](http://courses.cs.deu.edu.tr/cse572/veri%20madencilikuygulamalar.doc), (Erişim Tarihi: 20.01.2004).
- Doğan, Nurhan Ve Kazım Özdamar (2003); “Chaid Analizi ve Aile Planlaması ile İlgili Bir Uygulama,” *Türkiye Klinikleri Tıp Bilimleri Dergisi*, Cilt 23, Sayı 5, s. 392-397.
- Dolgun, M. Özgür (2008); “Türkiye’deki Hastanelerin Veri Madenciliğiyle Gruplandırılması,” *XI. Biyoistatistik Kongresi*, 27 – 30 Mayıs, Malatya.
- Durmuş, Mustafa Seçkin ve Serdar İplikçi (2007); “Veri Kümeleme Algoritmalarının Performansları Üzerine Karşılaştırmalı Bir Çalışma,” *Akademik Bilişim 2007*, 31 Ocak - 02 Şubat, Kütahya.
- Durukan, M. Banu (1997); “Hisse Senetleri İMKB’de İşlem Gören Firmaların Sermaye Yapısı Üzerine Bir Araştırma: 1990 – 1995,” *İMKB Dergisi*, Yıl 1, No 3, s. 75-91.
- Edelstein, Herb (2000); *Mining Large Database - A Case Study*, Two Crows Corporation.
- Elder, J. F. ve D. Pregiben (1995); “A Statistical Perspective on KDD,” *The 1st International Conference on Knowledge Discovery and Data Mining*, 20-21 Ağustos, Montreal, s. 87-93.
- Erdem, O. Ayhan ve Emre Uzun (2005); “Yapay Sinir Ağları İle Türkçe Times New Roman, Arial ve Elyazısı Karakterleri Tanıma,” *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, Cilt 20, Sayı 1, 13-19.

- Erken, Arzum (1998); “Başlıca Fiyat Bazlı Oranların Hisse Senedi Getirisi Üzerindeki Etkileri ve İstanbul Menkul Kıymetler Borsası’nda Bir Uygulama,” Yayınlanmamış Yüksek Lisans Tezi, Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Fayyad, Usama, Gregory Piatetsky-Shapiro ve Padhraic Smyth (1996); “From Data Mining to Knowledge Discovery in Databases,” *AI Magazine*, Cilt 17, Sayı 3, s. 68-70., s. 37-54.
- Fayyad, Usama (1997); “Data Mining and Knowledge Discovery in Databases: Implications for Scientific Databases,” *Ninth International Conference on Scientific and Statistical Database Management*, 11-13 Ağustos, Olympia, s. 2-11.
- Francis, Louise (2006); “Martian Chronicles: Is Mars Beter Than Neural Networks,” <http://www.salford-systems.com/doc/03wf027.pdf>, (Erişim Tarihi:23.03.2006).
- Fu, Yongjian (1997), “Data Mining Tasks, Techniques and Applications,” No: 6648/97, *IEEE*, s. 18-20.
- Ghosh, A. ve B. Nath (2004); “Multi-Objective Rule Mining Using Genetic Algorithms,” *Information Sciences*, Cilt 163, Sayı 1-3, s. 123-133.
- Grobler, B. R., T.C. Bisschoff ve K.C. Moloi (2002); “The Chaid-Technique and the Relationship between School Effectiveness and Various Independent Variables,” *International Studies in Educational Administration*, Cilt 30, Sayı 3, s. 44-56.
- Gürsakal, Necmi (2001); *Sosyal Bilimlerde Araştırma Yöntemleri*, VIPAŞ AŞ, Yayın No: 54, Bursa.
- Haley, Barbara J. (1999); *The Benefits of Data Warehousing at Whirlpool*, Idea Group Inc, Hershey, PA, USA.
- Han, J. ve M. Kamber (2001); *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, USA.
- Hand, David, Heikki Manila ve Padhraic Symth (2001); *Principles of Data Mining*, The MIT Press, London.
- Haris, Milton ve Artur Raviv (1991); “The Theory of Capital Structure,” *The Journal of Finance*, Cilt 46, Sayı 1, s. 297-355.
- Hastie, T., R Tibshirani, ve J. Friedman (2001); *The Elements of Statistical Learning; Data Mining, Inference and Prediction*, Springer Series in Statistics, USA.
- Haughton, Dominique ve Samer Oulabi (1999); “Direct marketing modeling with CART and CHAID,” *Journal of Direct Marketing*, Cilt 11, Sayı 4, s. 42-52.

- Hegland, Markus (2001); *Data Mining Techniques*, Acta Numerica, Cambridge University Press, s. 313–355.
- Huang, Guihai ve Frank M. Song (2006); “The Determinants of Capital Structure: Evidence From China,” *China Economic Review*, Cilt 17, Sayı 1, s. 14– 36.
- Istaitieh, Abdulaziz ve Jose M. Rodriguez-Fernandez (2006); “Factor-Product Markets and Firm’s Capital Structure: A Literature Review,” *Review of Financial Economics*, Cilt 15, Sayı 1, s. 49–75.
- İlhan, Sevinç (2005); “Introduction to Data Mining,” Kocaeli Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Ana Bilim Dalı Ders Notu, Kocaeli.
- İmamoğlu, Tuba (2005); “*Veri Madenciliğinde Karar Ağaçları ile Bir Öğrenci Ders Başarısı Tahmin Aracı*,” Yayınlanmamış Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli.
- İMKB (2007); www.imkb.gov.tr, (Erişim Tarihi:06.04.2007).
- İnan, Onur (2003); “*Öğrenci İşleri Veritabanı Üzerinde Veri Madenciliği Uygulamaları*,” Yayınlanmamış Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Işık, Nihat S. (2003); “*Karar Sürecinde Bilgi Sistemi Uygulamaları ve Dağıtım Kanallarının Belirlenmesinde Coğrafi Bilgi Sistemi Kullanımı*,” Yayınlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- ITRG Info-Tech Research Group (2003); *Data Mining: Tools, Techniques And ROI*, ITRG Press, London.
- Kale, Jayant R., Thomas H. Noe ve Gabriel G. Ramirez (1991), “The Effect of Business Risk on Corporate Capital Structure: Theory and Evidence,” *The Journal of Finance*, Cilt 46, Sayı 5, s. 1693-1715.
- Kleissner, Charly (1998); “Data Mining for Enterprise,” *31st Annual Hawaii International Conference on System Scienes*, 1060–3425/98, s. 295–304.
- Korkmaz, Turhan, Ali Sait Albayrak ve Abdülmecit Karataş (2007), “Hisse Senetleri İMKB’de İşlem Gören KOBİ’lerin Sermaye Yapısının İncelenmesi: 1997–2004 Dönemi,” *İktisat İşletme ve Finans Dergisi*, Yıl 22, s. 79-95.
- Kovalerchuk, B. ve E. Vityaev (2002); *Data Mining in Finance: Advances in Relational and Hybrid Methods*, Kluwer Academic Publishers, USA.
- Koyuncugil, Ali Serhan (2006); “*Bulanık Veri Madenciliği ve Sermaye Piyasasına Uygulanması*,” Yayınlanmamış Doktora Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

- Koyuncugil, Ali Serhan (2007); *Borsa Şirketlerinin Sektörel Risk Profillerinin Veri Madenciliğiyle Belirlenmesi*, Sermaye Piyasası Kurulu Araştırma Raporu, Araştırma Dairesi, Ankara.
- Koyuncugil, Ali Serhan ve Nermin Özgülbaş (2006); “İMKB’de İşlem Gören KOBİ’lerin Finansal Başarısızlığına Etki Eden Faktörlerin Veri Madenciliği İle Belirlenmesi,” *3. KOBİ ve Verimlilik Kongresi*, 17–18 Kasım, İstanbul.
- Kuzey, Cemil (2000); “*Sayısal Karar Destek Sistemleri Veri Ambarları*,” İstanbul Üniversitesi Sosyal Bilimler Enstitüsü İşletme Ana bilim Dalı Ders Notu, İstanbul.
- Larose, Daniel T. (2005); *Discovering Knowledge in Data: An Introduction to Data Mining*, John and Wiley Sons Incorporated, USA.
- Laura, Galguera, David Luna ve M. Paz Mendez (2006); “Predictive Segmentation in Action: Using CHAID to Segment Loyalty Card Holders,” *International Journal of Market Research*, Cilt 48, Sayı 4, s. 459- 479.
- Liu, Tong (1999); “*The Determinants of Corporate Capital Structure: Evidence from Listed Companies in China, Presented in Partial Fulfillment of The Requirements*,” Yayınlanmamış Yüksek Lisans Tezi, Concordia Üniversitesi, Kanada.
- Margaritis, Dimitris ve Maria Psillaki (2007); “Capital Structure and Firm Efficiency,” *Journal of Business Finance and Accounting*, Cilt 34, Sayı 9 ve 10, s.1447–1469.
- McCarty, John A. ve Manoj Hastak (2007), “Segmentation Approaches in Data-Mining: A Comparison of RFM, CHAID and Logistic Regression,” *Journal of Business Research*, Cilt 60, Sayı 6, s. 656–662.
- Moss, L.T. ve S. Atre (2003); *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*, Addison-Wesley Publishing, USA.
- Oğuzlar, Ayşe (2003); “Veri Önişleme”, *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Sayı 21, s. 67-76.
- Olafsson, Sigurdur, Xiaonan Li ve Shuning Wu (2008); “Operations Research and Data Mining,” *European Journal of Operational Research*, Cilt 187, Sayı 3, s.1429- 1448.
- Özekes, Serhat (2003); “Veri Madenciliği Modelleri ve Uygulama Alanları,” *İstanbul Ticaret Üniversitesi Dergisi*, Yıl 2, Sayı 3, s. 65–82.
- Özkan, Yalçın (2006); *Veri Tabanı Sistemleri*, Alfa Yayınları, İstanbul.

- Piatetsky-Shapiro, Gregory (1991); "Knowledge Discovery in Real Databases: A Workshop Report," *AI Magazine*, Cilt 11, Sayı 5, s. 68-70.
- Pastor, Lubos ve Pietro Veronesi (2003); "StockValuation and Learning About Profitability," *The Journal of Finance*, Cilt 58, Sayı 5, s. 1749-1790.
- Piramuthu, Selwyn (2004); "Evaluating Feature Selection Methods for Learning in Data Mining Applications," *European Journal of Operational Research*, Cilt 156, Sayı 2, s. 483-494.
- Roiger, R. J. ve M. W. Geatz (2003); *Data Mining A Tutorial-Based Primer*, Addison Wesley, USA.
- Simoudis, Evangelos (1996); "Reality Check For Data Mining," *IEEE Expert: Intelligent Systems and Their Applications*, Cilt 11, Sayı 5, s. 26-33.
- Sogorb-Mira Francisco (2005); "How SME Uniqueness Affects Capital Structure: Evidence from 1994–1998 Spanish Data Panel," *Small Business Economics*, Cilt 25, Sayı 5, s. 447–457.
- SPSS (1999); *AnwerTree Algorithm Summary*, SPSS White Paper, USA.
- SPSS (2002); *AnwerTree 3.1 User's Guide*, SPSS Inc., USA.
- Sun, Jie ve Hui Li (2008); "Data Mining Method for Listed Companies, Financial Distress Prediction," *Knowledge-Based Systems*, Cilt 21, Sayı 1, s. 1-5.
- Tang, Chun-Hung (Hugo) ve SooCheong (Shawn) Jang (2007); "Revisit to The Determinants of Capital Structure: A Comparison Between Lodging Firms and Software Firms," *International Journal of Hospitality Management*, Cilt 26, Sayı 1, s.175-187.
- Thomas, Lyn. C. (2000); "A Survey of Credit and Behavioral Scoring: Forecasting Financial Risk of Lending to Consumer," *International Journal of Forecasting*, Cilt 16, Sayı 2, s. 149-172.
- Toktaş, Peral ve Melek Başak Demirhan (2004); "Risk Analizinde Veri Madenciliği Uygulamalar," *YA/EM Yöneylem Araştırması/Endüstri Mühendisliği- XXIV Ulusal Kongresi*, 15- 18 Haziran, Adana.
- Tunay, K. Batu (2001); "Türkiye’de Paranın Gelir Dolaşım Hızlarının Mars Yöntemiyle Tahmini," *ODTÜ Gelişme Dergisi*, Cilt 28, Sayı 3-4, s. 431-454.
- Türe, Mevlut, Füsün Tokatlı ve İmran Kurt, (2008); "Using Kaplan-Meier Analysis Together With Decision Tree Methods (C&RT, CHAID, QUEST, C4.5 and ID3) In Determining Recurrence-Free Survival of Breast Cancer Patients," *Expert Systems With Applications*, Article in Press.

- Usgurlu, Ümit B. (2007); “Veri Tabanı, Veri Madenciliği, Veri Ambarı, Veri Pazarı,” www.Mail.Baskent.Edu.Tr/~20396676/0302/Bil483/Hw2.Pdf, (Erişim Tarihi: 08.03.2007)
- Vahaplar, Alper (2003); “*Bir Coğrafi Veri Madenciliği Uygulaması*,” Yayınlanmamış Yüksek Lisans Tezi, Ege Üniversitesi. Fen Bilimler Enstitüsü, İzmir.
- Van Diepen, Merel ve Philip Hans Franses (2006); “Evaluating Chi-Squared Automatic Interaction Detection,” *Information Systems*, Cilt 31, Sayı 8, s. 814-831.
- Yıldırım, Pınar, Mahmut Uludağ ve Abdülkadir Görür (2008); “Hastane Bilgi Sistemlerinde Veri Madenciliği”, *Akademik Bilişim 2008*, 30 Ocak - 01 Şubat, Çanakkale.
- Ziarko, W. (1991); *The Discovery, Analysis, and Representation of Data Dependencies in Databases, Knowledge Discovery in Databases*, AAAI/MIT Press, s.195-212.