



Munich Personal RePEc Archive

**Machine Learning Techniques in
Determining the Dynamics of Corporate
Financial Distress: An Empirical
Treatment and a Comparative Analysis
of Financial and Non-Financial Micro
Data of the Turkish Private Sector**

Cakir, Murat

CENTRAL BANK OF THE REPUBLIC OF TURKEY

December 2005

Online at <https://mpra.ub.uni-muenchen.de/55975/>

MPRA Paper No. 55975, posted 16 May 2014 17:33 UTC

**FİRMA BAŞARISIZLIĞININ DİNAMİKLERİNİN
BELİRLENMESİNDE MAKİNA ÖĞRENMEİ TEKNİKLERİ:
AMPİRİK UYGULAMALAR VE KARŞILAŞTIRMALİ ANALİZ**

Murat ÇAKIR

Danışman
Prof. Dr. Halil Altay Güvenir

Uzmanlık Yeterlilik Tezi

Türkiye Cumhuriyet Merkez Bankası
İstatistik Genel Müdürlüğü
Ankara, Aralık 2005

ÖNSÖZ

Ekonomik krizler sırasında ve sonrasında, büyük sayıda firmanın iflas etmesiyle, firma başarısızlığının dinamiklerinin araştırılması finans kuramının üzerinde önemle durduğu konulardan biri durumuna gelmiştir. Büyük bunalım sonrasında başlayan ve günümüze kadar giderek artan bir ilgiyle sürdürülen firma başarısızlığı çalışmaları ve oluşturulan literatür, çoğunlukla, finans kuramı bağlamında tahvil yatırımcılarının risk analizlerinin incelenmesi amacına odaklanmıştır. Finansal otoritelerin başarısızlık çalışmalarıysa, özellikle merkez bankalarının reel sektöre verdiği kredilerin risklerinin ölçülmesi çerçevesinde geliştirilmiştir. Ekonomik politikaların reel sektörde yarattığı risk etkilerinin incelenmesinde de firma başarısızlığı kuramından yararlanılmasıyla firma ve sektör risk profillerinin elde edilmesinin olanaklı olduğu düşünülmektedir.

Bu çerçevede, firma iflas maliyetlerinin de göz önünde bulundurulmasıyla, bilişimdeki en önemli gelişmelerden biri olarak kabul edilen makina öğrenmesi yöntemleri kullanılarak firma başarısızlığının dinamikleri ampirik olarak incelenmiş ve kullanılabilir bir model geliştirilmesi amaçlanmıştır.

Bu amaçla hazırlamış olduğum çalışmamda bana her konuda destek olan İstatistik Genel Müdürü Cahit Özçet'e, İstatistik Genel Müdür Yardımcısı Dr. Zerrin Gürgenci'ye, Reel Sektör Verileri Müdürü Metin Öner'e, Müdür Yardımcıları Serap Çelen ve Emel Dinçer'e, metinle ilgili her türlü eleştiri ve katkılarını esirgemeyen Müdür Yardımcısı Gülin Küçükay'a, tezin hazırlanmasında büyük emekleri geçen değerli hocam Prof. Dr. Halil Altay Güvenir'e, her türlü desteğini hiçbir zaman eksik etmeyen aileme ve dostlarıma teşekkürü borç bilirim.

Murat ÇAKIR

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	i
İÇİNDEKİLER	ii
TABLO LİSTESİ	vi
ŞEKİL LİSTESİ	vii
KISALTMA LİSTESİ	viii
SEMBOL LİSTESİ	ix
EK LİSTESİ	x
ÖZET	xi
ABSTRACT	xiii
GİRİŞ	1

BİRİNCİ BÖLÜM

FİNANSAL BASKI VE FİNANSAL BAŞARISIZLIK	5
1.1. Finansal Baskı ve Firma Başarısızlığı	5
1.2. Finansal Baskı Sürecinin Aşamaları ve Firma Başarısızlığı	7
1.2.1. Finansal Baskı	8
1.2.1.1. Finansal Baskı Süreci	11
1.2.1.1.1. Acze Düşme	11
1.2.1.1.2. İflas	11
1.2.1.1.3. Tasfiye	12
1.2.2. Firma Başarısızlığı	13
1.3. Finansal Baskının Nedenleri ve Teşhisi	14
1.3.1. Finansal Baskının Nedenleri	15
1.3.2. Finansal Baskı Göstergeleri	16
1.4. Finansal Baskı, Firma Performansı ve Verimsizlik Kayıpları	17
1.4.1. Finansal Baskı ve Firma Performansı	17
1.5. Finansal Baskı Maliyetleri	18
1.5.1. Finansal Baskı Maliyet Türleri	21
1.6. Firma Başarısızlığı Tahmin Literatürü	23

1.7. Finansal Başarısızlığı Modellemedeki Güçlükler.....	28
1.8. Finansal Tabloların Bilgi İçeriği ve Finansal Başarısızlık Modellerindeki Değişkenler	32
1.8.1. Finansal Tablolardaki Bilgi	32

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ VE MAKİNA ÖĞRENMESİ YÖNTEMLERİ.....	35
2.1. Veri Madenciliği ve Makina Öğrenmesi Kuramı	35
2.1.1. Girdi, Kavram, Örnek ve Öznitelikler	38
2.1.2. Veri Seti	39
2.1.3. Örneklem Büyüklüğü.....	41
2.1.4. Eksik ve Hatalı Değerler	42
2.2. Mantıksal Makina Öğrenmesi Algoritmaları.....	44
2.2.1. Karar Ağaçları	45
2.2.2. Kapsama Algoritmaları ve Sınıflandırma Kuralları	47
2.2.3. Karar Ağaçları ve Kurallar	49
2.2.4. Karar Ağaçlarından Kural Öğrenme.....	50
2.2.5. Kurallar ve Karar Listeleri.....	51
2.2.6. Kapsama Algoritmaları ve Eksik Değerler.....	51
2.2.7. Yararlı ve Yararsız Kurallar	51
2.2.8 Parçasal Karar Ağaçları	52
2.3. Makina Öğrenmesi ve Sınıflandırma	53
2.3.1. Sınıflandırma Doğruluğu	54
2.3.2. Sınıflandırma Maliyetleri	54
2.3.3. Sınıflandırmada Fayda ve Maliyet Duyarlılığı	55
2.3.3.1. Maliyete Duyarlı Öğrenme	56
2.3.3.2. Maliyete Duyarlı Öğrenmede Öznitelik Bağımlılığı	58
2.3.3.3. Öğrenmede Fayda Maksimizasyonu.....	61
2.3.3.4. Fayda Doğruluğu	63
2.3.3.5. Karma Optimizasyon.....	64
2.4. Kurallar ve Kural İlginçliği.....	66

ÜÇÜNCÜ BÖLÜM

WEKA PLATFORMU VE BCFP	68
3.1. Makina Öğrenmesi ve Firma Başarısızlığı Analizi	68
3.2. WEKA Platformu	69
3.2.1. WEKA Platformu Öğrenme Algoritmaları	71
3.2.1.1. Kapsama Algoritmaları	71
3.2.1.1.1. ZeroR	71
3.2.1.1.2. OneR	72
3.2.1.1.3. DecisionTable	73
3.2.1.2. Karar Ağacı Algoritmaları	77
3.2.1.2.1. DecisionStump	77
3.2.1.2.2. j48.J48	77
3.2.1.3. Parçasal Karar Ağaçları	81
3.2.1.3.1. j48.PART	81
3.2.2. Maliyete Duyarlı Öğrenme	81
3.2.2.1. Meta Sınıflandırma	82
3.2.2.1.1. CostSensitiveClassifier	82
3.2.2.1.2. MetaCost	83
3.3. Öznitelik İzdüşümü ve BCFP	84
3.3.1. Bilgi Gösterimi	86
3.3.2. Eğitim	86
3.3.3. Sınıflandırma	87
3.3.4. BCFP'nin Oluşturduğu Kurallar	89

DÖRDÜNCÜ BÖLÜM

MAKİNA ÖĞRENMESİ YÖNTEMLERİ UYGULAMALARI - AMPİRİK ANALİZ VE YORUMLAR	92
4.1. Veri Setinin Yapısı	92
4.1.1. Veri Seti	92
4.1.2. Örneklem Seçimi ve Örneklem Büyüklüğü	93
4.1.3. Firmaların Hukuksal Durumları	93
4.1.4. Verinin Hazırlanması ve Standart Form	93
4.1.5. Bilgi Gösterimi	94
4.1.6. Sınıf Etiketlerinin Belirlenmesi	95
4.1.7. WEKA Alt Örneklem Büyüklükleri	98

4.1.8. Örneklerin Sınıflar Bazında Sektörel Dağılımı	98
4.1.9. Özniteliklerin Sınıf Yığılma Dağılımları.....	99
4.1.10. Uç ve Eksik Değerler	100
4.1.11. Veri Setiyle İlgili Kısıtlar	100
4.2. WEKA Analizleri.....	100
4.2.1. Öğrenme Algoritmalarının Sonuçları.....	100
4.2.2. Meta Sınıflandırma Sonuçları	103
4.2.2.1. Maliyete Duyarlı Öğrenme ve Meta Sınıflandırma Yöntemleri	103
4.2.2.2. Sınıflandırma Sonuçları	104
4.2.2.3. Meta Sınıflandırıcılar ve Alt Örneklemeler (10 Set).....	106
4.2.2.4. CostSensitiveClassifier ve MetaCost	109
4.2.2.5. Sınıflandırma Doğruluğu ve Maliyet Verimliliği: Senaryo Analizleri.....	109
4.2.2.6. Değerlendirme ve Yorumlar	116
4.2.2.7. Senaryo Analizi Sonuçları	117
4.2.2.8. WEKA Platformu Genel Sonuçları	119
4.3. BCFP Analiz ve Yorumları	125
4.3.1. BCFP Veri Seti.....	125
4.3.2. BCFP Algoritmasında Fayda-Maliyet Bilgisi.....	125
4.3.3. Duyarlılık Analizi Sonuçları	126
4.3.4. Fayda Doğruluğu	129
4.3.5. Hedef Sınıflandırma Doğruluğu	129
4.3.6. Optimal Fayda Maliyet Bileşimi ve Toplam Fayda	130
4.3.7. Genel Sonuçlar	131

BEŞİNCİ BÖLÜM

SONUÇ VE ÖNERİLER.....	132
KAYNAKÇA	135
EKLER	142

TABLO LİSTESİ

Sayfa No

Tablo 3.1. WEKA Platformu Öğrenme Algoritmaları	70
Tablo 3.2. Örnek Karar Tablosu.....	74
Tablo 4.1. Veri Setinin Genel Yapısı	95
Tablo 4.2. Sınıfların Sektörel Dağılımı	98
Tablo 4.3. Öğrenme Algoritmalarının Sonuçları (1:10).....	101
Tablo 4.4. Öğrenme Algoritmalarının Sonuçları (1:20).....	102
Tablo 4.5. CostSensitiveClassifier Sonuçları (1:10).....	103
Tablo 4.6. CostSensitiveClassifier Sonuçları (1:20).....	104
Tablo 4.7. MetaCost Sonuçları (1:10)	105
Tablo 4.8. MetaCost Sonuçları (1:20)	106
Tablo 4.9. Alt Örneklem İstatistikleri.....	107
Tablo 4.10. Genel Sınıflandırma Doğruluğu.....	107
Tablo 4.11. Hedef Sınıflandırma Doğruluğu.....	108
Tablo 4.12. Ortalama Maliyetler	108
Tablo 4.13. Ortalama Maliyet-1	110
Tablo 4.14. Ortalama Maliyet-2.....	111
Tablo 4.15. Ortalama Maliyet-3.....	112
Tablo 4.16. Ortalama Maliyet Verimliliği.....	113
Tablo 4.17. Maliyet Duyarlılığı-1	114
Tablo 4.18. Hedef Sınıflandırma Duyarlılığı	115
Tablo 4.19. Maliyet Duyarlılığı-2	115
Tablo 4.20. Genel Sonuçlar	116
Tablo 4.21. Senaryo Analizi Sonuçları	117
Tablo 4.22. Genel Sınıflandırma Doğruluğu ve Fayda-Maliyet Duyarlılığı.....	126
Tablo 4.23. Hedef Sınıflandırma Doğruluğu ve Fayda-Maliyet Duyarlılığı.....	129

ŞEKİL LİSTESİ

	<u>Sayfa No</u>
Şekil 1.1. Finansal Baskı Maliyetleri.....	19
Şekil 2.1. Karar Ağaçlarının Genel Yapısı.....	47
Şekil 2.2. Sabit Maliyet Tablosu	57
Şekil 2.3. Değişken Maliyet Tablosu	58
Şekil 2.4. Banka Kredileri ve Değişken Fayda-Maliyet Tablosu	59
Şekil 2.5. Firma Başarısızlığı ve Değişken Fayda-Maliyet Tablosu.....	61
Şekil 2.6. Sabit Fayda Tablosu	62
Şekil 2.7. Değişken Fayda Tablosu.....	63
Şekil 2.8. Sabit Fayda-Maliyet Tablosu.....	65
Şekil 2.9. Değişken Fayda-Maliyet Tablosu	65
Şekil 3.1. BCFP'nin Üçgen Dağılım Varsayımı	90
Şekil 4.1. Genel ve Hedef Sınıflandırma Doğruluğu ve Optimal Fayda- Maliyet Bileşenleri.....	127
Şekil 4.2. Genel Sınıflandırma Doğruluğu ve Optimal Fayda Maliyet- Bileşenleri	128
Şekil 4.3. Hedef Sınıflandırma Doğruluğu ve Optimal Fayda Maliyet- Bileşenleri	128

KISALTMA LİSTESİ

BCFP	: Benefit Maximizing Classifier Based Feature Projection (Fayda Maksimizasyonu Tabanlı Öznitelik İzdüşümü)
BMFT	: Benefit Maximizing Feature Treshold (Fayda Maksimizasyonu Tabanlı Öznitelik İzdüşümü)
IRIL	: Interactive Rule Interestingness Learning (Etkileşimli Kural İlginçliği Öğrenme)
BIS	: Bank of International Settlements (Uluslararası Ödemeler Bankası)
WEKA	: Waikato Environment for Knowledge Analysis (Waikato Bilgi Analiz Platformu)
NACE (Rev.1)	: Nomenclature Generale des Activités Economiques dans les Communautés Europeennes (Avrupa Birliği Tarafından Benimsenen Sektör Sınıflandırma Sistemi)

SEMBOL LİSTESİ

C_i	: i Sınıfı
C_j	: j Sınıfı
D_L	: Borçlu Firmanın Piyasa Değeri
D_U	: Borçsuz Firmanın Piyasa Değeri
F_k	: k Özniteliği
$g_{t+j+1,t+j}$	: t+j Döneminden t+j+1 Dönemine Mutlak Net Sermaye Büyüme Oranı
k.a.	: Maliyet/Fayda Kısıtı Altında
lb	: Aralık Alt Sınırı
Γ_i	: Set_i Örneklem Büyüklük Katsayısı
MAX(.)	: Maksimizasyon Amaç Fonksiyonu
MIN(.)	: Minimizasyon Amaç Fonksiyonu
$\mu(\alpha)$	: Birinci Tip Hata Maliyeti
$\mu(\beta)$	: İkinci Tip Hata Maliyeti
$\mu_{A(k)}$	: k Özniteliği İçin A Sınıfının Aritmetik Ortalaması
$\mu_{B(k)}$	: k Özniteliği İçin B Sınıfının Aritmetik Ortalaması
$\mu_{AB(k)}$	: k Özniteliği İçin Sınıf Ortalamalarının Aritmetik Ortalaması
N_c	: Aralıktaki Örneklerden c Sınıf Etiketli Olanların Sayısı
$\bar{O}_k$	: k Özniteliği
P_i	: i Öncülü
P_t	: t Dönemindeki Kâr/Zarar
q_f	: f Özniteliği İçin Sorgulama Örneğinin Değeri
$\zeta(Set_i)$	: Set_i Örneklem Büyüklüğü
S_{Tahmin}	: Tahmin Edilen Sınıf Değeri
$S_{Gerçek}$	: Gerçek Sınıf Değeri
Θ	: Mantıksal İşleç
T	: Eşik Değeri
ub	: Aralık Üst Sınırı
V_i	: Aralığın i Sınıfına Verdiği Oy Sayısı
V_p	: Aralığın p Sınıfına Verdiği Oy Sayısı

EK LİSTESİ

Sayfa No

Ek 1. İmalat Sanayii Alt Sektörleri (NACE (Rev.1) Kodları).....	143
Ek 2. Öznitelikler, Açıklamaları ve Aldıkları Değerler	144
Ek 3. BCFP İlginç ve Anlamlı Kurallar.....	153

ÖZET

Yakın dönemlerde yaşanan finansal krizler ve özellikle büyük ölçekli firma iflasları, bankacılık sektörünün ve finansal otoritelerin finansal sektörle birlikte reel sektörü de takip etmesine ve dikkatlerini firma riskliliğine çevirmelerine yol açmıştır. Bu bağlamda Uluslararası Ödemeler Bankası (BIS), bankaların içsel derecelendirme sistemleri geliştirmeleri gerekliliğini bir ilke kararı olarak BASEL II kriterleri arasına dahil etmiştir. Böylece, içsel derecelendirme ve risk kriterlerinin yerel ülke bankalarınca dikkate alınarak, Avrupa Birliği ülkelerinde 2007 sonuna kadar, Türkiye’de ise en geç 2008 Ocak tarihinden itibaren uygulamaya konması hedeflenmektedir.

Reel sektör riskliliğinin incelenmesinde firmaların finansal ve operasyonel bilgileri, risk analizinin girdisini oluşturmaktadır. Bu bilgilerin, gerek toplulaştırılarak sektör eğilimleri, gerekse firma bazında incelenmesiyle firma finansal davranışları hakkında bilgi edinmek olanaklıdır.

Firma bazında riskliliğin belirlenmesi, finans kuramında finansal baskı ve firma başarısızlığı kavramlarıyla incelenmektedir. Bu iki başlık altında, özellikle Büyük Bunalım sonrasında önemli çalışmalar yapılmıştır. İlk dönem çalışmaları, riskli ve sağlıklı firmaların finansal göstergeleri arasındaki farklılıklara yoğunlaşırken, özellikle donanım ve yazılım kapasitesindeki ciddi düzeyde artışla birlikte, son 20 yılda bilişim sistemleri firma başarısızlığı tahmin sürecinin en önemli bileşenlerinden biri haline gelmiştir. Özellikle, makina öğrenmesi yöntemleri başlığı altında ele alınan yöntemlerle firma başarısızlığının tahmin edilmesinde önemli yol kat edilmiştir. Bu bağlamda, firma ve sektör risklilik analizinde makina öğrenmesi yöntemlerinin önemli bir potansiyel taşıdığı banka kredi portföylerinin yönetimi ve finansal otoritelerin politika kararlarının biçimlendirilmesinde yardımcı bir araç olarak kullanılabileceği düşünülmektedir.

Bu alıřma, finansal baskı ve firma bařarıszlıęı kuramları ve makina ğrenmesi yöntemlerini ele aldıktan sonra ampirik uygulamalarla firma bařarıszlıęını incelemektedir. Özellikle, maliyete duyarlı sınıflandırma ve sonuçları üzerinde önemle durulmuřtur. Kullanılan veri seti, Merkez Bankası Reel Sektör Verileri Müdürlüęü veri tabanından elde edilmiřtir.

Anahtar Kelimeler: Finansal Baskı, Firma Bařarıszlıęı, Makina ğrenmesi, Sınıflandırma ve Maliyete Duyarlı ğrenme.

ABSTRACT

Recent financial crises and especially large corporate bankruptcies, have led bank managements and financial authorities to follow and monitor both financial and real sector risks, and to focus on firm failures. Bank of International Settlements, has therefore, taken the decision to include the necessity for banks to employ internal rating systems among BASEL II criteria. Thus, risk assessment and internal rating systems criteria would be made operational by the individual European Union banking systems, by the end of 2007, and January 2008 in Turkey, at the latest.

Financial and operational information of the firms, makes up the input to the risk analysis. This information can be aggregated to portray the sectoral trends, and/or focused upon on a firm basis to understand firms' financial behaviours. Finance theory summarizes firms' risks under financial distress and firm failure. There have been a myriad of works under these two headings, particularly in the United States, after the Great Depression. While early studies have focused upon the differences in the financial ratios of financially sound and failed firms, especially with the advances in computing capacity, the last two decades have witnessed an increasing use of machine learning methods in the failure prediction. Therefore, machine learning methods can be considered as having great potential in failure prediction and being good candidates as decision aids for policy-making.

This study considers financial distress and firm failure on theoretical grounds, gives a compact but elaborate explanation of machine learning schemes, and analyzes the results of these schemes run with data obtained from the database of Real Sector Data Division of the Central Bank. Cost sensitive learning was given special attention in the analysis.

Keywords: Financial Distress, Firm Failure, Machine Learning, Classification and Cost Sensitive Learning.

GİRİŞ

Finansal krizler sırasında ve sonrasında yaşanan gelişmelerin, finansal ve reel sektöre olumsuz etkileri geçtiğimiz yüzyıl boyunca gözlenmiştir. Yakın dönemlerde yaşanan krizlerde, finans sektörü kadar her ölçekte reel sektör firmasının finansal baskıya girmesi ve özellikle büyük ölçekli firmaların iflasları, firma bazında ve sektörler genelinde risklerin incelenmesi ihtiyacını doğurmuştur.

Sözü edilen riskler, çeşitli menfaat gruplarını farklı açılardan ilgilendirmektedir. Bu menfaat grupları, firmanın hissedarları, alacaklıları, tedarikçiler, müşteriler, çalışanlar, kredi ilişkisi içinde bulunduğu bankalar ve finansal otoritelerdir. Her menfaat grubunun risk algılaması ve getiri beklentileri birbirinden farklıdır. Getiri beklentilerini şekillendiren ölçüt risk algılaması olduğundan, bu grupların firmaya yönelik tutum ve davranışlarını riskliliğin ölçülmesi biçimlendirmektedir.

Bankalar doğrudan, finansal otoriteler ise bankacılık sistemi üzerinden dolaylı olarak kredi mekanizmasıyla oluşan risk grupları içinde en büyük riskleri taşımaktadırlar¹. Bu bağlamda, özellikle bankaların kredi portföylerinin risk profillerini, finansal otoritelerin ise ekonomi politikalarının firma ve sektör bazında riskliliğe etkilerini takip etmeleri gerekmektedir.

Bankalar tarafından riskliliğin takibi, küresel ekonomilerdeki kırılganlığın artması, finansal istikrarsızlık ve krizlerle birlikte BIS'in gündemine alınmış ve banka içsel derecelendirme sistemlerinin gerekliliği BASEL II kriterleri arasına dahil edilmiştir. BASEL II'de öngörülen risk ölçüm ve içsel derecelendirme sistemlerinin Avrupa Birliği ülkelerinde 2007 yılı içinde, Türkiye'de ise Ocak 2008 tarihinden itibaren uygulamaya konması hedeflenmektedir.

¹ Bu risklerin boyutu yanında karmaşıklığı da ayrı bir sorun oluşturmaktadır.

Firma riskliliği ya da firma başarısızlığının tahminine yönelik çalışmaların geçmişi Büyük Bunalım'a (1929-1939) kadar uzanmaktadır. İlk dönem başarısızlık çalışmalarında, analiz yöntemleri ve bilgi işleme altyapısındaki kısıtlar nedeniyle, ilgi odağı sadece sağlıklı ve sağlıklı olmayan firmaların mali oranları arasındaki farklılıkların karşılaştırılmasıyla sınırlı kalmıştır. Aradan geçen 40 yıla yakın bir süre boyunca üzerinde fazla durulmayan başarısızlık tahmin çalışmaları, 1960'lı yılların sonlarından itibaren yeniden finans araştırmalarının gündemine alınmış ve gelişen altyapıyla birlikte daha karmaşık yöntemlerin geliştirilmesi olanaklı olmuştur. İkinci dönem çalışmaları olarak adlandırılabilen bu çalışmalarda, genellikle matematiksel ve istatistiksel yöntemler kullanılarak finansal oranlar arasındaki ilişkiler modellenmeye ve firma davranışları açıklanmaya çalışılmıştır. Son yirmi yılda gelişen analiz yöntemleri, bilgi işleme, yazılım ve donanım altyapısındaki ilerlemelerle, matematiksel ve istatistiksel yöntemlerle yapılan uygulamalar daha da karmaşılaşmıştır. Bilgi teknolojilerindeki en önemli gelişmelerden sayılan makina öğrenmesi yöntemleri aracılığıyla parametrik olmayan modellerin geliştirilmesi, ayrıca, finansal bilgilerin yanında niteliksel bilgilerin de bu modellerde kullanılabilmesi olanaklı olmuştur.

Firmaların içsel ve dışsal etmenlere bağlı olarak gelişen davranışları ve bu davranışlarının somut bir yansıması olarak faaliyetleri ve finansal yapılarındaki gelişmeler, firmanın durumunu yansıtan finansal tablolar ve finansal tablo dipnotları olarak gözlemlenmekte, bunun yanında, modern yönetsel yöntemleri uygulayan firmalarda, niteliksel bilgiler içeren daha ayrıntılı faaliyet raporlarına da rastlanmaktadır².

Firma göstergelerinin incelenmesi ve bu göstergelerin tüm firmalar arasında genel bir eğilim sergileyip sergilemediğinin sorgulanması yoluyla finansal yapı ve davranış örüntü ya da şablonlarının çıkarılması, finansal baskının etkilerinin anlaşılmasını sağlayacaktır. Bununla birlikte, firma davranışları, genellikle karmaşık süreçler sonrasında gözlemlenir ve bu davranışların hangi nedenden kaynaklandığını ilk bakışta söylemek kolay

² Bu tür bilgiler arasında firma yöneticilerinin yaşları, firmanın piyasa tarafından nasıl algılandığı vb. bilgiler yer almaktadır.

değildir. Firmayla doğrudan ya da dolaylı ilişkisi olan tüm menfaat grupları bu davranış örüntülerini değişik ölçülerde ve olumlu ya da olumsuz etkilerler. Örneğin, firmanın sermayedarları ve alacaklıları arasında sürekli bir menfaat çatışması ve rekabet vardır ve bu rekabetin olumsuz etkileri³, finansal baskı altındaki bir firmada en sert şekliyle gözlemlenecektir.

Finansal baskı ve firma başarısızlığının dinamiklerinin anlaşılması, sektör ya da firma bazında riskliliğin açıklanabilmesi açısından önemlidir. Firma başarısızlığının analizi, firmaların finansal ve operasyonel bilgiler yanında, niteliksel bilgilerinin de incelenmesini gerektirmektedir. Bunlardan standart ve göreceli olarak nesnel olan finansal bilgiler diğerlerinden daha fazla bilgi içermektedir. Operasyonel ve niteliksel bilgilere kolaylıkla ulaşılamamakta, ulaşılan bilgilerse firmalar arasında standart bir yapı sergilememekte, bu bağlamda, bu bilgilerin kullanımı zorlaşmaktadır.

Çalışmamızda, finansal baskının nasıl gözlemlendiği ve firmanın finansal göstergelerine nasıl yansıdığı incelenmektedir. Bu çerçevede, firma göstergeleri makina öğrenmesi algoritmaları aracılığıyla incelenerek, firma başarısızlığının öngörülmesine yönelik modellerin karşılaştırılması ve kullanılabilecek modeller geliştirilmesi amaçlanmıştır. Tezin amaçları:

- Finansal baskı ve firma başarısızlığının temel altyapısı incelenerek, genel bir firma başarısızlık ölçütü geliştirmek⁴,
 - Makina öğrenmesi yöntemlerinin incelenmesiyle, ampirik analizlerde kullanılmak üzere mantıksal algoritmalar arasından bir yöntemi belirlemek ve
 - Ampirik uygulamalar sonucunda makina öğrenmesi yöntemlerinden kullanılabilir bir model oluşturma potansiyelinin varlığını belirlemek
- olarak özetlenebilir.

Tez yukarıdaki amaçlar doğrultusunda beş bölümden oluşmuştur. Birinci Bölümde, finansal baskı kuramı ve firma başarısızlığının ayrıntıları incelenmiştir.

³ İddia sahipleri arasındaki rekabetin yönetsel verimliliği arttırmak gibi olumlu yönlerinin de olduğu iddia edilmektedir.

⁴ Dördüncü bölümde, sınıf etiketlerinin belirlenmesi başlığı altında incelenmektedir.

İkinci Bölümde, makina öğrenmesi yöntemleri, örneklerle incelenmiş, özellikle maliyete ve faydaya duyarlı sınıflandırma yöntemlerinin ayrıntılarına değinilerek açıklamaları yapılmıştır.

Üçüncü Bölümde, makina öğrenmesi çalışmalarında sıklıkla başvurulanan Waikato Bilgi Analiz Platformu (WEKA) öğrenme algoritmaları ile faydaya duyarlı öğrenme için geliştirilen Fayda Maksimizasyonu Tabanlı Öznitelik İzdüşümü (BCFP) algoritması anlatılmıştır.

Dördüncü Bölümde, Üçüncü Bölümde açıklamaları yapılan WEKA platformu algoritmaları ve BCFP algoritmasıyla yapılan ampirik uygulamaların sonuçları sunulmaktadır.

Son bölümde, tez ana hatlarıyla özetlenmiş ve öneriler sunulmuştur.

BİRİNCİ BÖLÜM

FİNANSAL BASKI VE FİNANSAL BAŞARISIZLIK

1.1. Finansal Baskı ve Firma Başarısızlığı

Piyasa ekonomilerinde, firmaların herhangi bir sektöre giriş ya da çıkışları, rekabet sürecinin temelini oluşturmaktadır. Rekabet süreci, yeterli sayıda firmanın sektörde kalmasını ve piyasa talebinin rekabetçi fiyatlarla ve verimli bir üretim süreciyle karşılanmasını sağlar. Diğer taraftan, firmanın giriş ya da çıkışı, her zaman firmanın fiziksel olarak sektöre dahil olması ya da sektörden ayrılması anlamına gelmeyebilir. Giriş ya da çıkış, faaliyetlerde artış ya da azalış, üretim sürecine ayrılan kaynakların artması ya da azalması ve/veya faaliyet kolunun değiştirilmesi biçiminde de gözlemlenebilir. Bu bağlamda rekabet süreci, verimli kaynaklar sektöre dahil olurken verimsizlerin sektörden dışlanması biçiminde algılanabilir. Örneğin, bazı ürünlere talebin azalması, üretime ayrılan kaynakların bir kısmının başka üretim süreçlerine kaydırılmasına ya da üretim tesisinin tamamen kapatılmasına yol açabilir. Büyük ölçekli firmalarda, çıkış süreci kaynakların üretim süreçlerindeki yerlerinin yeniden yapılandırılmasıyla tanımlanabilir. Bu bağlamda, piyasa ekonomileri ve rekabetçi yapı verimsiz süreçlerden verimli olanlara doğru bir kaynak akışı ya da hareketiyle nitelendirilebilir. Kuramsal olarak, tam rekabetçi bir piyasada, çıkışın gerçekleştiği durumların çok azında, hatta seyreklikle, finansal baskı, temerrüt ya da acze düşme, iflas ve firmanın faaliyetini tamamen durdurması beklenir (Hashi, 1997).

Modern firmalar, firma varlıkları üzerinde iddia sahibi farklı grupların menfaatlerini düzenleyen resmi ve örtülü bir anlaşmalar ağıyla tanımlanabilir. Menfaat grupları ya da iddia edenler, hükümet, bankalar, teminatlı ve teminatsız alacaklılar, çalışanlar, tahvil alacaklıları, müşteriler, tedarikçiler, ve yöneticiler ve hissedarlardır. Bu anlaşmalar ağı, gelişmiş piyasa ekonomilerinde, mülkiyet haklarının bir parçasıdır. Söz konusu anlaşmaların

işleyişi, piyasa ekonomilerinde, finansal sistem ve finansal piyasalar aracılığıyla gerçekleşir. Finansal kurum ve piyasalar, ekonomik birimlerin performanslarıyla ilgili bilgileri sunar, piyasa katılımcılarının tepkilerini yansıtır ve ekonomik birimlerin faaliyetlerinin düzenli bir biçimde yürütülmesini sağlarlar (Hashi,1997).

Gelişen piyasalarda, finansal baskı kendini düşüşe geçen piyasa fiyatlarıyla gösterirken, olası mekanizmaların harekete geçmesini de sağlar. Diğer taraftan, birleşme ve devralma faaliyetleri finansal baskı ve düşen fiyatlara bağlı olarak yoğunlaşmaktadır. Bu durum özellikle, piyasa katılımcılarının, firmanın finansal problemlerini, verimsiz üretim ve yönetsel sistemler ya da artık talep edilmeyen ürünlerin üretilmesi bağlamında geçici bir durum olarak algılamasından kaynaklanmaktadır. Devralma mekanizması aracılığıyla, kuramsal olarak, finansal baskı altındaki firmaların, yeni sahiplerinin kontrolünde, doğru ürün yelpazesini daha verimli bir biçimde üretmesi sağlanabilecektir. Bu bağlamda, devralma süreci, atıl ya da verimsiz kaynakların sistemden etkin olarak çıkarılmasını sağlayacaktır (Whitekar, 1999).

Diğer taraftan, baskı altındaki firma, alacaklılarıyla, borçlarını yeniden yapılandırmak ve finansal yapısını düzeltmek amacıyla resmi ya da gayri resmi görüşmelere başvurabilir. Yeniden yapılandırma programları, genellikle, işten çıkarma ve zarar eden faaliyetlerin durdurulmasını da içerecek şekilde, firmanın yeniden yapılandırılmasını öngörür. Burada da kaynakların sektörden çıkması söz konusudur. Son olarak, başvurulacak başka bir yol kalmadığında, finansal baskı altındaki firmanın tasfiyesi, yani fiziksel varlığının ortadan kalkması bir seçenek olarak ortaya çıkmaktadır.

Gelişmiş piyasa ekonomilerinde, geçmişte finansal baskıya giren firmalar ve bunlar arasından iflas edenlerin sayısı oldukça az olmuştur. Gelişmekte olan ülkelerdeyse, geçmiş dönemlerde, her ölçekten çok sayıda firma finansal baskı sürecine girmiştir. Finansal baskıya giren firmaların tasfiye edilmesi ya da tamamen dışlanması rekabet sürecinin doğal bir sonucu olarak görülebilir. Bununla birlikte, böyle firmaların tasfiyesine izin verilmesi veya zorlanması, potansiyel bir felaketle sonuçlanabilir: Üretim

kapasitesinde önemli ölçüde bir azalma olabilir ya da bu kapasite tamamen ortadan kalkabilir. Ayrıca, işsizlik sigortası ve sosyal güvenlik ödemeleri devlet bütçesine ek bir yük getirecek ve makroekonomik politikaları olumsuz yönde etkileyecektir. Gelişen ekonomilerde, bu süreçlerin herhangi bir şekilde olumsuz sonuçlanması sosyal refah açısından arzu edilmez ve menfaat gruplarının finansal baskının olumsuz sonuçlarını hoşgörmesi ve katlanması beklenemez⁵ (Hashi, 1997).

Kriz dönemlerinde, özellikle finansal yapıları bozuk ve problemlili işletmelerin finansal güçlükler yaşadıkları, hatta iflas ettikleri bilinmektedir. Bununla birlikte, bir firmanın finansal baskıya girmesi için ekonomik bir krizin yaşanması gerekmez. Ekonominin genel gidişatının iyi olduğu dönemlerde bile, firmaların çağdaş bilimsel yöntemlerle yönetilmemesi finansal problemlerle karşılaşmaları için yeterlidir. Hatta bazen bilimsel yöntemler bile firmanın iflas etmesini önlemeye yetmeyebilir⁶.

1.2. Finansal Baskı Sürecinin Aşamaları ve Firma Başarısızlığı

Firma başarısızlığı, finansal ve operasyonel baskı alt başlıkları altında incelenmektedir. Bu bağlamda, finansal ve operasyonel baskı sürecinin kuramsal altyapısının tanımlanması gerekir. Finansal baskı, finansman kuramında, firma değerlendirme ve optimal sermaye yapısının belirlenmesi başlıkları altında ele alınmaktadır. Bu bağlamda, nakit akımı yaratma ve borç ödeme gücü gibi daha çok firmanın borç-sermaye yapısını belirleyen konular üzerinde yoğunlaşmaktadır⁷. Operasyonel baskı analizinde, finansal göstergelerden çok yönetsel ve çoğunlukla niteliksel etmenler ele alınmaktadır.

Finansal baskı süreci dinamik ve çoğunlukla uzun bir süreçtir ve firmanın sermaye yapısı, yatırım politikaları ve performansı üzerinde etkilidir

⁵ Çalışmamızın kapsamı dışında olduğu için finansal baskının ekonomiye olumsuz etkilerine ayrıntılı bir biçimde değinilmemiştir.

⁶ Örneğin aktif büyüklüğü 125 milyar ABD doları, bilanço dışı pozisyonuysa 1.25 trilyon ABD dolarına ulaşan Long Term Capital Management adındaki büyük fon Fisher Black gibi Nobel İktisat Ödülü almış kişiler tarafından yönetilmesine rağmen iflasın eşiğine gelmiş, FED mali sistemde oluşabilecek bir çöküntüyü önlemek için bu firmaya büyük miktarda kaynak aktarmak zorunda kalmıştı (Perold, 1999).

⁷ Finansal başarısızlığı tahmin çalışmalarında, genellikle bu dar kapsamdan çıkılarak, firmanın hem niteliksel hem de niceliksel özellikleri ele alınmaktadır. Çalışmada, çok sayıda standart niteliksel değişkenin elde edilmesi olanaklı olmadığından, sadece firma yaşı, İMKB kotasyon bilgisi, sektör bilgisi gibi niteliksel değişkenler kullanılabilmektedir. Doğrudan kullanılabilecek niteliksel değişkenler yerine, operasyonel baskının etkilerinin dolaylı olarak sorgulanabileceği finansal değişkenlere başvurulmuştur.

(Kahl, 2002). Bir firma finansal baskıya girmeye başladığı zaman, tehlikeli bir bölgeye girmeye başlamış demektir. Baskının fark edilmesiyle birlikte, firmanın faaliyetleriyle ilgili önlemlerin alınması gerekir. Baskının erken aşamada anlaşılması ve bu önlemlerin daha çabuk alınması, firmanın tehlikeli bölgeden olabildiğince erken çıkarılmasını ve sürecin firmaya daha az zarar vermesini sağlayacaktır (Whitekar, 1999; Güvenir, 2003). Diğer taraftan, baskının geç fark edilmesi hasarın derinleşmesine ve belki de firmanın kurtarılması zor bir aşamaya gelmesine neden olacaktır.

Finansal baskı süreci ve aşamaları aşağıda açıklanmaktadır.

1.2.1. Finansal Baskı

Literatürde, finansal baskının tanımı, geniş bir yelpaze içinde yer almaktadır. Dinamik ve uzun bir süreç olan finansal baskı süreci, sağlıklı bir firmanın finansal yapısının kısa süreli ve aşırı büyük tek bir olay ya da birbirini takip eden olaylar zinciri ya da uzun süre boyunca tekrarlanan olumsuz olaylar sonucunda belirli bir eşik noktasını geçmesi ile başlar. Sürecin başlangıcı olarak kabul edilen nokta veya aralığın nerede ve özelliklerinin neler olduğu hakkında kesin bir bilgi bulunmamaktadır. Bu aşamada gözlemlenebilen finansal baskı altındaki firmanın finansal sorunlar yaşamaya başlamasıdır. Bunun yanında, finansal baskının, hissedilmeye başladığı andan itibaren geçici veya sürekli olup olmaması, firmanın sürekliliğini belirleyecektir. Sürecin aşamalarının birbirine geçtiği bölgelerde net ve kesin bir ayırım yapmak olanaklı değildir ve sonraki aşamalar, genellikle önceki aşamaların sonucuyken, önceki aşamalar, aynı zamanda sonraki aşamaların bir parçasıdır. Dolayısıyla, finansal baskı aşamaları kesin çizgilerle birbirinden ayrılamamaktadır.

Finansal baskı altındaki firmalar erken ve ileri aşamalarda farklı özellikler göstermektedirler. İlk aşamadaki firmalarda, kilit yönetimdeki hızlı değişim, müşteri portföyünden önemli müşterilerin çıkması, faaliyet zararları, aralıklı nakit açıkları ve ticari alacakların tahsilinde sorunlar gözlemlenmektedir.

Finansal baskının orta aşaması, sürekli faaliyet zararları, borç ödemelerinde ek vade ya da erteleme talebi, borç sözleşme koşullarının ihlali, önemli tedarikçilere borç ödemelerinde aksama ve tedarikçilerin vadelerinde kısılma gibi sorunlarla özdeşleştirilmektedir. Bu aşamada, nakit yönetimi günlük faaliyetlerden daha önemli hale gelmeye ve öne çıkmaya başlar.

Geç aşamadaysa, firmalar sürekli faaliyet zararları yaşarken, nakit açıkları büyür ve borç sözleşmelerinin ihlali kronik bir sorun olmaya başlar. Şüpheli alacakların tahsili neredeyse olanaksızlaşır ve nitelikli çalışanların istifalarında artma gözlenmeye başlar.

Finansal baskı literatürde:

- Firma varlıklarının değerinin, belirli bir eşik düzeyinin altına düşmesi (Purnanandam, 2005)
 - Nakit çıkışları ve kâr paylarının ertelenmesi ya da azaltılması (Jaggi ve Lee, 2002),
 - Acze düşme, yani borç ödemelerinin vadelerinde yapılamaması (Purnanandam, 2005),
 - Firma likit varlıklarının borç ödemelerini karşılamaması,
 - Dönen varlıklar düzeyinin, vadesi gelen krediler, diğer borç ödemeleri ve yatırım giderlerini karşılamaması (Mc Mahon ve Reese, 2003),
 - Faaliyet nakit girişlerinin kısa vadeli borçları karşılamaması (Wruck, 1990),
 - Verimsizlik kayıplarına yol açan ve sürgit düşük nakit girişleri ya da nakit çıkışlarıyla özdeşleştirilen durum (Mc Mahon ve Reese, 2003) ve
 - Firmanın, borç ödeme araçlarından yoksun kalması (Ross ve Westerfield, 1996)
- olarak tanımlanmaktadır.

Bu tanımlardan, finansal baskının, borç ödeme gücü ve nakit akımlarıyla özdeşleştirilebileceği ve acze düşme, iflas ve tasfiyenin finansal baskının aşamaları olduğu söylenebilir.

En genel şekilde finansal baskıya giren firmalar, borç ödeme sorununu iki yoldan birine başvurarak çözmek durumundadırlar. Birinci yol, sorunu iflasa başvurmadan çözmektir. Bu durumda, firmanın varlıkları ya da yükümlülüklerinin veya her ikisinin birden yeniden yapılandırılması söz konusudur. İkinci yolsa, daha sert bir çözüm yolu olan iflastır. Firmanın iflasa başvurması durumunda, menfaat gruplarının firma üzerindeki güç dengeleri ve sorumluluk dereceleri değişecektir. İflas eden firma ya tasfiye edilir ya da yeniden yapılandırılır. Buradaki yeniden yapılandırma iflas etmeden yeniden yapılandırmadan farklıdır; sürecin sonunda, firmanın sermaye yapısı alacaklılar yönünde değişecek ve alacaklılar firmanın yeni sahipleri olacaktır. Finansal baskı altındaki firmaların diğer seçenekleri, birleşme, borcun yeniden yapılandırılması ve gönüllü tasfiye olarak sıralanabilir (Gilbert ve diğerleri, 1990).

Finansal baskı altında firma çeşitli düzeylerde verimsizlik kayıplarına uğrayacaktır. Örneğin, ampirik çalışmalar finansal baskı altında firmaların, ekonomik ve/veya sektörel krizlerde piyasa paylarının önemli bir kısmını rakiplerine kaptırdıklarını göstermektedir (Purnanandam, 2005).

Finansal baskı süreci, farklı menfaat grupları açısından maliyetli bir süreçtir. Temel olarak, doğrudan ve dolaylı maliyetler olarak iki grup altında incelenen baskı maliyetleri, finansal baskının aşamasına ve menfaat gruplarına göre değişiklik gösteren göreceli maliyetlerdir.

Finansal baskı bağlamında, finansal baskıdan kaynaklanan verimsizlik kayıplarının, finansal baskı maliyetleri ile bu maliyetlerin ölçeğinin ve finansal baskının sermaye yapısı üzerine etkilerinin bilinmesi önemlidir.

1.2.1.1. Finansal Baskı Süreci

1.2.1.1.1. Acze Düşme

Acze düşme, finansal baskı sürecinin aşamalarından biridir ve literatürde:

- Firma varlıklarının değerinin, borçların değerinin altına düşmesi (Purnanandam, 2005),
- Firmanın dönen ve duran varlıkları toplamının vadesi gelen borçların toplamının altına düşmesi (Mc Mahon ve Reese, 2003) ve
- Borçların vadesi geldiğinde ödenememesi (Mc Mahon ve Reese, 2003) olarak tanımlanmaktadır.

Finansal yapının sürekli bozulmasıyla acze düşme gerçekleşecek, bu sürecin devam etmesi durumundaysa, büyük olasılıkla alacaklılar iflas sürecini başlatacaktır. Alacaklıların iflas sürecine başvurması, firmanın finansal durumunun nasıl algılandığıyla ilintilidir. Firmanın, borçlarını uzun süre ödememesi, borçlarını yeniden yapılandırma talebi, borçların büyüklüğü gibi etmenler bu algılamayı olumsuz etkileyen faktörlerdir.

Altman (1983) finansal baskıyı, stok ve akım tabanlı acze düşme riskiyle ilişkilendirmektedir. Stok tabanlı acze düşme firmanın borçları varlıklarından fazla olduğunda gerçekleşir. Akım tabanlı acze düşme ise, firmanın nakit girişlerinin borç ödemelerini karşılayamaması olarak tanımlanmaktadır.

1.2.1.1.2. İflas

Firma iflası, genellikle finansal baskıdan kaynaklanıyorsa da, sürecin kendisi hukuksaldır (Dietrich, 1984). Sürecin başlatılması farklı menfaat gruplarının farklı talepleriyle gerçekleşir. Alacaklıların iflas sürecini başlatmaları genellikle borç ödemeleriyle ilgili sorunlardan kaynaklanmaktadır. Diğer taraftan, sermayedarların firmayı çeşitli nedenlerle iflas ettirmesi söz konusudur. Örneğin, kötü niyetli iflas olarak da

nitelendirilebilecek bir iflas türünde firma sahipleri, finansal yapısı sağlam ve sorunsuz bir firmayı iflas ettirebilirler. Ayrıca, faaliyet merkezinin bulunduğu ülkenin iflas mevzuatının sermaye yeterliliğiyle ilgili ölçütleri firma iflaslarında bir etmen olarak karşımıza çıkmaktadır. Sermaye yitikleşmesi ölçütü bu kısıtlardan en önemlisidir.

İflas kavramı literatürde:

- Zorunlu ödemelerin yerine getirilememesi ya da borçların vadesinde ödenememesi, (Haugen ve Senbet 1978),
- Firmanın varlıklarının defter değerinin borçlarının defter değeri altına düşmesi ve
- Firma sahipliğinin, sermayedarlardan alacaklılara geçmesi (Haugen ve Senbet 1988) olarak betimlenmektedir.

Birinci grup altındaki iflas türü likidite iflasıdır. İkinci grupta incelenen iflas türüyse, borç ödeme gücü zafiyeti kaynaklı iflas olarak adlandırılır. Bu iki iflas türünü birbirinden tamamiyle ayırmak olanaklı değildir; azalan likidite düzeyi borç ödeme gücünün zafiyetinden kaynaklanabilir ya da borç ödeme gücünün azaldığına yönelik bazı işaretler aslında likidite kısıtlarından kaynaklanabilir. Söz konusu iflas türlerinden likidite iflasına diğerinden daha sık rastlanmaktadır. Borç ödeme zafiyetinden kaynaklanan iflas sermaye kayıplarının telafi edilmesiyle önlenabilir. Fakat, likidite kısıtı nedeniyle borç ödeme güçlüğü yaşıyorsa burada alacaklılar ikna edilemediği sürece iflas kaçınılmaz olabilir (Back ve diğerleri, 1996).

1.2.1.1.3. Tasfiye

Sermaye bütçelemesi kararı olarak tasfiye, firma varlıklarının satış yoluyla elden çıkarılması ve firmanın fiziksel varlığının sona ermesidir (Haugen ve Senbet, 1978). Tasfiyenin genel özellikleri:

- Tasfiye kararının, firma varlıklarının tasfiye değerinin, faaliyete devam halindeki piyasa değerini geçtiğinde alınması (Haugen ve Senbet, 1978),

- Firmanın tasfiye edilmesi durumunda varlıkların ayrı ayrı ya da firmanın bir bütün halinde satılması ve elde edilen satış gelirinin alacaklılar ve hissedarlar arasında mutlak öncelik⁸ sırasına göre dağıtılması (Hashi, 1997) ve

- Genellikle, bütünüyle ya da tek tek firma varlıklarının elden çıkarılmasıyla hedeflenen düzeyde gelir elde edilememesi (Mc Mahon ve Reese, 2003)

olarak sıralanabilir.

1.2.2. Firma Başarısızlığı

Finansal baskı aşamalarının tanımlamalarından da anlaşılacağı gibi, literatürde, baskının kendisi ve aşamaları için farklı tanımlar arasında bazı benzerlikler ve çakışmalar görülmektedir. Bu bağlamda, finansal baskı tanımlarına dayalı bir modelleme yapmak oldukça zor görünmektedir. Söz konusu modellemenin, alana özgü bir yapı taşıması, kullanıcıların özgül taleplerini karşılaması gerekliliği, finansal baskının finansman kuramının dar bir alanında incelenmiş olması ve halihazırda üzerinde tam bir fikir birliğine varılmış bir finansal baskı kuramının eksikliği, araştırmacıları, bu dar çerçeveden çıkararak başarısızlığın her yönüyle incelendiği bir alanda çalışmaya yönlendirmiştir. Bu tanım alanı, firma başarısızlığı olarak adlandırılmaktadır.

Firma başarısızlığı, finansal ve operasyonel baskı başlıkları altında, niceliksel ve niteliksel tüm özelliklerin ayrı ayrı ve bir arada analizini olanaklı kılmaktadır. Başarısızlığın modellenmesinde en önemli aşama başarısız firmanın tanımlanmasıdır. Beaver (1966) başarısızlığı, acze düşme, iflas ve büyük meblağda nakit kredi kullanımı ile ilintilendirmektedir.

Deakin (1972) başarısız firmayı, acze düşmüş, iflas etmiş ya da tasfiye edilmiş olarak tanımlamaktadır.

⁸ Mutlak öncelik kuralı, tasfiye durumunda, hissedarlar ve yöneticilerin, genellikle, firmanın değerinin borçlarının altında kalması olasılığına bağlı olarak, süreç sonunda hemen hemen hiç birşey alamayacakları anlamına gelmektedir.

Blum (1974) başarısız firmaları, borçlarını vadesinde ödemeyen, alacaklılarla borç yapılandırmasına giden ya da iflas eden firmalar arasından seçmiştir.

Lau (1987), başarısızlığı, kâr paylarında azalma, acze düşme, gözaltı aşaması, iflas ve tasfiye ile betimlemektedir.

Ward (1994), kâr paylarında yarı yarıya azalma, acze düşme ve gönüllü ya da zorunlu gözaltı aşamasını finansal başarısızlık olarak tanımlamıştır.

Tanımlamalardan anlaşıldığı gibi, finansal başarısızlığın tam ve ortak bir tanımı bulunmamaktadır. Başarısızlığın tanımı, araştırmalar arasında farklılık göstermektedir. Tanımlamaların hemen hepsinde finansal başarısızlık, acze düşme ve iflasla ilintilendirilmektedir. Bununla birlikte, iflas başlığında da ele alındığı gibi, büyük bir olasılıkla finansal baskıdan kaynaklanan iflasın, homojen bir tanımı olmaması (Jones, 1987), bu ölçütün iyi bir başarısızlık ölçütü olamayacağını göstermektedir. Ayrıca, tanımlamalarda söz edilen olayların bazılarının gözlemlenmesi de olanaklı olmayabilir. Bu kısıt, veri setindeki firma sayısının azalması ya da örneklemin küçülmesi demektir ki, makina öğrenmesi algoritmalarıyla ilgili bölümde de açıklandığı gibi bilgi kayıplarına yol açacaktır. Dolayısıyla, firma başarısızlığı tanımının olabildiğince genel olması tercih edilmelidir⁹.

1.3. Finansal Baskının Nedenleri ve Teşhisi

Gelişmiş ekonomilerde finansal baskının firma ya da yönetimden kaynaklandığı söylenebilir. Diğer taraftan, gelişmekte olan ekonomilerde durum bundan daha karmaşıktır. Aynı anda çok sayıda firmanın acze düşmesinin nedeni sadece içsel nedenler olmayabilir. Dışsal şoklar ya da hükümet politikaları da finansal baskı sürecini tetikleyebilir. Makroekonomik politikalar, ticari serbestleşme ve finansal piyasaların derin olmaması da finansal güçlükler yol açabilir. Üretim süreçlerindeki uyarlamalar, kaynak edinimi, ürün yelpazesinin seçimi ya da pazarlama stratejileri birçok firma için

⁹ Başarısızlık tanımı, ampirik analiz bölümünde ayrıntılarıyla yapılmaktadır.

güç ve bu firmaların değişimlere kısa sürede tepki vermeleri olanaklı olmayabilir. Olası mülkiyet değişikliği ya da mülkiyet dönüşüm sürecinin kendisi ve yeni yatırımlar için gerekli fonların bulunamaması olasılığı sorunun daha da derinleşmesiyle sonuçlanabilir (Hashi, 1997).

1.3.1. Finansal Baskının Nedenleri

Finansal baskının nedenleri, dışsal ve içsel olarak iki başlık altında ele alınabilir. Dışsal nedenler, firmanın kontrolü dışındaki etmenlerdir ve iki grup altında incelenebilirler. Bunlar:

- Makroekonomik nedenler, ekonomik durgunluk, döviz kurlarındaki aşırı değişimler, işsizlik oranları, arz ve talep şokları, teknoloji ve hukuksal çerçevedeki anlık ya da sürekli değişimlerdir.
- Sektörel nedenler, aşırı kapasite, küresel arz ve talep kaymaları, ihracat kısıtlamaları, çevresel kısıtlamalar, teknolojik altyapıdaki değişimlerdir.

İçsel nedenler, kontrolü büyük oranda firmanın kendi elinde olan etmenlerdir. İçsel nedenler finansal ve operasyonel nedenler olarak iki başlık altında incelenmektedir. Bunlar:

- Finansal nedenler, vadesi aşırı uzatılmış, düşük teminat gösterilmiş borçlar, düşük nakit girişleri ya da nakit çıkışları, planlanmamış yatırım harcamaları, işletme sermayesi kısıtları, zayıf finansal ve yönetsel raporlama sistemleri, stoklar ya da sabit varlıklardaki aşırı büyüme, aşırı kredi borç yükü, yüksek kaldıraç oranı gibi finansal etmenlerdir.
- Operasyonel nedenler, müşteri portföyünden önemli bir müşterinin çıkması ya da eklenmesi, aşırı büyüme, hatalı yapılandırma girişimleri, yönetsel verimsizlik, yolsuzluk, yetki devri ve üst ve orta düzey yönetimdeki ihtilaflar ve vizyon eksikliği gibi etmenlerdir.

Özellikle firma başarısızlığında işaret edildiği gibi, finansal baskının nedenleri arasında operasyonel etmenler de yer almaktadır. Ayrıca, dışsal nedenler, finansal baskıya yol açan etmenler arasında gösterilmektedir. Buradan hareketle, literatürde çizilen çerçevenin, finansal baskı ve

başarısızlık tanımlarında da ortak bir noktada buluşmadığını söylemek olanaklıdır.

1.3.2. Finansal Baskı Göstergeleri

Finansal baskının tanımlanmasında diğer bir önemli unsur, baskı göstergeleridir. Aiyabei (2000), baskı göstergelerini:

- Üretim hattının kapatılması,
- Faaliyet zararları,
- İşten çıkarmalar,
- Üst düzey yönetimin istifası,
- Halka açık firmalar için, hızla düşen hisse senedi fiyatları¹⁰ ve
- Borç anlaşmalarının ihlal edilmesine yol açabilecek kadar aşırı

borç yükü

olarak tanımlamaktadır. Mc Mahon ve Reese (2003) ise, finansal baskı göstergelerini:

- Piyasa payındaki azalma ve satışlardaki düşmeye bağlı olarak faaliyet giderlerinin karşılanamaması,
- Aşırı kapasiteyle üretim ve saldırgan ve sürdürülemez fiyatlama stratejileri,
- Ticari borçların vadesinde ödenmemesi ve ek vade talebi ve
- Faiz oranları ve tüketici talebindeki olumsuz gelişmelere, rekabet pozisyonundaki hızlı değişimlere ve teknolojiye gelişmelere ayak uydurulamaması

gibi baskı altındaki firmalarda genellikle ampirik olarak gözlemlenen, işaretler olarak betimlemektedir.

¹⁰ Mükemmel piyasalarda, yatırımcıların riski neredeyse anında fiyatlara yansıttıkları kabul edilmektedir. Mükemmel piyasalar, işlem maliyetlerinin sıfır ya da sıfıra yakın olduğu ve piyasada faaliyet gösterenlerin fiyatları belirleyici değil kabul eden olduğu piyasadır. Bu piyasalarda resmi otoritelerin işlemler ve fon transferiyle ilgili müdahalesi yoktur ve piyasa şartları oyuncular tarafından belirlenir. Mükemmel piyasada finansal araçların fiyatları gerçek değerlerini ve yatırım aracıyla ilgili tüm bilgileri olduğu gibi yansıtır. Finansal araçla ilgili yeni gelişmeler de anında fiyatlara yansıtılır; bu tür piyasalar aynı zamanda verimli piyasalar olarak da adlandırılır. Verimli bir piyasada finansal aracın fiyatını etkileyecek herhangi bir bilgi kısa sürede piyasalara yansır ve piyasa oyuncuları bu bilgilere dayanarak aşırı kâr elde edemez (Rose, 1994).

1.4. Finansal Baskı, Firma Performansı ve Verimsizlik Kayıpları

Finansal baskının firmanın performansına olumlu ve olumsuz olmak üzere iki yönde etkisi olduğu kabul edilmektedir. Baskının olumsuz etkisi firmanın müşteri, tedarikçi ve alacaklılarıyla olan ilişkilerinde yarattığı verimsizlik, olumlu etkisiyse, yönetsel disipline bağlı olarak özellikle operasyonel anlamda verim artışlarıyla tanımlanmaktadır.

Finansal baskıdan kaynaklanan verimsizlik kayıpları:

- Firmaların müşteriler, tedarikçileri ve çalışanlarıyla ilişkilerinde bozulma,
 - Borç sözleşmelerindeki ek kısıtlar ve
 - Risk algılamasındaki olumsuzluk ve dışsal kaynak maliyetlerindeki artış
- olarak üç başlık altında toplanabilir.

Verimsizlik ve performans kayıpları, finansal baskı maliyetlerinin altyapısını oluşturmaktadır.

1.4.1. Finansal Baskı ve Firma Performansı

Finansman kuramı, finansal baskıyı sermaye yapısının belirlenmesinde önemli bir etmen ve yüksek maliyetli bir olgu olarak betimlemektedir. Baskı, menfaat gruplarını firma faaliyetlerine zarar verecek davranışlarda bulunmaya itmesi olasılığından dolayı maliyetli bir olay olarak görülmektedir. Bu davranışlar firmanın risk algılamasını olumsuz yönde etkiler ve menfaat gruplarıyla olan ilişkilerin maliyetini artırır (Opler ve Titman, 1994).

Bazı araştırmalar, finansal baskının firma performansına olumlu etkileri bulunduğu ve yapısal düzeltmeleri teşvik ettiğini öne sürmektedir (Brown ve diğerleri, 1992; Gilson, 1989). Bu önermeye göre, finansal baskı altında yöneticilerin firma değerini artırmaya yönelik kararlar alma eğilimi artmaktadır. Bu tür kararların olumlu sonuçlanması, baskının verimliliği artırdığı önermesini doğrulayacaktır (Opler ve Titman, 1994). Diğer taraftan,

finansal baskı altında, müşteriler ve rakiplerden kaynaklanan verimsizlik kayıplarının firma açısından maliyetli olacağı da açıktır.

Küçük ölçekli firmalar finansal açıdan oldukça kırılgan olup, aynı zamanda müşteri ve rakip kaynaklı verimsizlik kayıplarına da daha duyarlıdır. Büyük firmalar finansal baskının getirdiği disiplinden daha çok yararlanırlarken, yönetim kaynaklı verim kayıplarına daha duyarlıdırlar.

Finansal baskı ve performans arasındaki ilişki, araştırma geliştirme giderleri yüksek ve/veya yoğunlaşmanın yüksek olduğu sektörlerde faaliyet gösteren firmalarda daha fazla öne çıkmaktadır. Bu bulgu, gözlenen verimsizlik kayıplarının maliyetleri düşürme amaçlı üretim azalışlarından çok, müşteri ve rakip kaynaklı olduğu önermesiyle tutarlıdır. Bu bağlamda, finansal baskının büyük maliyetli kayıplara yol açarken, verimliliği artırmaya yönelik olarak faaliyetlerde yapılması gereken değişiklikleri de eş zamanlı olarak tetiklediği söylenebilir.

1.5. Finansal Baskı Maliyetleri

Finansman kuramına göre, baskı maliyetleri firmanın piyasa değerini etkiler. Bu önerme, finansal baskının yüksek borçlanma ya da kaldıraç oranlarından kaynaklandığını öne sürmektedir. Yüksek kaldıraç oranından, firmaların iki tür yarar sağladığı söylenebilir. Bunlar, yatırımın getirisi ve vergi avantajındaki artışlardır.

Finansal baskı maliyetli bir süreçtir. Yatırımcıların risk algılaması, kaldıraç oranı yüksek firmalarda finansal baskı olasılığına bağlı olarak olumsuz yönde değişecek ve risk primi olarak borçlu firmaların cari piyasa fiyatlarına yansıtılacaktır. Borçlu firma değeri üç ana başlık altında incelenebilir. İlk bileşen, firmanın borçsuz değeridir. Diğer bileşenler, borçlanmanın vergi avantajı ve borçlanmadan kaynaklanan finansal baskı maliyetleridir. Bu bağlamda, borçlu firmanın piyasa değeri:

$$D_L = D_U + BD(VA) - BD(FBM) \quad (1.1)$$

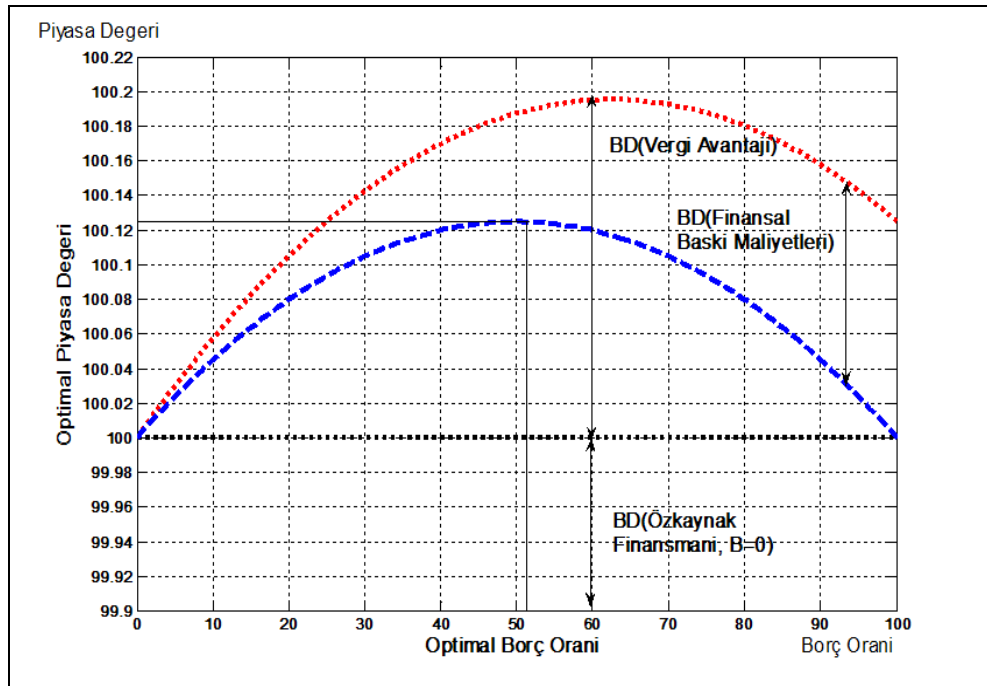
olarak ifade edilebilir.

D_L , borçlu firma değeri, D_U , borçsuz firma değeri, $BD(VA)$, vergi avantajının beklenen değeri ve $BD(FBM)$ ise finansal baskı maliyetlerinin beklenen değeridir.

Vergi avantajı (VA), borçlanmadan kaynaklanan faiz ve anapara ödemelerinin gider kabul edilmesinden dolayı, vergi öncesi kârlarda meydana gelen düşme ve bu bağlamda ödenecek vergi yükümlülüklerindeki azalmadır.

Finansal baskı maliyetleri (FBM), baskı olasılığı ve baskı gerçekleştiğinde katlanılacak maliyetlerin bir fonksiyonudur¹¹. Bu fonksiyonun biçimiyle ilgili ayrıntılı bir bilgi bulunmamaktadır.

Borçlu firmanın değeri, borçsuz firmanın piyasa değeri, veri dönemde sağlanması beklenen vergi avantajı ve finansal baskı maliyetlerinin beklenen değerinden oluşmaktadır. Borç kullanımı vergi avantajı sağlamakla birlikte, aşırı düzeyde borçlanma finansal baskı riskini de beraberinde getirir. Finansal baskı maliyetlerinin beklenen değeri borçlu firmanın değerini azaltır.



Şekil 1.1: Finansal Baskı Maliyetleri

¹¹ $BD(FBM) = P(FB) \times C(FB)$ olarak formüle edilebilir. $P(FB)$, finansal baskının gerçekleşme olasılığı, $C(FB)$ ise finansal baskının maliyetidir.

Aynı ilişki Şekil 1.1’de sunulmaktadır. Finansal baskı maliyetlerinin bugünkü değeri kaldıraçla birlikte yükselirken optimal bir sermaye yapısı (borç oranı) elde etmek olanaklıdır.

Firmanın finansal baskıyı aşamaması ve borç ödemelerini yerine getirememesi durumunda finansal baskı maliyetleri gerçekleşecektir. Finansal baskı maliyetleri, finansal baskının gerçekleşmesi durumunda ortaya çıkan maliyetlerdir (Morris, 1982).

Baskı maliyetlerinin belirleyenleri ve borçla ilişkisi sermaye yapısının belirlenmesinde önemlidir. Bununla birlikte, maliyetlerle ilgili kesin bir tanım vermek zordur. Baskı maliyetlerinin genel özellikleri aşağıda sunulmaktadır:

- Tasfiye sürecinin gecikmesiyle, stoklar ve sabit varlıkların bir bölümü ya da tamamı bozulabilir ve çürüyebilir. İflas davaları uzun yıllar sürebilir; bu dönem boyunca makina parkı kullanılamaz hale gelebilir, binalar yıpranabilir ve stoklar bozulabilir. Bunların hepsi firmanın değerinde erimeye neden olan maliyetlerdir.

- İflas sürecinde avukat ve mahkeme giderleri ve yönetim giderleri firma değerinin çok büyük bir bölümünü eritebilir.

- Firma iflas ettiğinde yöneticiler ve çalışanlar genellikle işlerini kaybeder. Bu bağlamda, firmanın varlığını kısa dönemde sürdürecektir, uzun dönemdeyse tehlikeye atacak girişimlerde bulunabilir. Örneğin, firma makinaların bakımını geciktirebilir, nakit sağlamak için değerli varlıkları piyasa değerinin altında elden çıkarabilir veya maliyetlerini düşürmek için ürün ya da hizmet kalitesini düşürür ki bu firmanın uzun dönem piyasa konumunu olumsuz yönde etkiler.

- Finansal baskı altındaki firmaların müşteri ve tedarikçileri potansiyel problemlerin farkındadır ve çoğunlukla firmayı daha da zor duruma sokacak önlemler alır ya da girişimlerde bulunurlar.

Finansal baskı maliyetleri, iddia sahiplerinin menfaatlerinin, firma faaliyet, yatırım ve finansman kararlarını olumsuz etkilediği ölçüde artacaktır. İddia sahipleri, piyasa değerini maksimize etme hedefini bir tarafa bırakıp kendi menfaatleri uğruna alacaklıların aleyhine davranışlarda bulunabilirler.

Bazı varlıkların değerleri finansal baskı sürecinde hemen hemen hiç düşmezken, diğerlerininki oldukça azalmaktadır. En büyük değer kayıpları maddi olmayan varlıklarda görülmektedir. Bu bağlamda, ampirik gözlemler, teknoloji firmaları, işgücü kalitesinin önemli olduğu hizmet firmaları ve hızlı büyüme gösteren firmalarda borçlanma oranlarının düşük olduğunu göstermektedir (Brealey ve Myers, 2000).

1.5.1. Finansal Baskı Maliyet Türleri

Finansal baskı maliyetleri, doğrudan ve dolaylı maliyetler olarak iki başlık altında incelenebilir. Doğrudan maliyetler, kanuni giderler ve vergi avantajının kaybı gibi maliyetlerden oluşmaktadır (Haugen, 1978).

Dolaylı maliyetler, finansal baskı olasılığından kaynaklanan faaliyet sorunlarına bağlı maliyetlerdir. Bu maliyetler firmanın arz ve talep fonksiyonlarındaki uzun vadeli kaymalardan doğmuştur ve bu kaymaların ölçülmesi bir yana, tanımlanması bile çok güçtür (Haugen ve Senbet, 1978).

Scherr (1988) baskı maliyetleriyle ilgili çalışmasında, literatürde ele alınan doğrudan ve dolaylı maliyetleri özetlemektedir. Doğrudan maliyetler şunlardır:

- **Nakit Giderler:** Avukatlara, hakemlere, yed-i emine ödenen ücretler, ekspertiz, mahkeme başvuru giderleri ve diğer mahkeme giderleri gibi üçüncü kişilere ödenen nakit giderlerdir.

- **Tasfiye Kayıpları:** Tasfiye durumunda firma varlıklarının satılması söz konusu olabilir. Finansal baskı altındaki bir firmanın varlıkları, satılması durumunda problemsiz bir firmanınkinden daha düşük gelir getirecektir. Tasfiye sürecinde yüklenilen ek maliyetler doğrudan ve birincil finansal baskı maliyetleridir.

Dolaylı maliyetler aşağıda sunulmaktadır:

- **Potansiyel Gelir Kayıpları:** Ön ödeme gerektiren ya da garantinin önemli olduğu satışlarda (örneğin, sermaye mallarının satışında), finansal baskı müşteriler açısından ürün değerinin azalması anlamına gelir. Müşteriler

bu değer kaybını dikkate aldıklarında ödemeye hazır oldukları fiyatı düşüreceklerdir. Fiyattaki bu ilk düşme yönetsel bir maliyettir. Baskı sürecinin daha geniş kitlelerce anlaşılması, firmanın satışlarında azalmaya ya da müşterilerin firma hizmet ya da mallarına olan fiyat tekliflerinde ek bir indirim yol açacaktır. Ek indirim talepleri baskıdan kaynaklanan maliyetlerdir.

- **Kaybedilen Potansiyel Kârlı Yatırım Fırsatları:** Finansal baskı altında firma yönetimleri, nakit varlıkları artırmak isteyecek, bu nedenle kârlı yatırım fırsatlarını kaçıracaklardır. Yönetimin dikkatini baskı sürecine yönlendirmesi bu fırsatların takip edilmesini güçleştirecek ve bu fırsatlar değerlendirilemeyecektir. Kaçırılan yatırım fırsatlarından dolayı uğranan değer kayıpları finansal baskı maliyetleridir.

- **Kayıp Vergi Avantajları:** Tasfiye halindeki bir firma, amortisman, yatırım muafiyeti ve benzeri nedenlerle önceden elde ettiği vergi avantajlarını kaybeder.

- **Maddi Olmayan Varlıklardaki Değer Kayıpları:** Finansal baskı altındaki firma, yetişmiş işgücü kaybı, tanıtım giderleri, peşin ödenmiş ürünlerden gelecekte elde edilecek gelirler ve araştırma geliştirme giderleri gibi masrafları geri döndüremediğinden bu gibi giderler finansal baskı maliyetleri olarak kabul edilmektedir.

- **Kaybedilen Yönetsel Zaman ve Enerji:** Firma yönetimi finansal baskıyı aşmaya çalışırken zaman ve enerjisinin çoğunu sorunu çözmeye harcadığından, daha verimli kullanılabilecekken kaybedilen zaman ve enerji finansal baskı maliyeti olarak ele alınabilir.

- **Tedarikçilerle Olan İlişkilerin Bozulması:** Ödeme gücü çeken bir firmanın tedarikçileri, satış koşullarını ağırlaştırabilirler. Firma finansal baskıyı atlattıysa bile, tedarikçilerin finansal baskının açığa çıkmasından sonraki yaklaşımları daha az işbirliği ve kolaylık içerebilir.

- **Hisse İhraç ve Pazarlaşma Maliyetleri:** Baskı altındaki halka açık bir firma, taze fon girişi sağlamak amacıyla yeni hisse ihracına başvurabilir. Bu durumda, hisse ihraç giderleri baskıdan kaynaklanan bir maliyet olarak görülebilir. Ayrıca baskı altındaki firmanın yeni hisse ihracı, piyasada olumsuz algılanabilir ve bu firmanın piyasa değerinde fazladan bir düşmeyle sonuçlanabilir. Kaynak yaratmak amacıyla başvurulabilecek diğer bir

seenek, bor szleşmesinin yenilenmesidir. Bor szleşmesinin yenilenmesinde, pazarlaşma ve borcun yeniden yapılandırılmasının ek bir maliyeti vardır.

- **İşgücü Maliyetlerinde Artış:** Baskı sürecindeki firmaların başvurduğu stratejilerden bir diğeri maliyetleri düşürebilmek amacıyla işgücünü azaltmaktır. Firma gereksiz gördüğü işgücünün işine son verirken, firmada kalanların yeni yapıya uygun işlere başlatılması söz konusu olacaktır. Bu durumda, başlangıçta işgücü verimi düşeceğinden üretim ve hizmet maliyeti artacaktır.

Maliyetlerin çeşitliliğinden dolayı bir ölçüde akışmadan söz edilebilir. Örneğın, vergi avantajlarındaki kayıplar, hem doğrudan hem de dolaylı maliyet özellikleri taşıyan maliyetlerdir.

Doğrudan ve dolaylı finansal baskı maliyetlerinin toplamı yüksek olabilir. Bu toplamın, özellikle tasfiye sürecinin uzun ve karmaşık olduğu büyük ölçekli firmalar için çok büyük olduğu düşünülmektedir (Brealey ve Myers, 2000).

1.6. Firma Başarısızlığı Tahmin Literatürü

Kredi derecelendirme modelleri ve bu bağlamda finansal başarısızlığı tahmin ihtiyacı, son dönemlerde gerçekleşen, özellikle büyük hacimli iflaslar ve Basel II Anlaşmasının getirdiğı kredi riskine bağılı sermaye yeterliliğı ölçütüyle birlikte önem kazanmaya başlamıştır (Altman, 2003).

Basel II bağlamında, birçok ülkede, bankalar ve finansal kurumlar kendi iç kredi derecelendirme modelini geliştirmiştir¹². Bu modellerle, her türden kredi varlığının temerrüde düşme olasılığı ya da kredi verenin temerrüt durumunda kaybını hesaplayan sistem ve süreçlerin geliştirilmesi amaçlanmaktadır. Basel II'nin yaklaşımı, bankacılık sisteminin ve ekonomi otoritelerinin, reel sektörün risk profillerinin yakından takibi ve finansal sistemin kredi varlıklarına duyarlılığını kontrol etmek amacını taşımaktadır (Altman, 2003).

¹² Benzeri süreç, birkaç yıldan bu yana ülkemizde de devam etmektedir.

1980'lerden önceki çalışmalarda, genellikle, ampirik yöntemler kullanılmış, bu çalışmalarda dikkatle seçilmiş finansal oranlar kullanılarak tahmin doğruluğunun artırılması hedeflenmiştir (Back ve diğerleri, 1996). Başarısızlık tahminine yönelik araştırmaların uzun zamandan beri devam ediyor olmasına rağmen, firmaların neden başarısız olduklarına dair ortak kabul gören kuramsal bir altyapı henüz tam olarak oluşturulamamıştır. Wilcox (1973), kuramsal çerçeve eksikliğinin, firmalarla ilgili yeterli bilgi elde edilememesine yol açtığını, dolayısıyla sağlıklı bir genelleme yapılamadığını öne sürmektedir. Bu bağlamda, değişkenlerin belirlenmesinde oranların kuramsaldan çok ampirik özellikleri öne çıkmıştır.

Başarısızlık tahmin modellerinde iki temel yaklaşım göze çarpmaktadır: İlk ve en çok kullanılan yaklaşım, başarısızlığı tahmin ve en düşük hatayla sınıflandırmada kullanılacak tahmin edicilerin (finansal oranlar, değişken, öznelilik, vb.) ampirik olarak aranmasıdır. Diğer yaklaşımlar, iyileştirilebilir tahmin doğruluğu sağlayacak istatistiksel modellerin bulunması üzerine yoğunlaşmıştır (Back ve diğerleri, 1996).

Çalışmaların bir kısmı finansal baskıyı muhasebe bilgisinin yararlılığını sorgulamak için kullanırken, diğer bir bölümü de en yüksek doğrulukla başarısızlık tahmini yapabilecek modelleri geliştirmeyi amaçlamıştır.

Başarısızlık tahmin literatürü çoğunlukla akademik çevreler tarafından oluşturulmuştur. Müşterilerinden kaynaklanan potansiyel risk pozisyonlarını azaltmak amacıyla, ticari bankalar, denetim firmaları, derecelendirme kuruluşları, yatırım firmaları gibi kuruluşlar bu araştırmaların yakın takipçisi olmuştur.

Bu bağlamda, geliştirilen modeller genel olarak şirket kazançlarının tahmini, iflasların tahminine yönelik oran ve ekonomik trend bazlı modeller olarak gruplandırılabilir. Bu modellerden şirket kazançlarının tahminine yönelik olanlar analiz metodu olarak istatistiksel zaman serisi ve diğer istatistiksel yöntemleri ve uzman görüşlerini kullanmaktadırlar (Kutman, 2001).

Reel sektör firmalarında başarısızlığın önceden tahmin edilmesi ihtiyacı ve başarısızlığın tahminine yönelik erken uyarı modellerinin tasarımı yolunda ilk çabalar, Amerika Birleşik Devletleri'nde yaşanan Büyük Bunalım sonrasında, 1930'lu yılların ortasında ortaya konmuştur. Bu alanda ilk çalışma Smith-Winakor tarafından 1935 yılında yapılmıştır. 1942'deki çalışmasında Merwin, başarısız firmaların finansal oranlarının başarılı firmalarinkinden farklı olduğunu bulmuştur (Sung ve diğerleri, 1999). 1960'lı yılların ortasına gelindiğinde, sonraki çalışmalara önemli bir temel oluşturacak olan Beaver'ın (1966) Tek Değişkenli Analizini görüyoruz. Beaver'a göre tek değişkenli analiz orta düzeyde bir tahmin gücüne sahiptir; bu analiz finansal baskının tahminine yönelik değişkenleri belirlemekte, fakat, ilintili riski ölçmeye yönelik bir ölçek sunamamaktadır. Neter (1967), başarısızlığın tahmininde çok değişkenli bir modellemenin daha doğru sonuçlar vereceğini savunmuştur (Sung ve diğerleri, 1999).

Finansal başarısızlığın modellenmesinde çığır açan ilk çalışma, Altman'ın (1968) beş değişkenli Çoklu Ayırım Analizidir. Bu çalışmada, modelin yüzde 90'ın üzerinde tahmin gücü olduğu bulunmuştur. Bununla birlikte, örneklemdaki firmaların ölçeklerinin oldukça küçük olması, bu çalışmanın genellenebilirliğine gölge düşürmektedir. Daha genellenebilir bir ayırım analizi modeli olan ZETA modeli, Altman ve diğerleri (1977) tarafından geliştirilmiştir. ZETA modeli değişik sektörlerden her ölçekteki firmaya uygulanabilmektedir. ZETA modeliyle birlikte, ayırım analizi birçok çalışmaya esin kaynağı olmuş, kısıtlayıcı yönlerine yönelik eleştirilere rağmen, yapılan çalışmalarda karşılaştırma amacıyla referans olarak kullanılmıştır.

Blum (1974) değişken seçimini firmanın varlıklarının finansal kaynaklar havuzu olduğu kavramına dayandırmakta ve başarısızlık riskini beklenen nakit akımları cinsinden ifade etmektedir. Deakin (1972), Beaver'ın (1966) kullandığı 14 oran arasında, model için en yüksek tahmin gücünü sağlayacak doğrusal bileşimi bulmaya çalışmıştır. Libby (1975), Deakin (1972) çalışmasındaki değişkenlerin bir kısmını kullanarak muhasebe oranlarının kredi verme sürecinde ne kadar yararlı olabileceğini belirlemeye çalışmıştır. Libby bu çalışmasında Temel Bileşen Analizini kullanarak 14

değişken arasında beş değişim kaynağı belirlemiştir. Deakin (1977), Libby'nin (1975) faktör analizi katkısını kullanarak 1972 çalışmasını genişletmiştir. Bu genişletmeyle Deakin, başarılı firmaların ne sıklıkta başarısız olarak sınıflandırıldığını ve bu sınıflandırmanın yapısını bulmaya çalışmış, denetçi görüşlerini model sonuçlarıyla karşılaştırarak tahmin gücünü belirlemeyi amaçlamıştır.

Edminster (1972) diğer bir ayırım analizi çalışmasında, oran analizinin tahmin gücünün artırılmasının analiz yöntemine ve değişken olarak kullanılan finansal oranlara bağlı olduğunu belirtmektedir. Wilcox (1971) çalışmasında Kumarbaz Yıkım Modeli'ni kullanarak, Beaver'ın (1966) çalışmasının sonuçlarını daha iyi açıklayacak bir kuramsal model geliştirmeyi amaçlamış ve başarısızlığın tahminine yönelik daha iyi tahmin ediciler bulmaya çalışmıştır. Wilcox'a göre birçok çalışmada yeterli düzeyde kavramsal bir çerçeve çizilmemiştir. Scott (1981), kendi kuramsal çerçevesi içinde, ampirik ayırım modellerini gözlemlenmiş tahmin gücü ve kuramsal uygunluk açısından karşılaştırmış ve ZETA modelinin en uygun çok boyutlu model olduğu sonucuna ulaşmıştır. Scott'a göre ZETA modeli yüksek ayırma gücüne sahiptir, basittir ve hem muhasebe ve piyasa, hem de kazanç ve borçluluk değişkenlerini içermektedir.

Literatürde, ayırım analizi dışında yöntemler kullanılarak geliştirilen modeller de bulunmaktadır. Meyer ve Pifer'in (1970) Regresyon Analizi, Ohlson'un (1980) Lojistik Regresyon Modeli (Logit) ve Zmijewski'nin (1984) Probit Analizi Modeli firmanın başarısızlık olasılığını hesaplamaya ve bazı özgül değişkenlerin başarısızlık olasılıkları üzerindeki doğrudan etkilerini ölçmeye çalışmıştır. Ayrıca, Martin'in (1977) Logit Regresyonu, Korobov ve Stuhr'un (1985) Arktanjan Regresyonu ve West'in (1985) Faktör-Lojistik Analizi gibi istatistiksel yöntemler literatürdeki diğer modeller arasında sayılabilir.

Her ne kadar açıklayıcı değişkenler ve sınıflandırma arasında bir bağ kurabiliyorsa da, ayırım analizi bir takım istatistiksel problemlerden muzdariptir (Altman, 1993). Bu problemlere çözüm getirebilecek iki yöntem, Karar Ağacı ve Yinelemeli Bölümleme olarak öne sürülmüştür. Karar ağacı

analizi ayrık alt-kümeler arasındaki entropiyi¹³ maksimize etmeyi amaçlarken, yinelemeli bölümlleme hatalı sınıflandırmanın beklenen maliyetini minimum düzeye indirmeye çalışır (Quinlan 1986, 1993). Frydman, Altman ve Kao (1985) başarısızlık sınıflandırmasında yinelemeli bölümlleme yöntemiyle bir ayırım ağacı oluşturmuştur. Messier ve Hansen (1988) karar ağacı tekniğini geliştirerek benzer bir başarısızlık tahmin modeli tasarlamıştır.

1980'lerin başlarında Sinir Ağı Sistemleri yapay zekanın başat araştırma yöntemlerinden biri olmuştur ve araştırmacılar başarısızlık tahmin modelleri de dahil olmak üzere bu yöntemleri sınıflandırma problemlerine uygulamışlardır. İflas tahmin modellerinde kullanılan sinir ağı çalışmalarının büyük çoğunluğunun performansı ayırım analizi, logit analizi, genetik algoritma, karar ağacı ve diğer yöntemlerin performanslarıyla karşılaştırılmaktadır. Bir grup çalışma, sinir ağlarının performansının diğer tekniklerin performanslarından biraz daha iyi olduğunu bulmuştur. Buna karşın, genellikle, yapay sinir ağı modellerinin sonuçları çelişkilidir ya da hiç sonuç bulunamamıştır.

Diğer bir grup çalışmada Genetik Algoritma, sınıflandırma problemini çözmekte kullanılmıştır. Genetik algoritmalar, geniş ve karmaşık yüzeyleri tarayabilen stokastik tekniklerdir. Genetik ve evrim kuramına dayalı olarak çalışan genetik algoritmalar, seleksiyon, çaprazlama ve mutasyon işlemlerini de kullanarak bir yapay yapılar kitlesini sürekli yenileyerek değiştirir. Back, Laitinen ve Sere (1996), başarısızlık tahmininde genetik algoritmaları kullanmış ve sonuçları sinir ağıninkilerle karşılaştırmış ve oldukça ümit verici sonuçlar bulunmuştur. Günümüzde genetik algoritma ve sinir ağının birlikte kullanıldığı modellere literatürde yoğun olarak başvurulmaktadır.

Kim, Chung ve Paradise (1997) Uzman Karar Verme Sistemine, modeller ve veri seti arasındaki iletişimi de ekleyerek yeni bir boyut getirmiştir. Bu çalışmada, daha güvenilir uzman sistemlerinin veya karar modellerinin geliştirilmesinin, değişik algoritmalar kullanılarak oluşturulan

¹³ Entropiyle bilgi değeri kastedilmektedir. Bilgi değeri, herhangi bir kararın alınmasıyla elde edilen bilginin ölçülen miktarı olarak tanımlanmaktadır.

stratejilerin ne kadar iyi modellendiğinin anlaşılabilmesiyle olanaklı olduğu öne sürülmektedir.

Sung, Chang ve Lee (1999) çalışmasında, çoklu ayırım analizi ve karar ağacı yöntemlerinin tahmin gücü karşılaştırılmış, her iki modelde de normal ekonomik şartlar ve ekonomik kriz dönemleri için yapılan modellerde kullanılan değişkenler arasında istatistiksel anlamlılık yönünden fark bulunmuştur. Dolayısıyla, normal ekonomik şartlar için tasarlanan bir model, kriz yaşanan bir ekonomideki firmalar için kullanılırsa doğru sonuçlar vermeyecek ya da böyle bir modelin tahmin gücü (doğruluğu) sorgulanabilecektir. Bunun yanında, bu çalışma karar ağacının firma başarısızlığını modellemede en yüksek yorumlama gücüne sahip yöntem olduğu sonucuna ulaşmıştır.

1.7. Finansal Başarısızlığı Modellemedeki Güçlükler

Finansal modellerin tasarımında birçok sorunla karşılaşılmaktadır. Bu sorunlar şu şekilde özetlenebilir:

- **Teknik Varsayımların İhlali:** Finansal oranlar genellikle normal değil, yatık dağılım gösterirler; verideki stokastik süreçleri teşhis etmek zordur; oran dağılımlarının normal olmaması ya da lognormal olmaları firma büyüdükçe oran bileşenleri üzerinde çarpan etkisi göstermektedir. Geleneksel yöntemler hata terimleriyle ilgili kısıtlar getirmektedir. Standart parametrik yöntemler genellikle kitleyle ilgili dağılımsal varsayımlar yapmaktadır. Geçmiş çalışmaların çoğunda finansal oran değişkenlerinin standart dağılım varsayımlarına uymadıkları bulunmuştur.

- **Çoklu Bağıntı:** Eğer bağımsız değişkenler arasında çoklu bağıntı varsa regresyon sonuçları yanlış olacaktır.

- **Modellemenin Gereksiz Tekrarı:** Modelleme sürecinin gerçekliğin karmaşık bir tanımlamasına dönüştüğü durumda çözüm, artık karar vermede yardımcı olmaktan uzaktır.

- **Davranışsal Değişkenler:** Davranışsal değişkenlere bağlı olarak, problemlerin kesin olarak tanımlanamaması, kesin bilimsel kuralların ve belirli

ve birçok durumda tatmin edici sonuçların bulunamaması, bazı oranların tam olarak ölçülememesi gibi sorunlarla karşılaşmaktadır.

- **Karar Vericilerin Rasyonel Olmamaları:** Modellemedeki genel varsayım karar vericilerin her zaman rasyonel davrandığıdır. Bununla birlikte, yönetim sürecinde alınan kararlar rastgele etkenler, yöneticilerin duygusal davranışları ve optimal çözümün ne olduğunun bilinmemesinden önemli derecede etkilenir.

- **Optimizasyonun Maliyeti:** Kararların kaotik bir çevrede alınması, bazı durumlarda zaman/maliyet analizine izin vermez.

- **Değişkenlerin Tutarsızlığı:** Yaratıcı muhasebe teknikleri ve bazen firmanın geçmişiyle ilintili olarak bazı değişkenlerin tutarsız olması olanaklıdır. Muhasebe bilgisine dayalı oran dönüşümleri belirsiz ilişkileri tanımlamakta yetersizdir. Bu varsayımlarla kısıtlanmayan parametrik olmayan yöntemlerse genellikle soruna özgüdür ve yorumlanması zor olabilmektedir (Jaggi ve Lee, 2002).

- **Fonksiyonel Biçim:** Geleneksel modellerle ilgili önemli bir problem, model için uygun bir fonksiyonel yapının bilinmemesidir. Belki de uzmanların üzerinde fikir birliğine vardıkları tek ortak nokta fonksiyonel yapının karmaşık olması¹⁴ gerektiğidir. Bununla birlikte, olabildiğince basit ve anlaşılır bir fonksiyonel yapının elde edilmesi de ideal modellemenin en önemli ön koşullarından biridir. Karmaşık modellerin geliştirilmesi, halihazırdaki tahmin tekniklerine bağlı olarak kısıtlanmaktadır. Genellikle, doğru fonksiyonel ilişkinin araştırılmasından çok tahmin tekniğinin kolaylığı modelin seçilmesinde belirleyici ölçüt olmuştur. Karmaşık problemlerin modellemesinde kullanılan bir diğer yöntem olan uzman sistemleri bazı problemlere başarıyla uygulanmışsa da bazı eksik yönleri bu sistemlerin kullanımını zorlaştırmaktadır. Bunlar, programlamanın ve sistemin işlerliğini sürdürebilmenin zorluğu (Feigenbaum kısıtı), uzman sisteminin tümevarımsal öğrenme ve tanım kümesindeki değişikliklere uyum sağlamanın zorluğu¹⁵ (öğrenme problemi) ve insan uzmanların uzman sistemin bilgi/sorgulama

¹⁴ Gerçek modelin ne olduğunu bilmek neredeyse imkansızdır, ama, finansal baskı ya da iflas kavramını yakından bilen uzmanlar bu ilişkinin karmaşık olduğunu söylemektedir. Bu bağlamda, kullanılan tahmin ediciler tek başına değil birarada tahmin değerini belirlemektedir.

¹⁵ Yeni bilgilerin sisteme öğrenme amacıyla sunulması ve duruma bağlı olarak sistemin yeniden programlanması ve eğitilmesi gerekmektedir. Bu eğitimin ne aralıkla yapılacağı ne kadar veri kullanılması gerektiği vb. birçok soru vardır.

tabanını oluřturması ve bunları EĞER-İSE kurallarına d n řt r lmesinin zorluęu (bilgi m hendislięi kısıtı) olarak sayılabilir (Dorsey ve dięerleri, 1993).

Finansal bařarısızlık modellerinin oluřturulmasında da benzer sorunlarla karřılařılmaktadır. Bu sorunlar ařaęıda sunulmaktadır:

- **Finansal Bařarısızlık Sınıf Deęiřkeninin Aldıęı Deęerler:**

 nceki  alıřmalardaki kısıtlardan biri iki deęerli bařarısızlık sınıf deęiřkenidir. İki deęerli finansal baskı sınıflandırması bu s recin ařırı basite indirgenmiř ifadesidir. Bu t rden bir sınıflandırmanın finansal bařarısızlık s reci altında yatan ger ek yapıyı tanımlaması g  t r. Finansal bařarısızlık, genellikle, g zlemlenemeyen bir s re tir. Firmalar sadece iflas eden ya da etmeyen olarak sınıflandırılmaz; veri anda, firma finansal bařarısızlık s recinin herhangi bir ařamasında olabilir. Bununla birlikte, firmanın bu s recin hangi ařamasında olduęu g zlemlenemedięinden bu řekilde bir basitleřtirilmeye gidilmiřtir. Finans literat r  finansal bařarısızlık s recinde farklı olayların farklı d zeyde finansal bařarısızlıęa iřaret ettięini ve firmaların bu olayların bir oęunu yařadıktan sonra iflas ettiklerini kabul etmektedir.

- ** rneklem Yapısı:** Beaver (1966), Deakin (1972) ve Blum (1974)

 alıřmaları  rneklemelerin heterojen olması dolayısıyla eleřtirilmektedir.  rneklemin heterojenlięi, firmaların hukuksal durumları ve sekt r gibi  l  tlere g re se ilmemiř olması durumunda ortaya  ıkar. İgili sınıflandırma kullanılan bir  alıřmada,  rneklem sekt r,  l ek ve hukuksal durum  l  tleri a ısından homojen olmalı ve k tle daęılımını temsil etmelidir (Ward, 1999). Eęer heterojen bir  rneklem kullanılıyorsa,  zellikle, bu     l  t  karřılayan deęiřken deęerlerinin de veri setine dahil edilmesi bu sorunun   z lmesine yardımcı olacaktır.

- **İstatistiksel Varsayımların İhlali:** Finansal bařarısızlık

 alıřmalarının  oęunluęunda, istatistiksel y ntemlerin  nemli varsayımları ihlal edilmiřtir.  oklu ayırım analizi, tahmin edicilerin rassal olarak se ildiklerini ve normal daęıldıklarını varsaymaktadır. Ayrıca doęrusal  oklu ayırım analizi grupların varyans-kovaryans matrislerinin birbirlerine eřit olduklarını varsayar (Ward, 1999). Ger ekte, finansal oran daęılımları bu varsayımları ihlal etmektedir.

- **Örneklem Büyüklüğü:** Noreen (1988), benzetimle elde edilen verileri kullanarak probit ve doğrusal regresyonun başarısızlık tahminindeki performansını ölçmüştür. 100 örnekle elde probit sonuçları 200 örnekle tam tersine dönmüştür. Örneklem büyüklüğüne bağlı olarak sonuçların değişkenlik göstermesi, tutarsızlık göstergesidir.

- **Örneklem Yanlılığı:** Başarılı ve başarısız firmaların rassal olarak seçilmemesi, katmanlama yanlılığı yaratacaktır (Ward, 1999). Örneklem büyüklüğünün yetersiz olduğu çalışmalarda başarısız firmalarla başarılı firmalar genellikle birebir eşleştirilmiştir. Dolayısıyla, kullanılan örneklem gerçekteki kütle dağılımını temsil etmekten uzaktır (Ward, 1999; Keasey ve Watson, 1991; Platt ve Platt, 2000).

- **Zaman Boyutunda Örneklem ve Sınama Örnekleminin Seçimi:** Başarısızlık çalışmalarında karşılaşılan diğer bir kısıt firmaların geniş bir zaman dilimi içinde, farklı dönemlerden, dönem gözetmeksizin seçilmesidir. Olasılıkla, örneklemin seçilmesinde kullanılan farklı dönemlerin ekonomik koşulları birbirinden farklıdır. Bu bağlamda, farklı dönemlerden örneklerin kullanılması örneklem içindeki farklılıkların artmasına ve tahmin doğruluğunun azalmasına ya da yanlış genellemeler yapılmasına yol açacaktır. Ayrıca, modellerde sınama örneklemelerinin oluşturulmasında da eğitim örnekleminin seçildiği dönem kullanılmıştır. Bu, dönemler arasında durağanlığın olduğu varsayıldığında kabul edilebilir bir durumdur. Aksi durumda, sınama örneklemini farklı bir dönemden seçilmelidir. Bu sadece çapraz doğrulama değil dönemler arası doğrulamanın yapılması gerekliliğine de işaret etmektedir (Ward, 1999; Keasey ve Watson, 1991).

- **Verisi Kullanılan Dönemler Arasındaki Farklar:** Tahmin modellerinin oluşturulmasında kullanılan verilerin seçildiği dönem önem taşımaktadır. Zaman içinde geliştirilen muhasebe uygulamaları ve mevzuat değişiklikleri, nakit akımları ve bilanço değerlerinin tanımlarının zaman içinde farklılaşmasına yol açmıştır.

1.8. Finansal Tabloların Bilgi İçeriği ve Finansal Başarısızlık Modellerindeki Değişkenler

1.8.1. Finansal Tablolardaki Bilgi

Finansal tablolar firmayla ilgili finansal bilgileri filtrelenmiş olarak sunmaktadır. Finansal kararlar, gelir ve giderlerin durumuna, bu gelir ve giderlerin zamanlamasına, varlık değerlemesine ve amortisman yöntemlerine bağlı olarak alınır. Firmalar çoğu zaman yanıltıcı finansal tablolar hazırlamak için yeterli nedenleri bulacaktır: gelir vergisi, temettüler vb. finansal raporlardaki rakamlara göre belirlenir. Örneğin, firma yatırımcıları teşvik edebilmek ya da finansal baskı yaşadığı durumlarda zaman kazanabilmek için kâr gösterebilir.

Finansal oranlar kullanılarak firmanın devamlılığı hakkında niteliksel yorumlar yapılmaktadır. Fakat firmaların hazırladıkları oranların sektörel bazda ya da tüm firmalar için genellenebilirliği tartışmalıdır. Dolayısıyla finansal oranlar firmanın ve faaliyet gösterdiği sektörün başka özellikleriyle bir arada değerlendirilmelidir; iş çevrimleri arasındaki farklılıklar, sermaye getirisi derecesi, rekabet düzeyi, gelir ve giderlerin değişkenliği, sektörün iş çevrimlerine bağımlılığı vb. bu türden önemli etkenlerdir. Finansal oranlar tek başlarına ele alındığında, firmanın çözümlmeye konu olan özelliklerini gerçekten irdelemek yerine daha çok bu çözümlmede sorunlu alanlara işaret eden ölçütlerdir. Morris'in (1997) belirttiği gibi, veri bir ekonomik olay farklı oran örüntüleri yaratabileceği gibi, veri bir oran örüntüsü birbirinden farklı ekonomik olayların sonucunda ortaya çıkmış olabilir. Bu bağlamda, firmanın devamlılığıyla ilgili yorumların birden çok oranın birbirleriyle ilişkisi ele alınarak yapılması gerekmektedir (Bernhardsen, 2001).

Finansal tablolardan elde edilen bilgiler dışsal kaynaklardan elde edilenlerle birleştirildiğinde firmanın durumu hakkında genel bir hüküm vermek olanaklı olabilmektedir. İstatistiksel başarısızlık tahmin modelleri finansal oranların, başarısızlığın tahmin edilmesinde referans olarak kabul edilebileceğini varsaymaktadır. Bu türden basitleştirmeler başarısızlık tahmin

modellerinin ne düzeyde başarılı olduklarını sorgulanabilir hale getirmektedir. Yine de istatistiksel modeller pratikte faydalı olarak kabul edilmektedir.

Finansal oranlar faaliyetlerden kaynaklanan boyut etkisini ortadan kaldırmak üzere düzey değerlerinin birbirine bölünmesiyle elde edilen yapay ölçütlerdir. Bu ölçütler ölçüme konu olan özelliği kabaca ölçerler ve bazı oranlar bazı firmalar için anlamsızdır. Bu bağlamda, anlamsız oranların aldıkları değerler de genellikle anlamsız olabilmektedir. Büyük örneklerle yapılan çalışmalar oranların aldıkları uç değerleri dikkate almamaktadır. Veri oranının istatistiksel anlamlılığı kesimin nereden yapıldığına bağlıdır ve dolayısıyla modele hangi oranların dahil edileceğini belirlemek zorlaşmaktadır. Dahası, herhangi bir değişkene uygulanacak optimal kesim noktası diğer değişkenlere uygulanan kesim noktalarına da bağlıdır (Bernhardsen, 2001).

Bazı modellerde gayri safi yurtiçi hasıla büyüme hızı gibi bazı dışsal değişkenler modele dahil edilmiştir. Bu etmen tüm firmalar için aynı olduğundan, dışsal değişkenlerin modele dahil edilmesinde biraz yaratıcı olmak yararlı olacaktır. Olası bir seçenek, her yıl için ortalama temerrüt sıklığının modele dahil edilerek, yüksek ve düşük risk taşıyan faaliyet dönemlerinin belirlenmesi ve başarısızlık sınıflandırma modellerindeki açıklama gücünün gözlemlenmesidir. Burada dikkat edilmesi gereken bir nokta, bu tür karma modellerin yüksek tahmin gücü elde edememiş olmasıdır. Diğer bir seçenekse, modelin kesim noktalarının belirlenmesinde öncül olasılıklarının ya da sınıf üyeliklerinin ve hatalı sınıflandırma maliyetlerinin de dikkate alınmasıdır (Altman, 2003).

Bir firmanın varlığını sürdürebilmesi, finansal açıdan ne kadar sağlıklı olduğuna, nakit akımı üretebilme kapasitesine, sermaye piyasalarında borçlanabilmesine, finansal kapasitesine ve nakit şoklarına karşı dayanabilme gücüne bağlıdır.

Firmanın finansal açıdan sağlıklı olup olmadığı çeşitli ölçütlerle belirlenebilir. Bir işletmenin finansal gücünü etkileyen etmenler:

- Sermaye yapısı ve yeterliliği,
- Faaliyet nakit akımları ve maliyet yapısı,
- Kazanç yeterliliği, yatırımın getirisi ve kârlılık,
- Likidite,
- Varlık dönüşümü,
- Stratejik pozisyon ve
- Piyasa payları

olarak sıralanabilir. Bu etmenlerin ölçülmesinde kullanılan finansal oranlar ve ortak büyüklük oranları, analizde kullanılan özneliliklerdir. Öznelilikler ve kısaca açıklamaları Ek 2’de verilmiştir.

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ VE MAKİNA ÖĞRENMESİ YÖNTEMLERİ

2.1. Veri Madenciliği ve Makina Öğrenmesi Kuramı

Bilgi çağında, her türlü karar biriminin en önemli girdisi bilgidir. Ekonomik birim olarak firma açısından rekabet avantajı sayılabilecek bilgiyi incelemek, modellemek ve karar destek sistemlerinde kullanabilmek için güçlü araçlara ihtiyaç vardır. Kuramsal tabana oturtulmuş bilgi destek sistemleri, bilişim teknolojilerinde, özellikle son yıllardaki hem yazılım hem de donanım açısından gelişmelere paralel olarak, büyük miktardaki verileri işleme ve çıktı olarak elde edilen bilgiyi karar destek sistemlerinde kullanabilme olanağına kavuşmuştur.

Bilişim teknolojilerinin analiz alanında, son yıllardaki en önemli gelişmelerden biri veri madenciliğidir. Veri madenciliği, büyük hacimlerdeki verilerin işlenmesi yoluyla bilgiye ulaşılması ve bu verilerden değerli ve anlamlı örüntülerin araştırılması olarak tanımlanabilir (Langley, 1996; Indurkha ve Weiss, 1998; Frank ve Witten, 2000). Bu bağlamda, veri madenciliği, karar destek sistemlerinin ihtiyacı olan bilgiyi üretmeye aday bir sistem olabilir.

İnsan ve makina uzman, veri madenciliğinin birbirini tamamlayan en önemli bileşenleridir. İnsan uzman, veri tabanının hazırlanması, problemin tanımlanması ve hedeflerin belirlenmesinde aktif olan etkidir. Makina uzman ise verileri, hedeflerle uyuşan örüntüleri saptamak amacıyla taramak ve bu örüntülere uygun kuralları öğrenmekten sorumludur. Veri madenciliğinde veri elektronik ortamda saklanmakta, verideki örüntüler otomatik olarak aranmakta, doğrulanmakta ve tahmin veya sınıflandırma amacıyla kullanılmak üzere kurallar öğrenilmektedir¹⁶. İnsan uzmanın bilgi

¹⁶ Kural öğrenme her algoritma için geçerli değildir.

çıkarma sürecindeki diğer bir sorumluluğuysa, makina uzmanının öğrendiği bilginin incelenmesi, sorgulanması, sınanması ve tutarsız ve anlamsız bilgilerin ayıklanması, sorgulama ve sinama sonucunda veri tabanının düzenlenmesi, hatalı bilgi ve verilerin düzeltilmesidir. Bununla birlikte, veri madenciliğinin asıl amacı, insan uzmanı, makina uzmanının bilgi işleme yeteneği ve kapasitesine dayanarak bu sürece olabildiğince az dahil etmek ve değerli bir kaynak olarak zamanın daha önemli alanlarda kullanılabilmesini sağlamaktır (Indurkha ve Weiss, 1998).

Makina öğrenmesi, yapay zekanın veri madenciliği ve istatistikle ilintili bir dalıdır. Veri madenciliğinde olduğu gibi, makina öğrenmesinde de, analize konu olan alana ait veriler, örnekler, daha açık bir ifadeyle, örneklere ait özellikleri tanımlayan öznitelik değerleri olarak sisteme iletilir ve makina uzman ya da programın bu verilere ait örüntüyü ya da yapıyı öğrenmesi istenir. Öğrenme sürecinin en önemli özelliği, veri setine eklenen her yeni bilgi için, makina uzmanının önceki veri yapılarını öğrenmiş olarak, örüntüdeki değişmelerin ayırımına varması ve bu değişimlerin getirdiği bilginin kural setine eklemesinin varsayılmış olmasıdır. Böylece, makina uzmanının yeni durumlarla karşılaştığında, en azından potansiyel performansının artırılması amaçlanmaktadır. Makina öğrenmesi yöntemleri, büyük veri setlerinde rahatlıkla kullanılabilen ve daha çok bilgiyi kolaylıkla işlemeye yatkın güçlü yöntemlerdir. Bu buluşsal tahmin yöntemleri, istatistiksel yöntemlerden farklı olarak, sıklıkla güçlü ampirik sonuçlar üretebilmektedir (Frank ve Witten, 2000).

Öğrenme yöntemlerinin çoğunda, “öğrenilen problemin” yapısal tanımları araştırılır (İkizler, 2002). Bu tanımlar, genellikle oldukça karmaşık ve karar ağaçlarının öğrendiğine benzer kural setleri biçiminde ifade edilmektedir. Diğer taraftan, örneğin yapay sinir ağları gibi yöntemler, böyle tanımlar yapmazlar¹⁷. Makina öğrenmesi uygulamalarında tanımlar,

¹⁷ Sinir ağları, öğrenilen bilgiyle ilgili açık yapısal tanımlamalar içermeksizin örnekleri sınıflandırır. Yapay sinir ağları, diğer öğrenme algoritmalarıyla karşılaştırıldığında, zaman zaman daha yüksek performans göstermekle birlikte, öğrenme sırasında özniteliklere atanan ağırlıkların kuramsal olarak yorumlanabilmesi olanaksızdır. Bu bağlamda, bu yöntemle öğrenilen kuralların verideki örüntüyü “ezberlediği” ya da özelleştiği ve sinama amacıyla kullanılan örneklem dışı örneklerin sınıflandırılmasında güvenilmez olduğu söylenmektedir. Özelleşme ya da aşırı öğrenme problemi, kuralın veri setinin üzerine tam oturması, yapışması ya da ezberlemesi olarak tanımlanabilir. Özelleşmiş kural, eğitim seti örneklerini yüksek bir doğrulukla sınıflandırırken, sinama setlerinde düşük bir performans sergileyecektir.

genellikle algoritmanın yeni örnekleri sınıflandırma performansından daha önemlidir; tanımların kuramsal anlamlılık koşulu sağlandığında, yüksek sınıflandırma performansı algoritmanın tercih edilmesinde ikinci düzeyde aranacak ölçüttür¹⁸ (Frank ve Witten, 2000).

Veri madenciliği ve makina öğrenmesi yöntemlerinin temel özellikleri şöyle özetlenebilir:

- Her iki yöntem de kuramsal olmaktan çok ampirik öğrenme/sınıflandırma¹⁹ problemleriyle ilgilenir.
- Verideki yapısal örüntüleri bulmak ve tanımlamak, bu veriyi açıklayacak aracı, yani algoritmayı ve ilintili kural kümesini (kümelerini) bulmak ve bu bilgiyi tahmin veya sınıflandırma amacıyla kullanmak makina öğrenmesinin asıl amacıdır.
- Veri seti örneklerden oluşmaktadır. Her iki yöntemin çıktısı, yeni örnekler üzerinde yapılan tahmin/sınıflandırma ya da öğrenilen kurallardır²⁰. Elde edilen bilginin, tanım alanı bağlamında kuramsallaştırılması önemli bir hedeftir.
- Makina öğrenmesi ve veri madenciliği yöntemlerinde, veriden bilgi çıkarma yarım ya da tam otomatik süreçlerdir.
- Karar destek sistemleri açısından, her iki yöntemde de, makina uzmanının bilgi işleme kapasitesine dayanılarak, insan uzmanının bu alandaki eksiklikleri giderilmek ve rekabet avantajı sağlamak amaçlanmaktadır.

Gerçek hayat problemleri, özellikle finansal riskin değerlendirilmesi bu türden bir altyapı gerektirir. Veri kendi başına bir rekabet avantajı sağlamazken, bu veriden çıkarılan bilgi bu avantajı sağlayabilir. Finansal risk analizinde, özellikle iflas tahmininde, veri çıkarma ve model inşası oldukça zordur. Finansal risk analiz problemleri genellikle karmaşık olarak

¹⁸ Elbette bu ölçüt tanım alanına, kullanıcıların tercihlerine ve öğrenme algoritmalarının özelliklerine bağlıdır. Örneğin, logit ve sinir ağları, kuramsal anlamlılığın aranmadığı sık frekans dağılımlı verilerin kullanıldığı tanım uzaylarında hızlı analiz yapabilmeyi olanaklı kıldığından kullanıcı tarafından tercih edilebilir.

¹⁹ Makina öğrenmesi ve veri madenciliğinde, sınıflandırma ve öğrenme (ya da kontrollü öğrenme) birbirinin yerine kullanılmaktadır. Metin içinde, her iki terim de aynı kavrama işaret etmek üzere kullanılmıştır.

²⁰ Bir örnek belirli sayıda özniteliğin aldığı değer ve sınıf etiketinin değeriyle tanımlanır.

nitelendirilebilir; bu modellerde niteliksel ve niceliksel bir çok etmen ele alınmalıdır²¹.

2.1.1. Girdi, Kavram, Örnek ve Öznitelikler

Makina öğrenmesinin girdisi sonlu sayıdaki örneklerin sonlu sayıda özniteliklerinin değerleri biçimindedir. Öğrenilecek şey “kavram tanımı” olarak adlandırılır. Ampirik öğrenme sürecinin çıktısı olarak kavram, anlaşılır, tartışılabilir ve gerçek durumlara uygulanabilir olmalıdır.

“Öğrenici” ya da makina uzmana sunulan bilgi, örnekler kümesi şeklindedir: her örneğin değişik özellikleri, özgül özneliliğin ölçek değerleriyle nitelendirilmektedir. Tanım alanına ait kuramsal tanımlama yapıldıktan sonra, bu alana ait örnekler geniş bir kitle içinden seçilir²². Örneklere ait özniteliklerin değerleri ölçülür ve sınıflandırma algoritmasına sunulur, yani iletilir.

Sınıflandırma öğrenmesi ya da sadece sınıflandırma algoritmaları, kendisinden beklenen şekilde etiketlenmemiş yeni örnekleri sınıflandırmak üzere, daha önce sınıf etiketi verilmiş örneklerden oluşan veri setindeki örüntüleri öğrenmeye çalışır. Sınıflandırma öğrenmesi, kontrollü (denetlenen) öğrenme olarak da adlandırılmaktadır. Bu yöntemde öğrenme, eğitim örneklerinin değerleri öğrenme amacıyla algoritmaya sunulduktan sonra kontrol altında çalışır (İkizler, 2002).

Sınıflandırmanın başarısı, elde edilen kavram tanımının sınıf etiketleri bilinen fakat öğrenme algoritmasına sunulmamış bağımsız bir veri seti üzerinde denenmesiyle sorgulanmaktadır²³. Sınama verisiyle elde edilen başarının oranı, kavramın ne kadar iyi öğrenildiğini gösteren nesnel bir ölçüttür.

²¹ Niteliksel etmenlerin kullanımı karmaşık modelleme yeteneğinin yanında sağlam bir kuramsal altyapının da kurulmasını gerektirir. Bu aşamada, insan uzmanın modelleme ve kuramsal altyapısı önemli bir öge olarak öne çıkmaktadır.

²² İdeal örneklem kitlenin kendisi olsa da, ampirik analizlerde kullanılmak üzere seçilen makina öğrenmesi algoritmalarının fiziksel kısıtlarına bağlı olarak veri seti, kitleyi temsil edecek büyüklükte bir boyuta indirgenmelidir. Çalışmamızda kullanılan veri seti tüm kitlenin yüzde 80'ini temsil etmektedir.

²³ Bu yaklaşıma, eğitim ve sınama setleri oluşturabilecek büyüklükte örneklemelerin bulunması durumunda başvurulmaktadır. Örneklem küçüğü olması durumunda, çapraz doğrulama yöntemiyle sınama tercih edilir.

Makina öğrenmesinin girdisi, birden fazla örnekten oluşan bir veri kümesidir. Her örnek öğrenilecek kavramın bağımsız bir temsilcisidir ve her biri bilinen öznitelik değerleriyle tanımlanmaktadır.

Makina öğrenmesi algoritmalarında kullanılan her bağımsız bireysel örnek, özniteliklerin aldığı değerlerle tanımlanır. Bu değerler, özniteliğe karşılık gelen büyüklüğün bireysel örnekler için ölçülmesiyle elde edilir. Öznitelikler, farklı yapı ve özelliklerde değerler alırlar (Frank ve Witten, 2000). Bunlar:

- Sürekli öznitelikler gerçel ya da tamsayı değerlerden herhangi birini almaktadır.
- Kategorik sınıflandırmayı olanaklı kılan nominal öznitelikler, önceden belirlenmiş sonlu sayıda olası değerden birini almaktadırlar²⁴.
- Ordinal büyüklükler kategorileri sıramalayı olanaklı kılan büyüklüklerdir. Bu büyüklükler için mesafe kavramı, yani farklı örnekler için özniteliklerin aldıkları değerlerin birbirine olan uzaklığı, söz konusu değildir. Tutarlılığın sağlanması koşuluyla, sıralamanın yönünü belirleyen bir kural yoktur.
- Aralık büyüklükler ya da uzanım aralıkları sınıanabilir, sabit ve eşit birimlerle ölçülebilen büyüklüklerdir.
- Oran büyüklükler, kendisi için bir sıfır noktası²⁵ önceden belirlenmiş büyüklüklerdir. Oran büyüklükler gerçel sayılarla ifade edilir ve bu büyüklüklerle matematiksel işlemler yapmak olanaklıdır.

2.1.2. Veri Seti

Bilişim sistemlerindeki gelişmeler, büyük miktarda verinin istenilen formatta saklanması ve işlenmesi konusunda ilerlemeyi de beraberinde getirmiştir.

Veri depolarında saklanan ham veri, kayıtlar ya da örnekler ve bu örneklere ait özniteliklerin değerlerinden oluşmaktadır. Bu veri analiz

²⁴ Bu değişkenler matematiksel işlemlerin konusu olamaz. Bu özellikte bir özniteliği kullanan bir kural sadece eşitlik ya da eşitsizliği sınar Sayısal özniteliklerin kullanıldığı öğrenme yöntemlerinde büyüklük (>), küçüklük (<) ve eşitlik (=) ya da eşitsizlik (≠) koşulları sınanır.

²⁵ Sıfır noktasının ne olacağı tanım alanına ve kullanıcıya bağlı olarak farklılıklar gösterebilir.

amacıyla kullanılmak istenildiğinde belirli işlemlerden geçirilir. Veri tabanı oluşturulurken değişik kaynaklardan elde edilen veriler, karar destek sistemlerinde kullanılmak üzere istenilen formatlara dönüştürülür; dönüştürülmüş veri arasında hatalı, yanlış ya da eksik olanlar varsa bu kayıtlar düzeltilir ya da veri tabanından tamamen silinir. Farklı veri tabanlarında oluşturulan veya diğer kaynaklardan elde edilen veriler tek bir merkezi veri tabanında bütünleştirilir. Bu aşamada oluşturulan veri deposu, standart formatta bir veri tabanına aktarılıp, analiz için gerekli olan veri setine ya da setlerine dönüştürülür (Indurkha ve Weiss, 1998).

Büyük ölçekli veri setleri, veri madenciliğinin belirleyici ve en önemli özelliklerinden birisidir (Indurkha ve Weiss, 1998; Frank ve Witten, 2000). Büyük miktarda verinin saklanma ve işlenebilme kapasitesi, daha derin çözümleme yapabilme ve karar verme sürecinde yardımcı olması açısından dikkate değer bir gelişme olarak nitelendirilebilir. Veri setinin ya da kullanılacak örneklemin büyüklüğü aşağıda sayılan nedenlerden ötürü önemlidir.

- **Eğitim ve Sınama Örneklemeleri:** Öğrenme algoritmasının eğitimi için eldeki verinin yanısıra sınama amacıyla bir miktar başka verinin de bulunması gerekmektedir. Veri setinin bu iki amaca da hizmet edecek büyüklükte olmaması durumunda, kısıtlı verinin istenilen etkinlikte kullanılamaması sorunu yaşanır. Veri seti büyük olduğundaysa bu sorun daha başından engellenmiş olur (Noreen, 1988; Back ve diğerleri, 1997).

- **İstatistiksel Anlamlılık:** İstatistiksel anlamlılık testleri, güven aralıkları sınama setlerinin sayısı ile doğru orantılı olduğundan, büyük setlerde daha anlamlı sonuçlar elde edilir (Indurkha ve Weiss, 1998).

- **Daha Yoğun Arama-Tarama Olanğı:** Bilgi teknolojilerinde arama, tek bir büyük adım olarak tanımlanamaz; arama, daha çok, olası örüntüyü bulmak üzere bir çok zıt seçenek arasında birbirini dengeleyici biçimde işleyen bir süreçtir. Böyle bir süreç hem dar, hem de geniş arama uzaylarında, çok sayıda örnekle etkin sonuçlar üretebilir. Veri setinin büyüklüğü özellikle düşük yoğunluklu uygulamalarda önem kazanmaktadır (Keasey ve Watson, 1991).

- **Daha Kontrollü Deneme:** Veri setinin boyutları büyüdükçe, sınıflandırma algoritmasının çok sayıda farklı sına setiyle sınaarak en iyi sınıflandırma performansının elde edilmesi olanaklı olabilir (Ward, 1999).

Bununla birlikte, büyük veri setleriyle sınıflandırma sonucunda daha küçük setlerle elde edilene göre daha yüksek bir sınıflandırma doğruluğu potansiyeli her zaman garanti edilemeyebilir. Herhangi bir yöntemin, büyük boyutlu bir veri setinin sınıflandırma potansiyelini kullanıp kullanamayacağının önceden belirlenmesi gerekir.

2.1.3. Örneklem Büyüklüğü

Veri setinin büyüklüğünün önemi, analizde kullanılacak örneklemin büyüklüğü söz konusu olduğunda da kendini göstermektedir. Veri setinin büyüklüğüne yönelik ölçütler aynı şekilde örneklem büyüklüğü için de geçerlidir. Özet olarak, örneklem ne kadar büyükse sınıflandırma için öğrenilen kuralların doğruluğu ve kitleyi temsil oranı o ölçüde artmaktadır²⁶. Sınıflandırma algoritmaları açısından ele alındığında, kullanılacak örneklemin, öncelikle kitledeki sınıf dağılımlarını olabildiğince yüksek doğrulukla temsil etmesi gerekmektedir²⁷. Bu bağlamda, en yüksek sınıflandırma doğruluğunun ya da en düşük sınıflandırma maliyetinin elde edilebilmesi açısından temsil oranı önemlidir (Keasey ve Watson, 1991).

Hatalı sınıflandırma maliyetlerine bağlı olarak, özgül bir sınıfın doğru sınıflandırılması diğerlerinden daha önemli olabilir. Eğer bu sınıfın dağılımdaki sıklığı daha fazlaysa, standart hata tabanlı algoritmalar özgül sınıfı tahmin etmekte ve toplam maliyeti azaltmakta başarılı olacaktır. Bu bağlamda, halihazırda kullanılan hata minimizasyonu tabanlı algoritmalar maliyete duyarlı sınıflandırma amacıyla kullanılabilirler. Diğer taraftan, ilgi sınıf ya da hedef sınıf örneklerinin kitle ve örneklem içindeki yoğunluğu düşükse, standart maliyete duyarlı yöntemlerin kullanılabilmesi amacıyla, örneklemdaki sınıf dağılımının değiştirilmesi gerekmektedir²⁸. Sınıf

²⁶ Kullanılan veri setinin toplam kitleyi yüzde 80 oranında temsil ediyor olması, bu bağlamda, istenilen doğrulukta kurallar öğrenilmesini ve temsil oranını sağlayacaktır.

²⁷ Çalışmada kullanılan örneklem sınıf ve sektör dağılımları açısından kitleyi temsil edecek şekilde oluşturulmuştur.

²⁸ Çalışmamızda WEKA platformundaki standart algoritmalarla maliyete duyarlı sınıflandırma yapılabilmesi amacıyla katmanlama yöntemleri kullanılmıştır.

dağılımlarını değiştirmekte katmanlama yöntemleri kullanılmaktadır (İkizler, 2002).

Katmanlama, sınıf dağılımını sınıflandırma maliyetleriyle orantılı olarak değiştirmektir. Hedef sınıfı hatalı tahmin maliyetlerine bağlı olarak iki tür katmanlama yöntemi kullanılmaktadır.

Düşük Yoğunluklu Örneklem²⁹: Önemli ya da hedef sınıfa ait örnekler korunurken, diğer sınıflara ait örneklerden bir bölümü rastgele seçilir. Bu yöntemde eğitim setinin boyutları küçüldüğü için sınıflandırma verimliliği düşecektir.

Yüksek Yoğunluklu Örneklem: Sınıf maliyeti düşük olan örnekler korunurken diğer sınıfın örnek sayısı artırılır. Bu yöntemde, veride kayıp olmamakla birlikte gereksiz örnek sayısı arttığı için eğitim süresi uzayacaktır.

Katmanlama yöntemleri, veri setindeki orijinal sınıf dağılımını bozmaktadır. Katmanlama yöntemleriyle yapılan düzenlemeler sonucunda veri setine ait birçok “ilginç” özellik büyük olasılıkla yok olacaktır. Bu nedenle, katmanlamayla elde edilen sınıflandırma modelleri ve kavram tanımları gerçeği yansıtmayacaktır.

2.1.4. Eksik ve Hatalı Değerler

Ham veri ilk olarak elde edildiğinde bazı veri alanları çok önemli görülmemiş ve boş bırakılmış olabilir. Verinin toplanmasındaki asıl amacı etkilememesi koşuluyla, bu hataların düzeltilmesine gerek olmayabilir. Bununla birlikte, aynı veri seti makina öğrenmesiyle analizde kullanıldığında hatalı ve eksik değerler büyük önem taşımaya başlar³⁰ (Langley, 1996; Frank ve Witten, 2000).

Veri setlerinde bir örneğin herhangi bir özneliliğinin değeri eksik olabilir. Bu durumda, özneliliğin aldığı değer bilinmiyor olabilir ya da bu değer

²⁹ WEKA platformundaki algoritmalarla yapılan maliyete duyarlı sınıflandırma analizlerinde kullanılan alt örneklemeler, bu yöntemle başvurularak oluşturulmuştur.

³⁰ Her özneliliğin alabileceği değerlerin dikkatle incelenmesi gerekir. Yazım hataları hatalı değerlerle sonuçlanacaktır. Özgün (ham) verilerin analiz için kullanılacak veri setine dönüşümünde de hatalar korunmuş olabilir. Veri hazırlanırken hatalı değerlerin bulunup düzeltilmesi ya da silinmesine yönelik testlerin yapılması gerekir.

hiç yoktur³¹ (Indurkha ve Weiss, 1998). Eğer herhangi bir nedenle ölçüm yapılmamışsa, bu durumda bilgi kaybından söz edilebilir. Bu bağlamda, eksik değerlerden kaynaklanan bilgi kaybı çeşitli yöntemlerle giderilmeye çalışılmaktadır.

Eksik değerın anlamlı olduđu durumlarda izlenebilecek bir yol, eksik değeri özniteliğın olası herhangi bir değeri olarak ele almaktır. Burada genel olarak izlenen üç yöntem vardır:

- Tüm eksik değerleri global bir sabit değeri ile değıştirmek,
- Eksik değeri özniteliğın tüm değerlerinin ortalamasıyla değıştirmek,
- Eksik değeri, eksik değeri bulunduđu örneğın sınıf etiketini taşıyan tüm örneklerin özgül öznitelige ait ortalamasıyla değıştirmek.

Eksik değeri yerine ikame edilen değeri gerçek bir değeri değildir. Eksik değeri herhangi bir değeri ile ikame edilmesi verinin yanlılığını özgül özniteliğın aldıđı değeri yönünde artırır. Böylece veri seti, sınıflandırmada en çok eksik değeri alan öznitelikleri kural öğrenmede öne çıkaracak şekilde homojenleşecektir.

Eksik değerlerle çözümlemenin diğeri bir yolu da veri setinden eksik değeri olan öznitelikleri ya da örnekleri dışlamak veya eksik değerlerin sınıflandırma sürecini etkilemediğı, mantıksal algoritmaları kullanmaktır.

Eksik değeri öznitelikleri dışlamanın, çoğunlukla çok katı bir yöntem olduđu söylenebilir. Eksik değeri alan öznitelikleri kritik bilgiler barındırabilir. Diğeri taraftan, bazen eksik değeri örneğeri öğrenmede herhangi bir rol oynamazlar. Böyle durumlarda, bu örneklerin tam örneklerden bir farkı yoktur ve öğrenmede kullanılmaları olanaklıdır (Frank ve Witten, 2000).

Örneğeri öznitelikleri değeri hatalı olabilir. Böyle örneğeri makina öğrenmesinde kullanılmamalıdır. Hatalı değeri düzeltilmesi olanaklı ise, bu örneğeri hatalı değeri düzeltildikten sonra kullanılabilir.

³¹ Özellikle finansal tablolarda bazı kalemlerin bulunmamasından dolayı bazı oranlar hesaplanmamaktadır. Bu o oranın sıfır ya da sonsuz olması anlamına gelmeyebilir.

Son olarak, herhangi bir örneğin veri setinde birden fazla tekrarlanması başka bir problem kaynağını oluşturur. Çoğu makina öğrenmesi algoritması, tekrarın yanlışlık yaratması nedeniyle, hatalı kurallar öğrenecektir (Indurkha ve Weiss, 1998).

2.2. Mantıksal Makina Öğrenmesi Algoritmaları

Mantıksal makina öğrenmesi algoritmaları, temel olarak karar ağaçları ve doğrudan kural öğrenen algoritmalar (kapsama algoritmaları) olarak iki başlık altında incelenebilir. Bunlara ek olarak, karar ağaçları ve kapsama algoritmalarının melezi olan parçasal karar ağaçları da mantıksal algoritmalarıdır.

Mantıksal Makina Öğrenmesi Algoritmaları:

- Alan uzmanı olmayanlar tarafından bile anlaşılabilir basit yapıda kurallar öğrenmeleri,
- Öznitelik önseçimine gerek duymadan sınıflandırma yapabilmeleri,
- Dağılım, eksik değer, istatistiksel varsayımlar gibi kısıtlara karşı dayanıklı ve duyarsız olmaları ve
- Karmaşık problemleri çözümlemedeki yüksek performansları, dolayısıyla tercih edilmektedir (Frank ve Witten, 2000).

Aslında, her iki başlık altındaki algoritmalar yapısal olarak birbirine benzemektedir. Karar ağaçlarından kural öğrenmek olanaklı olduğu gibi, kurallardan karar ağacı öğrenmek de aynı şekilde olanaklıdır.

Mantıksal algoritmaların diğer bir ayırıcı özelliği ise kural bağımlılığıdır. Kural tablolarının öğrendiği kurallar birbirinden bağımsız iken, kural listelerinin öğrendiği kurallar belirli bir sırayı takip etmektedir. Çalışmada kullanılan mantıksal makina öğrenmesi algoritmalarının kuramsal açıklamaları aşağıda sunulmaktadır.

2.2.1. Karar Ağaçları

Karar ağaçları, sınıflandırma sürecinde özniteliklerin aldığı değerlerin mantıksal sorgulamalardan geçirilerek sınanması ile yukarıdan-aşağıya bir yol izlenmesi ve bu yaklaşımda başlangıcın ağacın köküne, sınıma koşullarının dal ayırımlarına ve son aşamada öğrenilen kuralların yapraklara benzemesi dolayısıyla ağaç olarak adlandırılmışlardır. Sınamanın başlangıcında, örneklerin ilk uğradıkları yer ağacın kökünü, dalların ayırım noktaları ağacın düğümlerini, sınıflandırmanın son bulunduğu yer ise ağacın yapraklarını oluşturur.

Karar ağaçlarının inşası yineleme ilkesine dayanır. Sınamanın başlangıcında her örnek ağacın köküne yerleştirilir. Kökten sonraki ilk düğümden en yüksek ayırımı yapan öznitelik değerine dayalı sınıma sorgulaması yapıldıktan sonra örnek, sırayla yönlendirildiği düğüme iletilir. Sırayla, her düğümden sınanan örnek son yaprağa ulaşana kadar sorgulamaya devam edilir. Öğrenme sürecinde, her yeni örnek için en yüksek ayırımı sağlayan öznitelik farklı düğümler ve sınıma sorgulamalarının oluşmasını sağlar. Bu bağlamda, karar ağacı dinamik bir öğrenme yöntemidir (Frank ve Witten, 2000).

Genellikle, bir düğümden sınanan özneliğin değeri sabit bir değerle karşılaştırılır³². Bunun yanında, bazı ağaçlar, iki özneliği birbiriyle veya bir ya da daha fazla öznelikten oluşan bir fonksiyonu karşılaştırmak amacıyla tasarlanabilir.

Nominal bir öznitelik için, alt düğümlerin sayısı genellikle özneliğin alacağı olası değerlerin sayısı kadardır. Bu durumda, her olası değer için bir dal çıkarılır ve aynı öznitelik ağacın sonraki düğümlerinde sınanmaz.

Bazı durumlarda, öznitelik değerlerinin dağılımı iki değer aralığında yoğunlaşmıştır ve ağaç değerin hangi alt kümede olduğuna bağlı olarak dallanır. Bu durumda, öznitelik aynı patika üzerinde birden fazla kez sınanabilir.

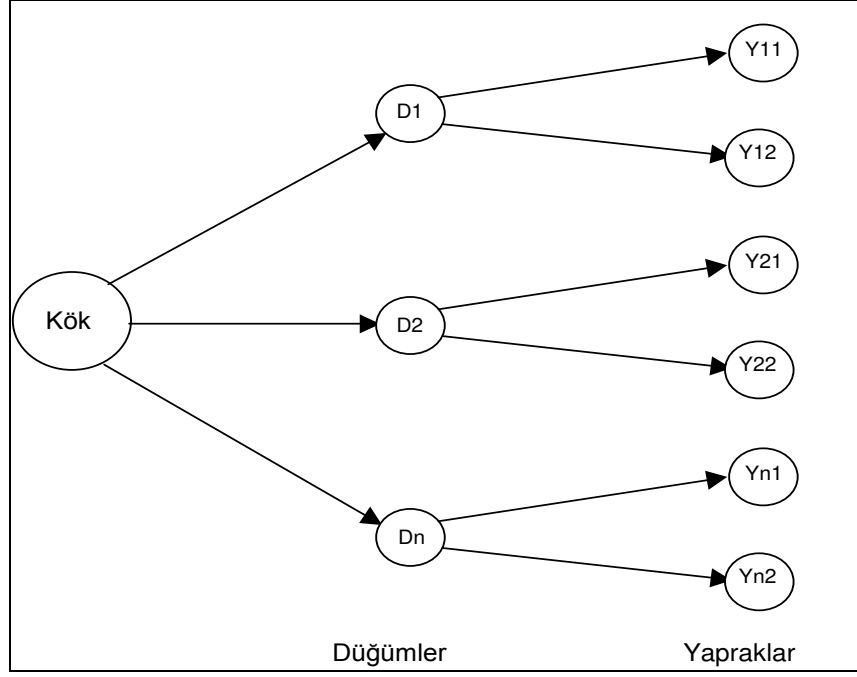
³² Bu sabit değer yeni örnek eklenene kadar oluşturulan ağacın düğümlerindeki sınıma sorgulamalarının sabit değeridir. Yeni örnek öğrenilirken sabit değer değişebilir.

Sayısal bir öznitelik için düğüm ikiye ayrılarak öznitelik değerinin sabit bir değerden küçük ya da büyük olup olmadığı sınanır ($<$ veya $>$). Tamsayı değerler alan bir öznitelik için düğümüne eşitlik ($=$) sorgulaması eklenerek üçüncü bir dal çıkarılabilir. Eşitliğin anlamlı olmadığı gerçel değerli öznitelikler tek bir sabit yerine bir aralık üzerinde sınanır. Bu sınamada düğümünden çıkan dal sayısı yine üçtür; öznitelik değerinin aralığın altında, üstünde ya da içinde olup olmadığı sorgulanır. Sayısal bir öznitelik, kökten yaprağa kadar herhangi bir patikada her yeni örnekle birlikte başka bir sabit kullanılarak defalarca sınanabilir.

Yaprak düğümleri ya da yaprak uçları o yaprağa ulaşan tüm örneklerle uygulanabilecek sınıfları, sınıflar kümesini ya da tüm olası sınıfların dağılımını verir. Sınıf etiketi bilinmeyen yeni bir örnek sınıflandırılırken, örneğe ait öznitelik değerleri birbirini izleyen bütün düğümlerde sınanır. Örnek tüm düğümlerdeki sınamaları geçer ve yaprağa ulaşırsa o yaprağın sınıf etiketine göre sınıflandırılır.

Karar ağacı oluşumu tüm yaprakların “saf” olması durumunda son bulmalıdır; yani o yaprağa ulaşan tüm örneklerin sınıf etiketleri aynı olmalıdır. Bununla birlikte, yaprakların saflaşması her zaman olanaklı değildir. Veri setinde, sınıf etiketleri dışında tüm öznitelik değerlerinin aynı olduğu iki örneğin bulunması durumunda karar ağacının inşası herhangi bir sonuç elde edilemeden durdurulur.

Karar ağacında Y yaprağına ulaşan, yani bu yaprakta sınıf etiketi alan, bir örnek bu şartların hepsini sağlamak zorundadır. Karar ağaçlarından kural öğrenilirken kökten yaprağa kadar bu hiyerarşik yapının izlenmesi gereklidir. Karar ağaçlarının öğrendikleri örüntüler bu sıranın korunması bağlamında katıdır. Ancak bazı kurallar (dallar) modelin aşırı özelleşmesini önlemek amacıyla budanır ve daha genel bir sınıflandırıcı elde edilebilir. Budama sürecinde daha genel bir sınıflandırıcı elde edilirken ve eğitim kümesi için sınıflandırmanın eğitim seti için doğruluğu azalırken, kesin olmamakla birlikte, sinama kümelerindeki tahmin doğruluğu ve genelleştirme potansiyeli artmaktadır (Quinlan, 1986). Karar ağaçlarının genel yapısı Şekil 2. 1’de sunulmaktadır.



Şekil 2.1: Karar Ağaçlarının Genel Yapısı

2.2.2. Kapsama Algoritmaları ve Sınıflandırma Kuralları

Karar ağaçlarındaki yukarıdan-aşağıya sorgulamanın tersine, sınıflandırma kural algoritmalarında aşağıdan-yukarı bir yol izlenerek, her sınıfa dahil olan örnekleri en fazla kapsayan özniteliklere bağlı kurallar öğrenilir ve bu süreçte kapsam dışı olan örnekler dışlanır. Kapsama algoritmaları, doğrudan kural öğrenen algoritmalarıdır.

Sınıflandırma kuralları, birbirine bağımlı³³ veya birbirinden bağımsız³⁴ olabilir. Bu kuralların her biri bağımsız bir bilgi parçası niteliği taşırlar ve sınıflandırmada her bir kural tek başına kullanılabilir. Burada, alan uzmanının tercihleri, kuralın kuramsal anlamlılığı, yeniliği ve ilginçliği gibi ölçütler hangi kuralın kullanılacağı konusunda yardımcı araçlardır. Bağımlı kuralların, birbirinden bağımsız olarak kullanılmaları söz konusu olamaz. Sınıflandırmada öğrenilen bütün kuralların aynı anda sağlanması koşulu aranır.

Kapsama algoritmaları, azami doğruluğu sağlayacak şekilde, oluşturulan kuralı her defasında yeniden sınama yoluyla çalışır. Kural algoritmalarında dikkat edilmesi gereken, eklenen kısıtın, bulunmak istenen

³³ Kural listeleri bağımlı kurallar öğrenen algoritmalara örnek oluşturmaktadır.

³⁴ Bağımsız kurallar, karar tablosu gibi algoritmaların öğrendiği kurallardır.

sınıf kapsamını artırırken diğerlerini dışlayacak şekilde, kuralın kapsamını daraltmasıdır.

Bir kuralın öncül veya önkoşulu, karar ağaçlarındaki düğümlerde yapılan sınamalarda da olduğu gibi bir dizi sınamadan oluşmaktadır. Soncul ya da son koşul, yaprak ya da kural tarafından kapsanan örneklere karşılık gelen sınıf etiketleri ya da sınıfların olasılık dağılımlarıdır.

Kural algoritmalarında önkoşullar, genellikle, mantıksal olarak VElenmiştir ve kuralın sağlanması bütün sınamalardan aynı anda geçilmesine bağlıdır. Diğer taraftan, birbirinden bağımsız kurallar kendi aralarında mantıksal olarak VEYAlanmıştır ve öncüllerden herhangi birinin doğrulanması durumunda örneğin sınıf etiketi, sonculun sınıf etiketi olacaktır³⁵.

Karar kurallarının karar ağaçlarına göre avantajlı yönleri şunlardır:

- Karar kuralları, özellikle başka kuralların kapsamadığı durumları kapsayan temel kuralın elde edilmesi olanaklıysa, karar ağaçlarına göre daha derli topludur ve anlaşılabilirliği daha kolaydır.
- Kapsama algoritmalarının çıktısı, her bir kural bağımsız bir bilgi parçası niteliği taşımaktadır. Yeni kurallar eski kural setine eklendiğinde, eski kurallar geçerliliğini yitirmezler. Karar ağaçlarındaysa veri setine eklenen her yeni örnekle, bütün ağacın yeniden şekillendirilmesi gerekebilir.

Diğer taraftan karar kurallarının karar ağaçlarına göre dezavantajlı yönleri de bulunmaktadır:

- Veri bir örnek, farklı kurallar tarafından farklı sınıf etiketleriyle etiketleniyorsa, genellikle bir çözüme ulaşmak olanaklı olmayabilir. Bu sorunu ele almanın bir yolu, böyle bir örneği sınıflandırmadan bırakmak ya da en çok tekrarlanan kuralı temel kural olarak kabul etmektir. Karar ağaçlarından elde

³⁵ Mantıksal VElenmiş kuralların genel yapısı aşağıdaki gibidir.

$P_1 \wedge P_2 \wedge P_3 \wedge \dots \wedge P_n \Rightarrow C_j$ $C_j = A, B, C, \dots, N$
biçimindedir. Burada P_i öncüllere, C_j ise sınıf özniteliklerine karşılık gelmektedir. Öncül (P_i), öznitelik (F_k) için eşitlik ($=$) ya da eşitsizlik ($<, >$) gibi koşulların belirli bir eşik noktası (T) için sağlandığı matematiksel ifadedir ($P_i \equiv F_k \ominus T$). Örneğin, firma oranları kullanılarak öğrenilen bir kural şu şekilde olabilir,
 $CariOran < 0.5 \vee LikiditeOrani > 1.73$
 $\vee StokDevirHizi > 3.07 \vee BankaKredileriNetSatislar > 1.87$
 $\Rightarrow$ **BAŞARISIZ** (n/N)

n/N kuralın ne kadar örnek tarafından sağlandığının bir ölçüsüdür. N : toplam örnek sayısı, n : kuralı sağlayan örnek sayısı.

edilen kurallarda genellikle böyle bir sorunla karşılaşılmamaktadır. Çünkü, bu kuralların yapısındaki gereksizlik³⁶ her türlü belirsizliği ortadan kaldırır.

- Kapsama algoritmalarının öğrendiği kuralların genel problemi, veriyi tamamıyla açıklamakla birlikte bağımsız, özellikle kirli, eğitim verilerinde doğru genelleme yapamama eğilimi taşımalarıdır.

Kapsama algoritmaları görünüşte uygulanması kolay yöntemlerse de, gerçek hayat problemlerinde, genellikle bunun tam tersi söz konusu olmaktadır. Genel kanı, her kuralın bağımsız bir bilgi parçası olduğu gerçekliğinin her tür sınıflandırma probleminde geçerli olmasıdır. Bununla birlikte, ikili sınıflandırma dışındaki problemlerde bu gerçeklik kuramsal geçerliliğini yitirir. Çoklu sınıflandırma problemlerinde karar kuralları, genellikle birbirinden bağımsız değildir ve kuralların uygulamasında sıralamaya uyulması bir zorunluluk halini almaktadır (Frank ve Witten, 2000).

2.2.3. Karar Ağaçları ve Kurallar

Karar ağaçlarından kural öğrenmek olanaklıdır. Genellikle, karar ağaçlarından öğrenilen kurallar, kapsama algoritmalarının doğrudan öğrendiği kurallardan daha karmaşıktır³⁷; bu durumda, gereksiz ya da anlamsız kuralların budanması gerekir.

Kurallar ve ağaçlar arasındaki tanımlamanın sadeliği ve açıklığı yönünden farklılıklar aşağıda özetlenmiştir. Bu farklılıklar:

- Kurallar bakışımı (simetrik) olabilirken, ağaçların her defasında en fazla bölüntülemeyi sağlayacak özneliği bulmak zorunda olması ve bu durumun aynı problemde öğrenilen kural setine göre daha büyük bir ağaç elde edilmesiyle sonuçlanması,
- İki'den fazla sınıfın söz konusu olduğu durumlarda, karar ağaçları bölüntülemenin saflığını artırmak üzere bütün sınıfları dikkate alırken, kapsama algoritmalarının bir defada diğer sınıfları dikkate almadan sadece tek bir sınıf üzerine yoğunlaşmasıdır.

³⁶ Gereksizlikle aşırı ayrıntılı ya da birbirini tekrar eden kurallar kastedilmektedir.

³⁷ Karar ağaçlarından öğrenilen kuralların açık olması kuralların herhangi bir sıra izlemek zorunluluğu olmadan uygulanmasını kolaylaştırır.

2.2.4. Karar Ağaçlarından Kural Öğrenme

Karar ağaçlarının oluşturduğu yapılardan öğrenilen kurallar her yaprak üzerinde, kökten o yaprağa kadar olan tüm sınamalar birleştirilerek elde edilir.

Öncelikle veri bir kuralda her koşul silinmek üzere denenir. Silme işlemiyle hangi eğitim örneklerinin kapsandığı belirlenir, özgün doğruluk performansı yeni kuralla yapılan sınıflandırmadan elde edilen doğruluk performansı ile karşılaştırılır. Eğer yeni kuralın doğruluk performansı, orijinalde fazladan bulunan koşul silindikten sonra iyileşiyorsa, özgül koşul silinebilir. Bu işlem, kısıtların silinmesiyle sınıflandırma doğruluğunda bir artış sağlanamayınca kadar devam ettirilir. Son aşamada, kuralların budanması sonlandığında birbirinin tekrarı kurallar da silinir³⁸.

Karar ağaçlarından kural öğrenmenin, gereksiz kuralları elemekte “açgözlü” olduğunu söylemek olanaklıdır ve bu süreç her zaman en gereksiz olduğu düşünülen koşulların silinmesini garanti etmeyebilir. Açgözlü yaklaşımlarda karşılaşılan önemli bir sorun bilgi işlem maliyetidir. Silinmeye aday her koşul için, yeni kuralın bütün eğitim örnekleri üzerinde sınanması gerekmektedir. Bu, karar ağaçlarından kural öğrenmenin çok yavaş olacağı anlamına gelir ki, daha hızlı ve doğrudan yöntemler varken bu yolla kural öğrenmek anlamsız olacaktır³⁹.

Karar ağacından elde edilen kurallar, ağacın kökünden yaprağına kadar her düğümde sına-eşik koşullarını sağlayan örnekler kullanılarak oluşturulur. Genel formu Şekil 2.1’de tanımlanan bir karar ağacından elde edilen kural(lar)ın genel yapısı aşağıdaki gibidir:

$P(K) \wedge P(D1) \wedge P(Y1) \Rightarrow \text{Sınıf (Y1'e kadar bütün öncülleri karşılayan örneklerin çoğunluğunun sahip olduğu sınıf etiketi)}$

³⁸ Örneğin, karar ağacından elde edilen iki kuraldan

Kural 1: $1.78 < \text{CariOran} < 2.13$ **İSE BAŞARISIZ** ve

Kural 2: $1.92 < \text{CariOran} < 2.01$ **İSE BAŞARISIZ**

kurallarından ikincisi gereksizdir ve budanmalıdır, çünkü, birinci kural tarafından kapsanmaktadır. Bu gereksizlik, birinci kuralın tanım aralığında aksi bir durumun sorgulanmaması durumunda geçerlidir. Örneğin,

Kural 2: $1.85 < \text{CariOran} < 1.97$ **İSE BAŞARILI**

olsaydı, öğrenilen karar ağacı yanlış ve anlamsız olurdu.

³⁹ Tek bir kural öğrenmek için tam bir karar ağacı oluşturmak gerçekten savurganca bir tutumdur. Yine de sürecin sayılan avantajlarından ödün verilmeden hızlandırılması olanaklıdır.

2.2.5. Kurallar ve Karar Listeleri

Kuralların uygulamasında herhangi bir “sıra” izlenmesi zorunluluğu yoktur; örneği kapsayan ve sorgulanan sınıflandırmayı sağlayan bir kural ya geliştirilir ya da böyle bir kural bulunmaz ve sınıflandırma yapılmaz. “Sıradan” bağımsız kurallar, her biri birbirinden bağımsız bilgi parçaları olarak, daha modüler bir yapı sergilemektedir. Bununla birlikte, birbirine zıt kuralların aynı sınıfı tahmin etmesi durumunda ne yapılması gerektiği pek de açık değildir. Böyle bir durumda özgül kural öğrenme algoritmasının veri setine uygulanması uygun olmayacaktır. Karar listelerinde, sıraya bağlı algoritmalar olmaları dolayısıyla, böyle sorunlarla karşılaşmaz. Karar listesinin sonuna eklenen kuralla, herhangi bir örnek sınıflandırılmadan bırakılmamış olur (Rivest, 2001).

Karar listesi öğrenmek için kullanılan kapsama algoritmalarının karar ağacı algoritmalarına üstünlüğü, son aşamalarda çoğu örneğin sınıflandırılmış (örnek kümesinden çıkarılmış) olması dolayısıyla daha az karmaşık olmaları ile karmaşık ve sorunlu örnekleri en son aşamaya kadar örnek kümesinde tutmaları ve bu örnekleri son ana kadar sınıflandırmamakta ısrarcı davranmalarıdır.

2.2.6. Kapsama Algoritmaları ve Eksik Değerler

Kapsama algoritmaları yardımıyla kural öğrenilirken, eksik değerler sanki herhangi bir sınama koşulunu geçmiyormuş gibi ele alınabilir. Bu yaklaşım, öğrenmeyi örnekleri sınıflandırmada başarılı olacağı bilinen sınamaları kullanmaya yönlendirdiği için özellikle “karar listesi” algoritmalarının tasarlanmasında kullanılmaya uygundur.

2.2.7. Yararlı ve Yararsız Kurallar

Herhangi bir kuralın ne düzeyde yararlı olduğunu belirlemenin bir yolu, rastgele seçilen başka bir kuralın veri kuralın doğruluk performansına eşit ya da daha iyi bir performans sergileyecek bir kural olup olmadığını sorgulamakla olanaklıdır.

Eğer genel olarak bir kurala yeni bir terim eklenmesi kuralı “mükemmel” bir duruma getiriyorsa, kuralın eğitim setine özelleşme tehlikesi baş gösterebilir. Öğrenmenin amacı, algoritmanın yeni örnekleri de başarılı olarak sınıflandıracak genel kurallar öğrenmek olduğu için, mükemmel kurallar öğrenmek çoğu zaman özel bir durum için geçerlidir ve bu durumdaki “mükemmel” kurallar şüpheyle karşılanmalıdır.

İyi kurallar geliştirilirken, öncelikle “mükemmel” kural(lar) öğrenilir. Bu kuralın, olasılıkla aşırı özelleşmesinden dolayı, eklenen son terimi budamak gereklidir. Eğer kural hala iyiye, son terimin silinmesiyle kuralın performansı düşene kadar silme işlemine devam edilebilir.

2.2.8 Parçasal Karar Ağaçları

Parçasal karar ağaçları Frank ve Witten (1998) tarafından geliştirilmiştir. Bu algoritmalar karar ağaçlarının geliştirdiği kurallarınkine eşdeğer ve birçok hızlı kural algoritmasından daha yüksek doğrulukta kurallar öğrenirler. Parçasal karar ağaçlarının diğer yöntemlere asıl üstün yanı, yüksek performansından çok basit olmasıdır. Karar ağaçları ve kural öğrenme yöntemlerini birleştirerek global optimizasyona gerek duymadan iyi kurallar elde edilmesini sağlar⁴⁰.

Parçasal karar ağaçları, karar ağaçlarıyla öğrenilen ağaçların budanmasıyla elde edilir. Karar ağacının budanması ağacın boyutunu küçülttüğünden, parçasal karar ağacıyla elde edilen kuralların sayısı karar ağacından okunarak elde edilenlere göre daha az olacak ve böylece öğrenme süresi kısılacaktır. Bununla birlikte, genellikle, eğitim örneklerinin sayısının azalmasıyla, elde edilen budanmış karar ağacının sınıflandırma doğruluğu da düşecektir. Büyük veri setlerinin öğrenilmesiyle bu sorun çözülebilmektedir.

Parçasal karar ağaçlarıyla kural öğrenilirken, her defasında, öğrenilen kuralın kapsadığı örnekler, örnek kümesinden çıkarılır ve bu şekilde sonuçta hiç örnek kalmayınca dek kural öğrenilmeye devam edilir.

⁴⁰ Parçasal karar ağaçları, karar ağacı ve kural öğrenme yöntemlerinin melezi öğrenme yöntemleridir. Bu algoritmalar, her ne kadar karar ağacı algoritmaları olarak görülseler de, çıktı açısından, doğrudan kural öğrenen algoritmalarıdır.

Böylece, budanmış bir karar ağacı oluşturulur, en geniş kapsamı olan yaprak kurala dahil edilir, ağaç tamamen silinir ve budanmış karar ağaçlarından kurallar elde edilir. Öğrenilen kurallar arasından en genel kuralın seçilmesiyle süreç sona erer.

Parçasal karar ağaçlarında, ağacın önceden budanmasıyla aşırı budama eğilimi en başından ortadan kaldırılır. Bu yöntemin karar ağacıyla birlikte kullanılması kural öğrenme sürecine esneklik ve sürat kazandırır (Frank ve Witten, 2000).

2.3. Makina Öğrenmesi ve Sınıflandırma

Makina öğrenmesinde, sınıflandırma algoritmaları, veri seti olarak öğrenme programına sunulan sınıf etiketli örnekleri kullanarak tanım alanına yönelik örüntüleri öğrenir. Sınıflandırma, örneklerden çıkarım olarak da bilinmektedir. Sınıflandırmada amaç, kavram tanımı elde edildikten sonra, daha önce algoritmaya tanıtılmamış örnekleri en yüksek doğrulukla etiketleyecek sınıflandırıcıyı geliştirmektir.

Sınıflandırmanın girdisi örnek seti, eğitim verisi ya da örnek uzayı olarak adlandırılır ve sınıf etiketlerinin sahip olduğu belirsiz olasılık dağılımıyla aynı olasılık dağılımına sahip olduğu varsayılır⁴¹.

Öznitelik değerleri sayısal ya da nominal olabilir. Sınıf öznitelikleri ise kategorik değerler almaktadır.

Tanım alanına bağlı olarak, ikiden fazla sınıf etiketi alan veri setlerinde, öğrenme problemi çoklu sınıflandırma problemine dönüşmektedir (İkizler, 2002).

⁴¹ 'Örneklem sınıf dağılımı kitlenin sınıf dağılımına eşittir' varsayımı.

2.3.1. Sınıflandırma Doğruluğu

Sınıflandırma doğruluğu, veri setiyle ilgili doğru tahminlerin oranı olarak tanımlanmaktadır ve modelle verinin ne derecede birbirleriyle uyumlu olduklarının bir göstergesidir⁴².

M sınıflandırma modelinin, S örnek setini D tanım alanında genel sınıflandırma ya da tahmin doğruluğu ($\text{Doğruluk}_{D(\text{Genel})}$):

$$\text{Doğruluk}_{D(\text{Genel})}(M, S) \equiv \frac{\text{Doğru Sınıflandırılan Örnek Sayısı}_{D(\text{Genel})}(M, S)}{\text{Toplam Örnek Sayısı}_{D(\text{Genel})}} \quad (2.1)$$

olarak tanımlanmaktadır. Tanım alanının özelliğine bağlı olarak öncelikli ya da hedef sınıfı tahmin doğruluğu⁴³ ($\text{Doğruluk}_{D(\text{Hedef})}$):

$$\text{Doğruluk}_{D(\text{Hedef})}(M, S) \equiv \frac{\text{Doğru Sınıflandırılan Örnek Sayısı}_{D(\text{Hedef})}(M, S)}{\text{Toplam Örnek Sayısı}_{D(\text{Hedef})}} \quad (2.2)$$

olarak tanımlanmaktadır. Genel ve hedef sınıfı tahmin doğruluğu metrikleri, farklı sınıflandırma algoritmalarının sınıflandırma verimliliklerinin karşılaştırılmasında ve veri öğrenme yönteminin veri seti içindeki verimliliğinin ölçülmesinde kullanılmaktadır.

2.3.2. Sınıflandırma Maliyetleri

Sınıflandırma algoritmalarının kullanımı bazı maliyetleri de beraberinde getirmektedir. İkizler (2002) bu maliyetleri aşağıdaki başlıklar altında özetlemektedir:

- **Hatalı Sınıflandırma Maliyetleri:** Bu tip maliyetler en önemli maliyetlerdir ve maliyete duyarlı öğrenme çalışmalarının önemli bir bölümü hatalı sınıflandırma maliyetlerini minimize etme yöntemleri üzerine yoğunlaşmıştır. Hatalı sınıflandırma maliyetleri sabit veya değişken olabilir.

⁴² Sınıflandırma algoritması sınıflandırma yaparken doğruluk maksimizasyonu amaç fonksiyonu altında çalışacaktır. Bu matematiksel olarak

MAX Doğruluk k.a. Maliyet Tablosu

biçiminde ifade edilebilir.

⁴³ Bu metrik, makina öğrenmesinde "Recall" olarak tanımlanmaktadır. Recall, herhangi bir sınıfa ait örneklerden doğru sınıflandırılanların, o sınıfa ait toplam örnek sayısına oranıdır.

Değişken maliyetler, özgün duruma, sınıflandırma zamanına⁴⁴ ve diğer vakaların sınıf etiketlerine bağlı olabilir. Halihazırda, değişken maliyetleri kullanmak üzere geliştirilen bir algoritma bulunmamaktadır. Böyle bir algoritmanın karmaşıklığı, özellikle veri setinin büyüklüğü arttıkça sınıflandırma algoritmasının işletim zamanı, donanım ve yazılım gerekleri ve bağlantılı maliyetleri de beraberinde getirmektedir.

- **Sinama Maliyetleri:** Sinama maliyetleri, standart sınamalar yapıldıktan sonra yüksek sınıflandırma doğruluğu sağlayan sınıflandırıcının elde edilememesi durumunda ortaya çıkmaktadır. Hatalı sınıflandırmanın maliyetinin sinama maliyetlerinden fazla olması durumunda, sınıflandırma doğruluğunu artıracak tüm sınamalara başvurulabilir. Sinama maliyetleri de sabit ya da değişken olabilir.

- **Eğitim Maliyeti:** Bazı örneklerin gerçek sınıf etiketlerinin belirlenmesi olanaklı olmayabilir. Böyle durumlarda öğrenme algoritmasının eğitim maliyetini minimize etmek üzere çalışması istenir.

- **Hesaplama Maliyeti:** Problemin boyutları ve karmaşıklığı, yer ve zaman gerekleri hesaplama maliyetlerini oluşturmaktadır.

- **Örnek Edinme Maliyeti:** Öğrenme için kullanılacak örneklerin edinilmesinin yani veri setini oluşturma maliyetleridir.

2.3.3. Sınıflandırmada Fayda ve Maliyet Duyarlılığı

Klasik sınıflandırma sistemleri bütün sınıflandırma hatalarını sınıflandırma doğruluğu ölçütüyle değerlendirir. Bununla birlikte, çoğu tanım alanında hataların farklı anlamları ve sonuçları olabilir. Örneğin, finansal başarısızlık analizinde, her yönüyle birbirinin aynı iki firmadan, “başarısız” olarak sınıflandırılan fakat “başarılı” olan bir firmaya yapmaktan kaçındığı yatırımın yatırımcıya maliyeti, “başarılı” olarak sınıflandırıldığı halde “başarısız” olan bir firmanın getirdiği maliyetten daha düşüktür. İlk durumda bu maliyet sadece gerçek olmayan fırsat maliyeti iken diğerinde gerçekten

⁴⁴ Tanım alanına bağlı olarak, sınıflandırmanın erken veya geç yapılması farklı maliyet rakamlarıyla sonuçlanabilir. Örneğin, finansal baskı yaşayan bir firmanın sınıf etiketinin, finansal baskının erken aşamalarında hatalı tahmin edilmesinin maliyeti, baskının ağırlaştığı geç aşamalarda hatalı tahmin edilmesinin maliyetinden daha düşüktür.

kaybedilen bir yatırıma karşılık gelmektedir. Bu gibi durumlarda, örnekleri en yüksek doğrulukla sınıflandırmak yeterli olmamaktadır. Bunun yerine, sınıflandırma programı, hatalı sınıflandırmanın maliyetlerini de göz önünde bulundurmalıdır⁴⁵. Burada dikkat edilmesi gereken bir konu, maliyete duyarlı sınıflandırmada, sınıflandırma doğruluğundan bir ölçüde feragat etmek durumunda kalılabileceğidir (İkizler, 2002).

İkili sınıflandırmada hatalı sınıflandırma maliyetleri, istatistikte sıklıkla kullanılan Birinci Tip ve İkinci Tip Hata maliyetlerinin toplamından oluşmaktadır. Bu iki hata türü, standart sınıflandırma algoritmalarında örneklerin sınıf etiketlerinin hatalı tahmin edilmesine karşılık gelmektedir. Maliyete duyarlı sınıflandırmada, tanım alanına, sınıf sayısına ve sınıf dağılımlarına bağlı olarak, sınıf etiketlerini hatalı tahmin etmenin bir maliyeti olması gerektiği, böylece algoritmaların daha yüksek doğrulukla sınıflandırma yapacak kuralları öğreneceği varsayılmaktadır. Diğer taraftan, doğru sınıflandırmanın ödüllendirilmesiyle aynı şekilde daha güçlü sınıflandırıcıların elde edilmesi söz konusu olabilir. Bu bölümde, maliyete ve faydaya duyarlı makina öğrenmesi ele alınmıştır.

2.3.3.1. Maliyete Duyarlı Öğrenme

Maliyete duyarlı sınıflandırma algoritmaları, sınıflandırma sürecinde maliyet bilgisinin kullanılmasını öngörmektedir. Maliyet bilgisi, verinin toplanması, özniteliklerin bulunması ve hatalı sınıflandırma gibi birçok başlığı barındırabilir. Bunlar arasında en önemlisi hatalı sınıflandırmanın maliyetidir (İkizler, 2002; Güvenir ve İkizler, 2003).

Hatalı sınıflandırma maliyetleri, tanım alanına, sınıf sayısına ve sınıf dağılımlarına bağlıdır. Tanım alanından kaynaklanan ilgi (hedef) sınıfın öncelikli olarak tahmin edilmesi gerekliliğine bağlı olarak, özellikle düşük yoğunluklu örneklemelerde birinci ve ikinci tip hata maliyetleri birbirinden farklı olarak gelişir. Örneğin, bankaların kredi verme kararlarında, krediyi geri ödemeyen bir müşterinin bankaya maliyeti sorunsuz bir müşteriye kredi

⁴⁵ Sınıflandırma algoritması sınıflandırma yaparken maliyet minimizasyonu amaç fonksiyonu altında çalışacaktır. Bu matematiksel olarak

MIN Maliyet k.a. Maliyet Tablosu biçiminde ifade edilebilir.

verilmemesinden doğan kayıp getiriden daha fazla olacaktır. Maliyetlerin bilinmesi, maliyete duyarlı analizi kolaylaştırmaktadır. Bununla birlikte, hatalı sınıflandırma maliyetlerinin büyüklükleri çoğunlukla belirsizdir. Bu bağlamda, hata maliyetlerinin bakışimsızlığı, gerçekçi ve kabul edilebilir bir durumdur (Güvenir ve İkizler, 2003).

Maliyete duyarlı sınıflandırmada, maliyet bilgisinin öğrenme algoritmasına iletilmesinde kullanılan araç maliyet tablosudur. Maliyet tablosunda birinci düzeyde, hatalı sınıflandırma maliyetleri sabittir⁴⁶. Sabit maliyet tablolarındaki hata maliyetleri, genellikle, kullanıcının deneme yanılma yoluyla atadığı büyüklüklerdir. Sabit maliyet tablolarındaki maliyet bilgisi veri setindeki farklı sınıf gruplarına ait tüm örnekler için geçerlidir ve sınıflandırma boyunca aynı bilgi kullanılmaktadır.

		TAHMİN EDİLEN	
		A	B
GERÇEK SINIF	A	0	M(A_B)
	B	M(B_A)	0
M(.) = Sabit			

Şekil 2.2: Sabit Maliyet Tablosu

M sınıflandırma modelinin, sabit maliyet bilgisiyle, S örnek setini, D tanım alanında hatalı sınıflandırmanın toplam maliyeti ($Maliyet_D^S(M,S)$):

$$Maliyet_D^S(M,S) \equiv \sum_1^{N(A_B)} M(A_B) + \sum_1^{N(B_A)} M(B_A) \quad (2.3)$$

olarak tanımlanmaktadır.

$M(A_B) \equiv$ A etiketli örneklerin B olarak sınıflandırılmasının maliyeti,
 $M(B_A) \equiv$ B etiketli örneklerin A olarak sınıflandırılmasının maliyeti,
 $N(A_B) \equiv$ A etiketli örneklerden B olarak sınıflandırılanların sayısı ve
 $N(B_A) \equiv$ B etiketli örneklerden A olarak sınıflandırılanların sayısıdır.

⁴⁶ Gerçek sınıf etiketinin A iken B olarak tahmin edilmesi durumunda yüklenilen hatalı tahmin maliyeti $M[A_B]$ ve gerçek sınıf etiketinin B iken A olarak tahmin edilmesi durumunda yüklenilen hatalı tahmin maliyeti $M[B_A]$ 'dir. Bu maliyetler bütün A iken B ya da B iken A olarak tahmin edilen örnekler için geçerlidir.

2.3.3.2. Maliyete Duyarlı Öğrenmede Öznitelik Bağımlılığı

Maliyete duyarlı sınıflandırmanın önemli bir özelliği farklı hatalı sınıflandırma maliyetlerine olan duyarlılığıdır. Maliyete duyarlı öğrenme çalışmalarının genelinde, hatalı sınıflandırma maliyetlerinin homojen olmayabileceği ve bu maliyetlerin de bireysel örnekler bağımlı olduğu gerçeği göz ardı edilmektedir. Fayda ve maliyetlerin homojen olmaması optimal sınıflandırma problemine yeni bir boyut getirmektedir. Sınıflandırma sürecinde toplam maliyeti minimize edecek ve/veya toplam faydayı maksimize edecek sınıflandırıcılar elde edilecektir. Öznitelik bağımlı maliyete/faydaya duyarlı sınıflandırma, sıradan maliyete/faydaya duyarlı sınıflandırma sürecinden farklıdır. Her bireysel örnek tek tek ele alınır ve çıktının göreceli önemine bağlı olarak her örneğe özgü kararlar elde edilir (Güvenir ve İkizler, 2003).

Özniteliğe bağımlı maliyete/faydaya duyarlı sınıflandırma problemlerinde, sabit veya statik bir fayda ya da maliyet bilgisinden söz edilemez. Bu bağlamda, her bireysel örneği sınıflandırmanın maliyeti ve/veya faydası bireysel örnekler için ayrı ayrı hesaplanır. Fayda ve maliyetlerin örnekler arasında homojen olmaması öznitelik bağımlılığı olarak adlandırılmaktadır.

Dinamik ya da değişken maliyet tablolarında, hatalı sınıflandırmanın maliyetleri, tanım alanı bağlamında belirlenen bir değişkenin fonksiyonu olarak tanımlanmaktadır (Şekil 2.3).

		TAHMİN EDİLEN	
		A	B
GERÇEK SINIF	A	0	$F[.](A_B)$
	B	$F[.](B_A)$	0
$F. = \text{Maliyet Fonksiyonu}$			

Şekil 2.3: Değişken Maliyet Tablosu

M sınıflandırma modelinin, değişken maliyet bilgisiyle, S örnek setini, D tanım alanında hatalı sınıflandırmanın toplam maliyeti (Maliyet_D^{tilde}(M,S)):

$$\text{Maliyet}_D^{\tilde{}}(M, S) \equiv \sum_1^{N(A_B)} F[.](A_B) + \sum_1^{N(B_A)} F[.](B_A) \quad (2.4)$$

olarak tanımlanmaktadır.

$F[.](A_B)$ = A etiketli örneklerin B olarak sınıflandırılmasının maliyet fonksiyonu,

$F[.](B_A)$ = B etiketli örneklerin A olarak sınıflandırılmasının maliyet fonksiyonu,

$N(A_B)$ = A etiketli örneklerden B olarak sınıflandırılanların sayısı ve

$N(B_A)$ = B etiketli örneklerden A olarak sınıflandırılanların sayısıdır.

Banka kredileri ve finansal başarısızlık tahmini, öznitelik maliyet bağımlılığına örnek alanlardır. Eğer fayda tablosu, verilen kredilerin nakit akımlarını yansıtacak şekilde düzenlenirse, her kredi başvurusu için farklı bir fayda tablosu söz konusu olacaktır. Bu maliyetler talep edilen kredinin büyüklüğüne bağlı olarak değişmektedir (Şekil 2.4).

		TAHMİN EDİLEN SINIF	
		Ödeme	Temerrüd
GERÇEK SINIF	Ödeme	$rxK(i)$	$-rxK(i)$
	Temerrüd	$-K(i)$	0
r = Kredi Faiz Oranı $K(i)$ = i Müşterisinin Kredi Talep Büyüklüğü			

Şekil 2.4: Banka Kredileri ve Değişken Fayda-Maliyet Tablosu

Banka kredileri tanım alanında, r bankanın talep ettiği kredi faiz oranını, $K(i)$ ise müşterinin talep ettiği kredi miktarını ifade etmektedir. Bu tabloya göre, eğer müşteri ödeme yapacak olduğu halde temerrüt olarak tahmin edildiyse bankanın kaybı sadece krediden elde edeceği faiz getirisi kadardır. Diğer taraftan, eğer müşteri temerrüt edecek olduğu halde “ödeme yapar” olarak tahmin edilirse bankanın kaybı verilen kredinin tamamıdır⁴⁷.

⁴⁷ Burada varsayım kredinin hiçbir şekilde geri dönmeyeceğidir. Eğer, temerrüdden geri kazanım varsa, yani herhangi bir şekilde (tasfiye, icra takibi vb yoluyla) banka verdiği kredinin bir bölümünü geri alırsa bu varsayım geçerliliğini yitirir. Bu durumda ortalama temerrüd olasılıkları ve temerrüdden geri kazanım oranlarının fayda tablosunun inşasında göz önüne alınması gerekir.

Banka kredileri tanım alanında, banka yetkilileri büyük miktartlı kredileri daha dikkatli takip edecektir. Çünkü büyük miktarlarda kredilerin getirisi ya da kaybı büyüktür. Dolayısıyla, bankanın her kredi başvurusundan elde edeceği fayda talep edilen kredinin miktarına bağımlıdır (Güvenir ve İkizler, 2003).

Firma başarısızlığı tanım alanında, fayda maliyet tablosunun elemanları kullanıcının sınıflara atadığı maliyetlere ve hedef sınıfa göre değişik değerler alabilmektedir⁴⁸. Tahvil yatırımcısı açısından, BAŞARILI olacağı tahmin edilerek, borç verilen firmanın BAŞARISIZ olması durumunda yatırımcının tüm yatırımını kaybetme riski söz konusu olabilir. Bu durumda hatalı sınıflandırma maliyeti, en azından, yatırımcının firmaya verdiği borç kadar olacaktır⁴⁹. Başarısızlık tahmini firmanın kredi talebi üzerine yapılmışsa analiz banka açısından yapılacaktır. Bu bağlamda, yukarıdaki şekliyle banka kredileri tanım alanında geçerli olan fayda-maliyet tablosu kullanılabilir. Eğer, başarısızlık tahmini politika yapıcılar tarafından yapılıyorsa bu durumda gerçek sınıf etiketinin BAŞARISIZ iken, firmanın BAŞARILI ya da BAŞARISIZ olarak tahmin edilmesi durumunda farklı bir fayda-maliyet profili oluşmaktadır. Başarısız sınıfı azınlık sınıfıysa hatalı tahmin durumunda hatalı sınıflandırma maliyeti yüksek olacağı için, politika yapıcıların hedef sınıfı BAŞARISIZ sınıfı olacaktır. Yani, başarısız firmaların en yüksek doğrulukla tahmin edilmesi arzu edilir. Bu bağlamda, gerçek sınıf etiketi BAŞARISIZ olan firma BAŞARILI olarak tahmin ediliyorsa ekonomi açısından en kötü durum firmanın iflas etmesi olabilir⁵⁰. Bu bağlamda, firmayı gerçek durumda BAŞARISIZ iken BAŞARILI olarak tahmin etmenin hatalı sınıflandırma maliyeti firmanın aktif/varlık büyüklüğü (-A(i)) olarak ele alınabilir. Firma sınıf etiketinin doğru tahmin edilmesi durumundaysa genel ekonomi açısından en azından firma büyüklüğü kadar fayda sağlar⁵¹.

⁴⁸ İflas ya da finansal baskı tanım alanında fayda-maliyet analizi, menfaat grubunun kayıp-kazanç profiline bağlı olarak değişiklik gösterir. Örneğin, maliyet, bir banka için verilen kredinin büyüklüğü iken, tahvil yatırımcısı için veri anda gelecekteki faiz ödemeleri ve anaparanın net bugünkü değeri, çalışanlar için gelecekte elde edecekleri maaş ve ücret gelirinin net bugünkü değeri ve reel ekonomi için ise firmanın büyüklüğü ve kayıp üretim kapasitesi vb. değişik menfaat gruplarının başarısızlık algılamasını etkileyen etmenlerdir.

⁴⁹ Aslında problem bundan daha karmaşıktır. Firma acze düştüğü zaman borçların tahsilinin gecikmesinden doğan zaman değeri kaybı, tasfiye durumundaki hukuksal maliyetler, vb. faktörler maliyet rakamının belirlenmesini güçleştirmektedir.

⁵⁰ Bunun ekonomiye maliyeti en az firmanın değeri kadardır.

⁵¹ Firmanın hayatta kalması doğrudan A(i) kadar bir fayda getiriyor olmasına rağmen, başarısızlık durumunda (örneğin firmanın iflası) ekonomiye bu büyüklükte bir maliyet getireceği için, A(i) kadar bir fayda kullanılabilir.

Ekonomi açısından bakıldığında firmanın iflası ya da faaliyetini sürdürmesi durumunda, farklı fayda-maliyet profilleri oluşturulabilir. Örneğin, politika yapıcılar istihdam kaybı, katma değer kaybı ve benzeri kriterleri tek tek ya da bir arada kullanabilirler. Eğer sınıflandırma algoritması, ikiden fazla değer alan bir finansal başarısızlık ya da baskı sürecini tahmin ediyorsa, bu bağlamda fayda-maliyet profili daha karmaşık olacaktır. Örneğin, finansal baskı sürecinin başında bir firmanın hatalı sınıflandırılma maliyeti, görece olarak, iflas eden bir firmanın hatalı sınıflandırma maliyetinden yüksektir.

		TAHMİN EDİLEN SINIF	
		BAŞARILI	BAŞARISIZ
GERÇEK SINIF	BAŞARILI	0	0
	BAŞARISIZ	-A(i)	A(i)
A(i) = Firma Varlıklarının Büyüklüğü Ekonomi politika yapıcıları açısından			

Şekil 2.5: Firma Başarısızlığı ve Değişken Fayda-Maliyet Tablosu

2.3.3.3. Öğrenmede Fayda Maksimizasyonu

Faydaya duyarlı sınıflandırma algoritmaları, klasik maliyete duyarlı algoritmalarından farklı olarak, hatalı tahminleri cezalandırmak yerine doğru sınıflandırmayı ödüllendirmeye dayalı bir öğrenme yapısı sunmaktadır⁵². Faydaya duyarlı sınıflandırmada öğrenme problemi, doğruluk maksimizasyonu amaç fonksiyonu yerine, fayda maksimizasyonu amaç fonksiyonuyla birlikte çözülmektedir⁵³. Doğru sınıflandırmanın faydası sabit bir değer ya da bir fonksiyon olabilir (İkizler, 2002; Güvenir, 2003).

Elkan (2001), faydaya duyarlı öğrenmede doğru sınıflandırmanın getireceği fayda kazançlarının önemini, faydaya dayalı hesaplamaların karşılaştırılabileceği doğal ortak bir taban bulunduğunu bu bağlamda hatadan kaçınmanın da kolaylaştığını ifade ederek vurgulamıştır. Fayda kavramının

⁵² Fayda maksimizasyonuna dayalı sınıflandırma için İkizler (2002), ayrıntılı bir literatür bilgisi ve kuramsal altyapı sunmaktadır.

⁵³ Sınıflandırma algoritması sınıflandırma yaparken fayda maksimizasyonu amaç fonksiyonu altında çalışacaktır. Bu matematiksel olarak

MAX Fayda k.a. Fayda Tablosu biçiminde ifade edilebilir.

gerçek hayat problemlerine uygulanması daha kolaydır. Örneğin, bir yatırımın net getirisinin fayda cinsinden ölçülmesi daha anlamlıdır: eğer tahmin yatırımcı açısından kârlı görülüyorsa faydası da pozitiftir. Aksi durumda fayda negatiftir. Negatif fayda, bilinen hatalı sınıflandırma maliyetine eşittir. Fayda bilgisi, öğrenme algoritmasına fayda tabloları aracılığıyla iletilir.

Fayda bilgisi, maliyet bilgisinde de olduğu gibi statik veya özneliğe bağımlı, yani dinamik olabilir. Statik fayda bilgisi Şekil 2.6'da görüldüğü gibi, sabit doğru sınıflandırma fayda bilgilerinden oluşmaktadır. Sabit fayda bilgisi, tüm örneklem için homojendir.

		TAHMİN EDİLEN SINIF	
		A	B
GERÇEK SINIF	A	B(A_A)	0
	B	0	B(B_B)
B(.) = Sabit			

Şekil 2.6: Sabit Fayda Tablosu

M sınıflandırma modelinin, sabit fayda bilgisiyle, S örnek setini D tanım alanında doğru sınıflandırmanın toplam faydası ($Fayda_D^{\bar{s}}(M,S)$):

$$Fayda_D^{\bar{s}}(M,S) \equiv \sum_1^{N(A_A)} B(A_A) + \sum_1^{N(B_B)} B(B_B) \quad (2.5)$$

olarak tanımlanmaktadır.

$B(A_A)$ = A etiketli örneklerin doğru sınıflandırılma faydası,
 $B(B_B)$ = B etiketli örneklerin doğru sınıflandırılma faydası,
 $N(A_A)$ = A etiketli doğru sınıflandırılanların sayısı ve
 $N(B_B)$ = B etiketli örneklerden doğru sınıflandırılanların sayısıdır.

Özniteliğe dayalı fayda bilgisi dinamik ve her örnek için özel ve tektir. Şekil 2.7. deki fayda tablosu, her örnek için belirli bir değişkenin fonksiyonu olarak ifade edilebilir.

		TAHMİN EDİLEN SINIF	
		A	B
GERÇEK SINIF	A	$B[.](A_A)$	0
	B	0	$B[.](B_B)$
$B. = \text{Fayda Fonksiyonu}$			

Şekil 2.7: Değişken Fayda Tablosu

Dinamik fayda bilgisiyle, doğru sınıflandırmanın toplam faydası ($\text{Fayda}_D^{\tilde{d}}(M,S)$):

$$\text{Fayda}_D^{\tilde{d}}(M,S) \equiv \sum_{1}^{N(A_A)} B[.](A_A) + \sum_{1}^{N(B_B)} B[.](B_B) \quad (2.6)$$

olarak tanımlanmaktadır.

$B[.](A_A)$ = A etiketli örneklerin doğru sınıflandırılma fayda fonksiyonu,

$B[.](B_B)$ = B etiketli örneklerin doğru sınıflandırılma fayda fonksiyonu,

$N(A_A)$ = A etiketli doğru sınıflandırılanların sayısı ve

$N(B_B)$ = B etiketli örneklerden doğru sınıflandırılanların sayısıdır.

2.3.3.4. Fayda Doğruluğu

Maliyet ve faydaya duyarlı algoritmaların tahmin doğruluğu, standart sınıflandırma doğruluğu metriğiyle ölçülemez. Çünkü, bu sınıflandırma algoritmalarının objektif fonksiyonları, standart algoritmaların maksimum sayıda örneği doğru sınıflandırma amacından farklı olarak, sınıflandırma hatalarını minimize etmeyi ya da toplam faydayı maksimize etmeyi amaçlamaktadır. Bu tür sınıflandırma problemlerinde sınıflandırma doğruluğu fayda veya maliyet doğruluğu kavramıyla temsil edilir.

Fayda doğruluğu, klasik sınıflandırma doğruluğunun daha genel bir formudur. Bu metrikler, iki sınıflandırıcının görelî fayda ya da maliyetlerinin karşılaştırmasını kolaylaştırır. Aynı zamanda, veri alanda algoritma verimliliğinin bir ölçütüdür.

Fayda doğruluğu, veri sınıflandırma probleminde elde edilen faydanın bu alan içinde elde edilebilecek azami faydaya görelî büyüklüğüdür. M sınıflandırma modelinin S örnek setini D tanım alanındaki fayda doğruluğu, doğru sınıflandırmadan elde edilen faydanın toplam elde edilebilir faydaya oranıdır (Güvenir ve İkizler, 2003). Fayda doğruluğu:

$$\text{FaydaDoğruluğu}_D(M, S) \equiv \frac{\text{Fayda}_D(M, S) - \min B_D(S)}{\max B_D(S) - \min B_D(S)} \quad (2.7)$$

olarak tanımlanmaktadır. $\text{Fayda}_D(M, S)$, M modeli tarafından D tanım alanında elde edilen fayda, $\min B_D$ ve $\max B_D$ sırayla D tanım alanında elde edilen en düşük ve yüksek faydaya karşılık gelmektedir⁵⁴. Bu metrik, elde edilen faydayı [0,1] aralığına taşır. Fayda tablosunun köşegen elemanları 1 ve diğer elemanları 0 ise tüm sınıfların önem ağırlıkları eşit ve hatalı sınıflandırmanın maliyeti sıfır olur. Bu özgül durumda, fayda doğruluğu standart hata-tabanlı sınıflandırıcıların sınıflandırma doğruluğuna eşit olur.

2.3.3.5. Karma Optimizasyon

Karma optimizasyonda maliyet ve fayda bilgileri aynı anda kullanılarak sınıflandırma yapılır. Öğrenme algoritması, veri fayda-maliyet bilgisi ışığında, net fayda maksimizasyon objektif fonksiyonunu sağlamak üzere çalışır. Bu bağlamda, fayda ve maliyet tabloları hem fayda hem de maliyet bilgisini aynı anda algoritmaya iletecek bir formatta tasarlanmıştır. Sabit fayda-maliyet tablosunun genel formatı Şekil 2.8’de sunulmaktadır.

⁵⁴ $\min B_D$ tüm örneklerin hatalı tahmin edilmesiyle elde edilen fayda (0) iken, $\max B_D$, tüm örneklerin doğru tahmin edilmesiyle elde edilen faydadır (1).

		TAHMİN EDİLEN SINIF	
		A	B
GERÇEK SINIF	A	B(A_A)	M(A_B)
	B	M(B_A)	B(B_B)
M(.) = Sabit B(.) = Sabit			

Şekil 2.8: Sabit Fayda-Maliyet Tablosu

Karma tablolar ya da fayda-maliyet tabloları, maliyet ve fayda tabloları birleştirilerek oluşturulur. Bu tablolar oluşturulurken, sabit ve değişken maliyet-fayda tablolarının tek tek bir arada ele alınması gerekmektedir⁵⁵.

Sabit maliyet ve fayda bilgilerinin kullanılmasıyla elde edilen sabit net fayda ($\text{NetFayda}_D^{\bar{d}}(M, S)$):

$$\text{NetFayda}_D^{\bar{d}}(M, S) \equiv \text{Fayda}_D^{\bar{s}}(M, S) + \text{Maliyet}_D^{\bar{s}}(M, S) \quad (2.8)$$

olarak tanımlanmaktadır. Sabit fayda ($\text{Fayda}_D^{\bar{d}}(M, S)$) ve sabit maliyet ($\text{Maliyet}_D^{\bar{d}}(M, S)$) bilgileri yukarıda tanımlanan bilgilerdir.

		TAHMİN EDİLEN SINIF	
		A	B
GERÇEK SINIF	A	B[.](A_A)	F[.](A_B)
	B	F[.](B_A)	B[.](B_B)
F. = Maliyet Fonksiyonu B[.](ii) = Fayda Fonksiyonu			

Şekil 2.9: Değişken Fayda-Maliyet Tablosu

Benzer şekilde, özniteliğe bağımlı yani dinamik fayda ve maliyet bilgilerinin birleştirilmesiyle, değişken fayda-maliyet tablosu oluşturulur.

⁵⁵ Sabit maliyet tabloları, sabit fayda tablolarıyla, değişken maliyet tabloları değişken fayda tablolarıyla toplanmalıdır.

Değişken fayda-maliyet bilgisiyle elde edilen değişken net fayda ($\text{NetFayda}_D^{\tilde{d}}(M, S)$):

$$\text{NetFayda}_D^{\tilde{d}}(M, S) \equiv \text{Fayda}_D^{\tilde{d}}(M, S) + \text{Maliyet}_D^{\tilde{d}}(M, S) \quad (2.9)$$

olarak tanımlanmaktadır. Değişken fayda ($\text{Fayda}_D^{\tilde{d}}(M, S)$) ve değişken maliyet ($\text{Maliyet}_D^{\tilde{d}}(M, S)$) bilgileri yukarıda tanımlanan bilgilerdir.

2.4. Kurallar ve Kural İlginçliği

Öğrenme algoritması aracılığıyla doğrudan öğrenilen, karar ağaçlarından doğrudan okunan ya da karar ağaçlarının budanmasıyla öğrenilen kuralların, alan uzmanı tarafından kullanılabilir olması önemlidir. Bu bağlamda kuralların sağlaması gereken özellikler aşağıda sunulmaktadır. Kurallar:

- **Basit olmalıdır:** Öğrenilen kurallar veri madenciliği alanı dışında konunun uzmanı olmayanlar tarafından da anlaşılabilir olmalıdır (Frank ve Witten, 2000).
- **Anlaşılır olmalıdır:** Kurallar karmaşık olmamalıdır. Karmaşık kurallar alan uzmanlarınca bile kolaylıkla anlaşılabilir.
- **Yapay ilişkilere işaret etmemelidir:** Olasılıkla öğrenme algoritması sadece verinin yapısından kaynaklanan kuramsal olarak anlamsız kurallar da öğrenecektir. Anlamsız kuralların ayıklanmasıyla bu soruna bir ölçüde çözüm getirilebilir. İdeal olarak, algoritmanın olabildiğince çok (az) anlamlı (anlamsız) kurallar öğrenmesi istenir.
- **Genellenebilir olmalıdır:** Kurallar sadece eğitim setine değil test setine yönelik de yüksek doğrulukla sınıflandırmayı olanaklı kılan kurallar olmalıdır. Özgül kuralların, verinin (eğitim setinin) özelleşmesi ya da aşırı öğrenilmesiyle sadece veriye yönelik kurallar öğrenilmesi riski vardır.

- **Kurallar ilginç olmalıdır:** Kural ilginçliği, tanım alanı uzmanı tarafından ilk bakışta görülemeyen, fakat tanım alanında anlamlı olan kuralların, algoritma tarafından öğrenilmesidir (Freitas, 1999).

ÜÇÜNCÜ BÖLÜM

WEKA PLATFORMU VE BCFP

3.1. Makina Öğrenmesi ve Firma Başarısızlığı Analizi

Firma başarısızlığı tahmin literatürü başlığı altında, önceki çalışmalarda kullanılan bazı modeller kısaca açıklanmıştır. İkinci Bölümde açıklanan mantıksal makina öğrenmesi yöntemleri ve modellemeyle ilgili güçlükler, analizde kullanılan algoritmaların seçiminde mantıksal yöntemleri öne çıkarmıştır. Firma başarısızlığının tahmini, doğası gereği bir sınıflandırma⁵⁶ problemi olduğundan, sınıflandırma yapan algoritmaların kullanılması önem kazanmaktadır.

Mantıksal makina öğrenmesi algoritmaları, genellikle basit yapıda ve anlaşılır kurallar ya da karar ağaçları öğrenmeleri, kısıtlayıcı varsayımlara karşı sağlam ve duyarsız olmaları, karmaşık problemlerin analizindeki yüksek performansları ve öznitelik önseçimine gerek duymadan sınıflandırma yapabilmelerinden dolayı tercih edilmektedir. Ayrıca, özellikle alan uzmanı olmayanlar tarafından anlaşılabilir basit kurallar öğrenebilmeleri, sınıflandırma yapabilen mantıksal öğrenme algoritmalarını kullanmamızda önemli bir etken olmuştur.

Mantıksal algoritmalar, İkinci Bölümde de açıklandığı gibi, karar ağaçları ve doğrudan kural öğrenen algoritmalar ve bu iki yöntemin melezleri olan parçasal karar ağaçlarıdır. Kural öğrenen algoritmalar ve karar ağaçları, basitten ve karmaşığa doğru bir yapıda ele alınabilir. Algoritmanın tasarımı, basit veya karmaşık olup olmayacağını belirleyen asıl ölçüttür. Bununla birlikte, alan uzmanlarının çoğunluğunun algoritmanın tasarım ayrıntılarını bilmesi beklenemeyeceğinden, öğrenilen kuralın karmaşıklığı, kullanıcıya,

⁵⁶ İkinci Bölüm'de de belirtildiği gibi öğrenme ve sınıflandırma kavramları aynı anlamda kullanılmıştır.

algoritmanın basitliği veya karmaşıklığıyla ilgili ipucu verebilir. Burada dikkate alınması gereken ölçüt, basit ve genellenebilir sonuçlar üreten algoritmaların tercih edilir olmalarıdır.

Çalışmamızda, veri madenciliği ve makine öğrenmesi araştırmalarında yaygın olarak kullanılan WEKA platformu⁵⁷ algoritmalarıyla, faydaya duyarlı sınıflandırma için geliştirilen BCFP algoritmasına başvurulmuştur. Her iki grup algoritmanın seçilmesinde de göz önüne alınan ölçüt öncelikle mantıksal algoritmalar olmalarıdır. Ayrıca, BCFP algoritmasının kullanılmasının amacı, faydaya duyarlı öğrenmenin başarısızlık analizine katkısının sorgulanmasıdır.

Bu bölümde, kullanılan mantıksal öğrenme algoritmaları, ampirik analizler ve sonuçları açıklanacaktır.

3.2. WEKA Platformu

WEKA platformunda, sınıflandırma, kümeleme, eşleme, öznitelik seçimi ve veri görüntüleme paketleri bulunmaktadır⁵⁸. Firma başarısızlığının tahmini, bir sınıflandırma problemi olduğu için WEKA sınıflandırma paketindeki sınıflandırma algoritmaları kullanılmıştır. WEKA sınıflandırma algoritmaları tek tek (çıplak) ya da birleştirilmiş (giydirilmiş-meta) olarak da kullanılabilirler.

Öğrenme algoritmalarının sınıflandırmada kullandıkları ölçütler, bu sınıflandırmanın sonuçlarını kullananlar için önemlidir. Algoritma sınıflandırma yaparken herhangi bir şekilde kullanıcıya anlamsız gelecek ölçütleri kullanıyorsa sınıflandırma anlamsız olacaktır. Bu bağlamda, kullanıcılara ya da alan uzmanlarına anlamlı gelecek sınıflandırma kurallarının öğrenilmesi makina öğrenmesi problemlerinin önemli bir bölümünün üzerinde yoğunlaştığı bir konu olagelmıştır. WEKA platformundaki

⁵⁷ WEKA platformu Java tabanlı bir makina öğrenmesi paketi olup, Yeni Zelanda Waikato Üniversitesi'nde geliştirilmiştir. WEKA platformu'nun en önemli özelliklerinden birisi açık kaynak kodlu olmasıdır. Bunun yanında, WEKA'nın bugüne kadar geliştirilmiş tüm sürümleri ve yardımcı malzeme WEKA'nın sitesinden ücretsiz olarak edinilebilmektedir (www.cs.waikato.ac.nz/~ml/weka/). WEKA ticari olarak geliştirilen diğer bir çok platformdan daha fazla öğrenme algoritmasını bünyesinde barındırmaktadır. WEKA platformunun çalışmamızdaki analizlerde kullanılmak amacı ile seçilmiş olmasının nedenleri arasında bilgi işlem maliyetinin yanında bu platformun diğer ticari yapay zeka platformlarına göre gelişim açısından daha dinamik bir yapı sergilemesi de önemli bir etken olmuştur.

⁵⁸ Bu paketlerdeki algoritmalar, araştırmaya konu olan problemin özelliğine bağlı olarak sınıflandırma, kümeleme ya da eşleme öğrenme algoritması olarak adlandırılırlar.

öğrenme algoritmalarının bir bölümü sadece sınıflandırma yapmakta bununla birlikte sınıflandırmanın hangi ölçütlere göre yapıldığı belirtilmemektedir. Diğer algoritmalar sınıflandırmanın hangi ölçütlere göre yapıldığını gösteren bir ya da daha fazla sayıda kural ya da karar ağacı öğrenmektedir⁵⁹. WEKA platformundaki öğrenme algoritmaları Tablo 3.1’de gösterilmektedir.

TABLO 3.1. WEKA PLATFORMU ÖĞRENME ALGORİTMALARI

Öğrenme Algoritması	WEKA Platformu Paket Adresi	Kural Üretiyor mu?	Analizde Kullanılıyor mu?
Majority / Average Predictor	weka.classifiers.ZeroR	Evet	Evet
1R	weka.classifiers.OneR	Evet	Evet
Naive Bayes	weka.classifiers.NaiveBayes	Hayır	Hayır
DecisionTable	weka.classifiers.DecisionTable	Evet	Evet
Instance Based Learner	weka.classifiers.IBk	Hayır	Hayır
C4.5	weka.classifiers.j48.J48	Evet	Evet
PART Rule Learner	weka.classifiers.j48.PART	Evet	Evet
Support Vector Machine	weka.classifiers.SMO	Hayır	Hayır
Linear Regression ^a	weka.classifiers.LinearRegression	Evet (Regresyon Denklemi)	Hayır
M5’ Model Tree Learner ^a	weka.classifiers.m5.M5Prime	Evet (Regresyon Denklemi)	Hayır
Locally Weighted Regression ^a	weka.classifiers.LWR	Evet (Regresyon Denklemi)	Hayır
One - level Decision Trees	weka.classifiers.DecisionStump	Evet	Evet

^a Bu algoritmalar sınıf değerlerinin sayısal olmasını gerektirmektedir. Çalışmamızda sınıf özneliğinin aldığı değerler nominal olduğundan bu algoritmalar analizde kullanılmamıştır.
Kaynak: Frank ve Witten, 2000

Analizlerde kullanılmak üzere, kural öğrenen altı WEKA platformu sınıflandırma algoritması seçilmiştir. Bunlar:

- ZeroR (weka.classifiers.ZeroR)
- OneR (weka.classifiers.OneR)
- DecisionTable (weka.classifiers.DecisionTable)
- DecisionTree⁶⁰ (weka.classifiers.j48.J48)
- PARTDecisionTree⁶¹ (weka.classifiers.j48.PART)
- DecisionStump (weka.classifiers.DecisionStump)

algoritmalarıdır.

⁵⁹ Burada ölçütler, öğrenilen kurallarda belirtilen özneliliklerin aldıkları değerlerin aralıkları ve/veya eşik noktalarıdır.

⁶⁰ WEKA platformunda karar ağacı öğrenmesi için kullanılan algoritma Ross Quinlan’ın “Programming for Machine Learning” isimli eserinde tanımlanan C4.5 algoritmasıdır (WEKA altında kullanılan adı ile j48.J48). Her ne kadar C5.0 adı altında ticari bir sürümü bulunsun da C4.5 algoritması da oldukça sağlam/güvenilir bir karar ağacı öğrenme yöntemidir.

⁶¹ Parçasal karar ağaçları Frank ve Witten (1998) tarafından geliştirilmiştir.

3.2.1. WEKA Platformu Öğrenme Algoritmaları

Bu çalışmada kullanılan öğrenme algoritmalarının seçilmesinde, kural veya karar ağacı öğrenebilen sınıflandırma algoritmaları olmaları temel ölçüt olmuştur. Aşağıda, WEKA platformundaki kapsama, karar ağacı ve parçasal karar ağacı sınıflandırma algoritmaları tanıtılmıştır.

3.2.1.1. Kapsama Algoritmaları

WEKA platformundaki, sınıflandırma yapan kapsama ya da kural öğrenme algoritmaları, basitten karmaşığa ZeroR, OneR ve DecisionTable olarak sıralanabilir. Bu bölümde, WEKA kapsama algoritmaları sunulmaktadır.

3.2.1.1.1. ZeroR

ZeroR kural öğrenen algoritmaların en ilkelidir ve ZeroR'ın kullanılmasının amacı diğer algoritmaların performans sonuçlarının karşılaştırılacağı bir sınıflandırma doğruluğu ve sınıf dağılımı bulmaktır: ZeroR'dan daha karmaşık algoritmaların daha iyi sınıflandırma doğruluğu elde etmesi beklenir. Bazı durumlarda daha karmaşık algoritmalar ZeroR'dan daha düşük performans gösterebilmektedir. Bu durum ciddi özelleşmeye işaret etmektedir. ZeroR özünde sadece çoğunluk sınıf dağılımını tahmin eder ve bu dağılım bir kural olarak kabul edilir. ZeroR'da sayısal sınıf öznitelikleri için aritmetik ortalama, nominal sınıf öznitelikleri içinse mod sınıfın aldığı değer olarak tahmin edilir. ZeroR bunun dışında herhangi bir kural öğrenmez. Bu sınıflandırmanın yorumu şu şekilde yapılabilir: ZeroR yeni herhangi bir örnek sunulduğunda özgül örneği, çoğunluk sınıfına ait olarak sınıflandırır⁶². Örnek ZeroR çıktısı aşağıdakine benzer bir kuraldır⁶³.

Sınıf: **BAŞARILI**

Sınıflandırma Doğruluğu: % 50.4425

Başarısız olanlar dahil, bütün örneklerin sınıf etiketi BAŞARILI olarak

⁶² Diğer algoritmalar da zaman zaman ZeroR'da olduğu gibi sadece çoğunluk sınıfını tahmin ederler. Bu durumda, öğrenilen kuralın yorumu ZeroR'ın yorumuyla aynı olacaktır.

⁶³ ZeroR ve diğer algoritmalarındaki örneklerin oluşturulmasında, sinama amacıyla çalıştırılan küçük ölçekli bir örnekleme elde edilen kurallar ve ağaçlar kullanılmıştır.

tahmin edilse bile algoritma sınıflandırmada % 50.4425 başarı sağlamıştır. Diğer bir deyimle, BAŞARILI olarak etiketlenmiş firmaların toplam firma sayısına oranı % 50.4425'tir. Bu bağlamda, örnekteki sınıflandırma doğruluğu, örneklerin sınıf etiketlerine göre örneklem içindeki dağılımına eşittir.

3.2.1.1.2. OneR

OneR algoritması⁶⁴, tek bir özniteliği sınamak üzere kural setleri biçiminde ifade edilen tek seviyeli karar ağaçları oluşturur. Basit ve düşük maliyetli bir yöntemdir ve genellikle “iyi” kurallar öğrenir. OneR yönteminde, birbirinden bağımsız özniteliklerden yalnız ve yalnız tek bir öznitelik sınanır ve bu öznitelik dallandırılır. Her dal özniteliğin farklı bir değerine karşılık gelmektedir. Eğitim setinde sınıf dağılımında en fazla bulunan sınıf kullanılır ve çoğunluk sınıf etiketini taşımayan örnekler hatalı örnekler olarak belirlenir. Her özniteliğin aldığı her değer için farklı bir kural seti oluşturulur. Daha sonra her özniteliğin oluşturduğu kural setinin hata oranları değerlendirilir ve en iyi kural seti seçilir. Bu kurallar, eğitim setindeki tüm örneklerin sınıflandırılması için öğrenilir ve kullanılır.

OneR ile modellemede eksik değerler ve sayısal özniteliklerin her ikisi ayrı ayrı ya da bir arada kullanılabilir. OneR algoritmasında eksik değer, başka bir öznitelik değeri olarak kabul edilir. Sayısal öznitelikler basit bir kesiklileştirme yöntemi kullanılarak nominale dönüştürülür; bu durumda çözümleme büyük sayılarda alt kategorilerle sonuçlanmaktadır. OneR yöntemi, bu durumda doğal olarak, örneklerin içinde bulundukları bölüntünün sınıf etiketine sahip olmaları için, veri setini çok sayıda alt kategoriye bölen öznitelikleri seçmeye çalışacaktır. Bu nedenle, bazı durumlarda özelleşme/aşırı öğrenme problemi ile karşı karşıya kalınmaktadır.

OneR algoritmasının öğrendiği kurallar tek bir özniteliğe dayalı olduğundan, sayısal (sürekli) öznitelikler için bu kurallar olasılıkla birden fazla eşik noktası içerecektir. Bu durumda, OneR daha genel şekilde tek bir

⁶⁴ Holte (1993).

özniteliğe ve bir ya da daha fazla eşik noktasına dayalı bir ya da daha fazla kural öğrenmektedir denilebilir⁶⁵.

Öğrenilen kuralın yapısı, kural öğrenen algoritmaların öğrendiği ya da karar ağaçlarından doğrudan okunan kurallara benzer; OneR algoritmasının öğrendiği kurallar Mantıksal VElenmiştir. Aşağıda, OneR'in öğrendiği kural ve nasıl yorumlandığına bir örnek sunulmaktadır.

DonenVarliklarAktifToplamiOrani:
< 0.6148 -> **BAŞARISIZ**
< 0.6440 -> **BAŞARILI**
< 0.6754 -> **BAŞARISIZ**
< 0.7145 -> **BAŞARILI**
< 0.7599 -> **BAŞARISIZ**
< 0.7746 -> **BAŞARILI**
< 0.8121 -> **BAŞARISIZ**
>= 0.8121 -> **BAŞARILI**
Sınıflandırma Doğruluğu: % 62.8319

DonenVarliklarAktifToplamiOrani özniteliğinin değer aralıklarına göre örnekler % 62.8319 sınıflandırmanın doğruluğuyla BAŞARILI veya BAŞARISIZ olarak sınıflandırılmıştır. Bu kuralları sağlayan yeni bir örneğin, % 62.8319 olasılıkla iki sınıf etiketinden birini alması beklenir. Örneğin, DonenVarliklarAktifToplamiOrani = 0.5325 olan bir örnek, BAŞARISIZ olarak sınıflandırılacaktır. Diğer taraftan, oran değeri [0.6754, 0.7145) aralığında olan bir başka örnek BAŞARILI olarak sınıflandırılmaktadır.

3.2.1.1.3. DecisionTable

Makina öğrenmesi çıktısının ifade edilmesinin en basit ve temel yolu çıktıyı girdiye benzer bir şekilde, yani karar tablosu biçiminde sunmaktır. Öğrenmede herhangi bir özniteliğin etkisi yoksa, anlamsız olduğu kabul edilen öznitelik değeri çıkarılmış tablo⁶⁶, karar tablosunun çıktısıdır. Algoritma, anlamsız özniteliği otomatik olarak dışlayacak, yani öğrenme sürecinde öznitelik seçimini kendisi yapacaktır. Karar tabloları, öğrenmede eksik değerli öznitelikleri kullanabilmektedir.

⁶⁵ OneR kuralının birden fazla eşik noktası öğrenmesi, bir ölçüde özelleşmeye işaret etmektedir.

⁶⁶ Söz konusu çıktı tablo, tüm veri setini tanımlayan bir kural setine dönüştürülebilir.

Karar tablolarının öğrenme sürecinde en önemli etken, algoritmaya özgü öznitelik seçimidir. Özünde karar tablosu algoritmalarının amacı kararın verilmesinde hangi özniteliğin seçileceğini yani anlamlı olduğunu bulmaktır. Bu seçim genellikle farklı öznitelik setleriyle oluşturulan sınıflandırıcıların çapraz doğrulama⁶⁷ performanslarının karşılaştırılmasıyla yapılır. En yüksek sınıflandırma performansını sağlayan öznitelik alt seti, kural öğrenmekte kullanılır. Çapraz doğrulama, bu tür sınıflandırıcılar için kullanımı kolay ve düşük maliyetli bir yöntemdir. Karar tablosundan çapraz doğrulama hatasının elde edilmesi tablodaki ölçütleri karşılayan sınıf sayısının değiştirilmesidir, çünkü tablonun yapısı örnek ekleme/çıkarma ile değişmez⁶⁸.

Karar tabloları öznitelik seçiminde sarma yaklaşımı kullanılmaktadır. Karar tabloları, genellikle, sınama verisi tablodaki herhangi bir ölçütü karşılamadığında, örneğe çoğunluk sınıfının etiketini verir. Bununla birlikte, herhangi bir örneğe en yakın örneğin sınıf etiketinin verilmesi de olanaklıdır.

WEKA platformundaki karar tablosu algoritması DecisionTable algoritmasıdır. Bir kapsama algoritması olan DecisionTable, bir ya da daha fazla kural öğrenir. Bunlar, İkinci Bölümde tanımlanan mantıksal VElenmiş kurallardır. Aşağıda bir DecisionTable örneği verilmektedir:

TABLO 3.2. ÖRNEK KARAR TABLOSU

KURAL	ÖZNİTELİKLER						SINIF
	CariOran	KısaVadeli Alacaklar AktifToplami Orani	KısaVadeli YabancıKaynaklar PasifToplami Orani	AktifDevirHizi	FaizveVergiOncesi KarDuranVarliklar	NetiSletmeSermayesi OzKaynaklar	
1	'(0.98267-inf)'	'(0.397032-inf)'	'(0.328213-inf)'	'(1.472272-inf)'	'(0.967182-inf)'	'(0.156145-inf)'	BAŞARILI
2	'(0.98267-inf)'	'(-inf-0.397032)'	'(0.328213-inf)'	'(1.472272-inf)'	'(0.967182-inf)'	'(0.156145-inf)'	BAŞARILI
3	'(0.98267-inf)'	'(0.397032-inf)'	'(-inf-0.328213)'	'(1.472272-inf)'	'(0.967182-inf)'	'(0.156145-inf)'	BAŞARISIZ
4	'(0.98267-inf)'	'(0.397032-inf)'	'(0.328213-inf)'	'(1.472272-inf)'	'(-inf-0.967182)'	'(0.156145-inf)'	BAŞARILI
5	'(0.98267-inf)'	'(-inf-0.397032)'	'(0.328213-inf)'	'(-inf-1.472272)'	'(-inf-0.967182)'	'(0.156145-inf)'	BAŞARILI
6	'(-inf-0.98267)'	'(-inf-0.397032)'	'(0.328213-inf)'	'(-inf-1.472272)'	'(0.967182-inf)'	'(-inf-0.156145)'	BAŞARISIZ
7	'(0.98267-inf)'	'(-inf-0.397032)'	'(0.328213-inf)'	'(1.472272-inf)'	'(-inf-0.967182)'	'(0.156145-inf)'	BAŞARISIZ
8	'(-inf-0.98267)'	'(-inf-0.397032)'	'(0.328213-inf)'	'(1.472272-inf)'	'(0.967182-inf)'	'(-inf-0.156145)'	BAŞARISIZ

Yukarıdaki kural tablosu % 70.7965 doğrulukla öğrenilmiştir. Elde edilen kural tablosundan mantıksal VElenmiş kurallar çıkarılabilir. Bu kurallar:

⁶⁷ Çapraz doğrulamada sabit sayıda "kata" (bölüntü) karar verilir (genellikle 10 kullanılır). N sayıda bölüntünün kullanıldığı durumda, veri seti, öğrenme algoritması tarafında rassal olarak birbirine eşit N parçaya bölünür. Her bir bölüntü sırasıyla sınama, kalan N-1 bölüntüyse eğitim için kullanılır. Bu yöntem N'li çapraz doğrulama olarak adlandırılır. Ulaşılan sonuçların ortalama değeri alınarak toplam sınıflandırma sonucu elde edilir.

⁶⁸ Yerel optimumda takılma olasılığı daha az olduğu için, öznitelik uzayı en iyi-en önce yöntemiyle taranır. En iyi-en önce, bir yapay zeka arama yöntemi olup, her aşamada bir sonraki aşamada en iyi parçasal çözümü arar.

Kural 1:

CariOran>0.98267 **VE**
KisaVadeliAlacaklarAktifToplamiOrani>0.397032 **VE**
KisaVadeliYabanciKaynaklarPasifToplamiOrani>0.328213 **VE**
AktifDevirHizi>1.472272 **VE**
FaizVeVergiOncesiKarDuranVarliklar>0.967182 **VE**
NetIsletmeSermayesiOzKaynaklar>0.156145
İSE BAŞARILI

Kural 2:

CariOran>0.98267 **VE**
KisaVadeliAlacaklarAktifToplamiOrani=<0.397032 **VE**
KisaVadeliYabanciKaynaklarPasifToplamiOrani>0.328213 **VE**
AktifDevirHizi>1.472272 **VE**
FaizVeVergiOncesiKarDuranVarliklar>0.967182 **VE**
NetIsletmeSermayesiOzKaynaklar>0.156145
İSE BAŞARILI

Kural 3:

CariOran>0.98267 **VE**
KisaVadeliAlacaklarAktifToplamiOrani>0.397032 **VE**
KisaVadeliYabanciKaynaklarPasifToplamiOrani=<0.328213 **VE**
AktifDevirHizi>1.472272 **VE**
FaizVeVergiOncesiKarDuranVarliklar>0.967182 **VE**
NetIsletmeSermayesiOzKaynaklar>0.156145
İSE BAŞARISIZ

Kural 4:

CariOran>0.98267 **VE**
KisaVadeliAlacaklarAktifToplamiOrani>0.397032 **VE**
KisaVadeliYabanciKaynaklarPasifToplamiOrani>0.328213 **VE**
AktifDevirHizi>1.472272 **VE**
FaizVeVergiOncesiKarDuranVarliklar=<0.967182 **VE**
NetIsletmeSermayesiOzKaynaklar>0.156145
İSE BAŞARILI

Kural 5:

CariOran>0.98267 **VE**
KisaVadeliAlacaklarAktifToplamiOrani=<0.397032 **VE**
KisaVadeliYabanciKaynaklarPasifToplamiOrani>0.328213 **VE**
AktifDevirHizi=<1.472272 **VE**
FaizVeVergiOncesiKarDuranVarliklar=<0.967182 **VE**
NetIsletmeSermayesiOzKaynaklar>0.156145
İSE BAŞARILI

Kural 6:

CariOran= ≤ 0.98267 **VE**
KisaVadeliAlacaklarAktifToplamiOrani= ≤ 0.397032 **VE**
KisaVadeliYabancıKaynaklarPasifToplamiOrani > 0.328213 **VE**
AktifDevirHizi= ≤ 1.472272 **VE**
FaizVeVergiOncesiKarDuranVarliklar > 0.967182 **VE**
NetIsletmeSermayesiOzKaynaklar= ≤ 0.156145
İSE BAŞARISIZ

Kural 7:

CariOran > 0.98267 **VE**
KisaVadeliAlacaklarAktifToplamiOrani= ≤ 0.397032 **VE**
KisaVadeliYabancıKaynaklarPasifToplamiOrani > 0.328213 **VE**
AktifDevirHizi > 1.472272 **VE**
FaizVeVergiOncesiKarDuranVarliklar= ≤ 0.967182 **VE**
NetIsletmeSermayesiOzKaynaklar > 0.156145
İSE BAŞARISIZ

Kural 8:

CariOran= ≤ 0.98267 **VE**
KisaVadeliAlacaklarAktifToplamiOrani= ≤ 0.397032 **VE**
KisaVadeliYabancıKaynaklarPasifToplamiOrani > 0.328213 **VE**
AktifDevirHizi > 1.472272 **VE**
FaizVeVergiOncesiKarDuranVarliklar > 0.967182 **VE**
NetIsletmeSermayesiOzKaynaklar= ≤ 0.156145
İSE BAŞARISIZ

Sınıflandırma Doğruluğu: % 70.7965

olarak oluşturulmuştur. DecisionTable'ın çıktısı olan karar tablosunun ya da karar tablosundan elde edilen kuralların yorumu diğer çok öncüllü kurallarda olduğu gibidir. Birinci kurala göre (Kural 1), öğrenme seti dışında incelenen yeni bir örnek veri yolda eğer koşulların tümünü birden sağlıyorsa % 70.7965 olasılıkla BAŞARILI olarak sınıflandırılması beklenir. Üçüncü kural için, aynı şekilde koşulların hepsini birden sağlayan örneklerin % 70.7965 olasılıkla BAŞARISIZ sınıf etiketi alması beklenir. Örnekten anlaşılacağı gibi, DecisionTable algoritmasıyla öğrenilen kuralların birbirinden bağımsız olarak yorumlanması olanaklıdır. Elbette, herhangi bir örnek sadece bir kuralı sağlayarak sınıflandırılıyorsa da, diğer kuralların koşullarını ihlal etmemesi gerekir. Dolayısıyla, kural bağımsızlığı, o kuralı sağlayan örneklerin tanımlandığı uzayda geçerlidir; böylece karar tabloları için kuralların tam bağımsızlığından söz etmek olanaksızdır.

3.2.1.2. Karar Ağacı Algoritmaları

WEKA platformunda, sınıflandırma yapan karar ağacı algoritmaları, basitten karmaşığa doğru DecisionStump ve j48.J48 yöntemleri olarak sıralanabilir. Aşağıda, WEKA karar ağacı algoritmaları tanıtılmaktadır.

3.2.1.2.1. DecisionStump

İlkel bir karar ağacı olan DecisionStump tek aşamalı bir karar ağacı yani karar kütüğü öğrenir. DecisionStump sınıflandırıcısında kurallar OneR'da olduğu gibi tek bir öznitelik ve en yüksek doğrulukla sınıf bölüntülemesini yapan eşik noktasını tanımlayan mantıksal koşullardan oluşur. DecisionStump'ın OneR'dan farkı, tek öz niteliğe dayalı en büyük bölüntülemeyi gerçekleştiren tek eşik noktasına göre kural öğrenmesidir. Örneğin,

FaizVeVergiOncesiKarDuranVarliklar ≤ 0.9672 **İSE BAŞARISIZ**
FaizVeVergiOncesiKarDuranVarliklar > 0.9672 **İSE BAŞARILI**
FaizVeVergiOncesiKarDuranVarliklar "?" **İSE BAŞARILI**
Sınıflandırma Doğruluğu: % 71.6814

karar kütüğünde, FaizVeVergiOncesiKarDuranVarliklar en büyük bölüntülemeyi gerçekleştiren öznitelik, 0.9672 eşik noktasıdır.

DecisionStump'la yapılan sınıflandırmaya göre sınıf bölüntülemesini en yüksek doğrulukla (% 71.6814) tek başına yapan öznitelik FaizVeVergiÖncesiKarDuranVarliklar oranı olarak belirlenmiştir. Örnekteki kurala göre, yeni bir örneğin faiz ve vergi öncesi kâr/duran varlıklar oranı 0.9672'ten büyük olduğunda söz konusu firma % 71.6814 doğrulukla **BAŞARILI**, aksi durumda **BAŞARISIZ** olarak sınıflandırılmaktadır. FaizVeVergiÖncesiKarDuranVarliklar öz niteliği, eksik değer alan örnekler çoğunlukla **BAŞARILI** olarak sınıflandırılmıştır.

3.2.1.2.2. j48.J48

Quinlan'ın (1993) C4.5 karar ağacı algoritması tabanlı j48.J48 WEKA sınıflandırıcısı, DecisionStump tek düzeyli karar ağacı algoritmasına göre daha karmaşık kurallar öğrenir.

j48.48 karar ağacı algoritmasıyla öğrenilen ağaç örneği:

```
FaizVeVergiOncesiKarDuranVarliklar <= 0.967114
| DonenVarliklarAktifToplamiOrani <= 0.382584: BAŞARISIZ (Kök)
| DonenVarliklarAktifToplamiOrani > 0.382584
| | FaizGiderleriNetSatislarOrani <= 0
| | | DuranVarliklarOzKaynaklarOrani <= 0.513832: BAŞARISIZ (Y-1)
| | | DuranVarliklarOzKaynaklarOrani > 0.513832
| | | | KisaVadeliYabancıKaynaklarYabancıKaynaklarToplamiOrani <=
0.941896: BAŞARILI (Y-2)
| | | | KisaVadeliYabancıKaynaklarYabancıKaynaklarToplamiOrani >
0.941896
| | | | | KisaVadeliYabancıKaynaklarYabancıKaynaklarToplamiOrani <=
0.987936: BAŞARISIZ (Y-3)
| | | | | KisaVadeliYabancıKaynaklarYabancıKaynaklarToplamiOrani >
0.987936
| | | | | FaizVeVergiOncesiKarDuranVarliklar <= 0.072391:
BAŞARISIZ (Y-4)
| | | | | FaizVeVergiOncesiKarDuranVarliklar > 0.072391: BAŞARILI
(Y-5)
| | FaizGiderleriNetSatislarOrani > 0: BAŞARISIZ (Y-6)
FaizVeVergiOncesiKarDuranVarliklar > 0.967114
| StokDevirHizi <= 2.74887
| | BrütSatisKariAktifToplami <= 0.228337: BAŞARILI (Y-7)
| | BrütSatisKariAktifToplami > 0.228337: BAŞARISIZ (Y-8)
| StokDevirHizi > 2.74887
| | VergiOncesiKarNetSatislar <= 0.015203
| | | VergiOncesiKarNetSatislar <= 0.008225: BAŞARILI (Y-9)
| | | VergiOncesiKarNetSatislar > 0.008225: BAŞARISIZ (Y-10)
| | VergiOncesiKarNetSatislar > 0.015203: BAŞARILI (Y-11)
```

olarak sunulmaktadır. İkinci bölümde anlatıldığı gibi karar ağaçlarından kural elde edilmesi olanaklıdır. Yukarıdaki örnekte öğrenilen ağaçtan elde edilen kurallar ve ağaçtaki yaprak karşılıkları:

Kural 1: (Kök)

```
FaizVeVergiOncesiKarDuranVarliklar <= 0.967114 VE
DonenVarliklarAktifToplamiOrani <= 0.382584
İSE BAŞARISIZ
```

Kural 2: (Y-1)

```
FaizVeVergiOncesiKarDuranVarliklar <= 0.967114 VE
DonenVarliklarAktifToplamiOrani > 0.382584 VE
FaizGiderleriNetSatislarOrani <= 0 VE
DuranVarliklarOzKaynaklarOrani <= 0.513832
İSE BAŞARISIZ
```

Kural 3: (Y-2)

FaizVeVergiOncesiKarDuranVarliklar ≤ 0.967114 **VE**
DonenVarliklarAktifToplamiOrani > 0.382584 **VE**
FaizGiderleriNetSatislarOrani ≤ 0 **VE**
DuranVarliklarOzKaynaklarOrani > 0.513832 **VE**
KisaVadeliYabanciKaynaklarYabanciKaynaklarToplamiOrani ≤ 0.941896
İSE BAŞARILI

Kural 4: (Y-3)

FaizVeVergiOncesiKarDuranVarliklar ≤ 0.967114 **VE**
DonenVarliklarAktifToplamiOrani > 0.382584 **VE**
FaizGiderleriNetSatislarOrani ≤ 0 **VE**
DuranVarliklarOzKaynaklarOrani > 0.513832 **VE**
KisaVadeliYabanciKaynaklarYabanciKaynaklarToplamiOrani > 0.941896 **VE**
KisaVadeliYabanciKaynaklarYabanciKaynaklarToplamiOrani ≤ 0.987936
İSE BAŞARISIZ

Kural 5: (Y-4)

FaizVeVergiOncesiKarDuranVarliklar ≤ 0.967114 **VE**
DonenVarliklarAktifToplamiOrani > 0.382584 **VE**
FaizGiderleriNetSatislarOrani ≤ 0 **VE**
DuranVarliklarOzKaynaklarOrani > 0.513832 **VE**
KisaVadeliYabanciKaynaklarYabanciKaynaklarToplamiOrani > 0.941896 **VE**
KisaVadeliYabanciKaynaklarYabanciKaynaklarToplamiOrani > 0.987936 **VE** FaizVeVergiOncesiKarDuranVarliklar ≤ 0.072391
İSE BAŞARISIZ

Kural 6: (Y-5)

FaizVeVergiOncesiKarDuranVarliklar ≤ 0.967114 **VE**
DonenVarliklarAktifToplamiOrani > 0.382584 **VE**
FaizGiderleriNetSatislarOrani ≤ 0 **VE**
DuranVarliklarOzKaynaklarOrani > 0.513832 **VE**
KisaVadeliYabanciKaynaklarYabanciKaynaklarToplamiOrani > 0.941896 **VE**
KisaVadeliYabanciKaynaklarYabanciKaynaklarToplamiOrani > 0.987936 **VE** FaizVeVergiOncesiKarDuranVarliklar > 0.072391
İSE BAŞARILI

Kural 7: (Y-6)

FaizVeVergiOncesiKarDuranVarliklar ≤ 0.967114 **VE**
DonenVarliklarAktifToplamiOrani > 0.382584 **VE**
FaizGiderleriNetSatislarOrani > 0
İSE BAŞARISIZ

Kural 8: (Y-7)

FaizVeVergiOncesiKarDuranVarliklar > 0.967114 **VE** StokDevirHizi
 ≤ 2.74887 **VE** BrütSatisKariAktifToplami ≤ 0.228337
İSE BAŞARILI

Kural 9: (Y-8)

FaizVeVergiOncesiKarDuranVarliklar > 0.967114 **VE** StokDevirHizi
 ≤ 2.74887 **VE** BrütSatisKariAktifToplami > 0.228337
İSE BAŞARISIZ

Kural 10: (Y-9)

FaizVeVergiOncesiKarDuranVarliklar > 0.967114 **VE** StokDevirHizi > 2.74887 **VE** VergiOncesiKarNetSatislar ≤ 0.015203 **VE**
VergiOncesiKarNetSatislar > 0.008225
İSE BAŞARISIZ

Kural 11: (Y-10)

FaizVeVergiOncesiKarDuranVarliklar > 0.967114 **VE** StokDevirHizi > 2.74887 **VE** VergiOncesiKarNetSatislar > 0.015203
İSE BAŞARILI

Kural 12: (Y-11)

FaizVeVergiOncesiKarDuranVarliklar > 0.967114 **VE** StokDevirHizi > 2.74887 **VE** VergiOncesiKarNetSatislar ≤ 0.015203 **VE**
VergiOncesiKarNetSatislar ≤ 0.008225
İSE BAŞARILI

Sınıflandırma Doğruluğu: % 67.6991

olarak verilmiştir. Örnekteki ağaç, % 67.6991 sınıflandırma doğruluğuyla öğrenilmiştir. Sınıflandırma doğruluğu, bütün kuralların veya ağacın tamamı için geçerlidir. Ağacın, dalları ya da elde edilen kurallardan herhangi birini dışlamak olanaklı değildir. Yeni bir örnek sınıflandırılırken, öğrenilen 12 kuralın herhangi birini sağlaması yeterlidir. Bununla birlikte, diğer dallarda öğrenilen kurallara aykırı bir sonuç alınması ya da hiç sınıflandırılmaması durumunda, ağacın yeni örnek(ler)le yeniden öğrenilmesi gereklidir.

3.2.1.3. Parçasal Karar Ağaçları

3.2.1.3.1. j48.PART

Bu algoritma j48.J48 karar ağaçlarının budanmasıyla oluşturulan kurallar öğrenir. j48.PART algoritmasının öğrendiği kurallar doğrudan kural öğrenen algoritmaların öğrendiği kurallara benzerlik gösterir. Örnek olarak, daha önce küçük bir veri setiyle çalıştırılmış j48.PART algoritmasının öğrendiği kural aşağıda verilmektedir.

FaizVeVergiOncesiKarDuranVarliklar > 0.9671 **VE**
StokDevirHizi > 2.7489 **VE**
VergiOncesiKarNetSatislar > 0.015203
İSE BAŞARILI

Sınıflandırma Doğruluğu: % 70.3540

j48.PART algoritmasının öğrendiği kurallar, birbirinden bağımsız kurallardır⁶⁹. Öğrenilen kurala göre, faiz ve vergi öncesi kârlarının duran varlıklarına oranı 0.9671'den büyük, stok devir hızı 2.7489'dan büyük ve vergi öncesi kârının net satışlarına oranı 0.0152'den büyük olan firmalar % 70.3540 doğrulukla **BAŞARILI** olarak sınıflandırılmıştır. Sorgulama için sunulacak yeni örnekler, koşulları sağlamaları durumunda **BAŞARILI** olarak sınıflandırılır. Örnekteki kurala dikkat edilirse, bu kuralın j48.J48 algoritmasının öğrendiği karar ağacından elde edilen kurallardan Kural 10'la aynı olduğu görülecektir. j48.J48 karar ağacının budanmasıyla öğrenilen j48.PART, Kural 10 ya da Yaprak 10'a kadar ulaşan dal dışındaki bütün dalları budamıştır.

3.2.2. Maliyete Duyarlı Öğrenme

Maliyete duyarlı çalışacak bir biçimde tasarlanmamış sınıflandırma algoritmaları maliyetin her sınıf etiketi için farklı olduğu tanım alanlarında doğru sınıflandırma yapamayabilir.

Karar ağaçları da dahil olmak üzere çoğu öğrenme algoritması gerçek anlamda maliyete duyarlı değildir. Bu bağlamda, algoritmalar çoğunlukla hatalı sınıflandırma maliyetlerinden bağımsız olarak aynı

⁶⁹ Kural bağımsızlığı birden fazla kural öğrenildiği durumda geçerlidir.

sınıflandırıcıyı tahmin ederler. Sınıflandırıcının öğrenme algoritması tarafından farklı maliyet durumlarına göre uyarlanmasıyla daha iyi bir sınıflandırma performansı elde edilebilir.

İkili sınıflandırmada öğrenme algoritmasını maliyete duyarlı hale getirmenin genel ve basit bir yolu, sınıfların örneklem içindeki dağılımını değiştirecek şekilde örneklem boyutlarını değiştirmektir. Belirli bir sınıf etiketini taşıyan örneklerin sayısı toplam örneklem içinde belirli bir oranda artırıldığında, öğrenme algoritması hata sayısını azaltmaya çalışıyorsa, bu sınıfla ilgili yapılan hatalar daha fazla cezalandırılacağı için, hata yapmaktan kaçınan bir sınıflandırma algoritması geliştirilecektir. Bu yöntemin kuramsal tanımlamaları İkinci Bölümde katmanlama yöntemleri başlığı altında yapılmıştır. Veri setindeki ilgi ya da hedef sınıf örneklerinin sayısına bağlı olarak, düşük ya da yüksek yoğunluklu örneklem yöntemlerinden herhangi birisi kullanılabilir⁷⁰ (Frank ve Witten, 2000).

3.2.2.1. Meta Sınıflandırma

WEKA platformunda maliyete duyarlı öğrenme meta sınıflandırıcılar aracılığıyla yapılmaktadır. Meta sınıflandırıcılar meta öğrenme yöntemlerinde (önyükleme, itirme ve yığma) olduğu gibi diğer sınıflandırma algoritmalarıyla kendi özelliklerini birleştirerek sınıflandırma yaparlar. Meta sınıflandırma, bir anlamda sınıflandırmanın sınıflandırmasıdır.

WEKA platformundaki maliyete duyarlı öğrenme yapan meta sınıflandırıcılar CostSensitiveClassifier ve MetaCost'tur.

3.2.2.1.1. CostSensitiveClassifier

CostSensitiveClassifier, taban sınıflandırıcısının isteğe bağlı olarak maliyete duyarlı sınıflandırma yapmasını sağlar. CostSensitiveClassifier için maliyet duyarlılığı, maliyet minimizasyonu seçeneğinin aktifleştirilmesiyle olanaklıdır. Aksi durumda, taban sınıflandırıcı maliyet bilgisine duyarlı olarak

⁷⁰ Bu çalışmada, hedef sınıf (BAŞARISIZ) örneklerinin toplam örnek sayısına oranı az olduğundan, düşük yoğunluklu örneklem yöntemine başvurulmuştur.

çalışmaz. Bu meta sınıflandırıcıyla maliyete duyarlı sınıflandırma iki farklı yöntemle yapılmaktadır:

- Ağırlıklandırma yöntemi, taban aldığı sınıflandırma algoritmasının üzerinde, onun içeriğini değiştirmeksizin çalışır. Bu yöntem, örnek veri setinin elemanlarının ağırlıklarını sınıflandırma maliyetlerine bağlı olarak değiştirmeye dayalıdır. Değiştirme işlemi, taban sınıflandırıcının eğitim örneği ağırlıklarını kullanabiliyorsa doğrudan, kullanamıyorsa veri setindeki örnekler tekrarlanarak dağılımların maliyetlere oranla değiştirilmesiyle yapılmaktadır. Temel sınıflandırma algoritması, değerleri maliyete göre düzenlenmiş olan eğitim veri seti üzerinde hatayı minimize edecek şekilde çalıştırıldığında, oluşturulan veri setinin özelliğine bağlı olarak maliyete duyarlı sınıflandırma yapacaktır.

- Minimum hatalı sınıflandırma maliyet yöntemi, olasılık hesaplama özelliğine sahip herhangi bir sınıflandırma algoritmasının üzerinde çalışır. Temel aldığı bu algoritmanın örnek veri setini kullanarak tahmin ettiği olasılık değerlerini maliyet minimizasyonu optimum sınıflandırma denkleminde yerine koyarak, doğrudan sınama örneklerinin optimum sınıflarını tahmin eder. Çalışmada kullanılan yöntemler mantıksal algoritmalar olduğundan, WEKA CostSensitiveClassifier meta sınıflandırıcısı ağırlıklandırma yöntemini kullanmıştır.

3.2.2.1.2. MetaCost

MetaCost, maliyet bilgisi değiştiğinde bütün alt sınıflandırıcıları yeniden çalıştırmak zorunda değildir. Yalnızca son öğrenme aşamasının yeniden çalıştırılmasıyla MetaCost, maliyet tablosundaki değişikliklere karşı daha esnek bir hale gelmektedir.

MetaCost, CostSensitiveClassifier'da olduğundan farklı olarak taban sınıflandırıcıyı otomatik olarak maliyete duyarlı hale getirmektedir⁷¹. MetaCost, taban sınıflandırıcısının öncelikle bir önyükleme meta öğrenicisine ve bu birleşik öğrenme modelinin en düşük maliyet kısıtıyla çalıştırılan CostSensitiveClassifier meta sınıflandırıcısına aktarılmasıyla elde edilen

⁷¹ MetaCost'ta taban sınıflandırıcının maliyete duyarlı hale getirilmesinde Pedro Domingos (1999) da anlatılan yöntem kullanılmaktadır.

sonuçlara yakın sınıflandırma sonuçları elde eder. MetaCost, CostSensitiveClassifier'dan farklı olarak, taban öğrenme algoritmasıyla tek bir maliyete duyarlı sınıflandırıcı oluşturup daha hızlı sınıflandırma yapabilmekte, eğer taban sınıflandırıcısı yorumlanabiliyorsa, yorumlanabilir sonuçlar üretebilmektedir. Ayrıca, MetaCost yönteminde, taban sınıflandırıcılarının dağılım algoritması olması önkoşulu aranmamaktadır. CostSensitiveClassifier içinse bu koşulun sağlanması öngörülmektedir. MetaCost, eğitim veri setini yeniden sınıflandırırken tüm önyükleme aşamalarını kullanır.

MetaCost, bir örneğin yeniden sınıflandırılmasında her eğitim örneğini içeren aşamalarının kullanılması durumunda, sınıflandırmada marjinal bir iyileşme sağlamaktadır.

MetaCost'un en temel özelliği herhangi bir sınıflandırma algoritmasını alıp, onun etrafına sarılarak maliyeti minimize edecek şekilde sınıflandırma yapmasını sağlamaktır. Bu nedenle ara-öğrenme yöntemlerinden biri olarak kabul edilir. MetaCost, tüm eğitim veri setinden geriye koyma yöntemiyle belirli sayıda alt setin oluşturulmasıyla başlar. Öncelikle, eğitim setinin çoklu önyükleme kopyaları oluşturulur ve MetaCost her bir kopya için sınıflandırıcılar öğrenir. Daha sonra, her bir eğitim örneği için bu modellerin oyları kullanılarak her sınıfın yaklaşık olasılıkları belirlenir. Bu olasılıklar kullanılarak, eğitim setindeki her örnek tahmin edilen optimal sınıf etiketiyle yeniden etiketlenir. Daha sonra sınıflandırıcı yeni eğitim seti üzerinde çalıştırılır. Sınıf değerleri maliyeti düşürecek şekilde değiştirilmiş olan eğitim seti üzerinde, taban alınan ve hata sayısını azaltmaya yönelik sınıflandırma algoritması tekrar çalıştırılır.

3.3. Öznitelik İzdüşümü ve BCFP

BCFP öznitelik izdüşümüne dayalı tümevarımsal bir sınıflandırma algoritmasıdır. Öznitelik izdüşümünde, tüm eğitim örneklerinin ayrı ayrı öznitelikler üzerine izdüşümü alınır. Her doğrusal öznitelik boyutunda aralıklar öncelikle bir nokta olarak belirlenir. Bu izdüşümler üzerine, her öznitelik için bir aralık inşa edilir. Bir aralık, nokta ya da gerçek anlamda

aralık olabilir. Nokta [aralık] tek bir öznitelik değerine karşılık gelirken, aralık özniteliğin birden fazla değerini içermektedir.

Öznitelik izdüşümü tabanlı bilgi gösterimi, BCFP algoritmasının her özniteliği ayrı ayrı ele almasına izin verir. Eksik değerli öznitelikler göz ardı edilir ve eğitim ve sorgulama süreçlerinde sadece bilinen değerler kullanılır. Sınıflandırma modeli, eğitim örneklerinin bilinen değerleri kullanılarak oluşturulur. Ayrıca, sorgulanan örneğin sınıf özniteliklerin bilinen değerleri kullanılarak tahmin edilir. Bu özelliği BCFP'yi eksik değerlere karşı dayanıklı kılar. Öznitelik izdüşümü tabanlı bilgi gösteriminin diğer bir avantajıysa öğrenilen kuralların tek bir özniteliğe ve o özniteliğin veri değerlerine dayanmasıdır. Dolayısıyla, öğrenilen kurallar basit ve gerçek bir uzman tarafından da doğrulanabilir, yorumlanabilir ve anlaşılabilir hale gelecektir.

BCFP algoritmasının girdisi, eğitim örnekleri setinden oluşmaktadır. Eğitim örneklerinden öğrenme yoluyla, BCFP örnek setinde yer alan sınıflandırma bilgisini inşa eder. Bu bilgi eğitim setinin her öznitelik boyutu üzerinde öznitelik aralıklarının ayrı ayrı izdüşümünün alınmasıyla ifade edilir. Her öznitelik boyutu için, benzer özellikler taşıyan izdüşüm noktaları aralıklar halinde gruplandırılır. Bu bağlamda, bir aralık, aynı sınıflandırma etiketleri öğrenen öznitelik setinin bir genellemesidir. BCFP algoritmasında sınıflandırma, her öznitelik için yapılan bireysel tahminler arasındaki oylamaya dayanmaktadır. Öğrenme ve sınıflandırma sırasında her özniteliğin diğerlerinden bağımsız katkısıyla, BCFP eksik öznitelik değerlerini tamamen yok sayarak doğal bir yaklaşım sergiler⁷².

Bir aralığın sınıf etiketlerine verdiği oylar o aralığa düşen aynı etiket taşıyan örnek sayısına ve fayda tablosuna bağlıdır. Her sınıfa yakın sayıda oy veren bir kural sorgulanan örneğin sınıflandırılmasında bir etkiye bulunmaz. Bu türden kurallar genellikle tanım alanı uzmanına “ilginç” gelmediği için modelden çıkarılabilir.

⁷² Verinin orijinal dağılımında bazı özniteliklerin değerleri eksik olabilir. Eksik değerlerin yapay bir şekilde ikame edilmesi yerine örneğin olduğu gibi kullanılması daha gerçekçi bir yaklaşımdır.

BCFP algoritmasının bilgi gösterimi ve eğitim ve sınıflandırma aşamalarıyla ilgili açıklamalar aşağıda yapılmaktadır.

3.3.1. Bilgi Gösterimi

Eğitim setindeki her örnek, nominal (kesikli) ve/veya doğrusal (sürekli) öznitelik değerleri ve bir sınıf etiketinden oluşmaktadır. BCFP algoritması, kavram tanımını öznitelik aralıklarıyla yapmaktadır. Bir aralık, nokta ya da uzanım olarak ifade edilebilir. Uzanım aralığı veri bir özniteliğin birbirini izleyen değerlerinden oluşan bir set iken, nokta aralık tek bir öznitelik değerine karşılık gelmektedir.

Uzanım aralıkları için, uzanım değerinin üst ve alt sınırları ve her sınıf için oylar belirlenir. Diğer taraftan, nokta aralıkların alt ve üst sınırları birbirinin aynıdır. Dolayısıyla, bir aralık ilk iki elemanı alt ve üst sınırları, diğer elemanlarıysa her sınıf için hesaplanan oy sayılarından oluşan bir vektör olarak ifade edilebilir⁷³.

3.3.2. Eğitim⁷⁴

BCFP eğitim sürecinde, her f özniteliği için örnekler, öznitelik izdüşümleri oluşturmak üzere f özniteliğinin değerine göre sıralanır. Öncelikle, her izdüşüm için bir nokta aralık oluşturulur. Veri eğitim örneği için, oluşturulan nokta aralığın alt ve üst sınırları birbirine eşittir. Normalleştirilmiş fayda tablosu veri iken p sınıfının oyuna (V_p) verilen başlangıç değeri aşağıdaki gibi ifade edilir.

$$V_p = \begin{cases} \frac{1}{N} \sum_{c=1}^k N_c \times NB[p,c] & \text{ise } N_p > 0 \\ 0 & \text{diğerde} \end{cases} \quad (3.1)$$

N aralıkta bulunan toplam örnek sayısı, N_c aralıkta bulunan örneklerden c sınıf etiketini taşıyanların sayısı ve $NB[p,c]$ ise bir c sınıfı örneğinin p olarak sınıflandırılmasının normalleştirilmiş faydasıdır. Diğer bir

⁷³ $\langle lb, ub, V_1, V_2, \dots, V_k \rangle$

k = tanım alanındaki sınıf sayısı

$V_i = C_i$ sınıfına aralığın verdiği oy

⁷⁴ BCFP eğitim algoritması için, Güvenir'e (2003) bakınız.

deyişle, V_p veri aralığındaki tüm örneklerin p sınıfı olarak tahmin edilmesinin ortalama faydası olarak ifade edilebilir. p sınıfından eğer hiçbir örnek veri aralığında gözlemlenemiyorsa o aralığın p sınıfı için oyu 0'dır. Her aralık için oylama gücünün eşit olabilmesi için, o aralığa ait oylar normalleştirilir ve oyların toplamı 1'e eşit olur.

$$\sum_{p=1}^k V_p = 1. \quad (3.2)$$

Eğitim örneğinin f özniteliğinin değeri bilinmiyorsa⁷⁵ ("?" ile ifade edilir), o öznitelik değeri dikkate alınmaz. Doğrusal öznitelikler için BCFP nokta aralıkları genellemeye çalışır. En yüksek oylar, aynı sınıfı gösteren ardışık nokta aralıklarıyla uzanım aralıkları oluşturulacak şekilde birleştirilir.

Kararsız, yani oyları her sınıfa eşit olarak dağıtılan bir aralık ilginç değildir ve dışlanmalıdır. Bir kural, oylarının standart sapması veri bir minimum eşik değerinin (S_{min}) üzerindeyse "kararlı"dır. BCFP'nin kullandığı eşik

$$S_{min} = \frac{1}{k-1} \sqrt{\frac{1}{k}} \quad (3.3)$$

olarak tanımlanmaktadır. Bu eşik, aralığın bir sınıfa 0 oy verdiği ve oylarını tüm sınıflara eşit oranda dağıttığı durumda standart sapmaya eşittir.

3.3.3. Sınıflandırma⁷⁶

BCFP algoritması sınıflandırma sürecinde, öncelikle her sınıfın oyları 0'a eşitlenir. Sınıflandırma ayrıca her özniteliğe göre bir ön sınıflandırma süreci içerir.

Her f özniteliği için ön sınıflandırma, q_f , f özniteliği için sorgulama örneğinin değeri q iken, q_f 'in hangi f öznitelik aralığında bulunduğu dair bir arama süreci içerir. f değerinin veri örnek için bilinmediği durumda ($f="?"$) o öznitelik oylamaya katılmaz. Dolayısıyla eksik değerli öznitelikler göz ardı edilir. Eğer q_f biliniyorsa, q_f 'in içinde bulunduğu aralık I aranır. q_f f öznitelik aralıklarından herhangi biri içinde bulunmuyorsa, bu durumda da f özniteliği

⁷⁵ Eksik ya da diğer adıyla bilinmeyen değerler "?" ile ifade edilir.

⁷⁶ BCFP sınıflandırma algoritması için, Güvenir'e (2003) bakınız.

oynamaya katılmamaktadır; eğitim setinde f özniteliğinin değeri gözlemlenmemektedir. q_f 'in dahil olduğu bir I aralığı bulunduğundaysa, I aralığının oyları f özniteliğinin verdiği oylardır. Eğitim sırasında bir aralığın oyları normalleştirildiği için her özniteliğin oylama gücü birbirine eşittir. Son olarak, en yüksek sayıda oyu alan sınıf, sorgulanan q örneği için tahmin edilen sınıf etiketi olarak verilir. Sınıf etiketiyle birlikte her sınıfın aldığı oy oranı tahminin güvenilirliğini ölçmek amacıyla kullanılır. Sınıflandırmanın fayda doğruluğu, normalleştirilmiş fayda tablosundan doğrudan elde edilir. q 'nın gerçek sınıfı q_c ve tahmin edilen sınıf p ise, doğruluk $NB[p, q_c]$ 'dir. Tahmin doğruluğunun doğru sınıflandırılmış örneklerin toplam örnek sayısına oranı olarak verilen klasik tanımı, fayda tablosunun (NB) köşegen elemanları 1 ve diğer elemanları 0 olan özel bir duruma karşılık gelmektedir.

Fayda tablosunun elemanları tanım alanı uzmanına ya da sadece uzmana anlamlı gelebilecek herhangi bir sayı olabilir. Ölçü biriminin etkilerinden kurtulmak üzere BCFP algoritması fayda tablosunun elemanlarını $[0,1]$ aralığına normalleştirir; doğru tahminin faydası 1, hatalı tahminin faydasıysa 0'a eşit olur.

BCFP algoritması, özellikle her özniteliğin diğerlerinden bağımsız olduğu tanım alanlarında uygulanabilir. Bu gerekliliğin, bazı tanım alanlarında özniteliklerin bağımsız olmadığı gerçeğine dayalı olarak BCFP'nin uygulanabilirliğini kısıtladığı düşünülebilir. Holte (1993) çoğu gerçek hayat sınıflandırma probleminde aslında özniteliklerin birbirinden bağımsız olduğunu belirtmiştir.

Finansal sınıflandırma alanında, firmanın finansal gücü hakkında, özellikle karmaşık ve zaman zaman birbiriyle tutarsız, çarpıtılmış ve yanlış finansal bilgilerin sunulduğu bir durumda, bir tahmin yürütmek zordur. Finansal analizciler, genellikle, firma hakkındaki ölçüm ve gözlemlere dayalı kararlar verir. Analizci, tahmin yapmak ya da tanı koymak için tüm faktörleri,

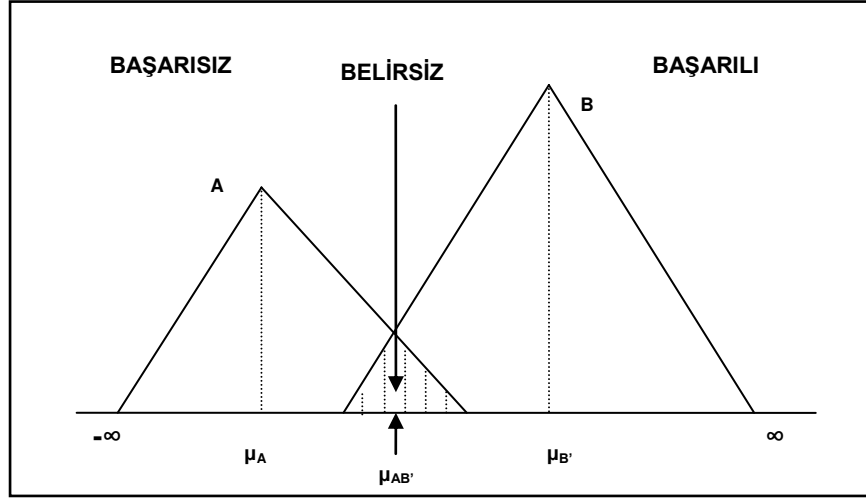
ne kadar istese de, öznel ve yanlı olarak değerlendirir⁷⁷. Dolayısıyla, daha karmaşık ve güvenilir yöntemler ve sistemler uzmanların veriyi ele alması ve daha iyi ve doğru tanılanması için gereklidir. Bilişim bilimi, makina öğrenmesi yöntemleri aracılığıyla bu karar verme sürecinde yardımcı olması amacıyla bazı karmaşık yöntemler önermektedir.

3.3.4. BCFP'nin Oluşturduğu Kurallar

BCFP fayda maksimizasyon amaç fonksiyonu kısıtı altında kural öğrenen bir algoritmadır. BCFP'nin kural öğrenme sürecinde, her öznitelik için optimal bir kesim noktası elde edilir. Bu kesim noktası, BCFP tarafından, veri setindeki sınıfların öznitelik dağılımlarının aritmetik ortalamalarının orta noktası olarak belirlenir ve kurallar oluşturulur. Kesim noktası, BCFP'nin veri öznitelik için oluşturduğu kuralın temel özelliğidir. Bu kurallar, BCFP'nin özgül bir niteliği olarak çözümleme süresince sabit kalırlar. Sınıflandırma sürecinde, özgül örneklerle doğru ya da hatalı sınıf etiketlerinin verilmesi ve fayda-maliyet kısıtları altında, örneklerin öznitelik kesim noktalarının neresinde kaldıklarına bağlı olarak sınıflandırma doğruluğu ve sınıflandırmanın toplam faydası elde edilir. Bu bağlamda, kuralların fayda-maliyet bilgisine duyarsız olduğu söylenebilir.

BCFP kurallarının sabit kurallar olması basit bir yaklaşım olarak düşünülebilir. Kuralların bu şekilde oluşturulması, algoritmanın daha verimli çalışması için, öznitelik değerlerinin sınıf dağılımlarının üçgensel bir yapı sergilediği varsayımına dayalıdır. Böylece BCFP, dağılımların nasıl bir yapı sergilediğiyle uğraşmadan, doğrudan sınıflandırma yapar. BCFP'nin kabul ettiği dağılım, Şekil 3.1'de sunulmaktadır.

⁷⁷ Burada, analizcinin uzman sistemlerine göre göreceli öznelliği söz konusudur. Finansal analizcinin öznelliği bilinçli ve istemli değildir. Analizci, herhangi bir sorgulamaya kendi teknik altyapısını ve olasılıkla duygularını da karıştırabileceği için çözümlemenin tarafsız, yansız ve nesnel ve duygulardan arınmış olması olasılığı bir uzman sisteme göre daha düşüktür. Bunun yanında, alan uzmanları çok büyük hacimde veriyi aynı anda analiz etmekte de büyük güçlük çekerler.



Şekil 3.1: BCFP'nin Üçgen Dağılım Varsayımı

μ_A , başarısız (A) sınıfının aritmetik ortalaması, μ_B , başarılı (B) sınıfının aritmetik ortalaması, μ_{AB} , ise μ_A ve μ_B 'nin aritmetik ortalamasıdır. BCFP'nin öğrendiği kuralların genel yapısı

EĞER $\ddot{O}_k \Theta \mu_{AB(k)}$ **İSE** sınıf C_m 'dir. $D_m = \% P$

olarak verilmektedir. Bu kuralda, $\ddot{O}_k$ k özniteliği, Θ $<$, $>$, $\leq$, $\geq$ ya da $=$ değerlerini alan sorgulama işleci, $\mu_{AB(k)}$ k özniteliği için sınıf ortalamalarının aritmetik ortalaması, C_m m sınıf değeridir. D_m ise genel sınıflandırma doğruluğu ya da BCFP analizinde fayda doğruluğuna karşılık gelmektedir.

İki sınıflı bir problemde, sayısal bir öznitelik için öğrenilen kural, kesim noktasının altında kalan örneklere bir sınıf etiketini, üstündekilere diğer sınıfın etiketini verir. Sayısal öznitelikler, kuramsal olarak eksi ve artı sonsuz arasında tüm değerleri alabilirler. Sayısal öznitelikler için sorgulama işleçleri eşitlik ($\leq$, $\geq$) ve eşitsizlik ($<$, $>$) değerlerini alırlar. BCFP, inceleme yılı ve firma yaşı gibi bazı kesikli öznitelikleri de sürekli olarak ele alır ve sürekli özniteliklerdekine benzer kurallar öğrenebilir. Örneğin, firma yaşı için 5.7 yıl gibi bir kesim noktasıyla kural elde edilebilir. Nominal ya da kesikli öznitelikler için eşitlik sorgulaması ($=$) söz konusudur.

Eksik değerlerin ele alınmasında BCFP, gerçek hayatta bazı örneklerin kimi öznitelik değerlerini almadığı varsayımını yapar. Buna göre, sınıflandırma sürecinde eksik değerler dikkate alınmaz. Bununla birlikte,

herhangi bir örneğin olabildiğince az sayıda eksik değer barındırması arzu edilir.

BCFP algoritmasının eleştirilmesi gereken bir yönü, tüm özniteliklerin sınıf dağılımını önceden varsayarak gerçek dağılımı göz ardı etmesidir. Tüm sınıflandırma algoritmalarında olduğu gibi, BCFP'de de arzu edilen, sınıflandırma sürecinde gerçek sınıf dağılımlarının dikkate alınmasıdır.

BCFP her bir öznitelik için birbirinden bağımsız kurallar öğrenir. Bu kurallardan herhangi biri, tek başına sınıflandırma amacıyla kullanılabilir, yani öğrenilen kurallardan sadece biri bile tüm veri setini sınıflandırmak için yeterlidir. Örneğin, aşağıdaki,

EĞER	$K_1: \bar{O}_1 \ominus \mu_{AB(1)}$	İSE	sınıf C_m 'dir.	$D_m = \% 78$
EĞER	$K_{75}: \bar{O}_{75} \ominus \mu_{AB(75)}$	İSE	sınıf C_m 'dir.	$D_m = \% 78$

kuralları arasında, herhangi biri tüm veri setini aynı doğrulukla (% 78) sınıflandırmaktadır. Bunun yanı sıra, öğrenilen diğer kuralların da sınıflandırma doğruluğu yukarıda öğrenilen iki kuralın sınıflandırma doğruluğunun aynı, yani % 78'dir. Kullanıcının amacına bağlı olarak, anlamlı ve bilinenler dışında ilginç gelen kurallardan herhangi biri sınıflandırma için kullanılabilir⁷⁸. Bu kurallardan hangisinin kullanılacağıyla ilgili önkoşul, kuralın kuramsal olarak anlamlı olmasıdır.

BCFP'nin öğrendiği kurallar, tüm veri setinin tek bir özniteliğe dayalı olarak sınıflandırılmasına odaklandığından, kuramsal olarak WEKA platformundaki OneR'in öğrendiği kurallara benzer. BCFP, OneR'dan farklı olarak öznitelik uzanım aralığını ikiye böler. Bununla birlikte, algoritma uzanım aralığını birden fazla kesim noktasıyla bölecek şekilde de değiştirilebilir, fakat bu durumda, kuralların basit, anlaşılır ve genel olması gibi istenen özellikleri kaybolacaktır⁷⁹.

⁷⁸ Öğrenilen kurallardan bazıları bilinen, diğerleri ilginç yani kuramsal olarak doğru ama beklenmedik, diğerleri ise kuramsal olarak anlamsız olacaktır.

⁷⁹ Çalışmamızda kullanılan BCFP algoritması uzanım aralıklarını ikiye bölecek şekilde tasarlanmıştır.

DÖRDÜNCÜ BÖLÜM

MAKİNA ÖĞRENMESİ YÖNTEMLERİ UYGULAMALARI - AMPİRİK ANALİZ VE YORUMLAR

Bu bölümde, çalışma konusu olan firma başarısızlığının dinamiklerini anlamak üzere yapılan ampirik analizler ele alınmaktadır.

4.1. Veri Setinin Yapısı

4.1.1. Veri Seti

Çalışmada kullanılan veri seti, Türkiye Cumhuriyet Merkez Bankası İstatistik Genel Müdürlüğü Reel Sektör Verileri Müdürlüğü veri tabanında 2001 yılı itibariyle bulunan 22,000 firmadan imalat sanayiinde faaliyet gösteren 5,636 firmanın finansal tabloları (Bilanço ve Gelir Tablosu) kullanılarak hazırlanmıştır. Bu firmalar, 5 - 13 yıllık sürekli verisi bulunan imalat sanayii firmaları arasından seçilmiştir. İmalat sanayii firmalarının seçilmesindeki amaç, imalat sanayii dışında faaliyet gösteren firmalarla bu firmalar arasında finansal tablolar açısından öngörülen yapısal farkların dışlanması ve daha homojen bir veri seti oluşturulmasıdır. 5,636 firma arasından rassal olarak seçilen⁸⁰ 4,000 firmanın verileri, makina öğrenmesi teknikleriyle analiz için gereken formatta kullanılmak üzere hazırlanmıştır. Çalışmada kullanılan öznitelikler Ek 2’de verilmektedir.

Modelin tasarımında, firma verileri yanında makroekonomik verilerin BCFP algoritmasıyla yapılan ön sınamalarında, sınıflandırma performansında herhangi bir artış sağlanamamıştır. Bu bağlamda, başlangıçta 194 öznitelikten oluşan veri setlerinden 29 makroekonomik öznitelik dışlanmış ve sadece firma verilerinin kullanılmasına karar verilmiştir. Makroekonomik verilerin ön sınamaları geçememesinin nedeni tam olarak bilinmiyorsa da,

⁸⁰ Rassal olarak seçilen 4,000 firmanın sektör ve sınıf dağılımları, kitlenin dağılımları ile aynıdır.

en azından finans kuramının öngördüğü biçimde, ekonomik politika etkilerinin firma mali tablolarına daha önceden yansımış olması olasılığı üzerinde durulabilir.

4.1.2. Örneklem Seçimi ve Örneklem Büyüklüğü

Örneklem seçiminin rassal olması, imalat sanayiinin alt sektörlerindeki firmaların toplam kitledeki temsil oranlarının korunması açısından önemlidir. Seçim yapılırken analizde kullanılan 4,000 firmanın alt sektörler göre dağılımının, tüm kitleyi oluşturan 5,636 firmanın alt sektör dağılımına yakın olmasına dikkat edilmiştir.

Örneklem büyüklüğü, makina öğrenmesi teknikleriyle analizde önemli bir etmendir. Örneklem ne kadar büyükse sınıflandırma için öğrenilen kuralların doğruluğu ve kitleyi temsil oranı o ölçüde artmaktadır. Kullanılan veri setinin kitle büyüklüğüne oranının % 71'i civarında olması dolayısıyla, yüksek doğrulukta genellenebilir kurallar öğrenilmesi beklenmektedir.

4.1.3. Firmaların Hukuksal Durumları

Veri setindeki firmalar özel sektörde faaliyet gösteren anonim ve limited şirketler arasından seçilmiştir. Anonim ve limited şirketlerin kullanılmasının nedeni, veri setinin daha homojen bir yapıda olmasını sağlamaktır. Ayrıca anonim ve limited şirketler kurumsal firma yapısını daha iyi yansıtmaktadır.

4.1.4. Verinin Hazırlanması ve Standart Form

Verinin analize hazırlanması aşamasında, veri tabanından firma verilerini almak üzere geliştirilen ve oranların otomatik olarak hesaplanmasını sağlayan standart bir form kullanılmıştır. Bu form, finansal oranların yanında ortak büyüklük özneliklerinin de hesaplanmasını eşanlı olarak sağlamaktadır.

Form yardımıyla hesaplanan öznitelik değerleri öncelikle toplulaştırılmış ve sınıflandırma algoritmalarında kullanılmak üzere, “.csv” ve “.arff” formatlarına dönüştürülmüştür⁸¹.

4.1.5. Bilgi Gösterimi

Öğrenme algoritmalarında kullanılan örnekler, özniteliklerin aldığı değerler ve sınıf etiketiyle gösterilmektedir.

Veri kullanımı kısıtları ve gizlilik kuralları çerçevesinde, ham veri setini oluşturan 4,000 firmanın kimlik bilgileri veri setinden çıkarılmıştır. Kimlik bilgisinin dışlanmasıyla, 4 yıldan daha fazla bilgisi olan herhangi bir firma için, yıl bilgisini de yeni bir öznitelik olarak kullanmak yoluyla, birden fazla genel örnek elde etmek olanaklı olmuştur. Böyle örnekler firma-yıl olarak tanımlanmaktadır. Örneğin, 5 yıllık bilgisi olan A firmasının, üç yıllık ufuk parametresi⁸² büyüklüğüne göre, t yılında BAŞARISIZ, t+1 yılında BAŞARILI sınıf etiketleri ve t ve t+1 yıllarındaki öznitelik değerleriyle, iki ayrı örnek elde edilebilir. $F(m; k) = m$ örneğinin k özniteliğinin aldığı değer, M örnek sayısı, K öznitelik sayısı ve $m = 1, 2, \dots, M$, $k = 1, 2, \dots, K$ olarak tanımlanırken, bu örnekler,

Örnek 1: $F(1; k)$, Sınıf : **BAŞARISIZ**, (Firma A, t)

Örnek 2: $F(1; k)$, Sınıf : **BAŞARILI**, (Firma A, t+1)

olarak gösterilebilir. Yıl bilgisi, herhangi bir öznitelik olarak öğrenmede kullanılmaktadır. Verilerin firma-yıl olarak düzenlenmesiyle, aynı veri setinden daha geniş bir eğitim seti oluşturmak olanaklı olmuştur. Böylece, her firma-yıl örnek birbirinden bağımsız firmaların bilgileri gibi eğitimde kullanılabilmiştir. Firma-yıl dönüşümüyle birlikte elde edilen veri seti 25,547 firma-yıl örnek⁸³ ve 165 öznitelikten oluşmaktadır. Firma-yıl örneklerden oluşan veri setinin genel görüntüsü, Tablo 4.1'deki gibidir.

⁸¹ BCFP algoritması .csv (comma separated values), WEKA algoritmaları da arff (attribute relation file format) formatında veri kullanmaktadır.

⁸² Ufuk parametresi ve sınıf etiketlerinin belirlenmesi, ileride anlatılmaktadır.

⁸³ 584 BAŞARISIZ ve 24,963 BAŞARILI firma-yıl örnekten oluşmaktadır.

TABLO 4.1. VERİ SETİNİN GENEL YAPISI

ÖRNEKLER	ÖZNİTELİKLER					SINIF
Örnek 1	F(1; 1)	F(1; 2)	F(1; 3)		F(1; K)	BAŞARILI
	F(2; 1)	F(2; 2)	F(2; 3)		F(2; K)	BAŞARISIZ
	F(3; 1)	F(3; 2)	F(3; 3)		F(3; K)	BAŞARILI
	F(4; 1)	F(4; 2)	F(4; 3)		F(4; K)	BAŞARILI
	F(5; 1)	F(5; 2)	F(5; 3)		F(5; K)	BAŞARISIZ
						
....						
Örnek M	F(M; 1)	F(M; 2)	F(M; 3)		F(M; K)	BAŞARILI

4.1.6. Sınıf Etiketlerinin Belirlenmesi

Sınıf etiketlerinin belirlenmesi analizcinin özgül tanımlamasına göre şekillenir ve herhangi bir veri seti bu tanımlamanın ölçütlerine göre etiketlendikten sonra sınıflandırma amacıyla kullanılacak duruma gelir. Sınıf etiketlerinin belirlenmesi verinin hazırlanmasındaki en önemli aşamadır. Standart veri hazırlama formundan aynı zamanda sınıf özneliliğinin aldığı değerlerin yani sınıf etiketlerinin belirlenmesinde de yararlanılmıştır.

Finansal başarısızlık literatüründeki çalışmaların çoğunluğunda kullanılan sınıf etiketi, BAŞARILI ya da BAŞARISIZ değerlerinden birini almaktadır. Burada, analizcilerin ve ilgili menfaat gruplarının üzerinde durdukları konu, başarıdan çok başarısızlığın tahmin edilebilmesidir. Bunun nedeni, İkinci Bölümde de anlatıldığı gibi, başarısız firmanın hatalı tahmin edilmesi durumunda bu maliyetin başarılı firmanın hatalı tahmin edilmesinin maliyetinden daha yüksek olmasıdır.

Finans kuramının öngördüğü şekilde belirli sayıda dönem boyunca üst üste zarar eden firmaların sermayesi yitikleşecek ve sermayenin toplam kaynaklar içindeki payı belirli bir oranın altına düştüğünde firmanın iflası istenebilecektir. Bu bağlamda, literatürdeki çalışmalardan bir bölümü sermaye yitikliğini başarısızlık ölçütü olarak kullanmıştır.

Sermaye yitikliğı, sadece üç yıl üstüste zarar bilgisi göz önünde bulundurulduğunda nesnel bir ölçüt olmayacaktır. Firmanın, zarar ettiği ilk yıldaki zararının ve eğer sonraki yıl kâr etmişse kârının büyüklükleri ve bu büyüklüklerin değişimi de sermaye yitikliğinin boyutu hakkında bilgi veren ölçütlerdir. Sorgulama dönemi boyunca dönemlerarası kâr ve zarar

rakamlarının karşılaştırılmasıyla mutlak büyüme oranı elde edilir. Mutlak büyüme oranı bilgisine ve kâr/zarar rakamlarına bakılarak, çalışmadaki başarısızlık ölçütü, firmanın üç yıl üst üste zarar etmesi ya da ilk yıl zarar, ikinci yıl kâr, üçüncü yıl zarar ederken, ilk yıldan ikinci yıla mutlak büyüme oranının ikinci yıldan son yıla mutlak büyüme oranından küçük olması olarak öngörülmüştür⁸⁴.

Sınıf etiketlerinin belirlenmesinde kullanılan temel iki ölçüt firma performans sorgulama ufku (n) ve analiz döneminin performans ufkuna uzaklığıdır (j). Sorgulama ufku, veri dönemin, kârlılık sorgulamasının yapıldığı 1, 2 ve/veya n dönemin ne kadar önünde olduğudur. Burada dikkat edilmesi gereken iki nokta, sorgulama ufkunun uzunluğu ($t+j$, $t+j+1$, ... , $t+j+n-1$; t analiz dönemi ve $t+j$ performans ufkunun başlangıç yılı iken n'nin büyüklüğü) ve analiz döneminin performans ufkuna uzaklığıdır (t analiz dönemi ve $t+j$ performans ufkunun başlangıç yılı iken j'nin büyüklüğü). Örneğin, sınıflandırma üç yıl öncesinden yapılıyorsa ($n=3$) veri t döneminde kâr sorgulaması $t+3$, $t+4$ ve $t+5$ yıllarında yapılacaktır⁸⁵. Bu örnekte j 3'e eşittir.

Finans kuramına göre sorgulama ufkunun uzun (n'nin büyük) olması sermaye yitikliğinin ne düzeye geldiğinin anlaşılması ve başarısızlık analizinde sınıfların belirlenmesi açısından gereklidir. Bu şekilde belirlenen sınıf etiketi daha kısa performans ufkusuyla belirlenen sınıf etiketine göre daha fazla bilgi içerir. Makina öğrenmesi açısından, n'nin büyük olması tahmin edilecek performans ufkunun uzun olması demektir⁸⁶ ki burada öğrenme

⁸⁴ $P_t > 0$ iken $P_{t+1} < 0$, $P_{t+2} > 0$ ve $P_{t+3} < 0$ olması durumunda firma t döneminde BAŞARISIZ olarak sınıflandırılmıştır. Burada bakılması gereken firmanın $t+1$ döneminde azalan net sermayesinin $t+2$ dönemindeki kârla ne düzeyde telafi edilebildiğidir. Eğer, $t+2$ döneminde kâra geçişteki büyümenin sağladığı iyileşme ($g_{t+2,t+1} \equiv t+1$ den $t+2$ 'ye olan mutlak büyüme oranı) $t+3$ 'teki sermaye düşüşünden ($g_{t+3,t+2} \equiv t+2$ den $t+3$ 'e olan mutlak büyüme oranı) az ise $t+2$ döneminde elde edilen kâr, sermayenin yitikleşmesini telafi edemeyecektir.

$P_t > 0$ iken $P_{t+1} < 0$, $P_{t+2} > 0$ ve $P_{t+3} < 0$

$g_{t+2,t+1} \equiv \text{abs}(\text{net sermaye}_{t+2}) / \text{abs}(\text{net sermaye}_{t+1}) + \text{abs}(\text{net sermaye}_{t+2})$

$g_{t+3,t+2} \equiv \text{abs}(\text{net sermaye}_{t+3}) / \text{abs}(\text{net sermaye}_{t+2}) + \text{abs}(\text{net sermaye}_{t+3})$

Örneğin, ufuk = 3 ve uzaklık = 1 (1yılönce3yıl) parametrelerine göre sınıf etiketi

Sınıft(P_{t+1} , P_{t+2} , P_{t+3}) =

eğer $P_{t+1} < 0$, $P_{t+2} < 0$ VE $P_{t+3} < 0$ VEYA $P_{t+1} < 0$, $P_{t+2} > 0$ VE $P_{t+3} < 0$ VE $g_{t+2,t+1} <$

$g_{t+3,t+2}$

ise **BAŞARISIZ**,

yoksa **BAŞARILI**

olarak verilmektedir.

⁸⁵ $P_t > 0$ iken $P_{t+3} < 0$, $P_{t+4} < 0$ ve $P_{t+5} < 0$ → **BAŞARISIZ**

⁸⁶ Burada t dönemine ait öznelitliklerin ($F(m,k)$ $m=1, 2, \dots, M$; $k=1, 2, \dots, K$; t inceleme dönemi) aldıkları değerler veri iken, öğrenme algoritmasının t döneminde firmanın ($t+1$, $t+2$, $t+3$ dönemlerinde) yukarıdaki ölçütlere bağlı olarak sınıf etiketi alıp alamayacağına yönelik kurallar öğrenmesi istenir.

algoritmasından beklenen, belirsizliğin arttığı olabildiğince uzun bir dönemi tahmin etmeye çalışmasıdır⁸⁷.

Benzer şekilde, analiz döneminin performans ufkuna uzaklığı da $(t+j)$ başlangıç dönemi iken j 'nin büyüklüğü) makina öğrenmesi açısından önemli bir parametredir. Bu parametrenin büyük olması halinde inceleme dönemi ve performans ufku arasındaki yıllara ait bilgi öğrenme algoritması tarafından kullanılamayacaktır⁸⁸.

Uzaklık ve ufuk parametreleri ve özniteliklerin inceleme döneminde aldığı değerler veri iken öğrenme algoritmasından beklenen, en az bilgiyle en kısa sürede⁸⁹, en iyi kuralları öğrenebilmesidir. Buna karşın, kuramsal açıdan makina öğrenmesi yöntemleri, istisnaları olmakla birlikte, ideal olarak olabildiğince çok bilgiyle eğitildiğinde iyi sınıflandırma kuralları öğrenmektedir.

Daha muhafazakâr analizciler, başarısızlık sınıf etiketlemesi yapılırken firmanın performans ufkunun daha dar tutulmasını savunmaktadırlar. Bu durum, üç yıldan daha az kâr/zarar bilgisine bakılarak etiketleme sonucunu doğurmaktadır. Örneğin, firma veri dönemden (t) sonraki iki $(n=2)$, hatta bir $(n=1)$ yıl dahi zarar etmişse “başarısız” olarak etiketlenmektedir⁹⁰.

Çalışmada kullanılan örnekler t zamanında kâr ederken, $t+1$, $t+2$ ve $t+3$ dönemlerinde zarar etmişse ya da t zamanında kâr ederken, $t+1$ döneminde zarar, $t+2$ döneminde kâr ve $t+3$ döneminde zarar etmiş ve $g_{t+2,t+1}$

⁸⁷ Firmanın t dönemine ait verileri kullanarak $t+j$, $t+j+1$, $t+j+2$, ... , $t+j+n-1$ dönemlerinde başarılı olup olamayacağını tahmin edilmesi demektir.

⁸⁸ Örneğin, ufuk = 3 ve uzaklık = 3 (3yılönce3yıl) parametrelerine göre sınıf etiketi

Sınıft(P_{t+3} , P_{t+4} , P_{t+5}) =
eğer $P_{t+3} < 0$, $P_{t+4} < 0$ VE $P_{t+5} < 0$ VEYA $P_{t+3} < 0$, $P_{t+4} > 0$ VE $P_{t+5} < 0$
VE $g_{t+4,t+3} < g_{t+5,t+4}$
ise **BAŞARISIZ**,
yoksa **BAŞARILI**

olarak verilmektedir.

Burada t dönemine ait özniteliklerin $(F(m,k))$ $m=1, 2, \dots, M$; $k=1, 2, \dots, K$; t inceleme dönemi) aldıkları değerler veri iken, öğrenme algoritmasının t döneminde firmanın $(t+3, t+4, t+5)$ dönemlerinde yukarıdaki ölçütlere bağlı olarak sınıf etiketi alıp alamayacağına yönelik kurallar öğrenmesi istenir. Bu analizde ise, $t+1$ ve $t+2$ dönemlerine ait bilgi, öğrenme algoritması tarafından kullanılmamaktadır.

⁸⁹ Burada söz konusu olan, öğrenme algoritmasının kural öğrenme maliyetini (zaman, veri temini, verilerin hazırlanması, algoritmanın kural öğrenme için kullandığı zaman vb.) en düşük seviyede tutmaktır. Bu bağlamda, finansal baskı analizi açısından, veri doğruluk düzeylerinde n ve j parametreleri ne kadar büyükse sınıflandırma o kadar başarılı, yani daha uzak dönemleri tahmin edebiliyor, denilebilir.

⁹⁰ İki yıl, $P_t > 0$ iken $P_{t+1} < 0$, $P_{t+2} < 0$ → **BAŞARISIZ**
Bir yıl, $P_t > 0$ iken $P_{t+1} < 0$, → **BAŞARISIZ**

$< g_{t+3,t+2}$ ise BAŞARISIZ olarak etiketlenmiştir. $g_{t+2,t+1}$, $t+1$ 'den $t+2$ 'ye olan mutlak net sermaye büyüme oranı, $g_{t+3,t+2}$, $t+2$ den $t+3$ 'e olan mutlak net sermaye büyüme oranı olarak tanımlanmıştır.

4.1.7. WEKA Alt Örneklem Büyüklükleri

WEKA platformunun yapısı nedeniyle büyük sayılarda örneklerden oluşan veri setleri kullanılamamaktadır⁹¹. Bu bağlamda, 4,000 firmanın finansal tablolarından elde edilen 25,547 firma-yıl örnek WEKA tarafından kullanılabilir boyutlara indirgenmiştir. Bu aşamada, büyük veri seti bölüntülenirken sınıf ve sektör dağılımlarının bozulmamasına dikkat edilmiştir. 1 yıl uzaklık ve 3 yıl ufuk parametre değerleriyle BAŞARILI örneklerin dağılımdaki oranı % 97.71, BAŞARISIZ örneklerin oranı % 2.29'dur. Bu oran alt bölüntülerde de olabildiğince korunmuştur. WEKA veri seti, doğrusal ve nominal olmak üzere toplam 165 öznitelik ve 4,272 (firma-yıl) örnekten oluşmaktadır.

4.1.8. Örneklerin Sınıflar Bazında Sektörel Dağılımı

Örneklerin sınıflar bazında sektörel dağılımı, Tablo 4.2'de sunulmaktadır.

TABLO 4.2. SINIFLARIN SEKTÖREL DAĞILIMI

		BAŞARILI Örnek Sayısı	BAŞARISIZ Örnek Sayısı	Toplam Örnek Sayısı	Sektör / Toplam (%)	BAŞARILI / Toplam (%)	BAŞARISIZ / Toplam (%)	BAŞARILI / Sektör (%)	BAŞARISIZ / Sektör (%)
1	DA	3,715	86	3,801	14.88%	14.54%	0.34%	97.74%	2.26%
2	DB	7,157	154	7,311	28.62%	28.02%	0.60%	97.89%	2.11%
3	DC	543	12	555	2.17%	2.13%	0.05%	97.84%	2.16%
4	DD	564	14	578	2.26%	2.21%	0.05%	97.58%	2.42%
5	DE	993	24	1,017	3.98%	3.89%	0.09%	97.64%	2.36%
6	DF	21	0	21	0.08%	0.08%	0.00%	100.00%	0.00%
7	DG	1,837	39	1,876	7.34%	7.19%	0.15%	97.92%	2.08%
8	DH	1,358	37	1,395	5.46%	5.32%	0.14%	97.35%	2.65%
9	DI	1,567	49	1,616	6.33%	6.13%	0.19%	96.97%	3.03%
10	DJ	2,590	56	2,646	10.36%	10.14%	0.22%	97.88%	2.12%
11	DK	1,704	42	1,746	6.83%	6.67%	0.16%	97.59%	2.41%
12	DL	1,213	22	1,235	4.83%	4.75%	0.09%	98.22%	1.78%
13	DM	1,165	36	1,201	4.70%	4.56%	0.14%	97.00%	3.00%
14	DN	536	13	549	2.15%	2.10%	0.05%	97.63%	2.37%
	Toplam			25,547	100.00%	97.71%	2.29%		

⁹¹ Çalışmanın yapıldığı dönemde bulunan en son WEKA platformu sürüm 3.2'dir. Java 1.2 ve WEKA 3.2 sürümleri ve kişisel bilgisayarların donanım kısıtları da göz önünde bulundurulduğunda büyük ölçekli ve çok sayıda öznitelikten oluşan örneklemelerin çözümleme amacı ile kullanılması olanaklı olmamıştır.

Çözümlemelerde kullanılan bilgisayar konfigürasyonu PIV 2.4 işlemci ve 2 Gb RAM'den oluşmaktadır.

Tablo 4.2 incelendiğinde, DF, DI, DL ve DM sektörleri dışındaki sektörlerde sınıfların sektörel dağılımının, tüm veri setindeki sınıf dağılımına yakın olduğu görülmektedir. Bu dağılım, veri seti oluşturulduktan sonra ortaya çıkan görüntüyü yansıtmaktadır. Böyle bir görüntünün elde edilmesinde, büyük olasılıkla, sektörlerin rassal seçim aracılığıyla temsil oranının korunmasının ve firma-yıl dönüşümünün rolü olduğu söylenebilir. İmalat sanayii alt sektörlerinin NACE (Rev.1) kodları ve karşılıkları, Ek 1’de verilmektedir.

4.1.9. Özniteliklerin Sınıf Yığılma Dağılımları

Öğrenme algoritmalarının en yüksek doğrulukla sınıflandırma (bölüntüleme) yapabilmesi için, özniteliklerin aldığı değerlerin BAŞARILI ve BAŞARISIZ sınıflarına göre dağılımlarının, yapısında en yüksek ayırık yığılma barındırması gerekmektedir. Ayırık yığılma, öznitelik sınıf dağılımlarının değer uzanımında en az kesişimle olabildiğince az belirsiz bölge bırakacak şekilde değerler almasıyla olanaklıdır. Çalışmada kullanılan öğrenme algoritmalarının, dağılım varsayımları yapmaması nedeniyle, ayırık yığılma yapısının sağlanabilmesi yeterlidir.

Veri setinin sınıf dağılımlarının normallik sınamaları, Jarque-Bera istatistiği kullanılarak yapılmıştır. Jarque-Bera istatistiği, herhangi bir dağılımın normalliğinin sınanması için kullanılabilir. Öznitelik değerlerinin toplam ve sınıflara göre dağılımları, normallik sınamalarını geçememiştir.

Ayırık yığılma sınamaları, Scott (1979) ve Freedman-Diaconis (1981) frekans aralığı genişlik ölçütleri kullanılarak çoklu histogram görüntüleme aracılığıyla yapılmıştır. Çoklu sınıf histogram dağılımları, özniteliklerin bazılarında ayırık yığılmanın sağlandığını, diğerlerindeyse belirsiz bölgenin oldukça büyük olduğunu göstermiştir. Bununla birlikte, bu çalışma bağlamında geliştirilen çoklu histogram görüntüleme sınamalarının istatistiksel bir dayanağı olmadığından, bu sınamalardan sadece hangi özniteliklerin en yüksek doğrulukla sınıflandırmayı sağlayacak bölüntülemeyi yapacağına dair bir fikir edinmek olanaklıdır. Bu bağlamda, bu sınamalar öznitelik önseçimi amacıyla kullanılmamıştır.

4.1.10. Uç ve Eksik Değerler

Çalışmada kullanılan öğrenme algoritmaları mantıksal algoritmalar ve bu bağlamda, eksik değerlere karşı sağlam oldukları söylenebilir. Bu koşulun sağlanabilmesi için eksik değerlerin toplam örneklem büyüklüğüne oranının küçük olması gereklidir. Aksi durumda, özgül özniteliğe ait bilgi kaybı öğrenme sürecini olumsuz etkileyecektir. Özellikle uç değerler ve sıfır bölme gibi sorunlar nedeniyle bazı öznitelik değerleri eksik değerle ikame edilmek zorunda kalınmıştır. Veri setindeki eksik değerlerin toplam veri noktası sayısına oranı % 2.9'dur.

Uç değerler, veri setinde özgül özniteliğin aldığı değer yönünde yanlılığa neden olmaktadır. Bu durumda, öğrenilen kural ya da karar ağaçları söz konusu öznitelik ve özgül özniteliğin aldığı değerler yönünde yanlı, dolayısıyla hatalı olacaktır. Mantıksal algoritmaların eksik değerlere olan duyarsızlığı ve sağlamlığı, uç değer alan veri noktalarının eksik değerle değiştirilmesini olanaklı kılmıştır. Böylece, İkinci Bölümde açıklanan eksik değer düzeltme yöntemlerinin uygulanması zorunluluğu yaşanmamıştır.

4.1.11. Veri Setiyle İlgili Kısıtlar

Veri seti, Birinci Bölümde anlatılan, zaman boyutunda örneklem yanlılığı ve verinin kullanıldığı dönem yanlılığı dışında herhangi bir yanlılık taşımamaktadır.

4.2. WEKA Analizleri

4.2.1. Öğrenme Algoritmalarının Sonuçları

Maliyete duyarlı olmayan sınıflandırma problemlerinde WEKA öğrenme algoritmaları çıplak olarak çalıştırılmıştır⁹². WEKA platformundaki ZeroR, OneR, DecisionTable, DecisionStump, j48.J48 ve j48.PART algoritmaları kullanılarak elde edilen sonuçlar Tablo 4.3'te verilmiştir.

⁹² Bu algoritmalar maliyet minimizasyonu objektif fonksiyonu kullanılmamaktadır. Çıktı olarak (sadece) yanlı sınıflandırılmış örneklerin toplam sınıflandırma maliyetleri hesaplanır. Bu nedenle çıplak sınıflandırma algoritmaları olarak adlandırılmaktadırlar.

TABLO 4.3. ÖĞRENME ALGORİTMALARININ SONUÇLARI (1:10)

Öğrenme Algoritmaları

Veri Seti: 1YP3YR_mik_1

İlişki: tcomb-weka.filters.AttributeFilter-V-R1-99,104-169

SINIF	YÜZDE DAĞILIM
BAŞARILI	97.66%
BAŞARISIZ	2.34%

Maliyet Tablosu	
0	1
10	0

Sınıflandırma İstatistikleri	ZeroR	OneR	DecisionTable	DecisionStump	j48.J48	j48.PART
Doğru Sınıflandırılan Örnekler (%)	97.66%	97.54%	97.59%	97.66%	97.43%	96.09%
Yanlış Sınıflandırılan Örnekler (%)	2.34%	2.46%	2.41%	2.34%	2.57%	3.91%
Toplam Maliyet	1000	1005	1003	1000	1001	995
Ortalama Maliyet	0.2341	0.2353	0.2348	0.2341	0.2343	0.2329
Ortalama Mutlak Hata	0.046	0.0246	0.0458	0.0455	0.045	0.0452
Görelî Mutlak Hata (%)	100.05%	53.51%	99.80%	98.99%	97.86%	98.34%
Toplam Örnek Sayısı	4272	4272	4272	4272	4272	4272
Ortalama Maliyet (2:1'e göre normalleştirilmiş) (**)	0.2341	0.2353	0.2348	0.2341	0.2343	0.2329
Sınıflandırma Tablosu						
Brinci Tip Hata (Gerçek Sınıf = BAŞARISIZ, Tahmin = BAŞARILI)	100	100	100	100	99	92
İkinci Tip Hata (Gerçek Sınıf = BAŞARILI, Tahmin = BAŞARISIZ)	0	5	3	0	11	75
"BAŞARISIZ" Sınıflandırma Doğruluğu (%)	0	0	0	0	1	8

(**) 2:1'e göre normalleştirme, 1:20 ile elde edilen toplam maliyetin 1:10 ile elde edilenle karşılaştırılabilmesi için 1:10 maliyet tablosu kullanılarak yapılmıştır.

Tablo 4.3'te çıplak olarak çalıştırılan altı algoritmanın da birbirine yakın doğrulukla sınıflandırma yaptığı görülmektedir. Toplam ve ortalama sınıflandırma maliyetleri de her algoritma için birbirine yakındır. Hedef sınıfının en yüksek doğrulukla sınıflandırıldığı iki algoritma j48.J48 ve j48.PART algoritmalarıdır (% 1 ve % 8).

Tablo 4.4'te, öncelikle 1:10⁹³ kısıtıyla çalıştırılan algoritmaların sınıflandırma doğruluğunun maliyete duyarlılığını ölçmek amacıyla, 1:20 maliyet kısıtı kullanılmıştır. 1:20 kısıtı altında çalıştırılan sınıflandırma sonuçları 1:10 kısıtı altındakilere yakın bulunmuştur. Maliyet bilgisinin değiştirilmesi bu algoritmaların sınıflandırma doğruluğunda herhangi bir iyileşme sağlamamıştır.

Toplam ve ortalama maliyetlerin önceki sınıflandırma sonuçlarına bağlı olarak değişiminin ölçülebilmesi için 1:20 hatalı sınıflandırma sonuçları 1:10 bilgisiyle yeniden hesaplanmıştır. Bu düzeltmeyle elde edilen sonuçlar 1:20 maliyet kısıtı altında yapılan sınıflandırmada j48.PART haricinde diğer

⁹³ 1:10 maliyet kısıtı altında maliyet bilgileri, $M(A_A) = 0$, $M(A_B) = 1$, $M(B_A) = 10$, $M(B_B) = 0$ olarak ifade edilmektedir.

$M(A_A) \equiv \text{Maliyet}(S_{\text{Tahmin}} = \text{BAŞARILI} \mid S_{\text{Gerçek}} = \text{BAŞARILI})$,
 $M(A_B) \equiv \text{Maliyet}(S_{\text{Tahmin}} = \text{BAŞARISIZ} \mid S_{\text{Gerçek}} = \text{BAŞARILI})$,
 $M(B_A) \equiv \text{Maliyet}(S_{\text{Tahmin}} = \text{BAŞARILI} \mid S_{\text{Gerçek}} = \text{BAŞARISIZ})$,
 $M(B_B) \equiv \text{Maliyet}(S_{\text{Tahmin}} = \text{BAŞARISIZ} \mid S_{\text{Gerçek}} = \text{BAŞARISIZ})$
 olarak tanımlanmaktadır.

algoritmaların sınıflandırma toplam ve ortalama maliyetlerinin aynı kaldığı ya da azaldığına işaret etmektedir.

TABLO 4.4. ÖĞRENME ALGORİTMALARININ SONUÇLARI (1:20)

Öğrenme Algoritmaları

Veri Seti: 1YP3YR_mik_1

İlişki: tcmb-weka.filters.AttributeFilter-V-R1-99,104-169

SINIF	YÜZDE DAĞILIM
BAŞARILI	97.66%
BAŞARISIZ	2.34%

Maliyet Tablosu	
0	1
20	0

Sınıflandırma İstatistikleri	ZeroR	OneR	DecisionTable	DecisionStump	j48.J48	j48.PART
Doğru Sınıflandırılan Örnekler (%)	97.66%	97.61%	97.66%	97.66%	97.33%	95.86%
Yanlış Sınıflandırılan Örnekler (%)	2.34%	2.39%	2.34%	2.34%	2.67%	4.14%
Toplam Maliyet	2000	2002	2000	2000	1957	1944
Ortalama Maliyet	0.4682	0.4686	0.4682	0.4682	0.4581	0.4551
Ortalama Mutlak Hata	0.046	0.0239	0.0452	0.0457	0.0441	0.0474
Görelî Mutlak Hata (%)	100.0514	51.9808	98.325	99.422	95.9454	103.1693
Toplam Örnek Sayısı	4272	4272	4272	4272	4272	4272
Ortalama Maliyet (2:1'e göre normalleştirilmiş) (**)	0.2341	0.2346	0.2341	0.2341	0.2310	0.2374
Toplam Maliyet (2:1'e göre normalleştirilmiş) (**)	1000	1002	1000	1000	987	1014
Sınıflandırma Tablosu						
Brinci Tip Hata (Gerçek Sınıf = BAŞARISIZ, Tahmin = BAŞARILI)	100	100	100	100	97	93
İkinci Tip Hata (Gerçek Sınıf = BAŞARILI, Tahmin = BAŞARISIZ)	0	2	0	0	17	84
"BAŞARISIZ" Sınıflandırma Doğruluğu (%)	0	0	0	0	3	7

(**) 2:1'e göre normalleştirme, 1:20 ile elde edilen toplam maliyetin 1:10 ile elde edilenle karşılaştırılabilmesi için 1:10 maliyet tablosu kullanılarak yapılmıştır.

Çıplak algoritmaların maliyet minimizasyonu amacını taşıyor olmasına rağmen, 1:20 kısıtı altında, düzeltilmiş maliyetlerin daha düşük olması ilginç bir sonuçtur. Yeni maliyet bilgisi altında çalıştırılan algoritmaların hedef sınıfı sınıflandırma doğruluğu ZeroR, OneR, DecisionTable ve DecisionStump algoritmaları altında değişmezken, j48.J48 algoritmasında artmış j48.PART için azalmıştır (% 3 ve % 7).

Çıplak öğrenme algoritmalarıyla elde edilen kurallar için dikkat edilmesi gereken, bu kuralların maliyet bilgisinden bağımsız öğrenilmiş olmalarıdır. Yani, kurallar çıplak algoritmalar için maliyete duyarsızdır. Bu durum her iki maliyet kısıtı altında çalıştırılan algoritmalar için de gözlemlenmiştir.

4.2.2. Meta Sınıflandırma Sonuçları

4.2.2.1. Maliyete Duyarlı Öğrenme ve Meta Sınıflandırma Yöntemleri

CostSensitiveClassifier ve MetaCost yöntemleriyle meta sınıflandırmada maliyete duyarlılığın ölçülmesinde maliyet bilgisinin etkisini sorgulamak amacıyla, çıplak sınıflandırıcılarda da kullanılan, iki farklı maliyet tablosu (bilgisi) kullanılmıştır (1:10, 1:20).

M(A_B) gerçek sınıf etiketi başarılı olan bir örneğin başarısız, M(B_A) ise gerçek sınıf etiketi başarısız olan bir örneğin başarılı olarak hatalı sınıflandırılmasının maliyetidir. Maliyetler tek bir örnek için geçerli ve bütün örnekler için aynıdır; maliyet tablolarının her ikisinde de M(A_B) 1 iken, M(B_A), 10 veya 20 değerlerini almaktadır.

WEKA platformu Meta sınıflandırıcılarından CostSensitiveClassifier ile veri maliyet kısıtları altında (1:10 ve 1:20) elde edilen sonuçlar Tablo 4.5'te sunulmaktadır.

TABLO 4.5. COSTSENSITIVECLASSIFIER SONUÇLARI (1:10)

Meta Sınıflandırma (CostSensitiveClassifier)

Veri Seti:1YP3YR_mik_1

İlişki:tomb-weka.filters.AttributeFilter-V-R1-99,104-169

SINIF	YÜZDE DAĞILIM
BAŞARILI	97.66%
BAŞARISIZ	2.34%

Maliyet Tablosu	
0	1
10	0

Sınıflandırma İstatistikleri	ZeroR	OneR (*)	DecisionTable	DecisionStump	j48.J48	j48.PART
Doğru Sınıflandırılan Örnekler (%)	97.66%	-	97.26%	96.54%	97.31%	95.22%
Yanlış Sınıflandırılan Örnekler (%)	2.34%	-	2.74%	3.46%	2.69%	4.78%
Toplam Maliyet	1000	-	999	1039	997	1032
Ortalama Maliyet	0.2341	-	0.2338	0.2432	0.2334	0.2416
Ortalama Mutlak Hata	0.0234	-	0.0274	0.0346	0.0269	0.0478
Görelî Mutlak Hata (%)	50.96%	-	59.63%	75.42%	58.61%	103.96%
Toplam Örnek Sayısı	4272	-	4272	4272	4272	4272
Ortalama Maliyet (2:1'e göre normalleştirilmiş) (**)	0.2341	-	0.2338	0.2432	0.2334	0.2416
Sınıflandırma Tablosu						
Brinci Tip Hata (Gerçek Sınıf = BAŞARISIZ, Tahmin = BAŞARILI)	100	-	98	99	98	92
İkinci Tip Hata (Gerçek Sınıf = BAŞARILI, Tahmin = BAŞARISIZ)	0	-	19	49	17	112
"BAŞARISIZ" Sınıflandırma Doğruluğu (%)	0	-	2	1	2	8

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından CostSensitiveClassifier sonuçları bulunmamaktadır.

(**) 2:1'e göre normalleştirme, 1:20 ile elde edilen toplam maliyetin 1:10 ile elde edilenle karşılaştırılabilmesi için 1:10 maliyet tablosu kullanılarak yapılmıştır.

4.2.2.2. Sınıflandırma Sonuçları

ZeroR ile her iki maliyet kısıtı altında da en yüksek doğrulukla sınıflandırma yapılmıştır. Tablo 4.5'te görüldüğü gibi, ZeroR algoritmasının toplam ve ortalama maliyetleri 1:10 maliyet kısıtı altında DecisionTable ve j48.J48 algoritmalarıyla elde edilenden daha yüksektir. Bu iki algoritma hedef sınıf örneklerini % 2 doğrulukla sınıflandırırken, çoğunluk sınıf örneklerinden sadece sırasıyla 19 ve 17 örneği başarısız olarak sınıflandırmış ve bu bağlamda toplam ve ortalama maliyetleri de düşmüştür. Diğer taraftan, j48.PART algoritmasıyla 1:10 kısıtı altında hedef sınıf örnekleri en yüksek doğrulukla (% 8) sınıflandırılmıştır. Bununla birlikte, aynı algoritma başarılı sınıf örneklerinden 112 tanesini hatalı sınıflandırdığından, toplam ve ortalama maliyeti DecisionTable ve j48.J48 algoritmalarından daha yüksektir. DecisionStump ile 1:10 kısıtı altında hedef sınıf örnekleri en düşük başarıyla sınıflandırılmıştır. Bu nedenle DecisionStump ile elde edilen toplam ve ortalama maliyetler diğer algoritmalarınkinden daha yüksek olmuştur.

Maliyete duyarlılığın ölçülmesi amacıyla oluşturulan diğer maliyet kısıtı (1:20) altında CostSensitiveClassifier meta sınıflandırıcısıyla çalıştırılan algoritmaların sonuçları Tablo 4.6'da sunulmaktadır.

TABLO 4.6. COSTSENSITIVECLASSIFIER SONUÇLARI (1:20)

Meta Sınıflandırma (CostSensitiveClassifier)

Veri Seti:

1YP3YR_mik_1

İlişki:

tcmw-veka.filters.AttributeFilter-V-R1-99,104-169

SINIF	YÜZDE DAĞILIM
BAŞARILI	97.66%
BAŞARISIZ	2.34%

Maliyet Tablosu	
0	1
20	0

Sınıflandırma İstatistikleri	ZeroR	OneR (*)	DecisionTable	DecisionStump	j48.J48	j48.PART
Doğru Sınıflandırılan Örnekler (%)	97.66%	-	95.55%	86.66%	94.24%	94.15%
Yanlış Sınıflandırılan Örnekler (%)	2.34%	-	4.45%	13.34%	5.76%	5.85%
Toplam Maliyet	2000	-	1957	2166	1899	1884
Ortalama Maliyet	0.4682	-	0.4581	0.507	0.4445	0.441
Ortalama Mutlak Hata	0.0234	-	0.0445	0.1334	0.0576	0.0585
Görelî Mutlak Hata (%)	50.96%	-	96.83%	290.48%	125.37%	127.40%
Toplam Örnek Sayısı	4272	-	4272	4272	4272	4272
Ortalama Maliyet (2:1'e göre normalleştirilmiş) (**)	0.2341	-	0.2404	0.3104	0.2409	0.2397
Toplam Maliyet (2:1'e göre normalleştirilmiş) (**)	1000	-	1027	1326	1029	1024
Sınıflandırma Tablosu						
İkinci Tip Hata (Gerçek Sınıf = BAŞARISIZ, Tahmin = BAŞARILI)	100	-	93	84	87	86
İkinci Tip Hata (Gerçek Sınıf = BAŞARILI, Tahmin = BAŞARISIZ)	0	-	97	486	159	164
"BAŞARISIZ" Sınıflandırma Doğruluğu (%)	0	-	7	16	13	14

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından CostSensitiveClassifier sonuçları bulunmamaktadır.

(**) 2:1'e göre normalleştirme, 1:20 ile elde edilen toplam maliyetin 1:10 ile elde edilenle karşılaştırılabilmesi için 1:10 maliyet tablosu kullanılarak yapılmıştır.

Daha yüksek maliyet kısıtı altında çalıştırılan algoritmalar arasında ZeroR ile en yüksek doğruluk ve en yüksek toplam ve ortalama maliyetlerle sınıflandırma yapılmıştır. ZeroR algoritmasının hedef sınıflandırma başarısı % 0'dır. DecisionTable, DecisionStump, j48.J48 ve j48.PART'ın hedef sınıflandırma oranları sırasıyla % 7, % 16, % 13 ve % 14 olarak gözlenmiştir. 1:20 maliyet kısıtı altında hedef sınıfı en yüksek doğrulukla (% 16) sınıflandıran algoritma DecisionStump algoritmasıdır. Aynı zamanda, bu algoritmanın aynı maliyet bilgisiyle toplam ve ortalama maliyetleri de diğerlerinden daha yüksek, genel sınıflandırma doğruluğuysa daha düşük olarak gerçekleşmiştir. Maliyete duyarlı sınıflandırmada kullanılan diğer meta sınıflandırıcı MetaCost'un 1:10 maliyet kısıtı altındaki sonuçları Tablo 4.7'de sunulmaktadır.

TABLO 4.7. METACOST SONUÇLARI (1:10)

Meta Sınıflandırma (MetaCost)

Veri Seti:

1YP3YR_mik_1

İlişki:

tcmb-weka.filters.AttributeFilter-V-R1-99,104-169

SINIF	YÜZDE DAĞILIM
BAŞARILI	97.66%
BAŞARISIZ	2.34%

Maliyet Tablosu	
0	1
10	0

Sınıflandırma İstatistikleri	ZeroR	OneR	DecisionTable	DecisionStump	j48.J48	j48.PART
Doğru Sınıflandırılan Örnekler (%)	97.66%	97.21%	97.66%	97.64%	93.82%	97.66%
Yanlış Sınıflandırılan Örnekler (%)	2.34%	2.79%	2.34%	2.36%	6.18%	2.34%
Toplam Maliyet	1000	1010	1000	1001	1047	1000
Ortalama Maliyet	0.2341	0.2364	0.2341	0.2343	0.2451	0.2341
Ortalama Mutlak Hata	0.0234	0.0279	0.0234	0.0236	0.0618	0.0234
Görelî Mutlak Hata (%)	50.96%	60.64%	50.96%	51.47%	134.54%	50.96%
Toplam Örnek Sayısı	4272	4272	4272	4272	4272	4272
Ortalama Maliyet (2:1'e göre normalleştirilmiş) (**)	0.2341	0.2364	0.2341	0.2343	0.2451	0.2341
Sınıflandırma Tablosu						
Brinci Tip Hata (Gerçek Sınıf = BAŞARISIZ, Tahmin = BAŞARILI)	100	99	100	100	87	100
İkinci Tip Hata (Gerçek Sınıf = BAŞARILI, Tahmin = BAŞARISIZ)	0	20	0	1	177	0
"BAŞARISIZ" Sınıflandırma Doğruluğu (%)	0	1	0	0	13	0

(**) 2:1'e göre normalleştirme, 1:20 ile elde edilen toplam maliyetin 1:10 ile elde edilenle karşılaştırılabilmesi için 1:10 maliyet tablosu kullanılarak yapılmıştır.

MetaCost, 1:10 ve 1:20 maliyet bilgileriyle çalıştırılmıştır. 1:10 maliyet bilgisiyle, j48.J48 dışında diğer algoritmaların sınıflandırma doğruluğu maliyet ve hedef sınıflandırma doğruluğu performansları birbirine yakın olarak gerçekleşmiştir. j48.J48 algoritmasının genel sınıflandırma performansı düşük ve maliyetleri yüksek iken hedef sınıflandırma doğruluğu % 13 olarak gözlemlenmektedir. OneR algoritması hedef sınıfı sadece % 1 başarıyla sınıflandırabilmiştir.

TABLO 4.8. METACOST SONUÇLARI (1:20)

Meta Sınıflandırma (MetaCost)

Veri Seti: 1YP3YR_mik_1

İlişki: tcmb-weka.filters.AttributeFilter-V-R1-99,104-169

SINIF	YÜZDE DAĞILIM
BAŞARILI	97.66%
BAŞARISIZ	2.34%

Maliyet Tablosu	
0	1
20	0

Sınıflandırma İstatistikleri	ZeroR	OneR	DecisionTable	DecisionStump	j48.J48	j48.PART
Doğru Sınıflandırılan Örnekler (%)	97.66%	97.21%	97.66%	93.94%	93.21%	97.66%
Yanlış Sınıflandırılan Örnekler (%)	2.34%	2.79%	2.34%	6.06%	6.79%	2.34%
Toplam Maliyet	2000	2000	2000	2083	1867	2000
Ortalama Maliyet	0.4682	0.4682	0.4682	0.4876	0.437	0.4682
Ortalama Mutlak Hata	0.0234	0.0279	0.0234	0.0606	0.0679	0.0234
Görelî Mutlak Hata (%)	0.509615	0.606442	0.509615	1.319903	1.477884	50.96%
Toplam Örnek Sayısı	4272	4272	4272	4272	4272	4272
Ortalama Maliyet (2:1'e göre normalleştirilmiş) (**)	0.2341	0.2364	0.2341	0.2629	0.2427	0.2341
Toplam Maliyet (2:1'e göre normalleştirilmiş) (**)	1000	1010	1000	1123	1037	1000
Sınıflandırma Tablosu						
Brinci Tip Hata (Gerçek Sınıf = BAŞARISIZ, Tahmin = BAŞARILI)	100	99	100	96	83	100
İkinci Tip Hata (Gerçek Sınıf = BAŞARILI, Tahmin = BAŞARISIZ)	0	20	0	163	207	0
"BAŞARISIZ" Sınıflandırma Doğruluğu (%)	0	1	0	4	17	0

(**) 2:1'e göre normalleştirme, 1:20 ile elde edilen toplam maliyetin 1:10 ile elde edilenle karşılaştırılabilmesi için 1:10 maliyet tablosu kullanılarak yapılmıştır.

1:20 maliyet kısıtı altında ZeroR, OneR, DecisionTable ve j48.PART kural öğrenme algoritmalarının genel sınıflandırma, maliyet ve hedef sınıflandırma performansları aynı kalırken, karar ağacı algoritmaları DecisionStump ve j48.J48'in hedef sınıfı öğrenme performansları marjinal düzeyde artmıştır (% 4 ve % 17). Karar ağaçlarından DecisionStump'ın toplam ve ortalama maliyetleri 1:10 kısıtı altındakine göre artarken j48.J48'in maliyetlerinde bir düşme gözlemlenmiştir.

4.2.2.3. Meta Sınıflandırıcılar ve Alt Örneklem (10 Set)

Çıplak sınıflandırıcılar meta sınıflandırıcılarla birleştirildiğinde, özünde maliyete duyarlı sınıflandırma yapacak şekilde tasarlanmadıkları için, maliyet bilgisi değişse de aynı kuralları öğrenmektedir. Bu bağlamda, kuralların maliyet bilgisine duyarlılığının sınanabilmesi amacıyla 4,272 örnekten oluşan veri seti sınıf dağılımları değişecek şekilde 10 alt örneklem halinde yeniden düzenlenmiştir. Elde edilen veri setlerinde indirgeme işlemi rassal olarak yapılmıştır⁹⁴. İndirgenmiş veri setlerinin maliyet duyarlılığını ölçmek amacıyla CostSensitiveClassifier ve MetaCost meta sınıflandırıcıları kullanılmıştır. Alt örneklemle ilgili istatistikler Tablo 4.9'da sunulmaktadır.

⁹⁴ Rassal elemelerde, her örneğe rassal sayılar atanmış ve belirli bir eşğin üzerindeki örnekler elenmiştir.

TABLO 4.9. ALT ÖRNEKLEM İSTATİSTİKLERİ

VERİ SETİ	SET 1	SET 2	SET 3	SET 4	SET 5	SET 6	SET 7	SET 8	SET 9	SET 10
ÖRNEKLEM BÜYÜKLÜĞÜ	2103	2336	2549	2754	2973	3186	3397	3609	3842	4069
ÖZNİTELİK SAYISI	165	165	165	165	165	165	165	165	165	165
MALİYET MATRİKSİ	0,1;10,0	0,1;10,0	0,1;10,0	0,1;10,0	0,1;10,0	0,1;10,0	0,1;10,0	0,1;10,0	0,1;10,0	0,1;10,0
SINIF DAĞILIMI (Örnek Sayısı)										
başarılı	2003	2236	2449	2654	2873	3086	3297	3509	3742	3969
başarısız	100	100	100	100	100	100	100	100	100	100
SINIF DAĞILIMI (Yüzde)										
başarılı	95.24	95.72	96.08	96.37	96.64	96.86	97.06	97.23	97.40	97.54
başarısız	4.76	4.28	3.92	3.63	3.36	3.14	2.94	2.77	2.60	2.46

İndirgenmiş alt örneklem, WEKA meta sınıflandırıcıları CostSensitiveClassifier ve MetaCost üzerine bindirilen DecisionStump, DecisionTable, j48.j48, j48.PART, OneR ve ZeroR çıplak sınıflandırıcılarıyla öğrenilmiştir⁹⁵.

Maliyete duyarlı genel ve hedef sınıflandırma doğruluğu sonuçları Tablo 4.10 ve Tablo 4.11’de sunulmaktadır.

TABLO 4.10. GENEL SINIFLANDIRMA DOĞRULUĞU (*) (%)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklem									
		SET 1	SET 2	SET 3	SET 4	SET 5	SET 6	SET 7	SET 8	SET 9	SET 10
costSensitive	DecisionStump	95,96	90,45	78,85	90,81	96,17	96,58	96,76	96,90	96,69	96,76
MetaCost	DecisionStump	96,00	91,90	95,57	96,11	96,03	96,74	97,03	97,17	97,24	97,52
costSensitive	DecisionTable	95,10	90,37	94,78	94,00	96,64	96,11	96,17	96,95	95,21	97,25
MetaCost	DecisionTable	96,15	95,16	95,80	96,30	96,47	96,58	96,88	97,09	97,11	97,49
costSensitive	J48	96,58	93,71	95,29	95,24	95,49	95,17	95,26	96,37	96,02	97,15
MetaCost	J48	89,82	88,66	89,64	90,09	91,29	92,09	92,96	92,66	92,69	93,88
costSensitive	J48PART	94,39	92,47	93,37	93,46	93,88	94,22	94,58	94,79	94,51	95,33
MetaCost	J48PART	87,45	85,36	85,25	86,24	87,66	87,85	88,22	90,05	90,06	90,07
costSensitive	OneR	-	-	-	-	-	-	-	-	-	-
MetaCost	OneR	95,00	94,48	94,98	95,35	95,89	95,79	96,41	96,70	97,03	97,15
costSensitive	ZeroR	95,24	95,72	96,08	96,37	96,64	96,86	97,06	97,23	97,40	97,54
MetaCost	ZeroR	95,25	95,72	96,08	96,37	96,64	96,86	97,06	97,23	97,40	97,54
	Maksimum	96,58	95,72	96,08	96,37	96,64	96,86	97,06	97,23	97,40	97,54
costSensitive	Minimum	94,39	90,37	78,85	90,81	93,88	94,22	94,58	94,79	94,51	95,33
	Maksimum	96,15	95,72	96,08	96,37	96,64	96,86	97,06	97,23	97,40	97,54
MetaCost	Minimum	87,45	85,36	85,25	86,24	87,66	87,85	88,22	90,05	90,06	90,07

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır.

10 alt örneklemle yapılan sınıflandırmalardan CostSensitiveClassifier meta sınıflandırıcısı altında en yüksek sınıflandırma doğruluğu ZeroR algoritmasıyla 4,069 örnekten oluşan Set 10’la (% 97.54) elde edilmiştir. Bu sınıflandırma doğruluğu aynı zamanda, veri setinin sınıf dağılımında çoğunluk (BAŞARILI) sınıfının toplam içindeki payıdır.

⁹⁵ OneR, bir dağılım sınıflandırıcısı olmadığından CostSensitiveClassifier meta sınıflandırıcısı ile sınıflandırma çözümleme sonuçları alınmamıştır.

CostSensitiveClassifier, hedef (BAŞARISIZ) sınıfı en yüksek doğrulukla sınıflandıran kuralları DecisionStump algoritmasıyla Set 3'le öğrenmiştir (% 52).

TABLO 4.11. HEDEF SINIFLANDIRMA DOĞRULUĞU^(*) (%)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklemeler									
		SET 1	SET 2	SET 3	SET 4	SET 5	SET 6	SET 7	SET 8	SET 9	SET 10
costSensitive	DecisionStump	15,00	21,00	52,00	12,00	0,00	1,00	0,00	0,00	1,00	0,00
MetaCost	DecisionStump	17,00	16,00	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
costSensitive	DecisionTable	41,00	11,00	5,00	8,00	0,00	2,00	4,00	1,00	7,00	2,00
MetaCost	DecisionTable	33,00	11,00	2,00	3,00	3,00	2,00	1,00	2,00	2,00	1,00
costSensitive	J48	43,00	16,00	8,00	1,00	8,00	9,00	8,00	4,00	6,00	2,00
MetaCost	J48	47,00	26,00	20,00	24,00	22,00	16,00	17,00	16,00	20,00	18,00
costSensitive	J48PART	36,00	12,00	18,00	9,00	11,00	11,00	14,00	7,00	7,00	9,00
MetaCost	J48PART	52,00	37,00	24,00	25,00	24,00	21,00	26,00	26,00	20,00	23,00
costSensitive	OneR	-	-	-	-	-	-	-	-	-	-
MetaCost	OneR	21,00	6,00	2,00	2,00	1,00	0,00	1,00	0,00	2,00	0,00
costSensitive	ZeroR	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
MetaCost	ZeroR	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Maksimum	43,00	21,00	52,00	12,00	11,00	11,00	14,00	7,00	7,00	9,00
costSensitive	Minimum	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Maksimum	52,00	37,00	24,00	25,00	24,00	21,00	26,00	26,00	20,00	23,00
MetaCost	Minimum	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır.

CostSensitiveClassifier ile meta sınıflandırmada DecisionStump algoritması dışındaki diğer yöntemlerde eğitim setinin ölçeği büyüdükçe genel sınıflandırma doğruluğu artarken, hedef sınıflandırma doğruluğunun azaldığı gözlemlenmektedir.

TABLO 4.12. ORTALAMA MALİYETLER^(*)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklemeler									
		SET 1	SET 2	SET 3	SET 4	SET 5	SET 6	SET 7	SET 8	SET 9	SET 10
costSensitive	DecisionStump	0,4042	0,3998	0,3809	0,3794	0,3411	0,3139	0,2973	0,2804	0,2650	0,2536
MetaCost	DecisionStump	0,3951	0,4045	0,3903	0,3656	0,3424	0,3151	0,2947	0,2776	0,2618	0,2460
costSensitive	DecisionTable	0,3019	0,4392	0,3876	0,3606	0,3364	0,3158	0,2926	0,2774	0,2657	0,2443
MetaCost	DecisionTable	0,3252	0,3913	0,3880	0,3540	0,3290	0,3110	0,2935	0,2735	0,2585	0,2440
costSensitive	J48	0,2782	0,3866	0,3719	0,3417	0,3236	0,3054	0,2911	0,2757	0,2600	0,2453
MetaCost	J48	0,3286	0,3985	0,3860	0,3475	0,3232	0,3164	0,2903	0,2829	0,2605	0,2426
costSensitive	J48PART	0,3300	0,4144	0,3558	0,3627	0,3306	0,3092	0,2820	0,2840	0,2728	0,2480
MetaCost	J48PART	0,3310	0,3891	0,4158	0,3827	0,3535	0,3446	0,3138	0,2840	0,2868	0,2696
costSensitive	OneR	-	-	-	-	-	-	-	-	-	-
MetaCost	OneR	0,3880	0,4174	0,3962	0,3667	0,3407	0,3245	0,2982	0,2823	0,2592	0,2497
costSensitive	ZeroR	0,4755	0,4281	0,3923	0,3631	0,3364	0,3139	0,2944	0,2771	0,2603	0,2458
MetaCost	ZeroR	0,4755	0,4281	0,3923	0,3631	0,3364	0,3139	0,2944	0,2771	0,2603	0,2458
	Maksimum	0,4755	0,4392	0,3923	0,3794	0,3411	0,3158	0,2973	0,2840	0,2728	0,2536
costSensitive	Minimum	0,2782	0,3866	0,3558	0,3417	0,3236	0,3054	0,2820	0,2757	0,2600	0,2443
	Maksimum	0,4755	0,4281	0,4158	0,3827	0,3535	0,3446	0,3138	0,2840	0,2868	0,2696
MetaCost	Minimum	0,3252	0,3891	0,3860	0,3475	0,3232	0,3110	0,2903	0,2735	0,2585	0,2426

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır.

MetaCost meta sınıflandırıcısıyla yapılan çözümlemelerde ZeroR algoritması, genel sınıflandırma doğruluğu Set 10'la en yüksek (% 97.54) algoritma olarak bulunmuştur. Hedef sınıflandırmada en yüksek sınıflandırma doğruluğu (% 52) ile j48.PART ve 2,103 örnekten oluşan Set 1'le elde edilmiştir.

Meta sınıflandırmalarda, ortalama maliyet örneklem büyüklüğüyle ters orantılı olarak değişmektedir. CostSensitiveClassifier'la en düşük ortalama maliyet DecisionTable ve Set 10'la (0.2440), MetaCost'la en düşük ortalama maliyet (0.2426) j48.48 ve Set 10'la elde edilmiştir (Tablo 4.12).

4.2.2.4. CostSensitiveClassifier ve MetaCost

Hedef sınıfın en yüksek doğrulukla tahmini her iki meta sınıflandırıcıyla % 52 olarak gerçekleşmiştir. MetaCost bu sınıflandırma doğruluğuna, daha küçük ölçekli bir veri seti olan Set 1 (2,103 örnek), 0.3310 ortalama maliyet ve j48.PART algoritmasıyla ulaşırken, CostSensitiveClassifier aynı performansı Set 3 (2,549 örnek), 0.3809 ortalama maliyet ve DecisionStump algoritmasıyla elde edebilmiştir. Bu bağlamda, CostSensitiveClassifier'ın MetaCost'a göre daha verimsiz bir sınıflandırma yaptığı düşünülebilir. Fakat, kural öğrenme algoritması olmasına rağmen j48.PART, birden fazla kural ve daha karmaşık bir örüntü elde etmiş ve bu açıdan daha basit ve anlaşılır kurallar öğrenen DecisionStump algoritmasından basitlik ölçütüne göre daha geride kalmıştır. Bu iki meta sınıflandırıcıdan hangisinin kullanılacağı, kullanım (kural) karmaşıklığının da bir maliyet etmeni olarak alınmasından sonra verilecek bir karardır⁹⁶.

4.2.2.5. Sınıflandırma Doğruluğu ve Maliyet Verimliliği: Senaryo Analizleri

Meta sınıflandırma yöntemleriyle maliyete duyarlı duruma getirilen öğrenme algoritmaları, genel ve hedef sınıflandırma doğruluğu ve ortalama maliyet ölçütleriyle karşılaştırılmıştır. Bununla birlikte, algoritmaların

⁹⁶ Basit ya da az öncüllü kurallar daha genel kurallardır ve anlamlı olduğu sürece karmaşık kurallara tercih edilirler.

değerlendirilmesinde sadece bu ölçütler yeterli değildir. Bu bağlamda, özellikle, hedef sınıflandırma ve maliyet verimliliğinin ölçülmesi amacıyla yedi adet senaryo oluşturulmuştur.

Senaryo 1. Ortalama Maliyet-1 (OM-1)⁹⁷: Ortalama maliyetler, örneklem büyüklüğü yerine en büyük örneklemin büyüklüğüne göre hesaplanmıştır.

TABLO 4.13. ORTALAMA MALİYET-1^(*)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklem									
		SET 1	SET 2	SET 3	SET 4	SET 5	SET 6	SET 7	SET 8	SET 9	SET 10
costSensitive	DecisionStump	0,2089	0,1942	0,1180	0,2163	0,2458	0,2433	0,2458	0,2458	0,2433	0,2458
MetaCost	DecisionStump	0,2042	0,2067	0,2411	0,2460	0,2460	0,2460	0,2460	0,2460	0,2460	0,2460
costSensitive	DecisionTable	0,1560	0,2298	0,2445	0,2371	0,2568	0,2519	0,2470	0,2543	0,2396	0,2519
MetaCost	DecisionTable	0,1681	0,2221	0,2443	0,2418	0,2418	0,2443	0,2467	0,2443	0,2443	0,2467
costSensitive	J48	0,1438	0,2101	0,2298	0,2470	0,2298	0,2273	0,2298	0,2396	0,2347	0,2445
MetaCost	J48	0,1698	0,2214	0,2362	0,2264	0,2313	0,2460	0,2436	0,2460	0,2362	0,2411
costSensitive	J48PART	0,1706	0,2295	0,2148	0,2369	0,2320	0,2320	0,2246	0,2418	0,2418	0,2369
MetaCost	J48PART	0,1711	0,2079	0,2399	0,2374	0,2399	0,2473	0,2350	0,2350	0,2497	0,2423
costSensitive	OneR	-	-	-	-	-	-	-	-	-	-
MetaCost	OneR	0,2005	0,2374	0,2472	0,2472	0,2497	0,2521	0,2497	0,2521	0,2472	0,2521
costSensitive	ZeroR	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458
MetaCost	ZeroR	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458	0,2458
	Maksimum	0,2458	0,2458	0,2458	0,2470	0,2568	0,2519	0,2470	0,2543	0,2458	0,2519
costSensitive	Minimum	0,1438	0,1942	0,1180	0,2163	0,2298	0,2273	0,2246	0,2396	0,2347	0,2369
	Maksimum	0,2458	0,2458	0,2472	0,2472	0,2497	0,2521	0,2497	0,2521	0,2497	0,2521
MetaCost	Minimum	0,1681	0,2067	0,2362	0,2264	0,2313	0,2443	0,2350	0,2350	0,2362	0,2411

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır.

Tablo 4.13'te maliyet rakamlarının set 10'daki 4,069 örneğin birim başına düşen maliyetleriyle karşılaştırılabilmesi amacıyla ortalama maliyet set 10'un örneklem büyüklüğüyle (4,069) normalleştirilmiştir. 10 veri seti arasında (ZeroR ile yapılan sınıflandırmalar dışında) elde edilen ortalama maliyetlerin en düşüğü Set 1'in ortalama maliyetleridir. Normalleştirme süreci, ortalama maliyetleri küçük ölçekli örneklemelerin lehine düşürmektedir.

Senaryo 2. Ortalama Maliyet-2 (OM-2)⁹⁸: Alt örneklemelerin birinci ve ikinci tip hataları, örneklem büyüklük katsayılarıyla düzeltilmiş ve ortalama maliyetler en büyük örneklemin büyüklüğüyle normalleştirilmiştir.

⁹⁷ Ortalama Maliyet-1 (OM-1)

$OM-1_i \equiv \{[(a_i|Set_i) \times \mu(\alpha)] + [(b_i|Set_i) \times \mu(\beta)]\} / \zeta(Set_{10})$; $i = 1, 2, \dots, 10$ olarak tanımlanmaktadır. $(a_i|Set_i)$, Set_i 'nin birinci tip hatası, $\mu(\alpha)$, birinci tip hata maliyeti, $(b_i|Set_i)$, Set_i 'nin ikinci tip hatası, $\mu(\beta)$, ikinci tip hata maliyeti, $\zeta(Set_{10})$, Set 10'un örneklem büyüklüğüdür (4,069).

⁹⁸ Ortalama Maliyet-2 (OM-2),

$\Gamma_i \equiv \zeta(Set_i) / \zeta(Set_{10})$; $i = 1, 2, \dots, 10$ iken,
 $OM-2_i \equiv \Gamma_i \times \{[(a_i|Set_i) \times \mu(\alpha)] + [(b_i|Set_i) \times \mu(\beta)]\} / \zeta(Set_{10})$ olarak tanımlanmaktadır. $\zeta(Set_i)$ i örnekleminin büyüklüğüdür.

TABLO 4.14. ORTALAMA MALİYET-2^(*)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklem									
		SET 1	SET 2	SET 3	SET 4	SET 5	SET 6	SET 7	SET 8	SET 9	SET 10
costSensitive	DecisionStump	0,4042	0,3382	0,1883	0,3195	0,3364	0,3107	0,2944	0,2771	0,2577	0,2458
MetaCost	DecisionStump	0,3951	0,3600	0,3848	0,3634	0,3367	0,3142	0,2946	0,2773	0,2605	0,2460
costSensitive	DecisionTable	0,3019	0,4002	0,3903	0,3504	0,3515	0,3217	0,2958	0,2868	0,2537	0,2519
MetaCost	DecisionTable	0,3252	0,3869	0,3899	0,3573	0,3309	0,3120	0,2955	0,2754	0,2587	0,2467
costSensitive	J48	0,2782	0,3660	0,3668	0,3649	0,3145	0,2903	0,2753	0,2702	0,2486	0,2445
MetaCost	J48	0,3286	0,3857	0,3770	0,3344	0,3165	0,3142	0,2917	0,2774	0,2501	0,2411
costSensitive	J48PART	0,3300	0,3998	0,3429	0,3500	0,3175	0,2963	0,2691	0,2726	0,2561	0,2369
MetaCost	J48PART	0,3310	0,3622	0,3829	0,3508	0,3283	0,3158	0,2815	0,2649	0,2645	0,2423
costSensitive	OneR	-	-	-	-	-	-	-	-	-	-
MetaCost	OneR	0,3880	0,4135	0,3947	0,3653	0,3417	0,3220	0,2991	0,2843	0,2618	0,2521
costSensitive	ZeroR	0,4755	0,4281	0,3923	0,3631	0,3364	0,3139	0,2944	0,2771	0,2603	0,2458
MetaCost	ZeroR	0,4755	0,4281	0,3923	0,3631	0,3364	0,3139	0,2944	0,2771	0,2603	0,2458
	Maksimum	0,4755	0,4281	0,3923	0,3649	0,3515	0,3217	0,2958	0,2868	0,2603	0,2519
costSensitive	Minimum	0,2782	0,3382	0,1883	0,3195	0,3145	0,2903	0,2691	0,2702	0,2486	0,2369
	Maksimum	0,4755	0,4281	0,3947	0,3653	0,3417	0,3220	0,2991	0,2843	0,2645	0,2521
MetaCost	Minimum	0,3252	0,3600	0,3770	0,3344	0,3165	0,3120	0,2815	0,2649	0,2501	0,2411

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır.

Tablo 4.14 alt örneklemdeki birinci ve ikinci tip hataların, örneklem büyüklüklerine göre düzeltilerek, Set 10'a göre normalleştirilmesiyle oluşturulmuştur. Düzenleme sonucunda gözlemlenen, uyarlanmış ortalama maliyetlerin örneklem büyüklüğüyle ters orantılı olarak değişmesidir. En yüksek uyarlanmış ortalama maliyetler Set 1'in ortalama maliyetleridir. Böyle bir maliyet yapısının oluşması, hedef hatalı sınıflandırma maliyetinin yüksek olmasından kaynaklanmaktadır. Ayrıca, küçük ölçekli veri setlerinde hedef sınıf örneklerinin örneklem içindeki oranının büyük olmasının da bu sonuca bir etkisi olmuştur.

Senaryo 3. Ortalama Maliyet-3 (OM-3_i)⁹⁹: En büyük örneklem birinci ve ikinci tip hataları, örneklem büyüklük katsayılarıyla düzeltilmiş ve ortalama maliyetler en büyük örneklem büyüklüğüyle normalleştirilmiştir.

⁹⁹ Ortalama Maliyet-3 (OM-3_i),

OM-3_i $\equiv \{[(a_i | \text{Set}_{10}) \times \mu(\alpha)] + [(b_i | \text{Set}_{10}) \times \mu(\beta)]\} / \zeta(\text{Set}_{10})$; $i = 1, 2, \dots, 10$ olarak tanımlanmaktadır.

TABLO 4.15. ORTALAMA MALİYET-3^(*)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklemeler									
		SET 1	SET 2	SET 3	SET 4	SET 5	SET 6	SET 7	SET 8	SET 9	SET 10
costSensitive	DecisionStump	0,4755	0,4281	0,3923	0,3631	0,3364	0,3139	0,2944	0,2771	0,2603	0,2458
MetaCost	DecisionStump	0,4759	0,4285	0,3927	0,3634	0,3367	0,3142	0,2946	0,2773	0,2605	0,2460
costSensitive	DecisionTable	0,4873	0,4387	0,4021	0,3721	0,3447	0,3217	0,3017	0,2840	0,2668	0,2519
MetaCost	DecisionTable	0,4774	0,4297	0,3938	0,3645	0,3377	0,3151	0,2955	0,2782	0,2613	0,2467
costSensitive	J48	0,4732	0,4260	0,3904	0,3613	0,3347	0,3123	0,2929	0,2757	0,2590	0,2445
MetaCost	J48	0,4665	0,4200	0,3849	0,3562	0,3300	0,3079	0,2888	0,2718	0,2553	0,2411
costSensitive	J48PART	0,4584	0,4127	0,3782	0,3500	0,3242	0,3026	0,2838	0,2671	0,2509	0,2369
MetaCost	J48PART	0,4689	0,4221	0,3869	0,3581	0,3317	0,3095	0,2903	0,2732	0,2567	0,2423
costSensitive	OneR	-	-	-	-	-	-	-	-	-	-
MetaCost	OneR	0,4879	0,4392	0,4025	0,3725	0,3451	0,3220	0,3020	0,2843	0,2670	0,2521
costSensitive	ZeroR	0,4755	0,4281	0,3923	0,3631	0,3364	0,3139	0,2944	0,2771	0,2603	0,2458
MetaCost	ZeroR	0,4755	0,4281	0,3923	0,3631	0,3364	0,3139	0,2944	0,2771	0,2603	0,2458
	Maksimum	0,4873	0,4387	0,4021	0,3721	0,3447	0,3217	0,3017	0,2840	0,2668	0,2519
costSensitive	Minimum	0,4584	0,4127	0,3782	0,3500	0,3242	0,3026	0,2838	0,2671	0,2509	0,2369
	Maksimum	0,4879	0,4392	0,4025	0,3725	0,3451	0,3220	0,3020	0,2843	0,2670	0,2521
MetaCost	Minimum	0,4665	0,4200	0,3849	0,3562	0,3300	0,3079	0,2888	0,2718	0,2553	0,2411

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır.

Tablo 4.15, Tablo 4.13'te karşılaşılan yanlışlığın (örneklem büyüklük etkisinin) giderilmesi amacıyla, her veri setindeki hatalı sınıflandırılan örnek sayısının örneklem büyüklükleriyle doğru orantılı ve Set 10'daki hatalı sınıflandırılan örnek kadar alınması ve Set 10'la normalleştirilmesiyle oluşturulmuştur. Bu düzeltme, özellikle j48.J48 gibi algoritmaların farklı maliyet bilgilerine duyarlılığını ölçmek amacıyla sınıf dağılımının değiştirilmesiyle elde edilen farklı ölçekteki alt örneklemelerin ortalama sınıflandırma maliyetlerinin karşılaştırılabilirliğini sağlamak amacıyla yapılmıştır. Tablo 4.15'te gözlemlenen, sınıf dağılımının hedef sınıf lehine değiştirilmesiyle, hedef sınıfı hatalı sınıflandırma maliyetlerinin yüksek olmasına bağlı olarak, ortalama sınıflandırma maliyetlerinin artmasıdır. Bununla birlikte, ortalama maliyetlerdeki artışın gerçek anlamda bir maliyet artışı olup olmadığı ancak hedef sınıflandırma doğruluğuyla birlikte ele alındığında anlaşılabilir. Amaç hedef sınıfı veri maliyet bilgisi kısıtı altında en yüksek doğrulukla sınıflandırmak olduğundan, ortalama maliyet bilgisi birim doğru sınıflandırılan hedef sınıfa ait örnek sayısı ile ilintilendirilmelidir (Tablo 4.16).

Tablo 4.15'te gözlemlenen, daha küçük örneklemdeki hatalar en büyük örneklemdeki (Set 10) hatalarla doğru oranda artırıldığında küçük

ölçekli örneklemelerin ortalama sınıflandırma maliyetleri artıyorsa da, örneklem içinde hedef sınıf örneklerinin oranı artırıldığında sınıflandırma maliyet verimliliğinin artmasıdır.

Senaryo 4. Ortalama Maliyet Verimliliği (OVM_i)¹⁰⁰: Senaryo 3'te hesaplanan ortalama maliyetler, her örneklemde elde edilen hedef doğru sınıflandırma performansına, yani doğru sınıflandırılan hedef sınıf örneklerinin sayısına bölünmüştür.

TABLO 4.16. ORTALAMA MALİYET VERİMLİLİĞİ^(*)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklemeler									
		SET 1	SET 2	SET 3	SET 4	SET 5	SET 6	SET 7	SET 8	SET 9	SET 10
costSensitive	DecisionStump	0,0317	0,0204	0,0075	0,0303	-	0,3139	-	-	0,2603	-
MetaCost	DecisionStump	0,0280	0,0268	0,1963	-	-	-	-	-	-	-
costSensitive	DecisionTable	0,0119	0,0399	0,0804	0,0465	-	0,1608	0,0754	0,2840	0,0381	0,1259
MetaCost	DecisionTable	0,0145	0,0391	0,1969	0,1215	0,1126	0,1575	0,2955	0,1391	0,1306	0,2467
costSensitive	J48	0,0110	0,0266	0,0488	0,3613	0,0418	0,0347	0,0366	0,0689	0,0432	0,1223
MetaCost	J48	0,0099	0,0162	0,0192	0,0148	0,0150	0,0192	0,0170	0,0170	0,0128	0,0134
costSensitive	J48PART	0,0127	0,0344	0,0210	0,0389	0,0295	0,0275	0,0203	0,0382	0,0358	0,0263
MetaCost	J48PART	0,0090	0,0114	0,0161	0,0143	0,0138	0,0147	0,0112	0,0105	0,0128	0,0105
costSensitive	OneR	-	-	-	-	-	-	-	-	-	-
MetaCost	OneR	0,0232	0,0732	0,2012	0,1863	0,3451	-	0,3020	-	0,1335	-
costSensitive	ZeroR	-	-	-	-	-	-	-	-	-	-
MetaCost	ZeroR	-	-	-	-	-	-	-	-	-	-
	Maksimum	0,0317	0,0399	0,0804	0,3613	0,0418	0,3139	0,0754	0,2840	0,2603	0,1259
costSensitive	Minimum	0,0110	0,0204	0,0075	0,0303	0,0295	0,0275	0,0203	0,0382	0,0358	0,0263
	Maksimum	0,0280	0,0732	0,2012	0,1863	0,3451	0,1575	0,3020	0,1391	0,1335	0,2467
MetaCost	Minimum	0,0090	0,0114	0,0161	0,0143	0,0138	0,0147	0,0112	0,0105	0,0128	0,0105

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır. Boş alanlar, hedef sınıf örneklerinden doğru sınıflandırılan örnek sayısının 0 olduğu alanlardır.

Tablo 4.16, Tablo 4.15'te elde edilen ortalama maliyetlerin doğru sınıflandırılan hedef sınıf örnek sayısına bölünmesiyle oluşturulmuştur¹⁰¹. Bu tablo hedef sınıf örneklerinin doğru sınıflandırılmasının¹⁰² ortalama maliyeti hakkında bilgi vermektedir. Tablo 4.16'da gözlemlenen, DecisionStump algoritması hariç, diğer algoritmalar MetaCost ve CostSensitiveClassifier meta yöntemleriyle birleştirildiğinde, en düşük örnek sayısına sahip olan Set 1'de, hedef sınıf örneklerinin en düşük maliyet verimliliğiyle sınıflandırılmış olmasıdır.

¹⁰⁰ Ortalama Maliyet Verimliliği (OVM_i),

$$OVM_i \equiv OM - 3_i / (A_{Hedef} | Set_i); i = 1, 2, \dots, 10$$

olarak tanımlanmaktadır. A_{Hedef} , hedef sınıflandırma doğruluğudur.

¹⁰¹ Bu düzeltme hedef doğru sınıflandırmanın maliyet verimliliğini ölçmek amacı ile yapılmıştır.

¹⁰² Hedef doğru sınıflandırmasındaki artış diğer sınıf aleyhine sınıflandırma doğruluğunun düşmesine yol açabilir; diğer sınıf örneklerinden hatalı sınıflandırılanların sayısı ve birim hatalı sınıflandırma maliyetleriyle doğru orantılı olarak ortalama maliyet yükselebilir.

Senaryo 5. Maliyet Duyarlılığı-1 (MD-1_{i,i+1})¹⁰³: Senaryo 1’de hesaplanan ortalama maliyetlerdeki değişme, her örneklemdaki hedef sınıf örneklerinin yüzde değişimine bölünmüştür.

TABLO 4.17. MALİYET DUYARLILIĞI-1(*)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklem								
		SET 1 - SET 2	SET 2 - SET 3	SET 3 - SET 4	SET 4 - SET 5	SET 5 - SET 6	SET 6 - SET 7	SET 7 - SET 8	SET 8 - SET 9	SET 9 - SET 10
costSensitive	DecisionStump	3,1	21,3	-33,7	-11,0	1,1	-1,3	0,0	1,5	-1,7
MetaCost	DecisionStump	-0,5	-9,6	-1,7	0,0	0,0	0,0	0,0	0,0	0,0
costSensitive	DecisionTable	-15,5	-4,1	2,5	-7,4	2,2	2,5	-4,3	8,8	-8,5
MetaCost	DecisionTable	-11,4	-6,2	0,8	0,0	-1,1	-1,3	1,4	0,0	-1,7
costSensitive	J48	-14,0	-5,5	-5,9	6,4	1,1	-1,3	-5,7	2,9	-6,8
MetaCost	J48	-10,9	-4,1	3,4	-1,8	-6,6	1,3	-1,4	5,9	-3,4
costSensitive	J48PART	-12,4	4,1	-7,6	1,8	0,0	3,8	-9,9	0,0	3,4
MetaCost	J48PART	-7,8	-8,9	0,8	-0,9	-3,3	6,3	0,0	-8,8	5,1
costSensitive	OneR	-	-	-	-	-	-	-	-	-
MetaCost	OneR	-7,8	-2,7	0,0	-0,9	-1,1	1,3	-1,4	2,9	-3,4
costSensitive	ZeroR	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
MetaCost	ZeroR	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	Maksimum	3,1	21,3	2,5	6,4	2,2	3,8	0,0	8,8	3,4
costSensitive	Minimum	-15,5	-5,5	-33,7	-11,0	0,0	-1,3	-9,9	0,0	-8,5
	Maksimum	0,0	0,0	3,4	0,0	0,0	6,3	1,4	5,9	5,1
MetaCost	Mininum	-11,4	-9,6	-1,7	-1,8	-6,6	-1,3	-1,4	-8,8	-3,4

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır. Büyük sayılar yüksek maliyet verimliliği anlamına gelmektedir.

Tablo 4.17’de sınıflandırma maliyet verimliliğindeki artış sunulmaktadır¹⁰⁴. Bu tablo hedef sınıf örneklerinin örneklem içindeki yüzde artışına bağlı olarak sınıflandırma maliyetlerindeki değişimi göstermektedir. Burada gözlemlenen, veri setinin ölçeği azalır ve hedef sınıf örneklerinin örneklem içindeki oranı artarken sınıflandırma maliyet verimliliğinin artmasıdır. CostSensitiveClassifier altında çalıştırılan DecisionStump algoritması Set 3’de en yüksek maliyet verimliliğine (21.3) ulaşmıştır.

Senaryo 6. Hedef Sınıflandırma Duyarlılığı (HDS_{i,i+1})¹⁰⁵: Her örneklemda doğru sınıflandırılan hedef sınıf örneklerinin sayısındaki değişme, her örneklemdaki hedef sınıf örneklerinin yüzde değişimine bölünmüştür.

¹⁰³ Maliyet duyarlılığı-1 (MD-1_{i,i+1})

$$MD-1_{i,i+1} \equiv \Delta(M_{ort}|Set_i, Set_{i-1}) / \Delta(P_{Hedef}|Set_i, Set_{i-1}) \\ \equiv [(M_{ort}|Set_i) - (M_{ort}|Set_{i-1})] / [(P_{Hedef}|Set_i) - (P_{Hedef}|Set_{i-1})]; i = 1, 2, \dots, 10$$

olarak tanımlanmaktadır. M_{ort} , ortalama maliyet, P_{Hedef} , hedef sınıfın alt örneklem içindeki payıdır.

¹⁰⁴ Tablo 6’daki maliyet duyarlılığı Tablo 2’deki maliyet rakamlarının hedef sınıf örneklerinin örneklem içindeki yüzde değişimine göre hesaplanmıştır.

¹⁰⁵ Sınıflandırma duyarlılığı (HDS_{i,i+1})

$$HDS_{i,i+1} \equiv \Delta(A_{Hedef}|Set_i, Set_{i-1}) / \Delta(P_{Hedef}|Set_i, Set_{i-1}) \\ \equiv [(A_{Hedef}|Set_i) - (A_{Hedef}|Set_{i-1})] / [(P_{Hedef}|Set_i) - (P_{Hedef}|Set_{i-1})]$$

olarak tanımlanmaktadır. A_{Hedef} , hedef sınıflandırma doğruluğu, P_{Hedef} , hedef sınıfın alt örneklem içindeki payıdır.

TABLO 4.18. HEDEF SINIFLANDIRMA DUYARLILIĞI^(*)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklemeler								
		SET 1 - SET 2	SET 2 - SET 3	SET 3 - SET 4	SET 4 - SET 5	SET 5 - SET 6	SET 6 - SET 7	SET 7 - SET 8	SET 8 - SET 9	SET 9 - SET 10
costSensitive	DecisionStump	12,7	86,7	-137,0	-44,9	4,4	-5,1	0,0	6,0	-6,9
MetaCost	DecisionStump	-2,1	-39,1	-6,8	0,0	0,0	0,0	0,0	0,0	0,0
costSensitive	DecisionTable	-63,3	-16,8	10,3	-29,9	8,9	10,3	-17,3	35,7	-34,4
MetaCost	DecisionTable	-46,4	-25,2	3,4	0,0	-4,4	-5,1	5,8	0,0	-6,9
costSensitive	J48	-56,9	-22,4	-24,0	26,2	4,4	-5,1	-23,1	11,9	-27,5
MetaCost	J48	-44,3	-16,8	13,7	-7,5	-26,7	5,1	-5,8	23,8	-13,8
costSensitive	J48PART	-50,6	16,8	-30,8	7,5	0,0	15,4	-40,5	0,0	13,8
MetaCost	J48PART	-31,6	-36,3	3,4	-3,7	-13,3	25,6	0,0	-35,7	20,7
costSensitive	OneR	-	-	-	-	-	-	-	-	-
MetaCost	OneR	-31,6	-11,2	0,0	-3,7	-4,4	5,1	-5,8	11,9	-13,8
costSensitive	ZeroR	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
MetaCost	ZeroR	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	Maksimum	12,7	86,7	10,3	26,2	8,9	15,4	0,0	35,7	13,8
costSensitive	Minimum	-63,3	-22,4	-137,0	-44,9	0,0	-5,1	-40,5	0,0	-34,4
	Maksimum	0,0	0,0	13,7	0,0	0,0	25,6	5,8	23,8	20,7
MetaCost	Minimum	-46,4	-39,1	-6,8	-7,5	-26,7	-5,1	-5,8	-35,7	-13,8

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır. Büyük sayılar yüksek maliyet verimliliği anlamına gelmektedir.

Tablo 4.18’de hedef sınıflandırma doğruluğunun hedef sınıfın örneklem içindeki yüzde değişimine duyarlılığı incelenmektedir. CostSensitiveClassifier altında çalıştırılan DecisionStump algoritması Set 3’de en yüksek hedef sınıflandırma verimliliğine (86.7) ulaşmıştır.

Senaryo 7. Maliyet Duyarlılığı-2 (MD-2_{i,i+1})¹⁰⁶: Senaryo 2’de hesaplanan ortalama maliyetlerdeki değişim, her örneklemdeki hedef sınıf örneklerinin yüzde değişimine bölünmüştür.

TABLO 4.19. MALİYET DUYARLILIĞI-2^(*)

Meta Sınıflandırıcı	Sınıflandırıcı	İndirgenmiş Örneklemeler								
		SET 1 - SET 2	SET 2 - SET 3	SET 3 - SET 4	SET 4 - SET 5	SET 5 - SET 6	SET 6 - SET 7	SET 7 - SET 8	SET 8 - SET 9	SET 9 - SET 10
costSensitive	DecisionStump	13,9	41,9	-44,9	-6,3	11,4	8,4	10,0	11,5	8,2
MetaCost	DecisionStump	7,4	-6,9	7,3	10,0	10,0	10,0	10,0	10,0	10,0
costSensitive	DecisionTable	-20,7	2,8	13,7	-0,4	13,2	13,3	5,2	19,6	1,3
MetaCost	DecisionTable	-13,0	-0,8	11,2	9,8	8,4	8,4	11,6	9,9	8,2
costSensitive	J48	-18,5	-0,2	0,6	18,9	10,7	7,7	2,9	12,8	2,8
MetaCost	J48	-12,0	2,4	14,6	6,7	1,0	11,5	8,3	16,2	6,2
costSensitive	J48PART	-14,7	15,9	-2,5	12,2	9,4	14,0	-2,1	9,8	13,2
MetaCost	J48PART	-6,6	-5,8	11,0	8,4	5,6	17,6	9,6	0,3	15,2
costSensitive	OneR	-	-	-	-	-	-	-	-	-
MetaCost	OneR	-5,4	5,3	10,1	8,8	8,8	11,8	8,6	13,4	6,7
costSensitive	ZeroR	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0
MetaCost	ZeroR	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0
	Maksimum	13,9	41,9	13,7	18,9	13,2	14,0	10,0	19,6	13,2
costSensitive	Minimum	-20,7	-0,2	-44,9	-6,3	9,4	7,7	-2,1	9,8	1,3
	Maksimum	10,0	10,0	14,6	10,0	10,0	17,6	11,6	16,2	15,2
MetaCost	Minimum	-13,0	-6,9	7,3	6,7	1,0	8,4	8,3	0,3	6,2

(*) OneR algoritması dağılım sınıflandırıcısı olmadığından, CostSensitiveClassifier sonuçları bulunmamaktadır. Büyük sayılar yüksek maliyet verimliliği anlamına gelmektedir.

¹⁰⁶ Maliyet duyarlılığı-2 (MD-2_{i,i+1})

$$MD-2_{i,i+1} = \Delta(M_{ort}|Set_i, Set_{i-1}) / \Delta(P_{Hedef}|Set_i, Set_{i-1})$$

$$= [(M_{ort}|Set_i) - (M_{ort}|Set_{i-1})] / [(P_{Hedef}|Set_i) - (P_{Hedef}|Set_{i-1})]$$

olarak tanımlanmaktadır. M_{ort} , ortalama maliyet, P_{Hedef} , hedef sınıfın alt örneklem içindeki payıdır.

Tablo 4.19, maliyet verimliliğinin hedef sınıf örneklerinin örneklem içindeki yüzde değişimine duyarlılığını göstermektedir. CostSensitiveClassifier altında çalıştırılan DecisionStump algoritması Set 3’de en yüksek maliyet verimliliğine (41.9) ulaşmıştır.

4.2.2.6. Değerlendirme ve Yorumlar

WEKA meta sınıflandırma sonuçları, genel ve hedef sınıflandırma doğruluğu, ortalama maliyet ve senaryo analizi sonuçları, Tablo 4.20 ve Tablo 4.21’de özet olarak sunulmaktadır.

TABLO 4.20. GENEL SONUÇLAR

Genel Sınıflandırma Doğruluğu (%)		En Yüksek	Veri Seti	En Düşük	Veri Seti
Meta Sınıflandırıcı	costSensitive	97.54	Set 10	78.85	Set 3
Algoritma		ZeroR		DecisionStump	
Meta Sınıflandırıcı	MetaCost	97.54	Set 10	85.25	Set 3
Algoritma		ZeroR		j48.PART	
Hedef Sınıflandırma Doğruluğu (%)		En Yüksek	Veri Seti	En Düşük	Veri Seti
Meta Sınıflandırıcı	costSensitive	52	Set 3	-	-
Algoritma		DecisionStump		-	
Meta Sınıflandırıcı	MetaCost	52	Set 1	-	-
Algoritma		j48.PART		-	
Ortalama Maliyet		En Yüksek	Veri Seti	En Düşük	Veri Seti
Meta Sınıflandırıcı	costSensitive	0.4755	Set 1	0.2443	Set 10
Algoritma		ZeroR		DecisionTable	
Meta Sınıflandırıcı	MetaCost	0.4755	Set 1	0.2426	Set 10
Algoritma		ZeroR		j48.J48	

- En yüksek genel sınıflandırma doğruluğu, her iki meta sınıflandırıcı için de ZeroR algoritmasıyla, Set 10’da elde edilmiştir (% 97.4). ZeroR, her durumda sınıf dağılımını tahmin ettiğinden, ikinci en yüksek sınıflandırma doğruluğu daha anlamlıdır. % 97.54’ten sonra en yüksek sınıflandırma doğruluğu, % 97.52 ile MetaCost-DecisionStump-Set 10 bileşimiyle elde edilmiştir (Tablo 4.10).

- En yüksek hedef sınıflandırma doğruluğu, MetaCost-j48.PART-Set 1 ve CostSensitiveClassifier-DecisionStump-Set 3 bileşimleriyle % 52 olarak bulunmuştur. Hedef sınıflandırma doğruluğu çalışma kapsamında öncelikli önem taşıdığından, diğer analizlerde bu iki üçlü bileşimin sağlanıp sağlanmadığına bakılmalıdır (Tablo 4.11).

- En yüksek ortalama sınıflandırma maliyeti, her iki meta sınıflandırıcıda da, ZeroR algoritması ve Set 1’de gerçekleşmiştir (Tablo 4.12).

TABLO 4.21. SENARYO ANALİZİ SONUÇLARI

Senaryo 1. Ortalama Maliyet-1 (OM-1 _i)		En Yüksek	Veri Seti	En Düşük	Veri Seti	En İyi Üçlü Bileşim
Meta Sınıflandırıcı	costSensitive	0.2568	Set 5	0.1180	Set 3	costSensitive- DecisionStump-Set 3
Algoritma		DecisionTable		DecisionStump		
Meta Sınıflandırıcı	MetaCost	0.2521	Set 10	0.1681	Set 1	
Algoritma		OneR		DecisionTable		
Senaryo 2. Ortalama Maliyet-2 (OM-2 _i)		En Yüksek	Veri Seti	En Düşük	Veri Seti	En İyi Üçlü Bileşim
Meta Sınıflandırıcı	costSensitive	0.4755	Set 1	0.1883	Set 3	costSensitive- DecisionStump-Set 3
Algoritma		ZeroR		DecisionStump		
Meta Sınıflandırıcı	MetaCost	0.4755	Set 1	0.2411	Set 10	
Algoritma		ZeroR		j48.J48		
Senaryo 3. Ortalama Maliyet-3 (OM-3 _i)		En Yüksek	Veri Seti	En Düşük	Veri Seti	En İyi Üçlü Bileşim
Meta Sınıflandırıcı	costSensitive	0.4873	Set 1	0.2369	Set 10	costSensitive- j48.PART-Set 10
Algoritma		DecisionTable		j48.PART		
Meta Sınıflandırıcı	MetaCost	0.4789	Set 1	0.2411	Set 10	
Algoritma		OneR		j48.J48		
Senaryo 4. Ortalama Maliyet Verimliliği (OVM _i)		En Yüksek	Veri Seti	En Düşük	Veri Seti	En İyi Üçlü Bileşim
Meta Sınıflandırıcı	costSensitive	0.3613	Set 4	0.0075	Set 3	costSensitive- DecisionStump-Set 3
Algoritma		j48.J48		DecisionStump		
Meta Sınıflandırıcı	MetaCost	0.3451	Set 5	0.0090	Set 1	
Algoritma		OneR		j48.PART		
Senaryo 5. Maliyet Duyarlılığı-1 (MD-1 _i ,i+1)		En Yüksek	Veri Seti	En Düşük	Veri Seti	En İyi Üçlü Bileşim
Meta Sınıflandırıcı	costSensitive	21.30	Set 2 - Set 3	-33.70	Set 3 - Set 4	costSensitive- DecisionStump-Set 3
Algoritma		DecisionStump		DecisionStump		
Meta Sınıflandırıcı	MetaCost	6.30	Set 6 - Set 7	-11.40	Set 1 - Set 2	
Algoritma		j48.PART		DecisionTable		
Senaryo 6. Hedef Sınıflandırma Duyarlılığı (HDS _i ,i+1)		En Yüksek	Veri Seti	En Düşük	Veri Seti	En İyi Üçlü Bileşim
Meta Sınıflandırıcı	costSensitive	86.70	Set 2 - Set 3	-137.00	Set 3 - Set 4	costSensitive- DecisionStump-Set 3
Algoritma		DecisionStump		DecisionStump		
Meta Sınıflandırıcı	MetaCost	25.60	Set 6 - Set 7	-46.40	Set 1 - Set 2	
Algoritma		j48.PART		DecisionTable		
Senaryo 7. Maliyet Duyarlılığı-2 (MD-2 _i ,i+1)		En Yüksek	Veri Seti	En Düşük	Veri Seti	En İyi Üçlü Bileşim
Meta Sınıflandırıcı	costSensitive	41.90	Set 2 - Set 3	-44.90	Set 3 - Set 4	costSensitive- DecisionStump-Set 3
Algoritma		DecisionStump		DecisionStump		
Meta Sınıflandırıcı	MetaCost	17.60	Set 6 - Set 7	-13.00	Set 1 - Set 2	
Algoritma		j48.PART		DecisionTable		

4.2.2.7. Senaryo Analizi Sonuçları

- **Senaryo 1. Ortalama Maliyet-1 (OM-1_i):** Uyarlanmış ortalama maliyet tablosunda (Tablo 4.13), en düşük maliyet, CostSensitiveClassifier-DecisionStump-Set 3 bileşimiyle 0.1180 olarak bulunmuştur. En yüksek maliyet, CostSensitiveClassifier-DecisionTable-Set 5 bileşiminde, 0.2568'dir. Bu senaryodaki dikkate değer sonuç, en düşük uyarlanmış ortalama maliyetin, en yüksek hedef sınıflandırma doğruluğuyla elde edilen CostSensitiveClassifier-DecisionStump-Set 3 bileşiminde gözlenmesidir.

- **Senaryo 2. Ortalama Maliyet-2 (OM-2_i):** Uyarlanmış ortalama maliyet tablosunda (Tablo 4.14), en düşük uyarlanmış ortalama maliyet

(0.1883), CostSensitiveClassifier-DecisionStump-Set 3 bileşiminde gözlenmiştir.

- **Senaryo 3. Ortalama Maliyet-3 (OM-3_i):** Uyarlanmış ortalama maliyetler arasında (Tablo 4.15), en düşük maliyet (0.2369), CostSensitive-j48.PART-Set 10 üçlü bileşimiyle elde edilmiştir. Senaryo 3'te elde edilen sonuç beklenen bir sonuçtur. Bunun nedeni, en büyük örneklemdaki birinci ve ikinci tip hataların örneklem büyüklük katsayılarıyla (Γ_i) düzeltilmesiyle ortalama maliyetlerde yukarı yanlılık oluşmasıdır.

- **Senaryo 4. Ortalama Maliyet Verimliliği (OVM_i):** Senaryo 3'te elde edilen yukarı yanlı maliyetlerin hedef sınıflandırma doğruluğuyla normalleştirilmesiyle elde edilen ortalama maliyet verimliliği sonuçları, hedef sınıflandırma doğruluğu sonuçlarıyla tutarlıdır. Yukarı yanlı maliyetlerle, CostSensitiveClassifier-DecisionStump-Set 3 bileşimiyle 0.0075, MetaCost-j48.PART-Set 1 bileşimiyle 0.0090 ortalama maliyet verimliliği elde edilmiştir. Bu analizde de en başarılı sonucun (0.0075) CostSensitiveClassifier-DecisionStump-Set 3 bileşimiyle elde edildiği gözlemlenmektedir.

- **Senaryo 5. Maliyet Duyarlılığı-1 (MD-1_{i,i+1}):** Bu ve sonraki iki senaryoda pozitif sonuçlar olumlu, negatif sonuçlar olumsuz olarak yorumlanmalıdır. Senaryo 5 sonuçları, en yüksek maliyet duyarlılığının (21.30), CostSensitiveClassifier-DecisionStump çiftiyle, Set 2-Set 3 geçişinde gerçekleştiğini göstermektedir.

- **Senaryo 6. Hedef Sınıflandırma Duyarlılığı (HDS_{i,i+1}):** Senaryo 6 sonuçlarında, hedef sınıflandırma doğruluğundaki en yüksek artışın (86.70), CostSensitiveClassifier-DecisionStump çiftiyle, Set 2-Set 3 geçişinde elde edildiği gözlemlenmektedir.

- **Senaryo 7. Maliyet Duyarlılığı-2 (MD-2_{i,i+1}):** Senaryo 5'teki sonuçlarla tutarlı bir biçimde, yukarı yanlı ortalama maliyetlerle elde edilen en yüksek maliyet duyarlılığı (41.90), CostSensitiveClassifier-DecisionStump çiftiyle, Set 2-Set 3 geçişinde gerçekleşmiştir.

4.2.2.8. WEKA Platformu Genel Sonuçları

- Meta sınıflandırıcı sınıflandırma ve hatalı sınıflandırmanın ortalama maliyet sonuçları maliyet bilgisine bağlı olarak değişmektedir. Bununla birlikte, bu değişme öğrenilen kural ve ağaçlara yansımamaktadır.

- Kural ve ağaçları maliyete duyarlı hale getirmek amacıyla katmanlama stratejileriyle indirgenmiş örneklemeler elde edilmiş, bu alt örneklemelerle farklı kurallar öğrenmek olanaklı olmuştur.

- Senaryo analizleriyle elde edilen sonuçlar, veri örneklem grupları arasında en verimli sınıflandırma yapan algoritmanın, CostSensitiveClassifier meta sınıflandırıcısıyla çalıştırılan, DecisionStump algoritması olduğunu göstermektedir. Bu algoritmayla en yüksek sınıflandırma ve maliyet performansı Set 3'le elde edilmiştir. Hedef sınıflandırma doğruluğunda aynı performansı gösteren MetaCost-j48.PART-Set 1 bileşimiyle senaryo analizlerinde aynı performans elde edilememiştir. WEKA meta sınıflandırıcılarıyla analizde en yüksek performansın elde edildiği CostSensitiveClassifier-DecisionStump-Set 3 bileşimiyle öğrenilen karar ağacı aşağıda sunulmaktadır.

yil <= 1996.5 : **BAŞARILI**

yil > 1996.5 : **BAŞARILI**

yil eksik : **BAŞARILI**

Sınıflandırma Doğruluğu = % 78.85

Hedef Sınıflandırma Doğruluğu = % 52.00

Bu karar ağacının finansal tanım alanında anlamsız olduğu çok açıktır! Bu durumda, sınıflandırma doğruluğu ve ortalama maliyet gibi standart performans ölçütlerinin tek başına bir şey ifade etmediği mantıksal algoritmalarda, öğrenilen kural ve ağacın ancak anlamlı olması durumunda bu ölçütlerin kullanılabileceği söylenebilir. Bu bağlamda, yapılması gereken en yüksek hedef doğruluğunu (% 52) sağlayan ikinci meta sınıflandırma bileşimi MetaCost-j48.PART-Set 1'in öğrendiği parçasal karar ağacının anlamlılığının sorgulanmasıdır. İkinci seçenekte en iyi ikinci hedef sınıflandırma doğruluğunu sağlayan bileşimin aranmasıdır. Bu seçeneğe başvurulduğunda, % 47.00 hedef sınıflandırma doğruluğuyla MetaCost-j48.J48-Set 1 bileşeninin öğrendiği karar ağacı ikinci kural setinin temelini

oluşturacaktır. Her iki bileşim de aynı veri setiyle öğrenildiği ve yapısal olarak j48.PART, j48.J48'in budanmasıyla elde edildiği için, benzer kurallar öğrenilmiş olması olasıdır.

MetaCost-j48.PART-Set 1'le öğrenilen parçasal karar ağacı ya da kurallar arasından, öncelikle yıl (dönem) bilgisi bulunan kurallar budanmıştır¹⁰⁷. Kalan kurallar arasından, özellikle, hedef sınıfla ilgili olanlar incelenerek anlamlı olup olmadıkları anlaşılmaya çalışılmıştır. Bu kurallar,

Kural 8:

CariOran $\leq$ 1.00226 **VE**
DönenVarliklarYabancıKaynakToplami $\leq$ 0.995192
İSE BAŞARISIZ, ve

Kural 18:

donemKar = H **VE**
NetSermayeDönemNetKz $>$ 94.660455 **VE**
KVDigerAlacaklar $\leq$ 0.070773 **VE**
BrütSatislar $>$ 1.001365
İSE BAŞARISIZ

olarak verilmektedir.

Aynı şekilde, MetaCost-j48.J48-Set 1'le öğrenilen karar ağacı, aşağıda sunulmuştur.

CariOran $\leq$ 1.001161: **BAŞARISIZ (Kök)**
CariOran $>$ 1.001161
| FaizGiderleriNetSatislarOrani $\leq$ 0.000061
| | KVBorcVeGiderKarsiliklari $\leq$ 0.180307: **BAŞARISIZ (Y-1)**
| | KVBorcVeGiderKarsiliklari $>$ 0.180307: **BAŞARILI (Y-2)**
| FaizGiderleriNetSatislarOrani $>$ 0.000061
| | GelecekYillaraAitGiderlerVeGelirTah $\leq$ 0.038505
| | | DigerUzunVadeliYabancıKaynaklar $\leq$ 0.002557
| | | | NetSermayeDönemNetKz $\leq$ 47.431776
| | | | | BankaKredileriAktifToplamiOrani $\leq$ 0.658647
| | | | | yıl $\leq$ 1996
| | | | | yıl $\leq$ 1989
| | | | | DönenVarliklarYabancıKaynakToplami $\leq$ 1.463209:
BAŞARILI (Y-3)
| | | | | DönenVarliklarYabancıKaynakToplami $>$ 1.463209

¹⁰⁷ Budamayla kastedilen, söz konusu kuralın anlamsız olması dolayısıyla kullanılmamasıdır. Aslında, son kural seti bütün yeni örnekler için geçerlidir ve herhangi bir kuralı sağlamaması durumunda yeni örneğin sınıflandırılması olanaklı değildir.

OlaganKarNetSatislar <= 0.137118: **BAŞARILI (Y-4)**
 OlaganKarNetSatislar > 0.137118: **BAŞARISIZ (Y-5)**
 yil > 1989: **BAŞARILI (Y-6)**
 yil > 1996
 GecmisYillarZararlari <= -0.036807
 NetIsletmeSermayesiOzKaynaklar <= 0.73408
 StokDevirHizi <= 2.413463: **BAŞARISIZ (Y-7)**
 StokDevirHizi > 2.413463
 GelecekAylaraAitGelirlerVeGiderTah <= 0.017237
BAŞARILI (Y-8)
 GelecekAylaraAitGelirlerVeGiderTah > 0.017237:
BAŞARISIZ (Y-9)
 NetIsletmeSermayesiOzKaynaklar > 0.73408:
BAŞARISIZ (Y-10)
 GecmisYillarZararlari > -0.036807
 FaizGiderleriNetSatislarOrani <= 0.152642
 yilKurtar <= 10: **BAŞARILI (Y-11)**
 yilKurtar > 10
 UVMaliBorclar <= 0.002413
 DigerKisaVadeliYabanciKaynaklar <= 0.000182
 DuranVarliklarOzKaynaklarOrani <= 0.848879
 CabukDegerlerNetSatislar <= 0.119312
 UVBorcVeGiderKarsiliklari <= 0.007593:
BAŞARILI (Y-12)
 UVBorcVeGiderKarsiliklari > 0.007593:
BAŞARISIZ (Y-13)
 CabukDegerlerNetSatislar > 0.119312:
BAŞARILI (Y-14)
 DuranVarliklarOzKaynaklarOrani > 0.848879
 KVTicariBorclar <= 0.143697:
BAŞARISIZ (Y-15)
 KVTicariBorclar > 0.143697:
BAŞARILI (Y-16)
 DigerKisaVadeliYabanciKaynaklar > 0.000182
 NetKarAktifToplamiOrani <= 0.015571:
BAŞARISIZ (Y-17)
 NetKarAktifToplamiOrani > 0.015571
 KVDigerBorclar <= 0.001027
 yilKurtar <= 21
 NetKisaVadeliDigerAlacaklarAktifToplami <= 0.001189: **BAŞARILI (Y-18)**
 NetKisaVadeliDigerAlacaklarAktifToplami > 0.001189: **BAŞARISIZ (Y-19)**
 yilKurtar > 21: **BAŞARISIZ (Y-20)**
 KVDigerBorclar > 0.001027:
BAŞARILI (Y-21)
 UVMaliBorclar > 0.002413: **BAŞARILI (Y-22)**
 FaizGiderleriNetSatislarOrani > 0.152642
 BrütSatislar <= 1.010213: **BAŞARILI (Y-23)**
 BrütSatislar > 1.010213

| | | | | | | | | | GelecekYillaraAitGelirlerVeGiderTah <= 0.000447
 | | | | | | | | | | MenkulKiymetler <= 0.016154:
BAŞARISIZ (Y-24)
 | | | | | | | | | | MenkulKiymetler > 0.016154: BAŞARILI (4.0)
 | | | | | | | | | | GelecekYillaraAitGelirlerVeGiderTah > 0.000447:
BAŞARILI (Y-25)
 | | | | | BankaKredileriAktifToplamiOrani > 0.658647
 | | | | | UVBorcVeGiderKarsiliklari <= 0.010909
 | | | | | OlagandisiGiderVeZararlar <= -0.004976
 | | | | | BrütSatisKariAktifToplami <= 0.363325:
BAŞARISIZ (Y-26)
 | | | | | BrütSatisKariAktifToplami > 0.363325: **BAŞARILI (Y-27)**
 | | | | | OlagandisiGiderVeZararlar > -0.004976
 | | | | | GecmisYillarZararlari <= -0.01132
 | | | | | yil <= 1993: **BAŞARILI (Y-28)**
 | | | | | yil > 1993: **BAŞARISIZ (Y-29)**
 | | | | | GecmisYillarZararlari > -0.01132: **BAŞARILI (Y-30)**
 | | | | | UVBorcVeGiderKarsiliklari > 0.010909: **BAŞARISIZ (Y-31)**
 | | | | | NetSermayeDönemNetKz > 47.431776
 | | | | | CabukDegerlerAktifToplami <= 0.090187:
BAŞARISIZ (Y-32)
 | | | | | CabukDegerlerAktifToplami > 0.090187
 | | | | | NetKarNetCalismaSermayesi <= 0.06446:
BAŞARILI (Y-33)
 | | | | | NetKarNetCalismaSermayesi > 0.06446: **BAŞARISIZ (Y-34)**
 | | | | | DigerUzunVadeliYabanciKaynaklar > 0.002557
 | | | | | FaizVeVergiOncesiKarDuranVarliklar <= 0.40255:
BAŞARISIZ (Y-35)
 | | | | | FaizVeVergiOncesiKarDuranVarliklar > 0.40255:
BAŞARILI (Y-36)
 | | | | | GelecekYillaraAitGiderlerVeGelirTah > 0.038505
 | | | | | KisaVadeliYabanciKaynaklarPasifToplamiOrani <= 0.297265:
BAŞARISIZ (Y-37)
 | | | | | KisaVadeliYabanciKaynaklarPasifToplamiOrani > 0.297265
 | | | | | NetSatislarStoklar <= 9.096956: **BAŞARILI (Y-38)**
 | | | | | NetSatislarStoklar > 9.096956: **BAŞARISIZ (Y-39)**

Karar ağacının öğrendiği kurallar arasından öncelikle yıl bilgisi bulunanlar elenmiştir. Kalan kurallar (yapraklar) arasında, hedef sınıfı ilgilendiren, ilginç ve/veya anlamlı kurallar,

Kural 1: (Kök)

CariOran <= 1.001161
İSE BAŞARISIZ

Kural 39: (Y-39)

CariOran > 1.001161 VE
FaizGiderleriNetSatislarOrani > 0.000061 VE
GelecekYillaraAitGiderlerVeGelirTah > 0.038505 VE
KisaVadeliYabancıKaynaklarPasifToplamıOrani > 0.297265 VE
NetSatislarStoklar > 9.096956
İSE BAŞARISIZ

olarak ağaçtan okunmuştur.

MetaCost-j48.PART-Set 1’le öğrenilen kurallar arasında, Kural 8’in ilk öncülü anlamlıyken, ikinci öncülü gerçekçi bir ölçüt değildir. Kural 8’e göre, her iki öncülün de sağlanması için, kısa vadeli yabancı kaynaklarla dönen varlıkların düzeyinin neredeyse aynı, uzun vadeli yabancı kaynaklarına toplam yabancı kaynakların % 7’si ya da daha altında olması gerekmektedir. Böyle bir finansman yapısı, sermaye ve duran varlıkların düzeylerinin aynı olmasını gerektirir. Uzun vadeli borç düzeyinin düşük ve kısa vadeli borç düzeyinin yüksek olmasıysa, sermaye düzeyinin de düşük olması durumunda, firmanın finansman riskini yükseltecektir. Bununla birlikte, bazı sektörlerdeki firmaların bu koşulları sağlayacak şekilde faaliyetlerini yürütmeleri olanaklıdır. Dolayısıyla, kuralın tamamen dışlanması da olanaklı değildir. Kural 18’deyse, NetSermayeDönemNetKz > 94.660455 dışındaki öncüller finansal olarak anlamlı değildir. Dolayısıyla, MetaCost-j48.PART-Set 1, uygulama seçenekleri arasından elenebilir.

İkinci en yüksek hedef sınıflandırma doğruluğunun sağlandığı MetaCost-j48.J48-Set 1 kural seti için durum biraz daha umut verici görünmektedir. MetaCost-j48.J48-Set 1’le öğrenilen kurallar arasında, Kural 1 finansal olarak tamamen anlamlıdır. Kural 39’un öncülleri arasında statik devir hızı da diyebileceğimiz, NetSatislarStoklar > 9.096956 dışında diğer bütün öncüller finansal olarak anlamlıdır. Ayrıca FaizGiderleriNetSatislarOrani > 0.000061 öncülünün eşik değerinin çok düşük olduğu göz önünde bulundurulduğunda, bu öncülün de fazla anlamlı olmadığını söylemek olanaklı olacaktır. Dolayısıyla Kural 39, seçenekler arasından çıkarılabilir.

Özet olarak,

- İki farklı grup meta sınıflandırmayla öğrenilen kural setlerinden, sadece MetaCost-j48.J48-Set 1’le öğrenilen Kural 1 finansal olarak hedef sınıflandırma için anlamlı bulunmuştur.

- MetaCost-j48.J48-Set 1 üçlü bileşimiyle elde edilen hedef sınıflandırma doğruluğunun % 47 olduğu düşünülürse, finansal analizin en temel ama ilginç olmayan parmak kurallarından “Cari Oran’ın 1’den düşük olması arzu edilmez” kuralıyla bile, WEKA meta sınıflandırma algoritmalarının yüksek bir sınıflandırma performansı elde edemediğini söylemek olanaklı olacaktır.

- Diğer taraftan, makina öğrenmesinin en temel kurallarından olan veri seti büyüklüğü ve tanım alanına yönelik entropi kazanımının özgül alana ait daha fazla bilgiyle edinilebileceği gerçeği, WEKA sınıflandırmalarında sadece temsil oranının sağlanabilmesi ilkesiyle göz ardı edilmiş ve hedef sınıf örneklerinin genel ve alt örneklerdeki payı düşük tutulmuştur. Bu bağlamda, hedef sınıf doğruluğundaki artışların, düşük ölçekli örnekler lehine gerçekleştiği gözlemlenmektedir. Örneklerdeki BAŞARISIZ örneklerin oranının sadece % 2.3 (Set 10’la Set 1 arasındaki fark) artırılmasıyla performansın iki grup meta sınıflandırmada (MetaCost-j48.J48-Set 1, MetaCost-j48.PART-Set 1) % 29 civarında artması iyi bir performans artışı olarak düşünülebilir.

- Veri setine yönelik koşulların¹⁰⁸ sağlanması durumunda, makina öğrenmesi uygulamalarının daha güçlü donanım ve yazılım desteğiyle kullanılabilir sonuçlar üreteceğini söylemek olanaklı olacaktır.

¹⁰⁸ Söz konusu koşullar, birinci bölümde modellemedeki güçlükler başlığı altında ele alınmaktadır.

4.3. BCFP Analiz ve Yorumları

4.3.1. BCFP Veri Seti

BCFP algoritması, WEKA platformunda olduğu gibi veri setinin boyutuna yönelik üst kısıtlar olmadan çalıştırılabilmektedir. Bu bağlamda, BCFP analizlerinde bütün veri seti (25,547 firma-yıl örnek) kullanılmıştır.

4.3.2. BCFP Algoritmasında Fayda-Maliyet Bilgisi

BCFP analizlerinde kullanılan fayda maksimizasyonu, karma optimizasyon bağlamında ele alınmaktadır. Orijinal algoritma, toplam fayda maksimizasyonu amaç fonksiyonu altında doğru sınıflandırmayı ödüllendirmek amacıyla, fayda-maliyet tablosundan doğru sınıflandırılan örnekler için pozitif fayda bilgilerini kullanacak şekilde tasarlanmıştır. Diğer taraftan karma optimizasyon, fayda-maliyet tablosunda, doğru sınıflandırmayı pozitif fayda bilgisiyle ödüllendirirken, hatalı sınıflandırmayı negatif maliyet bilgisiyle cezalandırmaktadır. Bu bağlamda, fayda-maliyet tablosunda hedef doğru sınıflandırmanın faydası ($B(B_B)$) pozitif, hedef hatalı sınıflandırmanın maliyeti ($M(B_A)$) negatif bir sayı olarak yer almaktadır.

Fayda-maliyet tablosunda, bireysel fayda-maliyet bilgilerinin büyüklükleri, incelenen sınıflandırma probleminin yapısına, özneliliklerin sınıflara göre yığılma dağılımına ve kullanıcının amaçlarına bağlı olarak değişebilir. Fayda maliyet bilgisinin oluşturulmasında dikkate alınması gereken, bireysel fayda ve maliyet bilgilerinin kullanıcıya ne ifade ettiği. Örneğin, BAŞARISIZ örneğin doğru sınıflandırılması hedefi anlamlı bir hedeftir. Diğer taraftan, BAŞARISIZ örneğin hatalı sınıflandırılması önemli bir maliyet etmenidir. Çalışma kapsamında, her iki ara hedef de önemlidir.

Fayda maliyet bilgisinin belirlenmesinde rehber bir çalışma bulunmadığından, optimal sınıflandırma sonuçlarının belirlendiği aralığa kadar bir dizi fayda-maliyet bilgisi kullanılmış ve deneme yanılma yoluyla elde edilen ilk fayda-maliyet tablosu, $B(A_A) = 1$, $B(B_B) = 1$, $M(A_B) = 0$ ve $M(B_A) = 0$ olarak belirlenmiştir.

Genel ve hedef sınıflandırma doğruluğunun, fayda ve maliyet bilgisine duyarlılığını ölçmek amacıyla, fayda-maliyet bilgileri sistemli olarak değiştirilmiş ve duyarlılık analizi yapılmıştır¹⁰⁹. BCFP algoritması duyarlılık analizi sonuçları Tablo 4.22 ve Tablo 4.23'te sunulmuştur.

4.3.3. Duyarlılık Analizi Sonuçları

Tablo 4.22'de sunulan genel sınıflandırma ya da fayda doğruluğunun¹¹⁰ fayda-maliyet bilgisine duyarlılık sonuçları, genel olarak fayda sabitken maliyet artırıldığında ve maliyet sabitken fayda artırıldığında fayda doğruluğunun azaldığını göstermektedir. Aynı şekilde, fayda ve maliyet aynı anda artırıldığında fayda doğruluğunun düştüğü gözlemlenmektedir. Ayrıca, fayda-maliyet-sınıflandırma doğruluğu uzayında daha düşük fayda ve yüksek maliyet düzeylerinde sınıflandırma doğruluğunun yüksek fayda ve düşük maliyet düzeylerindekiyle yüksek olduğu bölgelere rastlanmaktadır¹¹¹. Örneğin, fayda-maliyet bileşimi (1.01–0.09) olan noktada sınıflandırma doğruluğu % 18.68 iken, bu oran (1.04-0.07) noktasında % 15.54'e düşmektedir. Burada, fayda-maliyet-fayda doğruluğu uzayında optimal bir fayda-maliyet bileşiminin elde edilmesi olasılığı gündeme gelmektedir.

TABLO 4.22. GENEL SINIFLANDIRMA DOĞRULUĞU VE FAYDA-MALİYET DUYARLILIĞI (%)

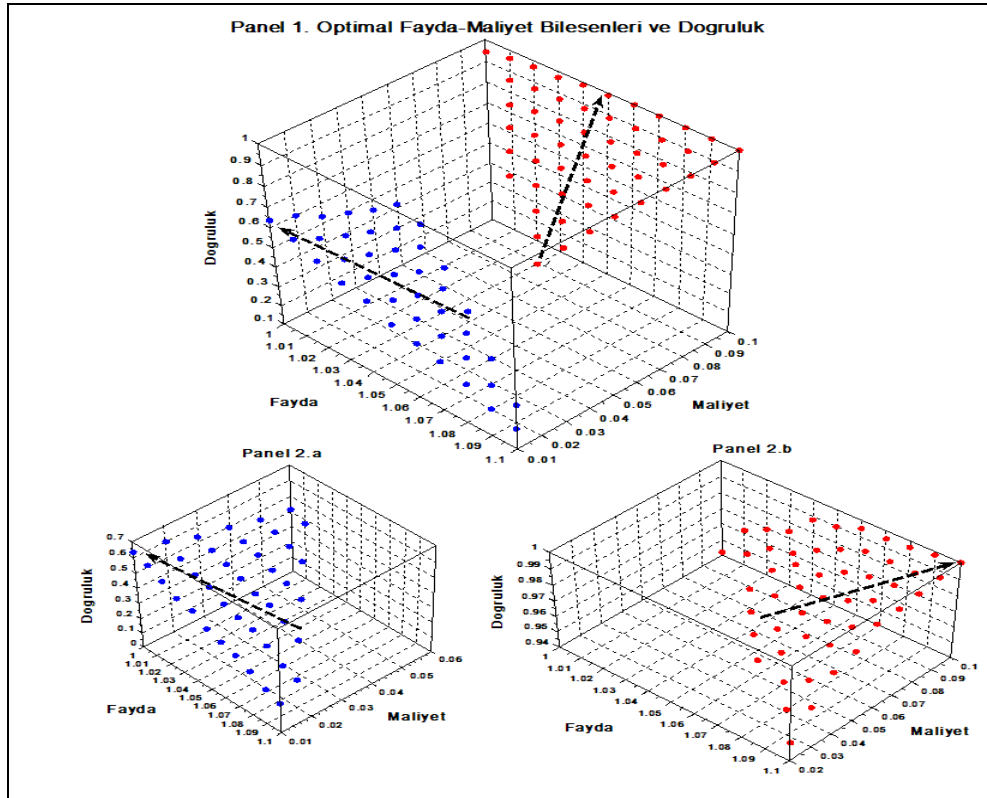
		"BAŞARISIZ"ı "BAŞARILI" Sınıflandırmanın Maliyeti									
		0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,10
"BAŞARISIZ"ı "BAŞARILI" Sınıflandırmanın Faydası	1,00	63,41	59,74	53,20	49,24	44,59	41,55	36,37	28,39	22,31	18,35
	1,01	60,06	54,22	49,75	45,53	41,44	37,36	30,25	23,96	18,68	15,39
	1,02	55,18	49,54	46,03	41,91	37,60	30,89	24,66	19,27	15,63	12,61
	1,03	50,11	46,60	41,81	37,14	32,56	26,16	20,31	15,89	12,28	10,13
	1,04	47,20	41,43	37,43	34,14	28,53	21,90	15,54	12,16	10,38	8,48
	1,05	41,51	37,73	35,35	28,72	21,76	15,42	12,25	10,28	9,03	7,05
	1,06	38,42	35,54	30,16	22,35	16,15	12,91	10,48	8,82	7,43	5,75
	1,07	35,21	30,41	23,16	17,40	12,75	10,81	9,09	7,40	5,94	5,02
	1,08	30,28	22,89	17,99	13,43	10,56	9,35	7,73	6,34	5,14	4,18
	1,09	24,25	18,86	14,62	11,32	9,36	7,83	6,70	5,16	4,38	3,55
	1,10	20,52	15,50	11,66	9,31	8,09	6,67	5,27	4,26	3,40	2,92

¹⁰⁹ Duyarlılık analizi için fayda-maliyet bilgisi tablosunda $B(A_A) = 1$ ve $M(A_B) = 0$ olarak sabit bırakılmış ve hedef doğru sınıflandırmanın faydası $B(B_B)_i = 1 + 0.01 \cdot i$ ve hedef hatalı sınıflandırmanın maliyeti $M(B_A) = -0.01 \cdot j$, ($i=0,1,2,...,10$, $j=1,2,...,10$) olarak değiştirilmiştir.

¹¹⁰ Faydaya duyarlı öğrenmede, genel sınıflandırma doğruluğu fayda doğruluğu olarak da ifade edilmektedir. Sonraki bölümlerde fayda doğruluğu, genel sınıflandırma doğruluğu yerine kullanılmıştır.

¹¹¹ Farklı fayda maliyet kısıtlarıyla elde edilen fayda doğruluğunun Şekil 4.2'de olduğu gibi görsel boyutta incelenmesiyle aynı sonuca ulaşılmaktadır. Bu alan üzerinde bazı bölgelerin diğerlerinden daha iyi sınıflandırma sonuçları verdiği görülmektedir.

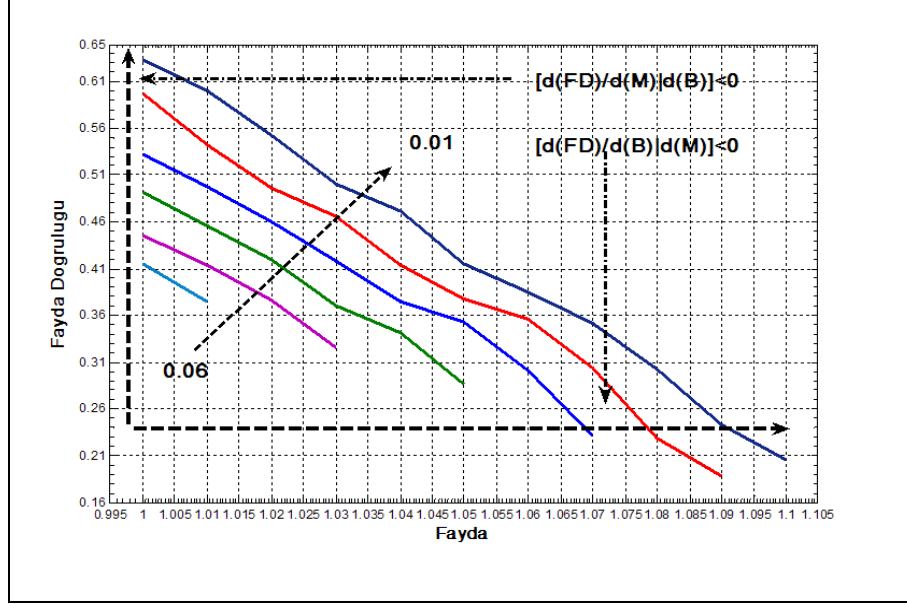
Optimal fayda-maliyet bilgisi tek başına nesnel olmaktan çok öznelir. Bu öznelik, kullanıcının amacına bağlı olarak nesnelleştirilebilir. Eğer, kullanım amacı özellikle hedef sınıflandırma doğruluğu yerine toplam fayda doğruluğunu artırmak, dolayısıyla genel sınıflandırma fayda maksimizasyonu olduğunda, optimal fayda-maliyet bileşimi fayda doğruluğu bağlamında ele alınmalıdır. Diğer yandan, kullanıcının özgül amacı genel doğruluk ve fayda maksimizasyonundan farklıysa, bu durumda, özgül hedefe yönelik sonucun iyileştirilmesi fayda maliyet bileşimini belirleyen kısıt olacaktır. Örneğin, yüksek hedef sınıflandırma doğruluğunu amaçlayan bir kullanıcı yüksek fayda-maliyet bileşimlerine yoğunlaşmalıdır.



Şekil 4.1: Genel Ve Hedef Sınıflandırma Doğruluğu ve Optimal Fayda Maliyet Bileşenleri

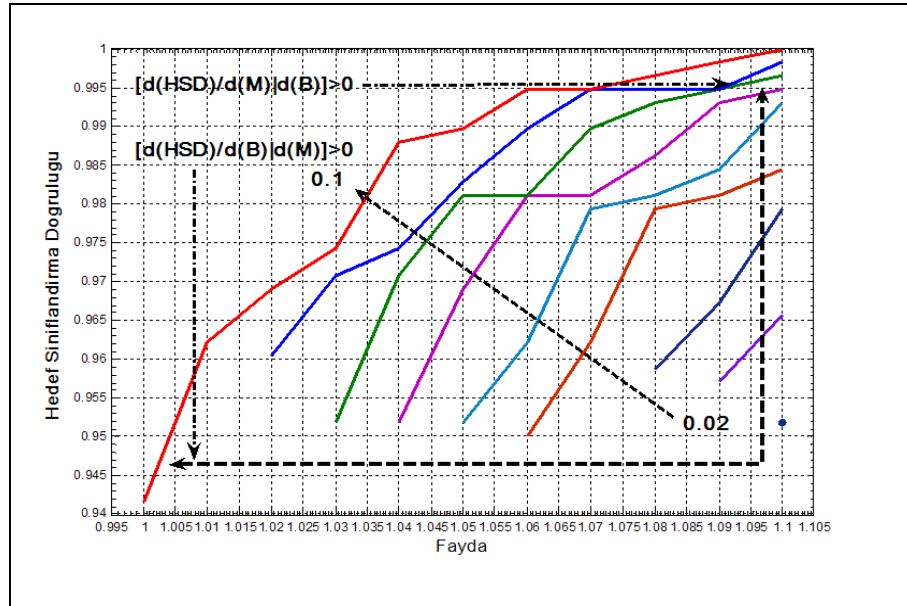
Şekil 4.1’de duyarlılık analizleriyle elde edilen genel sınıflandırma ya da fayda doğruluğu ve hedef sınıflandırma doğruluğunun fayda-maliyet bilgisine duyarlılığı görüntülenmektedir. Üst panelde (Panel 1), her iki performans ölçütü bir arada sunulmaktadır. Alt panelde, solda genel sınıflandırma ya da fayda doğruluğu (Panel 2.a), sağda ise hedef sınıflandırma doğruluğu (Panel 2.b) ayrıştırılmış olarak sunulmaktadır.

- Beklendiği gibi, hedef sınıflandırma ve fayda doğruluğunun maksimum olduğu alanlar ayrıktır ve her ikisi de fayda ve maliyete farklı şekilde duyarlıdır.



Şekil 4.2: Genel Sınıflandırma Doğruluğu ve Optimal Fayda Maliyet Bileşenleri

- Fayda doğruluğu maliyet ve faydayla doğru orantılı olarak azalmaktadır ($[d(FD)/d(M)]d(B)] < 0$, $[d(FD)/d(B)]d(M)] < 0$). Sabit fayda düzeylerinde maliyet, sabit maliyet düzeylerindeyse fayda artarken fayda doğruluğu azalmaktadır (Şekil 4.2).



Şekil 4.3: Hedef Sınıflandırma Doğruluğu ve Optimal Fayda Maliyet Bileşenleri

• Hedef sınıflandırma doğruluğu maliyet ve faydayla doğru orantılı olarak artmaktadır ($[d(HSD)/d(M)]d(B)]>0$, $[d(HSD)/d(B)]d(M)]>0$). Sabit fayda düzeylerinde maliyet, sabit maliyet düzeylerindeyse fayda artarken hedef sınıflandırma doğruluğu artmaktadır (Şekil 4.3).

4.3.4. Fayda Doğruluğu

Fayda-maliyet-fayda doğruluğu uzayında en yüksek genel sınıflandırma doğruluğu (1.00-0.01) fayda-maliyet bileşiminde elde edilmiştir (% 63.41). Kullanıcının ara hedeflerinden birisi genel sınıflandırma doğruluğuyorsa, daha düşük fayda ve daha düşük maliyet bilgisi düzeylerindeki fayda-maliyet bileşimleri daha optimaldir. Örneğin, 1.08 sabit fayda düzeyinde (1.08-0.01) bileşimiyle elde edilen sınıflandırma doğruluğu (% 30.28), (1.08-0.02) bileşimiyle elde edilen sınıflandırma doğruluğundan (% 22.89) daha yüksektir.

4.3.5. Hedef Sınıflandırma Doğruluğu

Hedef sınıflandırma doğruluğu uzayında en yüksek sınıflandırma doğruluğu % 100 (1.10-0.10) bileşimiyle elde edilmiştir. Bu bağlamda, (1.10-0.10) bileşiminin altında optimal bir sınırdan söz etmek olanaklı değildir. Bunun yerine, değişen fayda-maliyet düzeylerinde marjinal sınıflandırma doğruluğunun en yüksek olduğu bölge ya da fayda-maliyet bileşimleri araştırılmalıdır.

TABLO 4.23. HEDEF SINIFLANDIRMA DOĞRULUĞU VE FAYDA-MALİYET DUYARLILIĞI (%)

		"BAŞARISIZ"ı "BAŞARILI" Sınıflandırmanın Maliyeti									
		0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,10
"BAŞARISIZ"ı "BAŞARILI" Sınıflandırmanın Faydası	1,00	65,64	68,90	75,26	77,84	81,96	83,33	85,05	88,83	92,61	94,16
	1,01	68,38	74,05	77,66	80,93	83,16	84,71	87,63	91,07	94,16	96,22
	1,02	73,37	78,01	81,10	83,16	85,74	87,63	91,24	93,99	96,05	96,91
	1,03	77,84	80,58	82,99	86,08	87,46	90,55	93,47	95,19	97,08	97,42
	1,04	80,07	83,16	85,74	87,29	89,35	92,61	95,19	97,08	97,42	98,80
	1,05	83,68	85,57	87,11	89,35	92,61	95,19	96,91	98,11	98,28	98,97
	1,06	85,57	86,94	89,00	92,10	95,02	96,22	98,11	98,11	98,97	99,48
	1,07	86,94	88,66	91,24	94,33	96,22	97,94	98,11	98,97	99,48	99,48
	1,08	89,18	91,58	93,99	95,88	97,94	98,11	98,63	99,31	99,48	99,66
	1,09	91,58	93,64	95,70	96,74	98,11	98,45	99,31	99,48	99,48	99,83
	1,10	92,61	95,19	96,56	97,94	98,45	99,31	99,48	99,66	99,83	100,00

Farklı fayda-maliyet bileşimleriyle elde edilen en yüksek marjinal sınıflandırma doğruluğuna, 1.00 fayda düzeyinde maliyetin 0.02'den 0.03'e artırıldığı noktada ulaşılmaktadır (% 75.26 - % 68.90 = % 6.36). Yine de bu fayda maliyet bileşimiyle elde edilen en yüksek marjinal sınıflandırma doğruluğu ulaşılabilecek en yüksek sınıflandırma doğruluğu düzeyinde değildir. Tablo 4.23 incelendiğinde yüksek sınıflandırma doğruluğuna ulaşılan düzeylerde marjinal sınıflandırma doğruluğunun azaldığı görülmektedir. Burada yine başvurulması gereken kullanıcının ne düzeyde bir sınıflandırma doğruluğunu kabul edilebilir bulduğudur. Eğer kullanıcı ortalama düzeyde¹¹² bir doğruluğu kabul ediyorsa, marjinal doğruluk bilgisine başvurarak optimal fayda-maliyet bileşimini seçebilir.

4.3.6. Optimal Fayda Maliyet Bileşimi ve Toplam Fayda

BCFP'nin en temel varsayımı, sınıflandırmayla en yüksek faydayı elde etmek olduğundan, genel ve hedef sınıflandırmayla elde edilen fayda bilgisinin incelenmesi ve en yüksek sınıflandırma faydasını sağlayan fayda-maliyet bileşimlerinin belirlenmesi gerekir. Genel ve hedef sınıflandırma doğruluğunda da olduğu gibi, her iki kategorinin belirlenmesinde elde edilen faydanın toplam fayda içindeki payının ters orantılı olarak değişmesi beklenir¹¹³.

Optimal fayda-maliyet bileşimi çiftlerinin belirlenmesinde sınıflandırma doğruluğu bilgisine başvurmak öznel, sınıflandırmadan elde edilen fayda bilgisinin kullanılmasıysa nesnel bir yoldur. BCFP algoritmasıyla, sınıflandırmanın toplam faydası maksimize edildiğinden bu kabul edilebilir bir nitelendirme olacaktır. Toplam faydanın sınıflararası dağılımının çıktı olduğu düşünülürse, fayda maksimizasyonunda müdahale edilebilecek tek parametrenin fayda-maliyet bilgisi olduğu görülecektir. Elbette bu önerme, fayda-maliyet bilgisinin öznel olduğu, yani kullanıcı tarafından belirlendiği durumlarda geçerlidir. Bu yaklaşım, BCFP analizlerinde kullanılan sabit fayda-maliyet tablolarını ortaya çıkarmaktadır. Bununla birlikte, özniteliğe

¹¹² Hatasız sınıflandırma doğruluğunun şüpheyle karşılanması, en azından modelin aşırı özelleşmiş olması olasılığının göz önünde bulundurulması gerekmektedir.

¹¹³ Burada dikkat edilmesi gereken nokta bu genellemenin özniteliklerin sınıf dağılımına ve fayda-maliyet bilgisine duyarlı olduğudur.

dayalı yani her örnek için değişen fayda-maliyet bilgisi (değişken fayda-maliyet tabloları) söz konusu olduğunda, kullanıcının tek seçeneği fayda-maliyet bilgisini oluşturan fonksiyonun yapısı ve hangi öznitelikleri kullanacağına karar vermektir¹¹⁴. Makina öğrenmesinin tanımı gereği, sınıflandırmanın ya da öğrenme sürecinin anlamlı değişkenleri kendi başına bulması öngörüldüğünden, değişken fayda maliyet tablolarının fonksiyonel yapısı ve parametrelerinin de otomatik olarak belirlenmesi beklenir. Böyle bir algoritmanın tasarlanması belki de, faydaya duyarlı öğrenmenin en karmaşık aşaması olacaktır. Sonuç olarak, kullanıcının amaçları doğrultusunda, optimal fayda-maliyet bileşimi optimal alanın herhangi bir yerinde bulunabilir.

4.3.7. Genel Sonuçlar

BCFP algoritması finansal başarısızlığın tahmini bağlamında ilk olarak bu çalışmada kullanılmıştır¹¹⁵.

BCFP analizinde kullanılan fayda-maliyet tablosu karma optimizasyon varsayımı altında oluşturulmuştur. Ayrıca, genel sınıflandırma ve hedef sınıflandırmanın doğruluğuna bağlı olarak fayda ve maliyet bilgileri uyarlanmış ve duyarlılık analizi yapılmıştır.

BCFP analizlerinde aynı veri seti için her biri tek bir özniteliğe dayalı 184 kural öğrenilmiştir. BCFP ile öğrenilen kurallardan hedef sınıfa yönelik ilginç ve anlamlı bulunanların listesi Ek 3'te verilmiştir.

BCFP algoritmasıyla elde edilen sonuçlara göre,

- Fayda ve maliyet etmenini göz önünde bulundurmaları,
- Öğrenilen kuralların daha basit, anlamlı ve anlaşılır olmaları ve
- Daha yüksek genel ve hedef sınıflandırma performanslarıyla

WEKA algoritmaları ve meta sınıflandırma yöntemlerine alternatif modeller geliştirilmesinde BCFP yöntemleri öne çıkmaktadır.

¹¹⁴ Örneğin, kullanıcı, amacı doğrultusunda, firma başarısızlığı bağlamında, hedef hatalı sınıflandırmanın maliyetini firma büyüklüğü olarak kabul edebilir. Bu durumda maliyet fonksiyonu

$$F_{ij}(B_A) = f(\text{Varlık Büyüklüğü}) = - \text{Varlık Büyüklüğü}$$

olarak şekillenebilecektir.

Halihazırda, özniteliğe dayalı fayda-maliyet bilgisi ile optimizasyon yapan algoritmalar geliştirilmediğinden, çalışmamızda, sabit fayda-maliyet tabloları kullanılmıştır. Değişken fayda-maliyet tablosu, her ne kadar daha gerçekçi bir yaklaşım ise de, böyle bir algoritmanın kullanılması büyük veri setlerinde bilgi işlem ve donanım maliyetleri açısından, halihazırda optimal değildir.

¹¹⁵ İkizler ve Güvenir (2003), kredi başvurularının değerlendirilmesinde fayda maksimizasyonuna dayalı benzer bir algoritma olan BMFI algoritmasını kullanmıştır.

BEŞİNCİ BÖLÜM

SONUÇ VE ÖNERİLER

Finansal baskı ve başarısızlığın tahmini, reel sektör riskliliğinin takibi açısından ele alınması gereken önemli bir konu olagelmıştır. 60 yıldan uzun bir süredir üzerinde önemle durulmuş olmasına rağmen hala üzerinde genel fikir birliğine varılmış bir finansal baskı ve firma başarısızlığı kuramı oluşturulamamıştır. Yapılan çalışmalar, genellikle, finansal ve/veya finansal olmayan firma bilgilerinin ampirik olarak incelenmesine yoğunlaşmıştır. Çalışmada çizilen çerçeve, firma bilgilerinin makina öğrenmesi yöntemleri aracılığıyla ampirik olarak incelenmesiyle başarısızlık dinamiklerinin anlaşılması üzerine kurulmuştur.

Finansal baskı ve firma başarısızlığı çalışmalarında ele alınan baskı ve başarısızlık tanımları incelenerek ve bu modellerde karşılaşılan sorunlar ele alınarak Reel Sektör Verileri veri tabanından elde edilen veri seti, varsayımları karşılayacak şekilde düzenlenmiştir. Ayrıca, literatürde sermaye yitikliği olarak ifade edilen başarısızlık ölçütü genişletilmiştir.

Makina öğrenmesi yöntemleri incelenmiş ve mantıksal algoritmalar kullanım kolaylığı ve diğer avantajları dolayısıyla seçilmiştir. Mantıksal öğrenme algoritmaları karar ağaçları, kapsama algoritmaları ve parçasal karar ağaçlarıdır.

Ampirik analizde, makina öğrenmesi çalışmalarında sıklıkla başvurulanan WEKA platformundaki kural öğrenen altı mantıksal sınıflandırma algoritmasıyla birlikte, faydaya duyarlı öğrenme için geliştirilen öznitelik izdüşümü algoritması BCFP kullanılmıştır. WEKA platformundaki kapsama algoritmaları, ZeroR, OneR ve DecisionTable, karar ağaçları, DecisionStump ve j48.J48, parçasal karar ağacıysa j48.PART algoritmasıdır. WEKA algoritmalarının veri kısıtları nedeniyle 25,547 firma-yıl örnekten oluşan veri seti öncelikle rassal eleme yöntemiyle 4,272 firma-yıl örnekten oluşan bir

boyuta indirgenmiştir. Bu veri setiyle yapılan, çıplak ve meta sınıflandırmalarla yüksek hedef sınıflandırma doğrulukları elde etmek olanaklı olmamıştır. Ayrıca, orijinal olarak maliyete duyarlı tasarlanan algoritmalar olmamaları dolayısıyla, iki farklı maliyet kısıtı altında da aynı sınıflandırma kuralları öğrenilmiştir. Bu bağlamda, öğrenmeyi maliyete duyarlı hale getirmek amacıyla büyüklükleri 2,103 ve 4,069 firma-yıl örnek arasında değişen 10 alt örneklem oluşturulmuştur. 10 alt örneklem maliyete duyarlı yöntemlerle öğrendiği kurallar 1:10 maliyet bilgisiyle elde edilmiştir. Maliyete duyarlı kural öğrenmede, değiştirilen parametreler örneklem büyüklükleri, meta öğrenme yöntemi ve öğrenme algoritmasının kendisidir.

Maliyete duyarlı öğrenme yöntemlerinin 10 alt örneklemle elde edilen genel ve hedef sınıflandırma performanslarıyla maliyet verimliliği bilgileri karşılaştırılarak en başarılı yöntemler bulunmaya çalışılmıştır. Ayrıca meta sınıflandırıcı-öğrenme algoritması-örneklem üçlülükleri arasında karşılaştırma yapılabilmesi amacıyla, senaryo analizleri yapılmıştır.

WEKA algoritmalarıyla öğrenilen kurallar arasından, hedef sınıf örneklerinin en yüksek doğrulukla sınıflandırıldığı meta-algoritma çiftlerinin farklı örneklem büyüklükleriyle maliyet ve sınıflandırma performanslarına bakılarak anlamlı ve ilginç olanlar ayıklanmıştır. WEKA öğrenme algoritmaları, maliyete duyarlı sınıflandırma ve senaryo analizlerinin sonuçlarına göre MetaCost-j48.J48-Set 1 ve MetaCost-j48.PART-Set 1 meta sınıflandırıcı-öğrenme algoritması-örneklem üçlülükleri en başarılı bileşimler olarak bulunmuştur. Performansın belirlenmesinde genel sınıflandırma, maliyet ve hedef sınıflandırma doğruluğu yanında öğrenilen kural ve ağacın anlamlı olması da önemli bir etmen olmuştur.

BCFP algoritmasıyla yapılan ampirik analizde, WEKA platformu analizlerinden farklı olarak, veri setinin tamamı kullanılmıştır. Dolayısıyla, BCFP faydaya duyarlı öğrenmede daha çok bilgiyle eğitilebilmiştir. BCFP algoritmasıyla öğrenmede fayda doğruluğu ve hedef sınıflandırma doğruluğunun fayda-maliyet bilgisine duyarlılığının ölçülmesi amacıyla fayda-maliyet bilgisi değiştirilerek duyarlılık analizi yapılmıştır. Fayda ve hedef sınıflandırma doğruluğunun maksimum olduğu alanların ayırık oldukları

gözlemlenmiştir. Ayrıca, kullanım amacına bağlı olarak değişen optimal fayda maliyet bileşimlerinin bulunabilirliği de gösterilmiştir.

Mantıksal makina öğrenmesi yöntemleriyle 3,715'i WEKA algoritmalarıyla, 184'üyse BCFP algoritmasıyla olmak üzere toplam 3,899 adet kural tam ve indirgenmiş veri setleriyle öğrenilmiştir.

Bu çalışmada elde edilen bilgiler ışığında,

- Makina öğrenmesi yöntemleriyle firma başarısızlığı analizinin olanaklı olduğu,
- Veri tabanına yönelik melez¹¹⁶ sorgulama yöntemleriyle birlikte, fayda ve maliyete duyarlı makina öğrenmesi yöntemleri geliştirilerek sınıflandırma kuralları öğrenilebileceği,
- Fayda ve maliyete duyarlı yöntemlerin öznitelik bağımlılığı göz önünde bulundurularak, her örneğe özel fayda-maliyet bilgisi yöntemlerinin kullanılmasıyla daha anlamlı sonuçlar elde edilebileceği, anlaşılmıştır.

Sonraki çalışmalarda,

- Veri tabanındaki firma bilgilerinin döviz pozisyon ve finansman giderleri bilgisiyle genişletilmesi yoluyla kur ve faiz riskliliğinin ölçülmesi,
- Niteliksel ve operasyonel bilgilerle birlikte, dışsal etmenlerin de eklenmesiyle çalışmanın genişletilmesi,
- WEKA platformu algoritmalarının açık kodları kullanılarak, standart öğrenme yöntemlerinin fayda ve maliyet bilgisine duyarlı hale getirilmesi,
- BCFP algoritmasının dağılım varsayımlarının değiştirilerek daha anlamlı bir model oluşturulması,
- BCFP algoritmasına ek olarak çalışma bağlamında geliştirilen BMFT algoritmasının kullanılarak çalışmanın genişletilmesi,
- Kural ilginçliği öğrenme algoritmaları kullanılarak öğrenilen kurallara otomatik olarak ilginçlik etiketlerinin verilmesi, bu bağlamda etkileşimli kural ilginçliği algoritması IRIL'in geliştirilmesi hedeflenmektedir.

¹¹⁶ Melez sorgulama yöntemleri standart veri tabanı sorgulama yöntemleriyle makina öğrenmesi algoritmalarını bir arada kullanan yöntemler olarak adlandırılabilir.

KAYNAKÇA

- Aiyabei, J. (2000). Financial Distress: Theory, Measurement and Consequence, Seminar Paper Presented at the Catholic University of Eastern Africa, Department of Commerce.
- Altman, E. Corporate Distress Prediction Models in Turbulent Economic and Basel II Environment, Erişim: 17 Mart 2003, <http://pages.stern.nyu.edu/~ealtman/Corp-Distress.pdf>
- Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, 23, 3, s. 589-609.
- Altman, E., Haldeman, R. ve Narayanan. P. (1977). ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance*, 1, 1, s. 29-54.
- Altman, E. (1993). Corporate Financial Distress and Bankruptcy. New York, NY: John Wiley and Sons.
- Altman, E. (1983), Corporate Financial Distress: A Complete Guide to Predicting, Avoiding and Dealing with Bankruptcy. Toronto: Wiley and Sons.
- Back, B., Laitinen, T., Sere, K. ve Wezel, M. (1996). Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis and Genetic Algorithms, *Turku Centre for Computer Science, Technical Report No 40*.
- Back, B., Laitinen, T., Hekanaho, J. ve Sere, K. (1997). The Effect of Sample Size on Different Failure Prediction Methods, *Turku Centre for Computer Science Technical Report, No 155*.
- Back, B., Laitinen, T. ve Sere, K. (1996). Neural Network and Genetic Algorithms for Bankruptcy Predictions. Proceedings of the 3rd World Congress on Expert Systems, Korea Expert Systems Association.
- Beaver, W. (1966). Financial Ratios as Predictors of Failures. In Empirical Research in Accounting, *Selected Studies Supplement to the Journal of Accounting Research*, 4, 1, s. 71-127.
- Bernhardsen, E. (2001). A Model of Bankruptcy Prediction, *Norges Bank Financial Analysis and Structure Department, Research Department, Working Paper, ANO 2001/10*.

- Blum, M. P. (1974). Failing Company Discriminant Analysis. *Journal of Accounting Research*, 12, 1, s. 1-25.
- Brealey, R.A. ve Myers, S.C. (2000). Principles of Corporate Finance. McGraw-Hill.
- Brigham, E.F. ve Gapenski, L.C. (1989). Intermediate Financial Management, 3e, Dryden Press.
- Brown, D., James, C. ve Rygaert, M. (1992). The Effects of Leverage on Operating Performance: An Analysis of Firms' Responses to Poor Performance. *Working Paper*, 92-99, The University of Florida.
- Buzzell, R.D. ve Gale, B.T. (1987). The PIMS Principles Linking Strategy to Performance, New York: The Free Press.
- Deakin, E. B. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10, 1, s. 167-179.
- Deakin, E. B. (1977). Business Failure Prediction: An Empirical Analysis. Financial Crises: Institutions and Markets in a Fragile Environment, (Editör: E. Altman ve A. Sametz), s. 117-138, New York, NY: John Wiley.
- Dietrich, J. R. (1984). Discussion of Methodological Issues Related to The Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22 (Supplement): s. 83-86.
- Domingos, P. (1999). MetaCost: A General Method for Making Classifiers Cost Sensitive, Proceedings of the International Conference on Knowledge Discovery and Data Mining, San Diego, CA.
- Dorsey, R., Edmister, R. ve Johnson, J.(1993). Bankruptcy Prediction Using Artificial Neural Systems, *Sponsored by The Research Foundation of the Institute of Chartered Financial Analysts*.
- Edminster, R.O. (1972). An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *Journal of Financial and Quantitative Analysis*, 7, 1, s. 1477-1493.
- Elkan, C. (2001). The Foundations of Cost-Sensitive Learning, Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, (IJCAI'01).
- Frank, E. ve Witten, I.H. (1998). Generating Accurate Rule Sets Without Global Optimization, Proceedings Fifteenth International Conference on Machine Learning (Editör: J. Shavlik), s. 144-151, Madison, WI.SF: Morgan Kaufmann.

- Freitas, A.A. (1999). On Rule Interestingness Measures, *Knowledge-Based Systems*, 12(5-6), s. 309-315).
- Freedman, D. ve Diaconis, P. (1981). On the Histogram as a Density Estimator: L₂ Theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, s. 453–476.
- Frydman, H., Altman, E. ve Kao, D. (1985). Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. *Journal of Finance*, 40, 1, s. 269-291.
- Gilbert, L.R., Menon, K. ve Schwartz, K.B. (1990). Predicting Bankruptcy for Firms in Financial Distress, *Journal of Business Finance and Accounting*, 17 (1), Spring, s. 161.
- Gilson, S.C. (1989). Management Turnover and Financial Distress, *Journal of Financial Economics*, No: 25, s. 241-262.
- Güvenir, H.A. (2003). Benefit Maximization in Classification on Feature Projections, BU-CE-0308, *Bilkent University Technical Report*.
- Hashi, I. (1997). The Economics of Bankruptcy, Reorganisation, and Liquidation, *Russian and East European Finance and Trade*, Vol. 33, no 4, s. 6-34
- Haugen, R. A. ve Senbet, L. W. (1978). The Insignificance of Bankruptcy Costs: To the Theory of Optimal Capital Structure, *Journal of Finance*, Vol. XXXIII. No 2.
- Haugen, R. A. ve Senbet, L. W. (1988). Bankruptcy and Agency Costs: Their Significance to the Theory of Optimal Capital Structure, *Journal of Financial and Quantitative Analysis*, Vol. 23, No 1.
- Holte, R.C. (1991). Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. *Machine Learning*, 11, s. 69-91.
- İkizler, N. (2002). Benefit Maximizing Classification Using Feature Intervals, BU-CE-0208, *Bilkent University Technical Report*.
- İkizler, N. (2001). Mining Interesting Rules in Bank Loans Data, Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks, (Editör: A. Acan, I. Aybay ve M. Salamah), Gazimagusa, T.R.N.C., 238-246.
- İkizler, N. ve Güvenir, H.A. (2003). Feature Dependency in Benefit Maximization: A Case Study in the Evaluation of Bank Loan Applications, Proceedings of the Twelfth Turkish Symposium on Artificial Intelligence and Neural Networks, Çanakkale, Turkey.

- İkizler, N. ve Güvenir, H.A. (2003). Maximizing Benefit of Classifications Using Feature Intervals, BU-CE-0301, *Bilkent University Technical Report*.
- Jaggi, B. ve Lee, P. (2002). Managers of Firms in Financial Distress Face Tricky Accounting Choices, *Lubin Working Research*, Lubin School of Business of Pace University, s. 3-4.
- Jones, F. L. (1987). Current Techniques in Bankruptcy Prediction. *Journal of Accounting Literature*, 6 (Spring), s.131-164.
- Kahl, M. (2002). Economic Distress, Financial Distress and Dynamic Liquidation, *Journal of Finance*, Vol. LVII, No.1.
- Keasey, K. ve Watson, R. (1991). Financial Distress Prediction Models: A Review of Their Usefulness. *British Journal of Management*, Vol. 2, s. 89-102.
- Kim, C. N., Chung, H. M. ve Paradise, D. B. (1997). Inductive Modeling of Expert Decision Making in Loan Evaluation: A Decision Strategy Perspective. *Decision Support Systems*, 21-2, s. 83-98.
- Korobow, L. ve Stuhr, D. (1985). Performance Measurement of Early Warning Models. *Journal of Banking and Finance*, 9, 2, s. 267-273.
- Kutman, Ö. (2001). Türkiye'deki Şirketlerde Erken Uyarı Göstergelerinin Araştırılması, *Doğuş Üniversitesi Dergisi*.
- Laitinen, E.K. ve Laitinen T. (2000). Bankruptcy Prediction Application of the Taylor's Expansion in Logistic Regression, *International Review of Financial Analysis* 9, s. 327-349.
- Langley, Pat. (1996). Elements of Machine Learning, Morgan Kaufman Publishers, Inc., San Fransisco, CA.
- Lau, A. Hing-Ling. (1987). A Five-State Financial Distress Prediction Model. *Journal of Accounting Research*, 25, Spring, s.127-138.
- Libby, R. (1975). Accounting Ratios and the Prediction of Failure: Some Behavioral Evidence. *Journal of Accounting Research*, 13, 1, Spring, s. 150-161.
- Martin, D. (1977). Early Warning of Bank Failure, A Logit Regression Approach. *Journal of Banking and Finance*, 1, 3, s. 249-276.
- Messier, W. F. ve Hansen, J. (1988). Inducing Rules for Expert System Development: An Example Using Default and Bankruptcy Data. *Management Science*, 34, 12, s. 1403-1415.

- Morris, J. R. (1982). Taxes, Bankruptcy Costs and the Existence of an Optimal Capital Structure, *The Journal of Financial Research*, Vol. V, No. 3.
- Morris, J. R. (1997). Early Warning Indicators of Corporate Failure. Ashgate Publishing Ltd., Hants.
- Noreen, E. (1988). An Empirical Comparison of Probit and OLS Regression Hypothesis Tests. *Journal of Accounting Research*, 26, Spring, s. 119-133.
- Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, 18, 1, s. 109-131.
- Opler, T.C. ve Titman, S. (1994). Financial Distress and Corporate Distress, *Journal of Finance*, Vol. XLIX, No. 3.
- Perold, A. F. (1999). Long-Term Capital Management, Case Study, Harvard Business School, 9-200-007, November 5.
- Platt, D.H. ve Platt, M. (2002). Predicting Corporate Financial Distress: Reflections On Choice-Based Sample Bias, *Journal of Economics and Finance*, Vol. 26, No. 2, Summer, s. 184-199.
- Purnanandam, A. (Ağustos 2005). Financial Distress and Corporate Risk Management: Theory and Evidence, *Ross School of Business Working Paper*.
- Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1, s. 81-106.
- Quinlan, J. R. (1993). C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo.
- Quinlan, J. R. (1996). Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence Research*, 4, s. 77-90.
- Reese, S. ve Mc Mahon, T. (2003) Valuing Your Company's Distressed Assets, Presentation on the First Annual Turnaround Management and Corporate Resucturing Summit, *Institutional Investors Seminars*.
- Rivest, R.L. (2001). Learning Decision Lists. *Sponsored Paper, NSF Grant DCR-8607494, ARO Grant DAAL03-86-K-0171 ve Siemens Corporation*.
- Rose, P.S. (1994). Money and Capital Markets, The Financial System in an Increasingly Global Economy, 5e, IRWIN, INC., United States.

- Ross, S.A., Westerfield, R.W. ve Jaffe, J.F. (1996). Corporate Finance, 4th. edition: Irwin, Homewood, IL 60430.
- Scherr, F.C. (1988). The Bankruptcy Cost Puzzle, *Quarterly Journal of Business and Economics*, Vol: 27, Issue 3, s. 147-162.
- Scott, D. W. (1979). On Optimal and Data-Based Histograms. *Biometrika* 66, s. 605–610.
- Scott, J. (1981). The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models. *Journal of Banking and Finance*, 5, 3, September, s. 317-344.
- Sung, T., Chang, N. ve Lee, G. (1999). Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction, *Journal of Management Information Systems*, Vol. 16, No. 1, , Summer, s. 63-85.
- Tucker, J., Neural Networks Versus Logistic Regression in Financial Modelling: A Methodological Comparison, Eriřim: 28 Kasım 2002, <http://citeseer.nj.nec.com/cache/papers/cs/73/http:zSzzSzwww.bioel.e.nuee.nagoya-u.ac.jpzSzwsc1zSzpaperszSzfileszSztucker.pdf/neural-networks-versus-logistic.pdf>.
- Ward, T.J. (1994). An Empirical Study of the Incremental Predictive Ability of Beaver's Naive Operating Flow Measure Using Four-State Ordinal Models of Financial Distress. *Journal of Business Finance and Accounting* (June), s. 547-561.
- Ward, T.J. (1999). A Review of Financial Distress Research Methods and Recommendations for Future Research, *Academy of Accounting and Financial Studies*, Vol. 3, No. 1, s. 160-178.
- Weiss, L. (1981). Bankruptcy Prediction: A Methodological and Empirical Update. *Freeman School of Business Working Paper*, Tulane.
- Weiss, S.M. ve Indurkha, N. (1998). Predictive Data Mining, Morgan Kaufman Publishers, Inc., San Fransisco, CA.
- West, R. G. (1985). A Factor-Analytic Approach to Bank Condition. *Journal of Banking and Finance*, 9, 2, s. 253-266.
- Whitaker, R.B. (1999). The Early Stages of Financial Distress, *Journal of Economics and Finance*, Vol. 23, No. 2, Summer, s. 123-133.
- Wilcox, J. (1971). A Simple Theory of Financial Ratios as Predictors of Failures. *Journal of Accounting Research*, 9, 3, Autumn, s. 389-395.

- Wilcox, J. (1973). A Prediction of Business Failure Using Accounting Data. *Empirical Research in Accounting*, 2, s. 163-179.
- Witten, I.H. ve Frank, E. (2000). Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufman Publishers, Inc., San Fransisco, CA.
- Wruck, K. (1990). Financial Distress: Reorganization and Organization Efficiency, *Journal of Financial Economics*, 27, s. 425.
- Zavgren, C. V., Dugan, M. T. ve Reeve, J. M. (1988). The Association between Probabilities of Bankruptcy and Market Responses: A Test of Market Anticipation. *Journal of Business Finance and Accounting*, 15, 1, Spring, s. 19-45.
- Zadrozny, B. ve Elkan, C., Learning and Making Decisions When Costs and Probabilities are Both Unknown, Eriřim: 08 Aralık 2003, <http://citeseer.nj.nec.com/cache/papers/cs/>
- Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22, 1, s. 59-82.

EKLER

D - İMALAT SANAYİİ ALT SEKTÖRLERİ - NACE (Rev.1) KODLARI

- DA: Gıda, Meşrubat ve Tütün Ürünleri Sanayii
- DB: Tekstil ve Tekstil Ürünleri Sanayii
- DC: Deri ve Deri Ürünleri Sanayii
- DD: Ağaç ve Ağaç Ürünleri Sanayii
- DE: Kağıt Hamuru, Kağıt, Kağıt Ürün.ile Yayımlar ve Basım Sanayii
- DF: Kömür, Rafine Petrol Ürünleri ve Nükleer Yakıt Üretim Sanayii
- DG: Kimya Sanayii
- DH: Kauçuk ve Plastik Ürünleri Sanayii
- DI: Diğer Metal Dışı Madenler Sanayii
- DJ: Metal Ana Sanayii ve İşlenmiş Metal Ürünleri Üretimi Sanayii
- DK: Makina ve Teçhizat (Başka Yerde Sınıflandırılmamış) Sanayii
- DL: Elektrikli ve Optik Aletler Sanayii
- DM: Ulaştırma Araçları Sanayii
- DN: Mobilya İmalatı; Başka Yerde Sınıflandırılmamış Diğer İmalatlar

ÖZNİTELİKLER, AÇIKLAMALARI VE ALDIKLARI DEĞERLER

ÖZNİTELİK	AÇIKLAMA	ALDIĞI DEĞER
NİTELİKSEL BİLGİLER		
borsa	İMKB'ye koteli olup olmadığını gösteren öz nitelik	Koteli, Koteli değil
yilKurtar	Firmanın kuruluşundan inceleme dönemine kadar olan zaman, firmanın yaşı.	Gerçel
sektor	Firmanın faaliyet gösterdiği sektör	DA, DB, DC, DD, DE, DF, DG, DH, DI, DJ, DK, DL, DM, DN
ozKaynak	Öz kaynakların negatif olup olmadığını gösteren değişken	E, H
netSermaye	Net sermayenin negatif olup olmadığını gösteren değişken. (Net Sermaye, öz kaynaklardan sermaye yedeklerinin çıkarılmayıyla elde edilir.)	E, H
donemKar	Dönem kârının negatif olup olmadığını gösteren değişken	E, H
yil	İnceleme dönemi	Tamsayı (1989-2001)
LİKİDİTE ORANLARI		
CariOran	Firmanın dönen varlıklarının kısa vadeli yükümlülüklerine oranıdır. Firmanın likidite durumunun kabaca göstergesidir.	Gerçel
LikiditeOrani	Firmanın dönen varlıklarından stoklar gibi kolayca nakite dönüştürülemeyen varlıkları ve diğer dönen varlıkları gibi kalemler dışlandıktan sonra kısa vadeli borçların ne düzeyde karşılanabileceğinin göstergesidir. Cari orana göre daha muhafazakar bir likidite ölçütüdür.	Gerçel
NakitOrani	Hazır değerler ve menkul kıymetler gibi en likit varlıkların kısa vadeli yükümlülüklerin karşılanmasında ne derece yeterli olduğunun göstergesidir.	Gerçel
StoklarDönenVarliklarOrani	İşletme sermayesinin ne kadarının stoklara bağlandığını ölçer. Dolayısıyla, dönen varlıkların likiditesinin göstergesidir. Cari oran ve likidite oranıyla birlikte incelenmelidir.	Gerçel

ÖZ NİTELİK	AÇIKLAMA	ALDIĞI DEĞER
StoklarAktifToplamiOrani	Stokların toplam varlıklar içindeki düzeyinin ölçüsüdür. Cari oran ve likidite oranıyla birlikte incelenmelidir.	Gerçel
StokBagimlilikOrani	Firmanın kısa vadeli borçlarının geri ödenebilmesinde, stok satışına ne derece bağımlı olduğunu gösterir. Cari oran ve likidite oranıyla bir arada incelenmelidir.	Gerçel
KisaVadeliAlacaklarDönenVarliklarOrani	Firmanın dönen varlıklarının ne kadarının ticari alacaklara bağlandığını gösterir. Alacak devir hızıyla birlikte değerlendirildiğinde, firmanın alacak tahsil politikasında ne kadar başarılı olduğunun bir göstergesidir.	Gerçel
KisaVadeliAlacaklarAktifToplamiOrani	Kısa vadeli alacakların toplam varlıklar içindeki düzeyinin ölçüsüdür.	Gerçel
CabukDegerlerStoklar	Stokların para ve süratle paraya çevrilebilir kıymetlere (çabuk değerler) dönüşüm hızını gösterir.	Gerçel
DönenVarliklarYabancıKaynakToplami	Firmanın yabancı kaynaklarının, ne kadarlık bir kısmının dönen varlıklarının finansmanına ayrıldığının bir ölçüsüdür.	Gerçel
HazirDegerlerMenkulKiymetlerAktifToplami	Firmanın likit varlıklarının bir ölçüsüdür.	Gerçel
HazirDegerlerMenkulKiymetlerNetSatislar	Satışların hazır değerler ve menkul kıymetlere dönüşüm hızını gösterir.	Gerçel
HazirDegerlerMenkulKiymetlerYabancıKaynakToplami	Likit kaynaklarla yabancı kaynakların ne ölçüde karşılanabileceğinin bir ölçüsüdür. (Nakit oranının benzeri)	Gerçel
NetCabukDegerlerStoklar	Stokların para ve süratle paraya çevrilebilir kıymetlere (çabuk değerler) dönüşüm hızını gösterir.	Gerçel
NetIsletmeSermayesiAktifToplami	Likiditenin bir ölçüsüdür	Gerçel
NetIsletmeSermayesiKisaVadeliYabancıKaynaklar	Likiditenin bir ölçüsüdür	Gerçel
NetIsletmeSermayesiOzKaynaklar	Likiditenin bir ölçüsüdür	Gerçel
KVTicariAlacaklarKVTicariBorclar	Firmanın alacak tahsili ve borçlanma politikalarını bir arada inceler. Vade uyumu olup olmadığının bir göstergesidir. 1'den küçük olması tercih edilir. Aksi durumda uzun vadeli ticari borçlanmayla birlikte incelenmesi gerekir.	Gerçel
FİNANSAL YAPI ORANLARI		
YabancıKaynaklarToplamiAktifToplamiOrani	Firmanın aktiflerinin finansmanında ne oranda yabancı kaynak kullanıldığının bir ölçüsüdür.	Gerçel
OzKaynaklarAktifToplamiOrani	Firmanın aktiflerinin finansmanında ne oranda öz kaynak kullanıldığının bir ölçüsüdür.	Gerçel
OzKaynaklarYabancıKaynaklarToplamiOrani	Finansman yapısının bir ölçüsüdür.	Gerçel

ÖZİNİTELİK	AÇIKLAMA	ALDIĞI DEĞER
Kısa Vadeli Yabancı Kaynaklar Pasif Toplamı Oranı	Yabancı kaynakların vade yapısıyla ilgili bir orandır. Oranın yüksek olması, varlıkların büyük bir bölümünün kısa vadeli yabancı kaynaklarla finanse edildiğini gösterir. Likit varlıkları fazla olan bir firma için yüksek olması fazla risk taşımaz. Dönen Varlık/Aktif Toplamı oranıyla bir arada incelenmelidir.	Gerçek
Uzun Vadeli Yabancı Kaynaklar Pasif Toplamı Oranı	Yabancı kaynakların vade yapısıyla ilgili bir orandır. Sermaye yoğun teknoloji kullanan sektörlerde bu oran daha yüksektir. Bu oranın yüksek olması ayrıca, firmanın varlıklarının finansmanı için daha uzun vadeli kaynaklara ulaşabildiğinin göstergesi de olabilir. Duran Varlık/Aktif Toplamı oranıyla bir arada incelenmelidir.	Gerçek
Uzun Vadeli Yabancı Kaynaklar Devamlı Sermaye Oranı	Firmanın finansmanında uzun vadeli kaynaklarının ne kadarının firma dışından sağlandığının bir göstergesidir. Öz Kaynaklar/Yabancı Kaynaklar oranının bir başka şekilde ifadesidir. Burada vade açısından ayrıntı görmek daha olanaklıdır.	Gerçek
Maddi Duran Varlıklar Öz Kaynaklar Oranı	Öz kaynakların ne kadarlık bir kısmının maddi duran varlıkların finansmanında kullanıldığını gösterir. Bu oran, sektörün yapısına göre farklılık gösterebilir.	Gerçek
Maddi Duran Varlıklar Uzun Vadeli Yabancı Kaynaklar Oranı	Uzun vadeli yabancı kaynakların ne kadarlık bir kısmının maddi duran varlıkların finansmanında kullanıldığını gösterir.	Gerçek
Duran Varlıklar Yabancı Kaynaklar Toplamı Oranı	Yabancı kaynakların ne kadarlık kısmının maddi duran varlıkların finansmanında kullanıldığını gösterir.	Gerçek
Duran Varlıklar Öz Kaynaklar Oranı	Öz kaynakların ne kadarlık kısmının duran varlıkların finansmanında kullanıldığını gösterir.	Gerçek
Duran Varlıklar Devamlı Sermaye Oranı	Öz kaynak ve uzun vadeli yabancı kaynak toplamının ne kadarlık bir kısmının duran varlıkların finansmanında kullanıldığını gösterir.	Gerçek
Kısa Vadeli Yabancı Kaynaklar Yabancı Kaynaklar Toplamı Oranı	Varlıkların finansmanında kullanılan yabancı kaynakların kadarlık bir kısmının kısa vadeli olduğunu gösterir. Sermaye yoğun işletmelerde bu oran daha düşüktür.	Gerçek
Banka Kredileri Aktif Toplamı Oranı	Varlıkların ne kadarlık bir kısmının banka kredileriyle finanse edildiğini gösterir.	Gerçek
Kısa Vadeli Banka Kredileri Kısa Vadeli Yabancı Kaynaklar Oranı	Firmanın kısa vadeli yabancı kaynaklarının ne kadarının banka kredileriyle karşılandığının bir ölçüsüdür.	Gerçek
Banka Kredileri Yabancı Kaynaklar Toplamı Oranı	Firmanın yabancı kaynaklarının ne kadarının banka kredileriyle karşılandığının bir ölçüsüdür.	Gerçek

ÖZ NİTELİK	AÇIKLAMA	ALDIĞI DEĞER
Banka Kredileri Net Satışlar	Bankalardan alınan kredilerin ne kadarının net satışlar tarafından karşılanabildiğinin bir ölçüsüdür.	Gerçek
Dönen Varlıklar Aktif Toplamı Oranı	Firmanın toplam varlıkları içinde likit varlıklarının bir ölçüsüdür.	Gerçek
Duran Varlıklar Aktif Toplamı	Firmanın toplam varlıkları içinde duran varlıklarının payını gösterir.	Gerçek
Maddi Duran Varlıklar Aktif Toplamı Oranı	Firmanın toplam varlıkları içinde maddi duran varlıklarının payını gösterir.	Gerçek
Cabuk Değerler Aktif Toplamı	Firma likiditesinin bir ölçüsü olan oran, çabuk değerlerin (para ve süratle paraya çevrilebilir kıymetler) toplam varlıklara oranıdır.	Gerçek
Kısa Vadeli Yabancı Kaynaklar Duran Varlıklar	Duran varlıkların ne kadarlık bir kısmının kısa vadeli yabancı kaynaklarla finanse edildiğinin bir göstergesidir. Duran varlıkların kısa vadeli yabancı kaynaklarla finanse edilmesi arzulanmayan riskli bir duruma işaret eder.	Gerçek
Kısa Vadeli Yabancı Kaynaklar Öz Kaynaklar	Firmanın finansman yapısı hakkında bilgi veren, mali yeterliliğinin araştırılmasına olanak tanıyan bir orandır.	Gerçek
Uzun Vadeli Yabancı Kaynaklar Öz Kaynaklar	Devamlı Sermayenin ne kadarının yabancı kaynaklardan oluştuğunun bir ölçüsüdür.	Gerçek
DEVİR HIZLARI		
Stok Devir Hızı	Stokların ne kadar bir süre içerisinde üretim faaliyetinde tükendiği ve satış hasılatına dönüştüğünü göstermektedir.	Gerçek
Alacak Devir Hızı	Alacakların tahsil kabiliyetinin bir ölçüsüdür.	Gerçek
Çalışma Sermayesi Devir Hızı	Firmanın çalışma sermayesi verimliliğinin bir ölçüsüdür.	Gerçek
Net Çalışma Sermayesi Devir Hızı	Firmanın net çalışma sermayesi verimliliğinin bir ölçüsüdür.	Gerçek
Maddi Duran Varlık Devir Hızı	Firmanın maddi duran varlık dönüşüm verimliliğinin bir ölçüsüdür. Firmanın maddi duran varlıklara aşırı bir yatırım yapıp yapmadığını, veya maddi duran varlıkların kapasitelerinin üzerinde kullanılıp kullanılmadığını gösterir.	Gerçek
Duran Varlık Devir Hızı	Firmanın duran varlık dönüşüm verimliliğinin bir ölçüsüdür.	Gerçek
Öz Kaynaklar Devir Hızı	Firmanın sermayesinin verimli kullanılıp kullanılmadığının bir göstergesidir. Oranın yüksek olması, sermayenin verimli kullanıldığını gösterebileceği gibi, öz kaynakların finansmanda yetersiz kaldığı anlamına da gelebilir.	Gerçek
Aktif Devir Hızı	Firmanın sahip olduğu varlıklarının verimliliğinin bir ölçüsüdür.	Gerçek
Cabuk Değerler Net Satışlar	Satışların para ve süratle paraya çevrilebilir kıymetlere (çabuk değerler) dönüşüm hızını gösterir.	Gerçek

ÖZNETELİK	AÇIKLAMA	ALDIĞI DEĞER
NetSatislarKisaVadeliYabancıKaynakToplami	Kısa vadeli yabancı kaynakların ne kadarının net satışlardan karşılanabileceğinin bir ölçütüdür.	Gerçel
NetSatislarStoklar	Stok devir hızının statik bir şekilde ifadesidir.	Gerçel
NetSatislarYabancıKaynakToplami	Yabancı kaynakların ne kadarının net satışlardan sağlanabileceğinin bir göstergesidir.	Gerçel
StoklarNetCalismaSermayesi	Net çalışma sermayesinin üretim faaliyetine dönüşme çabukluğunun bir göstergesidir.	Gerçel
StoklarNetSatislar	Stok devir hızının statik bir şekilde ifadesidir.	Gerçel
KARLILIK ORANLARI		
NetKarOzKaynaklarOrani	"Yatırımın Getirisi" olarak da bilinen oran, yatırım verimliliğinin ölçüsüdür. Firma sahipleri tarafından sağlanan sermayenin bir birimine isabet eden kar nisbetini ortaya koyar.	Gerçel
NetKarKisaVadeliYabancıKaynak	firmanın kısa vadeli yükümlülüklerinin ne kadarının net karla karşılanabileceğinin ölçüsüdür.	Gerçel
VergiOncesiKarOzKaynaklarOrani	Net Kâr/Öz Kaynaklar Oranı ile bir arada ele alındığında, vergi yükümlülüklerinin, yatırım verimliliğini ne ölçüde azalttığının bir ölçüsüdür.	Gerçel
FaaliyetKariOzKaynaklar	Firma sahipleri tarafından sağlanan sermayenin bir birimine isabet eden esas faaliyetlerden elde edilen kar nisbetini ortaya koyar.	Gerçel
FaizVeVergiOncesiKarPasifTOrani	Firmanın kaynaklarının ne kadar verimli kullanıldığının bir ölçüsüdür.	Gerçel
NetKarAktifToplamiOrani	Varlıkların kârlılık veriminin ölçüsüdür.	Gerçel
NetKarDönenVarliklar	Dönen varlıkların verimliliğinin bir ölçütüdür.	Gerçel
NetKarNetCalismaSermayesi	Net çalışma sermayesinin verimliliğinin bir ölçüsüdür.	Gerçel
FaaliyetKariFaaliyetinGerçeklestirilmesindeKulvarlikOrani	Faaliyetin gerçekleşmesinde kullanılan varlıkların kârlılık veriminin ölçüsüdür.	Gerçel
FaaliyetKariAktifToplami	Firmanın varlıklarının ne kadar verimli kullanıldığının bir göstergesidir. Firmanın varlıklarının esas faaliyetlerden elde edilen getirisini ölçer.	Gerçel
FaaliyetKariNetIsletmeSermayesi	Firmanın net işletme sermayesinin ne kadar verimli kullanıldığının bir göstergesidir. Firmanın net işletme sermayesinin esas faaliyetlerden elde edilen getirisini ölçer.	Gerçel
BirikmeliKarlilikOrani	Firma kârlarının sermaye dönüştürülme oranıdır.	Gerçel
FaaliyetKariNetSatislarOrani	Satışların kâr verimliliğinin ölçüsüdür.	Gerçel
BrütSatisKariNetSatislarOrani	Satışların kâr verimliliğinin ölçüsüdür.	Gerçel

ÖZNİTELİK	AÇIKLAMA	ALDIĞI DEĞER
BrütSatisKariAktifToplami	Firmanın toplam varlıklarının ne kadarının brüt satış karına (firmanın net satışlarından maliyetleri düşüldükten sonra elde edilen kar) dönüştürülebildiğinin bir ölçüsüdür.	Gerçel
NetKarNetSatislarOrani	Satışların kâr verimliliğinin ölçüsüdür.	Gerçel
SatılanMalinMaliyetiNetSatislarOrani	Satışlar içinde üretim maliyetlerinin payını ölçer.	Gerçel
FaaliyetGiderleriNetSatislarOrani	Satışlar içinde faaliyet giderlerinin payını ölçer.	Gerçel
FaizGiderleriNetSatislarOrani	Satışları içinde finansman giderlerinin payını ölçer.	Gerçel
FaizGiderleriHazirDegerlerMenkulKiymetler	Faiz yükümlülüklerinin en likit varlıklar olan hazır değerler ve menkul kıymetlerle ne oranda karşılanabileceğinin ölçüsüdür.	Gerçel
FaizVeVergiOncesiKarDuranVarliklar	Firmanın duran varlıklarının ne kadar verimli kullandığının bir ölçüsüdür.	Gerçel
FaizVeVergiOncesiKarNetSatislar	Satışlar karşılığında vergi yükümlülükleri ve faiz giderleri ödenmeden önce elde edilen karın ölçüsüdür.	Gerçel
FaizVeVergiOncesiKarOzKaynaklar	Firma sahipleri tarafından sağlanan sermayenin bir birimine isabet eden vergi yükümlülükleri ve faiz giderleri ödenmeden önce elde edilen kar nisbetini ortaya koyar.Ekonomik rantabiliteyi sermaye açısından ele alan bir ölçüttür.	Gerçel
FaizVeVergiOncesiKarFaizGiderleriOrani	Vergi yükümlülükleri ve faiz giderleri ödenmeden önce elde edilen kârların, faiz giderlerinin ne kadarını karşıladığının bir ölçüsüdür.	Gerçel
NetKarVeFaizGiderleriFaizGiderleriOrani	Faiz giderleri ödenmeden önce elde edilen kârların faiz giderlerinin ne kadarını karşıladığının ölçüsüdür.	Gerçel
FaizGiderleriYabancıKaynakToplami	Yabancı kaynak kullanım maliyetinin bir ölçüsüdür.	Gerçel
NetKarFaizGiderleri	Finansman giderleri karşılandıktan sonra, firmanın faiz ödeme gücünün ölçüsüdür.	Gerçel
OlaganKarNetSatislar	Firmanın olağan faaliyetlerinden elde ettiği kârın net satışlara oranıdır.	Gerçel
VergiOncesiKarAktifToplami	Varlıkların verimliliğinin bir ölçüsüdür.	Gerçel
VergiOncesiKarKisaVadeliYabancıKaynaklar	Kısa vadeli yabancı kaynakların ne kadar verimli kullanıldığının bir göstergesidir.	Gerçel
VergiOncesiKarNetSatislar	Satışlar karşılığında vergi yükümlülükleri ödenmeden önce elde edilen karın ölçüsüdür.	Gerçel
FinansalGelirlerNetSatislar	Firmanın faaliyetlerinin dışında finansman faaliyetlerinden elde ettiği gelirleriyle net satışları arasındaki ilişkinin bir ölçüsüdür.	Gerçel
FinansalGiderlerNetSatislar	Firmanın finansman giderlerinin net satışlarına oranıdır	Gerçel

ÖZNETELİK	AÇIKLAMA	ALDIĞI DEĞER
VERİMLİLİK ORANLARI		
NetKısaVadeliTicariAlacaklarAktifToplami	Kısa vadeli ticari alacaklar ve kısa vadeli ticari borçlar arasındaki farkın varlık toplamına oranıdır	Gerçel
NetUzunVadeliTicariAlacaklarAktifToplami	Uzun vadeli ticari alacaklar ve uzun vadeli ticari borçlar arasındaki farkın varlık toplamına oranıdır	Gerçel
NetKısaVadeliDiğerAlacaklarAktifToplami	Kısa vadeli diğer alacaklar ve kısa vadeli diğer borçlar arasındaki farkın varlık toplamına oranıdır	Gerçel
NetUzunVadeliDiğerAlacaklarAktifToplami	Uzun vadeli diğer alacaklar ve uzun vadeli diğer borçlar arasındaki farkın varlık toplamına oranıdır	Gerçel
NetKısaVadeliTicariAlacaklarNetOzKaynaklar	Kısa vadeli ticari alacaklar ve kısa vadeli ticari borçlar arasındaki farkın net sermayeye (sermayeden yapay olarak oluşan sermaye artışı (sermaye yedekleri) çıkarıldıktan sonraki fark) oranıdır	Gerçel
NetUzunVadeliTicariAlacaklarNetOzKaynaklar	Uzun vadeli ticari alacaklar ve uzun vadeli ticari borçlar arasındaki farkın net sermayeye oranıdır	Gerçel
NetKısaVadeliDiğerAlacaklarNetOzKaynaklar	Kısa vadeli diğer alacaklar ve kısa vadeli diğer borçlar arasındaki farkın net sermayeye oranıdır	Gerçel
NetUzunVadeliDiğerAlacaklarNetOzKaynaklar	Uzun vadeli diğer alacaklar ve uzun vadeli diğer borçlar arasındaki farkın net sermayeye oranıdır	Gerçel
NetKısaVadeliTicariAlacaklarNetSatislar	Kısa vadeli ticari alacaklar ve kısa vadeli ticari borçlar arasındaki farkın net satışlara oranıdır	Gerçel
NetUzunVadeliTicariAlacaklarNetSatislar	Uzun vadeli ticari alacaklar ve uzun vadeli ticari borçlar arasındaki farkın net satışlara oranıdır	Gerçel
NetKısaVadeliDiğerAlacaklarNetSatislar	Kısa vadeli diğer alacaklar ve kısa vadeli diğer borçlar arasındaki farkın net satışlara oranıdır	Gerçel
NetUzunVadeliDiğerAlacaklarNetSatislar	Uzun vadeli diğer alacaklar ve uzun vadeli diğer borçlar arasındaki farkın net satışlara oranıdır	Gerçel
NetSermayeAktifToplami	Sermayeden yapay olarak oluşan sermaye artışı (sermaye yedekleri) çıkarıldıktan sonraki farkın varlıklara oranıdır. Sermaye yitiliğinin gerçekçi bir şekilde ölçülebilmesi için net sermayenin hesaplanması gerekir.	Gerçel
NetSermayeNetSatislar	Sermayeden yapay olarak oluşan sermaye artışı (sermaye yedekleri) çıkarıldıktan sonraki farkın net satışlara oranıdır.	Gerçel
NetSermayeDönemNetKz	Sermayeden yapay olarak oluşan sermaye artışı (sermaye yedekleri) çıkarıldıktan sonraki farkın dönem kârına oranıdır.	Gerçel

ÖZİNİTELİK	AÇIKLAMA	ALDIĞI DEĞER
ORTAK BÜYÜKLÜK ORANLARI		
DonenVarliklar		Gerçel
HazirDegerler		Gerçel
MenkulKiymetler		Gerçel
KVTicariAlacaklar		Gerçel
KVDigerAlacaklar		Gerçel
Stoklar		Gerçel
YillaraYayginInsveOnarimMaliyetleri		Gerçel
GelecekAylaraAitGiderlerVeGelirTah		Gerçel
DigerDönenVarliklar		Gerçel
DuranVarliklar		Gerçel
UVTicariAlacaklar		Gerçel
UVDigerAlacaklar		Gerçel
MaliDuranVarliklar		Gerçel
MaddiDuranVarliklar		Gerçel
MaddiOlmayanDuranVarliklar		Gerçel
OzelTükenmeyeTabiVarliklar		Gerçel
GelecekYillaraAitGiderlerVeGelirTah		Gerçel
DigerDuranVarliklar		Gerçel
KisaVadeliYabancıKaynaklar		Gerçel
KVMaliBorclar		Gerçel
KVTicariBorclar		Gerçel
KVDigerBorclar		Gerçel
KVAlinanAvanslar		Gerçel
YillaraYayginInsveOnarimHakedisleri		Gerçel
OdenecekVergiVeDigerYükümlülükler		Gerçel
KVBorcVeGiderKarsiliklari		Gerçel
GelecekAylaraAitGelirlerVeGiderTah		Gerçel
DigerKisaVadeliYabancıKaynaklar		Gerçel
UzunVadeliYabancıKaynaklar		Gerçel
UVMaliBorclar		Gerçel
UVTicariBorclar		Gerçel
UVDigerBorclar		Gerçel
UValinanAvanslar		Gerçel
UVBorcVeGiderKarsiliklari		Gerçel

ÖZİNİTELİK	AÇIKLAMA	ALDIĞI DEĞER
GelecekYillaraAitGelirlerVeGiderTah		Gerçel
DigerUzunVadeliYabancıKaynaklar		Gerçel
OzKaynaklar		Gerçel
OdenmisSermaye		Gerçel
SermayeYedekleri		Gerçel
KarYedekleri		Gerçel
GecmisYillarKarlari		Gerçel
GecmisYillarZararlari		Gerçel
DönemNetKari		Gerçel
BrütSatislar		Gerçel
SatisIndirimleri		Gerçel
SatislarinMaliyeti		Gerçel
BrütSatisKariVeyaZarari		Gerçel
FaaliyetGiderleri		Gerçel
FaaliyetKariVeyaZarari		Gerçel
DigerFaalolaganGelirVeKarlar		Gerçel
DigerFaalolaganGiderVeZarar		Gerçel
FinansmanGiderleri		Gerçel
OlaganKarVeyaZarar		Gerçel
OlagandisiGelirVeKarlar		Gerçel
OlagandisiGiderVeZararlar		Gerçel
DonemKarVeyaZarari		Gerçel
DonemKariVergiDigYasalYükkars		Gerçel
DonemNetKarVeyaZarari		Gerçel
sınıf		BAŞARILI, BAŞARISIZ

BCFP İLGİNÇ VE ANLAMLI KURALLAR**Kural 1:**

FaizGiderleriYabancıKaynakToplami > 0.14005
İSE BAŞARISIZ

Kural 2:

HazirDegerlerMenkulKiyametlerAktifToplami < 0.07921
İSE BAŞARISIZ

Kural 3:

NetIsletmeSermayesiKisaVadeliYabancıKaynaklar < 1.52567
İSE BAŞARISIZ

Kural 4:

VergiOncesiKarKisaVadeliYabancıKaynaklar < 0.28187
İSE BAŞARISIZ

Kural 5:

NetKisaVadeliTicariAlacaklarAktifToplami < 0.04913
İSE BAŞARISIZ

Kural 6:

NetUzunVadeliTicariAlacaklarAktifToplami < -0.0047
İSE BAŞARISIZ

Kural 7:

NetUzunVadeliTicariAlacaklarNetSatislar < -0.01480
İSE BAŞARISIZ

Kural 8:

NetSermayeAktifToplami < 0.25678
İSE BAŞARISIZ

Kural 9:

DuranVarliklar > 0.31386
İSE BAŞARISIZ

Kural 10:

MaliDuranVarliklar > 0.01961
İSE BAŞARISIZ

Kural 11:

MaddiOlmayanDuranVarliklar > 0.00403
İSE BAŞARISIZ

Kural 12:

DigerKisaVadeliYabanciKaynaklar > 0.00402
İSE BAŞARISIZ

Kural 13:

UzunVadeliYabanciKaynaklar > 0.08404
İSE BAŞARISIZ

Kural 14:

UVMaliBorclar > 0.04399
İSE BAŞARISIZ

Kural 15:

OzKaynaklar < 0.36845
İSE BAŞARISIZ

Kural 16:

DönemNetKari < 0.05559
İSE BAŞARISIZ

Kural 17:

SatisIndirimleri > 0.0322
İSE BAŞARISIZ