

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



HALKA AÇIK FİNANS DIŐI ŐİRKETLERDE
SÜREKLİLİK RİSKİNİN MAKİNE ÖĐRENMESİ İLE ÖNGÖRÜLMESİ

DOKTORA TEZİ

Gürkan SAYIN

Muhasebe ve Finans Yönetimi Anabilim Dalı

Muhasebe ve Denetimi Bilim Dalı

Tez Danışmanı: Prof. Dr. Gülümser Ünkaya

Őubat, 2019

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



HALKA AÇIK FİNANS DIŐI ŐİRKETLERDE
SÜREKLİLİK RİSKİNİN MAKİNE ÖĐRENMESİ İLE ÖNGÖRÜLMESİ

DOKTORA TEZİ

Gürkan SAYIN
(Y1414.650006)

Muhasebe ve Finans Yönetimi Anabilim Dalı

Muhasebe ve Denetimi Bilim Dalı

Tez Danışmanı: Prof. Dr. Gülümser Ünkaya

Şubat, 2019





14/03/2019

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ
DOKTORA TEZ ONAY BELGESİ

Enstitümüz Muhasebe Anabilim Dalı, Muhasebe ve Denetim Doktora Programı Y1414.650006 numaralı öğrencisi Gürkan SAYIN' ın "HALKA AÇIK FİNANS DIŞI ŞİRKETLERDE SÜREKLİLİK RİSKİNİN MAKİNE ÖĞRENMESİ İLE ÖNGÖRÜLMESİ" adlı doktora tez çalışması Enstitümüz Yönetim Kurulunun 13/02/2019 tarih ve 2019/03 sayılı kararı ile oluşturulan jüri tarafından *oybirliği* ile Doktora tezi olarak *Kabul* edilmiştir.

	Unvan- Ad-Soyad	İmza
Danışman	Prof. Dr. Gülümser ÜNKAYA	<i>[Signature]</i>
Üye (TİK)	Dr. Öğr. Üyesi Günay Deniz DURSUN	<i>[Signature]</i>
Üye (TİK)	Prof. Dr. Mustafa Nafiz DURU	<i>[Signature]</i>
Üye	Prof. Dr. Fatma PAMUKÇU	<i>[Signature]</i>
Üye	Prof. Dr. Rahmi YAMAK	<i>[Signature]</i>

Tezin Savunulduğu Tarih : 14/03/2019

Sosyal Bilimleri Enstitüsü Yönetim Kurulu'nun tarih ve sayılı kararı ile onaylanmıştır.

Prof. Dr. Ragıp Kutay KARACA

Enstitü Müdürü



YEMİN METNİ

Doktora tezi olarak sunduđum ‘‘Halka Aık Finans DıŐı Őirketlerde Srekliplik Riskinin Makine đrenmesi İle ngrlmesi’’ adlı tezin, proje safhasından sonulanmasına kadarki btn srelerde, bilimsel ahlak ve geleneklere aykırı dŐecek bir yardıma baŐvurulmaksızın yazıldıđını ve yararlandıđım eserlerin kaynaklarda gsterilenlerden olduđuunu, bunlara atıf yapılarak yararlanılmıŐ olduđuunu belirtir ve onurumla beyan ederim.

Grkan SAYIN
19 Mart 2019



ÖNSÖZ

Tez çalışmalarında bana yol gösteren ve her zaman destek olan saygıdeğer hocam Prof. Dr. Gülümser Ünkeya'ya sonsuz teşekkür ederim.

Şubat 2019

Gürkan SAYIN





İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	vii
İÇİNDEKİLER	ix
KISALTMALAR	xi
ÇİZELGE LİSTESİ.....	xiii
ŞEKİL LİSTESİ.....	xv
ÖZET.....	xvii
ABSTRACT	xix
1. GİRİŞ	1
2. KAVRAMSAL ÇERÇEVE	5
2.1 Bağımsız Denetim	5
2.2 Bağımsız Denetçi	5
2.3 Denetim Raporu	5
2.4 Halka Açık Finans Dışı Şirketler	5
2.5 İşletmenin Sürekliliği	6
2.6 Süreklilik Riski	6
2.7 Finansal Risk	7
2.8 Finansal Süreklilik Riski	7
2.9 Makine Öğrenmesi	8
2.9.1 Makine Öğrenmesi kavramları.....	8
2.9.1.1 Değişken Önemi.....	8
2.9.1.2 Şartlı Önemlilik.....	8
2.9.1.3 Karışıklık Matrisi	8
2.9.1.4 Model İçi Performans (In-Sample Fit).....	9
2.9.1.5 Çapraz Doğrulama (Cross Validation).....	9
2.9.1.6 Aşırı Uygunluk (Overfitting)	9
2.9.1.7 Logit fonksiyonu	9
2.9.1.8 Probit fonksiyonu	9
2.9.1.9 Özyinelemeli Bölümlendirme (Recursive Partitioning).....	10
2.9.2 Makine Öğrenmesi modelleri.....	10
2.9.2.1 Sınıflandırma Aracı.....	10
2.9.2.2 Çoklu Ayraç Analizi	10
2.9.2.3 Lojistik Model.....	10
2.9.2.4 Sınır Ağları.....	11
2.9.2.5 DVM	11
2.9.2.6 Karar Ağacı	11
2.9.2.7 Karar Ormanı	11
3. LİTERATÜRÜN İNCELENMESİ.....	13
3.1 Makine Öğrenmesi Modelleri	13
3.1.1 Tek değişkenli doğrusal modeller	13
3.1.2 Çok değişkenli doğrusal modeller.....	15
3.1.3 Çok değişkenli doğrusal olmayan modeller.....	20

3.2 Modelleme Esnasında Karşılaşılan Sorunlar.....	25
4. ÇALIŞMADA KULLANILAN MAKİNE ÖĞRENMESİ MODELLERİ	29
4.1 Çoklu Ayraç Analizi.....	29
4.2 Lojistik Model.....	31
4.3 Sinir Ağları.....	33
4.4 Destek Vektör Makineleri	36
4.5 Karar Ağaçları	40
4.5.1 Breiman Ağacı.....	44
4.5.2 Hothorn Ağacı	45
4.6 Karar Ormanları	46
4.6.1 Breiman Ormanı.....	47
4.6.2 Hothorn Ormanı	48
5. HALKA AÇIK FİNANS DIŞI ŞİRKETLERDE SÜREKLİLİK RİSKİNİN MAKİNE ÖĞRENMESİ İLE ÖNGÖRÜLMESİ	49
5.1 Araştırmanın Amacı	49
5.2 Araştırmanın Önemi	50
5.3 Araştırmanın Yöntemi.....	50
5.3.1 Verinin hazırlanması	50
5.3.2 Değişken sayısının azaltılması	52
5.3.3 Model oluşturma ve seçme.....	56
5.3.4 Çapraz Doğrulama.....	56
5.3.5 Model eğitimi ve optimizasyonu rutini	57
5.4 Araştırmanın Ana Kütlesi ve Örneklemi.....	57
5.5 Araştırmanın Kısıtları.....	59
5.5.1 Verinin hacmi	59
5.5.2 Modelleme araçlarının teorik altyapısı.....	60
5.6 Araştırmanın Bulguları.....	60
5.6.1 Çoklu Ayraç Analizi.....	60
5.6.2 Lojistik Model	65
5.6.3 Sinir Ağları	68
5.6.4 Destek Vektör Makineleri	76
5.6.5 Breiman Ağacı.....	81
5.6.6 Breiman Ağacı modelinde “Karmaşıklık Parametresi” etkisi.....	83
5.6.7 Hothorn Ağacı	85
5.6.8 Breiman Ormanı	87
5.6.9 Hothorn Ormanı	89
5.7 Araştırmanın Sonuçları	91
5.7.1 Model performanslarının karşılaştırılması	92
5.7.2 Modellerin kuvvetli ve zayıf yönlerinin karşılaştırılması	93
5.7.3 En başarılı model ve bu modelin süreklilik riski öngörüsü.....	96
6. SONUÇ VE ÖNERİLER	99
KAYNAKLAR.....	105
ÖZGEÇMİŞ.....	115

KISALTMALAR

BDY	: Bağımsız Denetim Yönetmeliği
BİST	: Borsa İstanbul A.Ş.
CAA	: Çoklu Ayraç Analizi
DVM	: Destek Vektör Makineleri
FAVÖK	: Faiz, Amortisman ve Vergi Öncesi Kar
Kurum	: Kamu Gözetimi, Muhasebe ve Denetim Standartları Kurumu
TDS	: Türkiye Denetim Standartları
TMA	: Türkiye Muhasebe Standartları



ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1 : Örnek Karışıklık Matrisi.....	9
Çizelge 4.1 : Karar Ağacı Modeli – Model İçi Performans.....	44
Çizelge 5.1 : Çalışmada Kullanılması Planlanan Finansal Oranlar.....	53
Çizelge 5.2 : Model Eğitimlerinde Kullanılacak “Özet” Finansal Oranlar.....	55
Çizelge 5.3 : Çalışmaya Konu Şirketlerin Dahil Olduğu Sektörler.....	58
Çizelge 5.4 : Değişkenlerin Sonuca Katkısı.....	61
Çizelge 5.5 : ÇAA Modeli, Tüm Değişkenler Kullanılmış.....	61
Çizelge 5.6 : Değişkenlerin Sonuca Katkısı.....	62
Çizelge 5.7 : ÇAA Modeli, Değişken Sayısı 7.....	62
Çizelge 5.8 : Değişkenlerin Sonuca Katkısı.....	63
Çizelge 5.9 : ÇAA Modeli, Değişken Sayısı 6.....	63
Çizelge 5.10 : Değişkenlerin Sonuca Katkısı.....	63
Çizelge 5.11 : ÇAA Modeli, 5 Değişkenli.....	64
Çizelge 5.12 : ÇAA Modellerinin Performans Karşılaştırması.....	64
Çizelge 5.13 : Değişkenlerin Sonuca Katkısı.....	65
Çizelge 5.14 : Lojistik Model – “Lasso” yöntemi.....	66
Çizelge 5.15 : Değişkenlerin Sonuca Katkısı.....	67
Çizelge 5.16 : Lojistik Model – “Ridge” yöntemi.....	67
Çizelge 5.17 : Lojistik Modeller’in Performans Karşılaştırması.....	68
Çizelge 5.18 : Değişkenlerin Sonuca Katkısı.....	69
Çizelge 5.19 : Sinir Ağı 1 – Model İçi Performans.....	70
Çizelge 5.20 : Değişkenlerin Sonuca Katkısı.....	71
Çizelge 5.21 : Sinir Ağı 2 – Model İçi Performans.....	72
Çizelge 5.22 : Sinir Ağı 3 – Model İçi Performans.....	72
Çizelge 5.23 : Değişkenlerin Sonuca Katkısı.....	73
Çizelge 5.24 : Sinir Ağı 4 – Model İçi Performans.....	74
Çizelge 5.25 : Sinir Ağı 5 – Model İçi Performans.....	74
Çizelge 5.26 : Sinir Ağı 6 – Model İçi Performans.....	75
Çizelge 5.27 : Sinir Ağları’nın Performans Karşılaştırması.....	76
Çizelge 5.28 : Parametre Optimizasyonu.....	77
Çizelge 5.29 : DVM Modeli – Model İçi Performans.....	78
Çizelge 5.30 : Breiman Ağacı – Model İçi Performans.....	82
Çizelge 5.31 : Değişkenlerin Sonuca Katkısı.....	82
Çizelge 5.32 : Aşırı Büyük Breiman Ağacı – Model İçi Performans.....	84
Çizelge 5.33 : Breiman Ağaçları Performans Karşılaştırması.....	84
Çizelge 5.34 : Değişkenlerin Sonuca Katkısı.....	86
Çizelge 5.35 : Hothorn Ağacı – Model İçi Performans.....	86
Çizelge 5.36 : Değişkenlerin Sonuca Katkısı “Doğru Sınıflama”.....	88
Çizelge 5.37 : Değişkenlerin Sonuca Katkısı “Yaprak Saflığı”.....	88
Çizelge 5.38 : Breiman Ormanı – Model İçi Performans.....	89
Çizelge 5.39 : Değişkenlerin Sonuca Katkısı.....	90

Çizelge 5.40 : Hothorn Ormanı – Model İçi Performans.....	91
Çizelge 5.41 : Modellerin En Başarılı Alternatiflerinin Karşılaştırılması	92
Çizelge 5.42 : Modellerin Kuvvetli ve Zayıf Yönleri.....	93



ŞEKİL LİSTESİ

Sayfa

Şekil 4.1 : Çoklu Ayraç Analizi'ne Örnek	30
Şekil 4.2 : Lojistik Model	33
Şekil 4.3 : Sınır Ağı	34
Şekil 4.4 : DVM Lineer Model	37
Şekil 4.5 : DVM Radyal Model	38
Şekil 4.6 : DVM Polinom Model	39
Şekil 4.7 : DVM Sigmoid Model	40
Şekil 4.8 : İki Grup Verinin Dağılımı	42
Şekil 4.9 : Karar Ağacı	43
Şekil 4.10: Karar Ağacı Ayrırımları	43
Şekil 5.1 : Değişkenlerin Sonuca “Bağımsız” Katkısı	54
Şekil 5.2 : Değişkenlerin Hedef Gruplar Bazında Dağılımı	55
Şekil 5.3 : Veri Kümesinin Dönemsel Dağılımı	59
Şekil 5.4 : 1. Model - 8 Değişken, 1 Saklı Katman, 5 Düğüm Noktası	69
Şekil 5.5 : 2. Model - 4 Değişken, 2 Saklı Katman, 5 Düğüm Noktası	71
Şekil 5.6 : 3. Model - 4 Değişken, 1 Saklı Katman, 5 Düğüm Noktası	72
Şekil 5.7 : 4. Model - 3 Değişken, 1 Saklı Katman, 4 Düğüm Noktası	73
Şekil 5.8 : 5. Model - 3 Değişken, 1 Saklı Katman, 3 Düğüm Noktası	74
Şekil 5.9 : 6. Model - 3 Değişken, 1 Saklı Katman, 2 Düğüm Noktası	75
Şekil 5.10: Borçlar/Aktifler – Aktif Karlılığı	79
Şekil 5.11: Borçlar/Aktifler – FAVÖK/Aktifler	80
Şekil 5.12: Breiman Ağacı Batıklık Öngörüsü	81
Şekil 5.13: Breiman Ağacı Batıklık Öngörüsü (Aşırı Büyük)	83
Şekil 5.14: Hothorn Ağacı Batıklık Öngörüsü	85
Şekil 5.15: En Başarılı Model	96



HALKA AÇIK FİNANS DIŐI ŐİRKETLERDE SÜREKLİLİK RİSKİNİN MAKİNE ÖĞRENMESİ İLE ÖNGÖRÜLMESİ

ÖZET

Bu çalışmanın, finans dışı Őirketlerin süreklilik risklerinin belirlenmesine, iki açıdan katkı yapması öngörülmektedir. Bunlar; Makine Öğrenmesi modellerinin süreklilik riski öngörüsüne yapabileceđi katkının ve bu modellerin pratikte uygulanabilirliklerinin irdelenmesidir.

Araştırmaya konu Makine Öğrenmesi modelleri sırasıyla Çoklu Ayraç Analizi, Lojistik Model, Sınır Ağları, Destek Vektör Makineleri, Karar Ağaçları ve Karar Ormanlarıdır. Çalışma kapsamında yapılan analizlere 1999-2016 döneminde Türkiye’de yerleşik, halka açık, finans sektörü dışındaki Őirketler dâhil edilmektedir.

Araştırmada R istatistik dilinden ve ilgili R Makine Öğrenmesi model kütüphanelerinden faydalanılmaktadır. Modellerin etkinlik ölçümleri örneklemin “model” ve “test” gruplarına ayrılması yöntemi (Çapraz Doğrulama) ile gerçekleştirilmektedir. İlk gruptaki veri kullanılarak modeller bina edilmekte, tamamen farklı örnekler içeren ikinci grup ile testler gerçekleştirilmektedir. Modelleme ve testler bu iki gurubun pek çok kere rastgele belirlenmesi ile tekrarlanmakta ve sonuçların istatistiki geçerliliđe erişmesi temin edilmektedir.

Bu çalışmanın sonucunda veriye uygulanan çok deđişkenli Makine Öğrenmesi modellerinin ortalama performansının %86 olduđu, en başarılı modelin %91, en zayıf modelin ise %81 düzeyinde bir başarı ile süreklilik risklerini öngörebildiđi tespit edilmektedir. Eldeki örneklem çerçevesinde, bir Breiman Karar Ağacı modelinin en başarılı süreklilik riski öngörüsü modeli olduđu ve bu modelin sade ve kolay anlaşılır yapısından ötürü pratikte de faydalı olabileceđi öngörülmektedir. Breiman Karar Ağacı modelinin her seviyedeki muhasebe bilgi kullanıcısı (yöneticiler, ortaklar, kredi kuruluşları, çalışanlar, potansiyel ortaklar, devlet ve toplum) tarafından kullanılması ve pek çok kredi riski takibi uygulamasına eklenmesi mümkün görünmektedir.

Anahtar Kelimeler: *Makine öğrenmesi modelleri, süreklilik riski öngörüsü, halka açık finans dışı Őirketler, Türkiye.*



GOING CONCERN PREDICTION VIA MACHINE LEARNING AMONG NON-FINANCE PUBLIC COMPANIES

ABSTRACT

This thesis aims to contribute to the going concern prediction studies through answering the questions of; can we use the “popular” Machine Learning tools in the process of going concern prediction effectively, and if so, which models are comparatively more practical to utilize in real world scenarios.

Machine Learning models used in this study are Multiple Discriminant Analysis, Logistic Models, Neural Networks, Support Vector Machines, Decision Trees and Random Forests. Data used to construct the models belong to non-finance Turkish companies traded publicly in the 1999-2016 period. R statistical language and related Machine Learning libraries are utilized to train the models. Performance of models are estimated by a commonly accepted method: cross-validation. In cross-validation, sample data is randomly divided into two independent groups. One of the groups is used to train the model and the other to test it. This process is repeated enough many times to reach statistical significance.

The study shows that models applied to the data attain an average prediction performance of 86%, where the best model in the study reaches a success level of 91% and the worst model remains at 81%. A Breiman Decision Tree appears to be the best model in predicting going concern risks. This renders the results to be far more interesting, since these models are the easiest to comprehend and build. It is foreseen that Breiman Decision Tree models can be utilized by any credit analyst at any level and can be incorporated to any credit analysis software without much effort.

Keywords: *Going concern prediction, machine learning models, non-finance listed companies, Turkey.*



1. GİRİŞ

İşletmelerin finansal süreklilik risklerinin öngörülebilmesi muhasebe bilgi kullanıcıları (yöneticiler, ortaklar, kredi kuruluşları, çalışanlar, potansiyel ortaklar, devlet ve toplum) açısından önemli bir ihtiyaç olarak ortaya çıkmaktadır. Bağımsız denetim raporları bu konuda faydalı olsa da her zaman ve her şirket için üretilmemektedirler. Herhangi bir şirket için bu raporun mevcudiyetinin finansal süreklilik risklerinin öngörülmesinde ne seviyede fayda sağlayabildiği ise uzun süredir araştırmalara konu olan ayrı bir meseleyi teşkil etmektedir. Bu ihtiyaçlara kısmi bir cevap vermeyi amaçlayan mevcut araştırmanın temel amacı halka açık finans dışı şirketlerin finansal süreklilik risklerinin çok değişkenli Makine Öğrenmesi modelleri ile objektif olarak öngörülmesidir. Araştırmaya konu Makine Öğrenmesi modellerinin performans, ön şartlar ve uygulanabilirlik açısından kıyaslanması ve ilgili modellerden benzer araştırmalarda ne seviyede faydalanılabileceğinin ortaya çıkarılması ise ikincil amacı oluşturmaktadır.

Bağımsız denetim raporlarının finansal süreklilik riski öngörü performansı kırk yıla yakın bir süredir pek çok araştırmaya konu olmaktadır. 1980'lerde finansal süreklilik riski üzerine yapılan ilk çalışmalarda kredi veren tarafların süreklilik risklerini genelde iyi tayin edebildiği ileri sürülse de, objektif bir model aracılığı ile bağımsız denetçilere bu konuda yardım edilebileceğini savunan makalelere de rastlanılmaktadır. Bu dönemde halka açık finansal verinin ağırlıklı olarak sadece performans tahmininde kullanıldığı ama bu verinin risk tayininde de kullanılması gerektiği ileri sürülmeye başlamakta, bu öneriler dönemin istatistiki yöntemleri kullanılarak oluşturulan modeller ile de desteklenmeye çalışılmaktadır.

1990'ların başında finansal süreklilik riski öngörüsü üzerine yapılan çalışmalarda bağımsız denetçi görüşlerinin başarılı sayılabileceği ve herhangi bir istatistiki yöntem ile bu kararların desteklenmesine gerek olmadığı savunulsa da takip eden çalışmalar bunun tam zıddını ileri sürmekte, bağımsız denetçilerin finansal süreklilik riski tayinlerinin oldukça zayıf olduğu ileri sürülmekte ve öncül çalışmaların ulaştığı sonuçlar olumsuz yönde güncellenmektedir. 1990'ların ortalarından itibaren ise

bağımsız denetçilerin sadece mesleki yargıya dayalı batıklık (işletmenin sürekliliği kabulü ortadan kalkmış) öngörüsü performansının %50'den daha düşük olduğu savı ağırlık kazanmaktadır. 1990'ların sonlarında bu başarısızlığın muhtemel nedenleri irdelenmekte ve bunların ağırlıklı olarak denetim sürecinin gelişmelere ayak uyduramaması yanında halka açık bilgiyi de tam olarak kullanamaması olduğu ileri sürülmektedir.

2000'li yılların başlarında yapılan çalışmalarda, denetim raporları ile süreklilik riski arasında doğrusal bir ilişkiden bahsedilebilse dahi, özellikle küçük bağımsız denetim firmalarının gri alandaki finansal süreklilik risklerini ekseriyetle göz ardı ettiği ileri sürülmektedir. Takip eden çalışmalar ise daha ziyade finansal süreklilik riski tayininde mesleki yargı yanında hangi istatistiki modellerin daha başarılı sonuçlar verebileceği üzerinedir. Mevcut çalışma da buna bir örnektir.

Süreklilik risklerinin denetçiler tarafından öngörülebilmesi üzerine yapılmış olan tüm bu çalışmalardan çıkarılabilecek en önemli sonuç, muhasebe bilgi kullanıcıları açısından Makine Öğrenmesi ve benzeri istatistiki modellerin, geliştirilmeleri halinde, faydalı olabilecekleridir. Bu modellerin parametrik olması veya olmaması (Makine Öğrenmesi modelleri) da önem arz etmemektedir, tamamı yerinde kullanıldığında faydalı sonuçlar üretebilmektedir. Bu yönde yapılan çalışmalarda, finansal süreklilik riski öngörüsü üzerine oluşturulmuş modeller ve sonuçlar derlendiğinde, genel olarak faydalı ve pratikte kullanılabilir oldukları ileri sürülmektedir.

Yakın dönemde bilişim teknolojisinde yaşanan gelişmeler süreklilik riski öngörüsünde Sinir Ağları, Karar Ağaçları, Karar Ormanları, Destek Vektör Makineleri ve benzeri pek çok işlem-yoğun, çok değişkenli ve doğrusal olmayan modelin denenmesini mümkün kılmaktadır. Hesaplanması kolay olduğu için günümüze kadar uygulama fırsatı bulmuş kimi istatistiki modeller, muhtemel doğrusal ilişkilerin ortaya çıkarılmasında başarılı olsalar da, doğrusal olmayan ilişkilerin keşfinde (teorik altyapıları gereği) yetersiz kalabilmekte, bu da çok değişkenli doğrusal olmayan Makine Öğrenmesi modellerinin denenmesini ve mümkünse pratikte de kullanılmasını zorunlu kılmaktadır.

Teknik gelişmelere paralel olarak süreklilik riski öngörüsü modellerinin de güncellenmesi gerekliliği açıktır. Son on yıl içinde öne çıkan bazı istatistiki modeller

(Karar Ağaçları örneğinde olduğu gibi) bu konuda oldukça ümit vericidir ve doğrusal olmayan ilişkilerin kolaylıkla modellenmesi yanında, verinin istatistiksel dağılımı üzerindeki geleneksel kısıtlamaları da hafifletilebilmektedir. Bu modellerin ekseriyeti Makine Öğrenmesi (Machine Learning) sınıfına girmekte ve eldeki veriyi kullanarak hem modeli hem de modelin parametrelerini kendi içinde belirleyebilmektedir.

Mevcut çalışma, yukarıdaki tarihi perspektife uygun olarak, farklı özelliklere sahip sekiz Makine Öğrenmesi modelinden finansal süreklilik risklerinin tespitinde faydalanmayı amaçlamakta ve bu modellerin kıyaslamalı etkinliklerini irdelemektedir.

Bu tez giriş ve sonuç bölümleri hariç, beş bölümden oluşmaktadır. Giriş bölümünü takip eden ikinci bölümde çalışmada başvuru alan temel kavramlara detaylı açıklamalar getirilmektedir. Üçüncü bölümde yer alan literatür incelemesi kendi içinde iki ana bölüme ayrılmıştır ve ilk bölümde Makine Öğrenmesi modellerinin ortaya çıkışı ve tarihi gelişimi özetlenirken ikinci bölümde bu modellerin pratik uygulamaları esnasında karşılaşılan sorunlar üzerine yapılmış çalışmalar derlenmektedir. Tezin dördüncü bölümünde mevcut araştırmada başvuru alan çok değişkenli Makine Öğrenmesi modellerinin teorik altyapıları, temel kullanım şekilleri ve dikkat edilmesi gereken yönleri ele alınmaktadır. Bu bölümde Çoklu Araç Analizi, Lojistik, Sinir Ağları, Breiman Ağacı, Hothorn Ağacı, Breiman Ormanı ve Hothorn Ormanı modelleri olmak üzere sekiz farklı modele değinilmektedir. Tezin beşinci bölümünde araştırmanın amacı, önemi, yöntemi, kısıtları, ana kümesi, seçilen örnekleme, bulguları ve sonuçları aktarılmaktadır. Bu bölümde, 1999-2016 döneminde Türkiye’de yerleşik, halka açık, finans sektörü dışındaki şirketlerden derlenen 106 örneğe bir önceki bölümde listelenen çok değişkenli Makine Öğrenmesi modelleri uygulanmaktadır. Uygulama R istatistik dili ve ilgili model kütüphaneleri aracılığı ile gerçekleştirilmektedir. Model değişkenleri örneklere ait işletme performansını, kaldıraç etkilerini ve likiditeyi gösteren finansal oranlardan oluşmaktadır. Ortaya çıkan modellerin pratikte gösterebilecekleri performans Çapraz Doğrulama yöntemi ile test edilmekte, modeller arası kıyaslamalar da bu sonuçlara dayanarak gerçekleştirilmektedir. Bölümün sonunda eldeki modeller sadece performans çerçevesinde değil pratikte uygulanabilirlik açısından da birbirleri ile karşılaştırılmaktadır. Sonuç olarak ise sade bir Breiman

Karar Ağacı modelinin eldeki veri çerçevesinde en başarılı model olduğu savunulmakta ve bu modelin pratikte de kolayca uygulanabileceği öngörülmektedir.



2. KAVRAMSAL ÇERÇEVE

Bu bölümde çalışmada başvurulan temel kavramlar açıklanmaktadır.

2.1 Bağımsız Denetim

Finansal tablo ve diğer finansal bilgilerin, finansal raporlama standartlarına uygunluğu ve doğruluğu hususunda, makul güvence sağlayacak yeterli ve uygun bağımsız denetim kanıtlarının elde edilmesi amacıyla, denetim standartlarında öngörülen gerekli bağımsız denetim tekniklerinin uygulanarak defter, kayıt ve belgeler üzerinden denetlenmesi ve değerlendirilerek rapora bağlanmasını ifade eder (BDY, 2015:1).

2.2 Bağımsız Denetçi

Bağımsız denetim yapmak üzere, 1/6/1989 tarihli ve 3568 sayılı Serbest Muhasebeci Mali Müşavirlik ve Yeminli Mali Müşavirlik Kanununa göre yeminli mali müşavirlik ya da serbest muhasebeci mali müşavirlik ruhsatını almış meslek mensupları arasından Kurum tarafından yetkilendirilen kişileri ifade eder (BDY, 2015:1).

2.3 Denetim Raporu

Denetim raporu, denetim kanıtlarının TDS çerçevesinde değerlendirilmesi sonucunda, belirlenen güvence seviyesine uygun şekilde oluşturulan denetçi görüşünün ve varsa dikkat çekilmek istenen diğer hususların kullanıcıların istifadesine sunulması amacıyla Kurum düzenlemelerine uygun olarak hazırlanan ve imzalayan denetim kuruluğu veya denetçi tarafından sorumluluğu üstlenilen belgedir. (BDY, 2015:5).

2.4 Halka Açık Finans Dışı Şirketler

Halka açık şirket, hisselerinin tamamı veya bir kısmı halka arz edilmiş şirket türüdür. Bu hisse senetlerinin BİST'de işlem görebileceği iki piyasa bulunmaktadır, bunlar:

Hisse Senetleri Piyasası ve Gelişen İşletmeler Piyasası'dır (URL-1, 2018).

Halka açık finans dışı şirketler ise, finansal piyasalarda aracılık işlemleri gerçekleştirilmeyen (banka, sigorta şirketi, tüketici finansmanı şirketi, finansal kiralama şirketi, faktöring şirketi ve benzeri dışında kalan), sermayesi paylara bölünmüş ve sermaye payları SPK Pay Tebliği (VII-128.1) çerçevesinde halka arz edilmiş, hisse senetleri Borsa İstanbul A.Ş.'de işlem gören ve Türk Ticaret Kanunu'nun Madde 329 ile tanımlanan anonim şirketleri ifade etmektedir.

2.5 İşletmenin Sürekliliği

Bu kavram, işletmelerin faaliyetlerini bir süreye bağlı olmaksızın sürdüreceğini ifade eder. Bu nedenle işletme sahiplerinin ya da hissedarlarının yaşam süreleriyle bağlı değildir. İşletmenin sürekliliği kavramı maliyet esasının temelini oluşturur. Bu kavramın, işletmeler açısından geçerliliğinin bulunmadığı veya ortadan kalktığı durumlarda ise, bu husus mali tabloların dipnotlarında açıklanır (URL-2, 2018).

Yönetim, finansal tabloları düzenlerken işletmenin faaliyetlerini süreklilik içerisinde devam etme yetisini değerlendirir. Yönetimin şirketi tasfiye etme veya ticari faaliyetini sona erdirmeye niyeti veya mecburiyeti yoksa finansal tablolar süreklilik kavramına göre düzenlenir. Yönetim, değerlendirmelerini yaparken işletme faaliyetlerinin sürekliliğine kuşku düşürecek önemli belirsizlikler taşıyan olaylar veya koşulların farkındaysa, bu belirsizlikleri açıklar. İşletme, finansal 7 tablolarını süreklilik esasına göre hazırlamaması halinde, bu hususu, finansal tablolarını düzenlemesi sırasında dayandırdığı temel ve işletmenin süreklilik arz etmeyecek şekilde değerlendirilmesinin nedeni ile birlikte açıklar (TMS 1, 2012:6-7).

2.6 Süreklilik Riski

Bağımsız denetçiler, topladıkları kanıtlar çerçevesinde, işletmenin sürekliliği üzerinde ciddi bir riskin oluşup oluşmadığını araştırırlar. İşletmede "İşletmenin Sürekliliği" kavramının ciddi seviyede tehlikeye girmiş olduğuna kanaat getirilirse, bu durum bağımsız denetim raporlarına yansıtılır ve okuyucunun dikkatine sunulur. İşletmenin sürekliliğini tehlikeye sokabilecek başlıca kaynaklar; tekrarlanan faaliyet zararları, çalışma sermayesi açıkları, vadesi gelen borçların ödenmesinde yaşanan güçlükler, önemli müşterilerin kaybı, sigortalanmamış doğal afetlerin yaşanmış olması, önemli yönetici veya kilit personel kaybı ve işletmenin faaliyetini tehlikeye

sokabilecek hukuki dava veya yaptırımlardır (Güredin, 2014:75-76).

2.7 Finansal Risk

Finansal risk işletmenin faaliyetlerini sürdürebilmesi için ihtiyaç duyduğu kaynakları sağlamada yaşayabileceği belirsizliklerdir. Bu belirsizlikler asıl olarak finansal yapının bir neticesidir. Finansal riskler; piyasa, likidite, kredi veya faaliyet kaynaklı olabilmektedir. Piyasa riski; döviz kuru, faiz oranı, menkul kıymetler ve mal fiyatlarındaki oynaklıklar, likidite riski (fonlama riski); ihtiyaç duyulan fonların zamanında ve istenen maliyetle elde edilememesi, kredi riski; şirketin finansal işlem gerçekleştirdiği karşı tarafın yükümlülüklerini yerine getirememesi ve faaliyet riski ise işlem süreçleri veya yönetimdeki yetersizlikler olarak tanımlanabilmektedir (Bolak, 2004:9).

Finansal riskleri kontrol edilebilirlik açısından ikiye ayırmak mümkündür; sistematik riskler (işletmenin kontrolü dışında) ve sistematik olmayan riskler (işletmenin kontrolünde olan). Finansal riskler ağırlıklı olarak işletmenin finansal yapısından kaynaklanmakta ve sistematik olmayan risklerden oluşmaktadırlar. Bir işletmenin kendini fonlama maliyetinin taşıdığı finansal riskler ile doğru orantılı olduğu kabul edilmektedir. İşletmenin finansal riskleri arttıkça fonlama için katlanacağı risk primi (risksiz bir işletmeye kıyasla katlanılan fazladan maliyet) de yükselmektedir. Bu çalışma çerçevesinde sistematik olmayan finansal riskler öngörülmeye çalışılmaktadır.

2.8 Finansal Süreklilik Riski

İşletmenin finansal faaliyetlerinden kaynaklı olarak ortaya çıkabilecek süreklilik risklerini ifade etmektedir. İşletmenin genel işleyişi çerçevesinde faaliyetlerini idame ettirebilecek yeterli fonlamayı yaratamaması, yüksek seviyede borçlanması ve borçlarını ödemede yaşayabileceği güçlükler (finansal riskler) bu tip risklere örneklerdir.

2.9 Makine Öğrenmesi

Makine Öğrenmesi bilgisayarlara, özel bir duruma spesifik bir program aracılığı olmadan, öğrenme yetisi kazandırmayı amaçlayan bilim dalıdır ve ağırlıklı olarak istatistik ve bilişim teknolojilerinden faydalanmaktadır (Samuel, 2000:206).

2.9.1 Makine Öğrenmesi kavramları

2.9.1.1 Değişken Önemi

Değişken Önemi, bir modele ait değişkenlerin eldeki sınıflandırma problemi çerçevesinde ne kadar belirleyici olduklarının göstergesidir. Hothorn ve diğ., (2006:652) eğer bahse konu değişkenin (örneklerin bir özelliği) önemi düşük ise (ilgili sınıflandırma problemine katkısı diğer değişkenlere kıyasla az ise) bu değişkenin modelden güvenle çıkarılabileceğini ve model performansının da bundan görece az etkileneceğini ileri sürmektedir. Herhangi bir modeldeki değişken sayısının uygun şekilde azaltılmasının Aşırı Uygunluk problemine karşı çok etkin bir yöntem olduğu savunulmaktadır (Platt ve Platt, 1990:31-32).

2.9.1.2 Şartlı Önemlilik

Değişken Önemi belirlenirken aradaki korelasyonlar da hesaba katıldığında Şartlı Önemlilik değerlerine ulaşılmaktadır. Şartlı Önemlilik hesaplamaları Değişken Önemi hesaplamalarına kıyasla daha işlem yoğun olsa da önemli kazanımlar sağlayabilmektedir. Bu yöntem ile yüksek korelasyona sahip değişkenler arasından en önemlisinin tayini mümkün olmakta ve ilişkili diğer değişkenler, sonuca tek başlarına etkileri yüksek görünse de, modelden güvenle çıkarılabilmektedir (Strobl ve diğ., 2009:1).

2.9.1.3 Karışıklık Matrisi

Karışıklık Matrisleri modellerin sınıflandırma performanslarını belirlemekte kullanılmaktadır. Model öngörüsü ile gerçek durumun örtüşme düzeyi model başarısının göstergesi olarak kabul edilmektedir.

Çizelge 2.1'de Karışıklık Matrisleri'ne bir örnek verilmektedir. Bu örnekte bir modelin eldeki veriyi A ve B sınıflarına ayırmaya çalıştığı varsayılmaktadır (Çizelge 2.1'deki satırlar). Model öngörüsünün ne derece başarılı olduğu, her bir sınıflandırma tipi için ayrı ayrı olmak üzere, gerçek durumla (Çizelge 2.1'deki sütunlar) kıyaslanmaktadır. Karışıklık Matrisi'nde öngörü ile gerçek durumun örtüştüğü

hücrelerin (Çizelge 2.1'deki diyagonal hücreler) toplamının (%49 + %48 = %97) model başarısı için bir gösterge olduğu kabul edilmektedir (Pearson, 1904:20-21).

Çizelge 2.1: Örnek Karışıklık Matrisi

		Gerçek	
		A	B
Öngörü	A	%49	%2
	B	%1	%48

2.9.1.4 Model İçi Performans (In-Sample Fit)

Model İçi Performans, bir modelin öngörü performansının yine model eğitimi için kullanılan veri ile test edildiği durumda ulaşılan başarı seviyesidir. Modelin pratikte nasıl bir başarı göstereceğinin, Çapraz Doğrulama performansı kadar, iyi bir indikatörü olarak kabul edilmemektedir (Platt ve Platt, 1990:31-32).

2.9.1.5 Çapraz Doğrulama (Cross Validation)

Çapraz Doğrulama model performansının bağımsız veri (modelleme esnasında kullanılmamış) ile gerçekleştirilen testler neticesinde belirlenmesidir. Çapraz Doğrulama performansı, modelin pratikte ulaşabileceği başarının, Model İçi Performans'a kıyasla, daha sağlıklı bir göstergesi olarak kabul edilmektedir (Jones, 1987:131).

2.9.1.6 Aşırı Uygunluk (Overfitting)

Aşırı Uygunluk, Model İçi Performans'ın Çapraz Doğrulama performansından ciddi oranda yüksek tespit edildiği durumdur. Model, eğitiminde kullanılan veriye, "aşırı uygun" hale gelmekte, bağımsız örneklerde başarısız sonuçlar üretmektedir (Platt ve Platt, 1990:31-32).

2.9.1.7 Logit fonksiyonu

Logit fonksiyonu (Berkson, 1994:357), Lojistik Model'lerde tüm bir değişken ekseninin (0,1) aralığına yansıtılmasını sağlamaktadır: $F(x) = \log(x / (1-x))$.

2.9.1.8 Probit fonksiyonu

Probit fonksiyonu (Bliss, 1934:38), kümülatif normal dağılımın tersidir ve "sürekli" bir fonksiyondur. Bu fonksiyon, Lojistik Model'lerde Logit fonksiyonunun alternatifidir.

2.9.1.9 Özyinelemeli Bölümlendirme (Recursive Partitioning)

Özyinelemeli Bölümlendirme, Makine Öğrenmesi'ne dahil ana model sınıflarından biridir ve örnek ana kümesini daha “saf” (aynı sınıfa ait örneklerden oluşan) alt gruplara ayırma amacını gütmektedir. Model, benzer bölümlendirme kurallarının elde edilen alt gruplara ardı ardına uygulanması ve önceden belirlenmiş kriterlere ulaşana kadar bunun tekrarlanması neticesinde oluşturulur (Therneau, 1983). Karar Ağaçları Özyinelemeli Bölümlendirme modellerine en iyi örneklerdendir.

2.9.2 Makine Öğrenmesi modelleri

Makine Öğrenmesi modelleri, Makine Öğrenmesi amacına yönelik olarak ortaya çıkarılmış modellerdir. Bu tezde kullanılan Makine Öğrenmesi modelleri şunlardır:

- Çoklu Ayraç Analizi
- Lojistik Model
- Sinir Ağları
- Destek Vektör Makineleri
- Karar Ağaçları
- Karar Ormanları

2.9.2.1 Sınıflandırma Aracı

Bu tez çerçevesinde, eldeki verinin sınıflara ayrılmasını amaçlayan Makine Öğrenmesi modellerine verilen genel addır.

2.9.2.2 Çoklu Ayraç Analizi

CAA bir sınıflandırma aracıdır ve ayırım için örneklerin sahip olduğu tüm özelliklerin tek tek dikkate alınması yerine, bunların doğrusal bir kombinasyonunun istenen sınıflandırmayı başarılı bir şekilde gerçekleştirmesini amaçlamaktadır (Altman, 1968).

2.9.2.3 Lojistik Model

Lojistik Model (Lojistik Regresyon Modeli) bir sınıflandırma aracıdır ve ismi ürettiği değerlerin (0, 1) aralığına kısıtlanmış olmasından gelmektedir. Modelin CAA'nden en önemli farkı, iki grup arasında sınıflandırma yaparken, yapılan sınıflandırmanın güvenilirliği hakkında da bir değer üretebilmesidir (Cox, 1958). Model, Logit bazlı ise “Logit Model”, Probit bazlı ise “Probit Model” olarak da adlandırılmaktadır.

2.9.2.4 Sinir Ağları

Sinir Ağları bir sınıflandırma aracıdır. Model, veri ve sonuç arasında birçok Lojistik Model'in paralel ve çapraz bağlanması sonucu ortaya çıkmaktadır (Widrow ve Hoff, 1960).

2.9.2.5 DVM

Destek Vektör Makineleri bir sınıflandırma aracıdır. DVM sınıflandırma yaparken elde edilen verinin özellikleri yanında bizzat örneklerin kendilerini de (özellik uzayındaki lokasyonlarını) dikkate alabilmekte ve bu sayede az sayıda örneğin referans alınması ile sınıflandırma öngörüsünün başarı ile yapılabilmesini mümkün kılmaktadır. Model, kural bazlı modelleme (örneklerin özelliklerine dayalı) yanında örnek bazlı modelleme (örneklerin özellik uzayındaki lokasyonuna dayalı) amacıyla da kullanılabilir (Cortes ve Vapnik, 1995).

2.9.2.6 Karar Ağacı

Karar Ağacı bir Özyinelemeli Bölümlendirme modelidir ve genellikle sınıflandırma aracı olarak kullanılsa da regresyon amacı ile de kullanılabilir. Karar Ağacı modellerinde veri yinelemeli olarak doğrusal ayrılar ile alt gruplara ayrılırken, elde edilen her bir alt gruptaki örneklemin mümkün olduğunca tek bir sınıfa ait olması (saflaşması) amaçlanmaktadır. (Breiman ve diğ., 1984).

2.9.2.7 Karar Ormanı

Karar Ormanları çok sayıda Karar Ağacı'nın bir araya gelmesiyle oluşmakta ve sınıflandırma aracı olarak kullanılmaktadır. Karar Ormanları'nın barındırdığı her bir Karar Ağacı elde edilen örneklemin ve değişkenlerin sadece bir kısmının kullanılması ile oluşturulmakta ve bu sayede elde edilen probleme farklı açılardan bakılabilmesi amaçlanmaktadır. Karar Ormanı tarafından gerçekleştirilen sınıflandırmalar onu oluşturan tüm Karar Ağaçları'nın sınıflandırma öngörülerinin bir araya getirilmesi ile belirlenmektedir (Breiman, 2001).



3. LİTERATÜRÜN İNCELENMESİ

Literatür incelemesi iki bölümden oluşmaktadır. İlk bölümde çalışmada faydalanılan Makine Öğrenmesi modellerinin ortaya çıkışına ve tarihi gelişimine odaklanılmaktadır. Literatür incelemesinin ikinci bölümünde ise bu modellerinin pratikte uygulanması esnasında karşılaşılan sorunlar üzerine yapılmış olan çalışmalara yer verilmektedir.

3.1 Makine Öğrenmesi Modelleri

Makine Öğrenmesi modelleri, tarihsel gelişime paralel olarak, üç bölümde ele alınmaktadır. İlk bölümde tek değişkenli doğrusal modellerin, ikinci bölümde bu modelleri bir adım öteye taşıyan çok değişkenli doğrusal modellerin ve son bölümde ise çok değişkenli doğrusal olmayan modellerin tarihi gelişimi incelenmektedir.

3.1.1 Tek değişkenli doğrusal modeller

Süreklilik riski öngörüsü üzerine yapılan çalışmaların geçmişi 1930'lara kadar (Bureau of Business Research, BBR, 1930:7) takip edilebilmektedir. İlk araştırma tek bir finansal oranın süreklilik riski öngörüsü üzerinde ne kadar etkili olabileceği üzerinedir. Bu çalışmada batık ve sağlıklı (süreklilik riski olmayan) şirketlerin finansal oranları karşılaştırılmakta ve muhtemel farklılıkların önemli bilgiler sağlayabileceği ileri sürülmektedir. BBR tarafından yapılan bu öncü çalışmada 29 endüstriyel firma ve pek çok finansal oran incelenmekte, sağlıklı ve batık şirketler için bu finansal oranların hangi ortalamalarda gerçekleştiği belirlenmekte, farklılıklar tespit edilmektedir. Çalışmanın sonucunda, iki finansal oranın diğer finansal oranlara nispetle daha başarılı göstergeler olduğu ve bunların arasından Çalışma Sermayesi/Toplam Aktifler oranının Cari Oran'dan daha etkili bir süreklilik riski göstergesi olduğu savunulmaktadır.

BBR nezdinde gerçekleştirilene benzer bir çalışma 1932'de Fitzpatrick tarafından tekrarlanmakta ve bu çalışmada 1920 ile 1929 arasında batmış olan 20 şirketin finansal oranları, aynı döneme ait 20 sağlıklı şirketin (benzer sektörlere ait) finansal oranları ile kıyaslanmaktadır. Yapılan çalışmada 13 finansal oran kullanılmakta ve

araştırmanın sonucunda süreklilik riski yaşayan şirketlerin finansal oranlarının, sağlıklı şirketlerin finansal oranlarına kıyasla, negatif ayrıştığı savunulmaktadır.

Bazı finansal oranların süreklilik riskini oldukça başarılı bir şekilde işaret edebilmesi benzer çalışmaları tetiklemiş görünmekte ve böyle bir çalışma da 1942'de Merwin'den gelmektedir. Merwin (1942) çalışmasında 1926-1936 yılları arasındaki on yıllık döneme ve bu dönemdeki küçük şirket iflaslarına odaklanmaktadır. Merwin, bundan önce gerçekleştirilen finansal orana dayalı süreklilik riski analizlerinin sadece büyük şirketler için geçerli olmadığını, küçük şirketlerin de benzer bir analize tabi tutulabileceğini ve başarılı süreklilik riski öngörülerini yapılabileceğini savunmaktadır (Merwin, 1942:1-6).

Süreklilik riski yaşayan şirket adedinin sağlıklılara oranla çok daha düşük olduğu kabul edildiğinde, sağlıklı şirketlerin finansal oran ortalamalarının hesaplanması ve süreklilik risklerinin bu ortalamalardan sapmalarından belirlenmesi akla yatkın bir süreç olarak ortaya çıkmaktadır. Bu mantık paralelinde, sağlıklı işletmeler için ortalama finansal oranların tespitine yönelik öncül çalışma Chudson (1945) tarafından gerçekleştirilmektedir. Çalışmanın sonucunda, beklentilerin aksine, genel geçer sağlıklı finansal oranların mevcut olmadığı ileri sürülmektedir (Chudson, 1945:1-16). Chudson analizinde, sadece sektör bazında birtakım aralıkların mevcudiyetinden bahsedilebileceğini (pek çok finansal oran için "sağlıklı" ortalamalar sektörden sektöre ciddi farklılıklar göstermektedir) savunmaktadır.

1930'lardan 1965'lere kadar yapılan çalışmaların genel amacı süreklilik riski öngörüsü için tek bir finansal oran aracılığı ile hesabı basit ve anlaşılması kolay bir yöntem belirlemektir (Belowary ve diğ., 2007:2). Tüm bu öncül çalışmalar ile bir sonraki adıma, birden çok finansal oranın aynı anda kullanılmasına, doğal bir hazırlık yapılmaktadır. Beaver'ın makalesi (1966:99-102) süreklilik riski öngörüsünde bir sonraki adıma geçiş için son basamağı hazırlamakta ve bu aşamadan sonra yapılması gerekeni de açıkça işaret etmektedir. Batık ve sağlıklı 79 çift şirketin finansal oranlarının karşılaştırıldığı çalışmada Lineer Ayraç Analizi (ÇAA'nin tek ayraçla oluşturulmuş hali) kullanılmaktadır. Bu çalışmada yanlış sınıflandırma oranını düşürmek için "Kesme Noktaları" (Cut-off Points) belirlenmekte, yani bazı "gri alanda kalmış" örnekler model öngörüsü dışında bırakılmaktadır. İlgili finansal oranlara uygulanan Kesme Noktaları az sayıda örnek için hiçbir sınıflandırılma yapılmaması sonucunu doğurmakla birlikte toplamda yanlış öngörü adedini

düşürmektedir. Gri alanda kalan örnekler tahmin dışı bırakılmakta, öngörü modeli bu örnekler hakkında yeteri kadar “kesin” bir kanaat üretememektedir. Daha sonra ÇAA modellerinde de kullanılacak olan Kesme Noktaları dışlanan örnek oranıyla tahmin gücü arasındaki dengeyi belirleyen optimizasyon parametreleri olarak ortaya çıkmaktadır. Aynı makalenin sonunda Beaver (1966:100-102), finansal oranların tek tek kullanımı yerine birlikte kullanılmalarının daha iyi sonuçlar vermesinin mümkün olabileceğini ileri sürmekte ve gelecek çalışmalara yön vermektedir. Bu döneme ait geç bir çalışma ise Wilcox’tan (1973:163) gelmekte ve temelde Beaver’ın (1966) analizlerini derinleştirmektedir.

3.1.2 Çok değişkenli doğrusal modeller

Pek çok finansal oranın süreklilik riski öngörüsüne, farklı seviyelerde de olsa, olumlu yönde katkı sağlaması mümkün görünmektedir. Bu görüşten yola çıkarak süreklilik riskini birden çok değişkenin birlikte kullanılması ile öngörmeyi amaçlayan Altman’ın çalışması (1968:568) ilk yaygın çok değişkenli doğrusal model olması sebebiyle bir dönemi açıp diğerini kapatmaktadır. Çalışma üretim firmalarına odaklanmaktadır. Çoklu Ayraç Analizi’ni kullanan Altman beş finansal faktörü farklı ağırlıklarla konsolide ederek bir “Z-score” üretmekte ve bu skor için de uygun Kesme Noktaları belirlemektedir. Altman makalesinde bu yöntemin süreklilik riskini bir yıl öncesinden %95 başarıyla (Model İçi Performans) öngördüğünü savunmakta, ancak iflas tarihinden uzaklaştıkça tahmin gücünün zayıfladığını da not etmektedir. Ortaya çıkan modelin süreklilik riski öngörü başarısı iki yıl içinde %72’ye, üç yıl içinde ise %48’e kadar gerilemektedir. Yöntemin bağımsız örnekler üzerindeki performansı (Çapraz Doğrulama performansı) ise %79 seviyesinde gerçekleşmektedir.

Altman’ın model olarak kullandığı Çoklu Ayraç Analizi şirketleri sektöre has karakteristik finansal oranları kullanarak “batık” ve “sağlıklı” (işletmenin sürekliliği kabulü geçerli) olmak üzere iki sınıfa ayırmayı hedeflemektedir. Modeli oluşturan örnekler üzerinden yapılan analizde öncelikle hangi finansal oranların kullanılacağına ve her birinin nasıl ağırlıklandırılacağına karar verilmektedir. Analizin devamında ise tespit edilen bu ağırlıkların ilgili finansal oranlarla çarpımı ve sonuçların bir araya getirilmesi ayraç skorunu üretmektedir. Skor, finansal riskin bir göstergesi olarak kabul edilmekte ve eğer belirli bir seviyenin (Kesme Noktası) üzerinde ise süreklilik riskinin yüksek ihtimal ile ortaya çıkmasını öngörülmektedir.

Kesme Noktası belirlenirken yanlış sınıflandırmanın düşük tutulması yanında çok fazla örneğin model dışında bırakılmamasına da gayret edilmektedir. Özet olarak; farklı finansal oranlar değişen ağırlıklarda bir araya getirilmekte, işletmeye farklı açılardan bakılabilmesi ve sürdürülebilirlik üzerine genel bir fikrin üretilebilmesi amaçlanmaktadır. Tahmin gücü yüksek finansal oranlar model içinde görece daha fazla ağırlığa sahip olmakta, nihai model performansı eldeki değişkenlerin öngörü performanslarının ağırlıklandırılmış doğrusal bir bileşkesi olarak ortaya çıkmaktadır.

Altman'la birlikte süreklilik riski öngörüsü üzerine yapılan çalışma sayısının hızla arttığı gözlenmektedir (Belowary, 2007:4). Sayının artışıyla, eldeki modelin süreklilik riski öngörüsünde başarılı sonuçlar üretmesi, farklı sektörler, ülkelere ve şirket büyüklüklerine uygulanabilir olması ve zaman içinde bu modelin yanına başka modellerin de eklenmesi etkili olmuş görünmektedir. Meyer ve Pifer (1970: 853) ÇAA modelini bankacılık sektörüne uygularken, Deakin (1972:178) modeli üretim yapmayan firmalara genellemekte ve her iki çalışma da olumlu sonuçlar alındığını ileri sürmektedir. Edmister (1972:1477-1478) ise modelin küçük şirketlerde, daha önceki araştırma sonuçları kadar, başarılı olmadığını savunmaktadır. Edmister bu durumu tek dönemlik finansalların yeteri kadar bilgi içermediği şeklinde yorumlamakta ve modelin birkaç dönemlik finansal veri ile beslediğinde daha başarılı performans gösterdiğini ileri sürmektedir. Altman ÇAA modelini demiryolu şirketlerine uygulamakta (1973:208-209) ve tekrar olumlu sonuçlar elde edildiğini savunmaktadır. Bu sırada model farklı araştırmacılar tarafından sürekli olarak test edilmektedir. Blum (1974:1) yaptığı çalışmada modelin süreklilik riski öngörüsünde oldukça başarılı olduğunu ileri sürmektedir. Sinkey (1975:21) ise süreklilik riskli taşıyan bankaların karakteristik özelliklerinin ortaya çıkarılmasında ÇAA modelinden faydalanılabileceğini savunmaktadır. Bu çalışmada artık ÇAA'nin başarısı sorgulanmamakta, model aracılığı ile en iyi finansal göstergelerin tespiti yapılmaya çalışılmaktadır. Pinches ve diğ. (1975:295) de aynı yolu takip ederek finansal oranların önemlilik sıralamasını ÇAA'ni kullanarak ortaya çıkarmayı hedeflemektedir. ÇAA modellerinde yeteri kadar önemli olmayan değişkenlerin belirlenmesi ve modelden çıkarılması Moyer (1977:11-12) tarafından ele alınmakta, bu makalede Altman'ın orijinal çalışması (1968) irdelenmekte ve gereğinden fazla değişken kullanıldığı için Aşırı Uygunluk sorunu (%95 Model İçi Performans'a karşın %79 Çapraz Doğrulama performansı) yaşandığını ileri sürmektedir. Modelin

çok dar alanlara uygulandığı örneklere de rastlamak mümkündür. Altman ve Loris'in (1976:1201-1202) masa-üstü aracı kurumlara yönelik çalışması buna bir örnektir ve çalışmanın neticesinde yine olumlu sonuçlar alındığı savunulmaktadır.

Altman ve diğ.'nin 1977'de (30-31) ele aldıkları makalede orijinal ÇAA modelini istatistiki açıdan daha kuvvetli temellere oturtmak için çaba harcadığı görülmektedir. Yeni modelin ortaya çıkarılmasındaki amaç orijinal modeldeki değişkenlerin teorik dağılımlarını normal dağılıma, şartlı oynaklıklarını da birbirlerine yakınlaştırmaktır. Model artık "doğrusal" değil, İkinci Derece Ayraç Analizi'dir (İDAA). Modelde yedi faktör kullanılmakta ve "Zeta Analizi" olarak adlandırılmaktadır. Yapılan çalışmada üretim ve perakende sektöründen şirketlere odaklanılmaktadır. Örnekleme her batık şirket için, benzer sektöre ve döneme ait, sağlıklı bir şirket de dahil edilmektedir. Bu modelin orijinal modele göre daha kuvvetli bir istatistiki temele sahip olduğu kabul edilmesine rağmen bağımsız test kümeleri üzerindeki başarısı, orijinal modele kıyasla, kayda değer bir ilerlemeye işaret etmemektedir.

Santomero ve Vinso (1977:185) orijinal modeli ticari bankalara uygularken, sınıflandırmanın ötesinde, Kesme Noktasına uzaklığı da dikkate alarak süreklilik riskini belirlemeye çalışmakta ancak model, yapısı gereği, böyle bir bilgiyi üretmeye çok uygun görülmemektedir. Araştırmacılar doğru yönde çaba sarfetmiş olmakla birlikte Kesme Noktasına uzaklığın yapılan öngörüye etkisini belirlemek ancak Lojistik Modeller ile mümkün olacaktır. Bu çalışma da, Beaver'ın (1966) yaptığı gibi, bir sonraki adıma işaret etmektedir.

ÇAA modelinin farklı ülke odaklı uygulamalarının pek çok araştırmaya konu olduğu görülmektedir. Altman ve Levallee (1980:147) yaptıkları çalışmada Çoklu Ayraç Analizi'ni Kanada'ya da uygulamakta ve pozitif sonuçların Birleşik Devletler'e has olmadığını savunmaktadır. Benzer çalışmalar Castagna ve Matolcsy (1981:23) tarafından Avusturalya için, Taffler (1982:342) tarafından İngiltere için, Sands ve diğ. (1983:24) tarafından tekrar Kanada için, Izan (1984:303) tarafından tekrar Avustralya için, Takahashi ve diğ. (1984:229) tarafından Japonya için, Taffler (1984:199) tarafından tekrar İngiltere için ve Skogsvik (1990:137) tarafından İsveç için tekrarlanmakta ve olumlu sonuçlar alındığı ileri sürülmektedir.

Model farklı sektörler için de defalarca uygulama alanı bulmuştur. Sharma ve Mahajan (1980:80) modeli perakendecilik sektörüne, Fulmer ve diğ. (1984:25) ticari bankalara, Rose ve Koları (1985:43) genel bankacılık sektörüne, Scaggs ve Crawford (1986:11) hava yollarına ve Wertheim ve Lynn (1993:529) ise hastanelere uyguladıklarını ve başarılı sonuçlar aldıklarını savunmaktadırlar.

Çoklu Ayraç Analizi'nin sıklıkla sınanması ve genelde de başarılı sonuçlar alınması onu önemli finansal değişkenlerin tespitinde başvurulan bir araç haline getirmiştir. Casey'nin (1980:603) çalışması bu durumun ilk örneklerindedir. Pettway ve Sinkey'in (1980:137) aynı yıl yaptıkları çalışma da süreklilik riski konusunda öncelikli ve önemli finansal göstergelerin tespitine yöneliktir. Chen ve Shimerda'nın (1981:51) yaptıkları çalışmada ise önemli finansal göstergeleri belirlemenin yanında gösterge adedi toplamının sınırlı tutulmasının da önemli olduğu vurgulanmakta ve bu yapılmadığında ise Aşırı Uygunluk sorunlarının yaşanabileceğini ileri sürülmektedir. Keasey ve Watson (1986:49) Çoklu Ayraç Analizi'ni kullanarak önemli finansal göstergelerin belirlenmesi yöntemini İngiltere'ye uygulamakta ve halka açık verinin süreklilik riski öngörüsü yapmaya yeterli olduğunu savunmaktadır. McNamara ve diğ. (1988:53) Birleşik Devletler özel sektörünü konu aldıkları ve ÇAA ile yaptıkları çalışmada, süreklilik riski öngörüsü için, sadece iki finansal gösterge ile %80'in üzerinde bir başarı sağlanabildiğini ileri sürmektedir. Ward (1994:547) benzer çalışmasında süreklilik riski öngörüsü konusunda en önemli değişkenin "Net Gelir + Amortisman" olduğunu ileri sürmektedir. Bu çalışma bünyesinde, amortismanına göre düzeltilmiş net gelirin diğer ana bilanço kalemlerine oranlanmasının en önemli göstergeleri oluşturduğu savunulmaktadır.

Çoklu Ayraç Analizi ile ilgili eleştiriler de yapılan araştırmalara konu olmuş görünmektedir. Dambolena ve Khoury (1980:1017) yaptıkları çalışmada çok temel bir noktayı, finansal oranların istikrarını sorgulamaktadır. Araştırmacılar makalelerinde finansal oran ortalamalarının yıllar içinde son derece dalgalı bir seyir izlediğini ve bu kadar oynak bir verinin güvenilir sonuçlar üretemeyeceğini ileri sürmektedir. El Hennawy ve Moris (1983:209) ise iflasın aleni hale gelmesinden ne kadar önceki verinin kullanılması gerektiğini sorgulamakta ve yaptıkları çalışmada bu dönemin beş yıla kadar uzayabileceğini savunmaktadır. Sung ve diğ. (1999:63) ise süreklilik riski öngörüsü için normal zamanlar ile kriz dönemlerinin ayrılması ve bunlar için farklı modeller kurulmasının daha doğru sonuçlar üretebileceğini ileri sürmektedir.

sürmektedirler.

Ohlson (1980) sıklıkla atıfta bulunulan öncül çalışmasında süreklilik riski öngörüsü yapmak için ÇAA yerine Lojistik Model'i tercih etmektedir. Lojistik Model ÇAA'ne kıyasla ön koşullar bakımından daha az kısıtlayıcıdır ve ÇAA'nin her örnek kitlesine göre değişen Z-Skoru yerine standart bir süreklilik riski göstergesi üretebilmektedir. Bu yöntemde de Kesme Noktası dışsal bir parametre olarak kullanılmakta ve yanlış sınıflandırma sorununu benzer şekilde hafifletmesi amaçlanmaktadır.

Logit Model'in ÇAA'ne göre daha az ön şart içermesi ve yüzde olarak bir süreklilik riski göstergesi üretebilmesi pek çok araştırmacıyı cezbetmiş görünmektedir. Modelin alanda kullanımına öncülük edenlerden biri olan Zavgren, 1985'te gerçekleştirdiği çalışmasında (s:19), Birleşik Devletler'deki endüstri şirketlerine yönelik oluşturduğu süreklilik riski öngörü modelinin başarılı sonuçlar ürettiğini savunmaktadır. Koh (1987:124) çalışmasında Logit yerine Probit fonksiyonunu kullanmakta ve bunun süreklilik riskini daha iyi yansıttığını ileri sürmektedir. Dambolena ve Shulman (1988:74) süreklilik riskinin en iyi nakit akımları aracılığı ile takip edilebileceğini ileri sürdükleri makalelerinde Logit Model'in başarılı sonuçlar ürettiğini savunmaktadırlar. Flagg ve diğ. (1991:67) de yaptıkları çalışma çerçevesinde süreklilik riskinin öngörülmesi konusunda Lojistik Model'in başarılı olduğunu ileri sürmektedirler. Laitinen (1991:649) makalesinde, yeni ve eski şirketlerin farklı süreklilik riskleri ile karşı karşıya olduğunu ve bunlar için farklı Logit Modeller'in kullanılması gerektiğini savunmaktadır. Westgaard ve Wijst (2001:338) ise gerçekleştirdikleri çalışmada modeli banka kurumsal kredi portföylerine uygulamakta ve yine başarılı sonuçlar aldıklarını ileri sürmektedirler.

Logit Model'in ülkeye özgü uygulandığı çalışmalar da bulunmaktadır. Keasey ve McGuinness (1990:119) çalışmalarında modeli İngiltere'de endüstri firmalarına uygulamakta ve başarılı sonuçlar aldıklarını savunmaktadırlar. Theodossiou da (1991:697) modeli Yunanistan'da test etmekte ve yine başarılı sonuçlar aldığını ileri sürmektedir.

Dar bir sektöre odaklanmış modellemeye iyi bir örnek Foreman'ın (2002:135-136) gerçekleştirdiği çalışmadır. Birleşik Devletler telekomünikasyon sektörünü konu alan bu çalışmada Logit Model'in başarılı ile uygulandığı savunulmaktadır.

Logit Model eldeki örnekleri ikiden fazla sınıfa da ayırabilmektedir. Genelde modellenin kullanıldığı çalışmalarda eldeki örneklerin sadece iki sınıfa (riskli/sağlıklı) ayrılması tercih edilse de Lau (1987:127-128) yaptığı analizde eldeki örnekleri beş sınıfa ayırmayı uygun görmekte ve bunun daha rafine sonuçlar verdiğini savunmaktadır. Johnsen ve Melicher ise (1994:269) benzer bir çalışma yapmakta ama beş yerine örnekleri üç sınıfa ayırmayı yeterli bulmaktadır.

Logit Model ile Çoklu Ayraç Analizi'nin süreklilik riski öngörüsü performanslarını kıyaslayan pek çok çalışmaya rastlanılmaktadır. Espahbodi (1991:53) yaptığı çalışmasında Logit Model ve ÇAA'nin süreklilik riski öngörüsünde performans olarak birbirine yakın olduğunu savunmaktadır. Joos ve diğ. (1998:59) ise Logit Model'in Çoklu Ayraç Analizi'nden daha başarılı olduğunu ileri sürmektedirler. Kahya ve Theodossiou (1999:323) her iki yöntemin de eksikleri olduğunu savunmakta ve zaman serisine bağlı başka bir yöntem önermektedirler. Bu konuda yakın tarihli bir diğer makale Jones ve Hensher'e (2004:1011) aittir ve bu çalışmanın sonucunda Logit Model'in Çoklu Ayraç Analizi'ne kıyasla daha başarılı olduğu savunulmaktadır.

Logit Model'in, Çoklu Ayraç Analizi'nde olduğu gibi, "önemli" değişkenlerin belirlenmesinde de kullanıldığı çalışmalara rastlanılmaktadır. Theodossiou (1993:441) modele birkaç döneme ait değişkenlerin dahil edilmesinin daha iyi sonuçlar verdiğini ileri sürmekte, trendin ve trendden sapmanın önemli olduğunu savunmaktadır. Platt ve diğ. (1994:491-492) ise yaptıkları çalışmada değişkenlerdeki enflasyon etkisini Logit Model aracılığı ile irdelemektedirler.

3.1.3 Çok değişkenli doğrusal olmayan modeller

Sinir Ağı modeli, doğrusal modellerden doğrusal olmayan modellere geçişi simgeleyen iyi bir örnektir. Sinir Ağları temelde basit Logit Modeller'in bir araya gelmesinden oluşmaktadır, ancak barındırdığı bu alt modellerin paralel ve seri olarak birbirine bağlanabilmesi ona doğrusal olmayan ilişkileri keşfetme imkanı sunmaktadır.

Sinir Ağları'nın kullanımı 1980'lerde başlasa da yoğun kullanımı 1990'lardır (Aziz ve Dar, 2006:19). Süreklilik riski öngörü modellemesinde Sinir Ağları'nın kullanımına ilk örneklerden biri ise Odom ve Sharda'nın (1990:63) çalışmasıdır. Bu çalışmada Altman'ın ilk kullandığı model olan ÇAA ve orijinal çalışmada (1968)

kullanılan faktörler baz alınmakta ve bu modelin sonuçları kıstas olarak kabul edilmektedir. Araştırmacıların yapmaya çalıştığı verinin değil modelin sorgulanmasıdır. Çalışmanın sonucunda Sinir Ağları'nın bağımsız örnekler üzerinde %82 oranında başarı sağladığı ileri sürülmektedir. Bu oran ÇAA'nin %74'lük bağımsız veri performansının üzerinde ve ciddi bir artışa işaret etmektedir. Benzer bir çalışma Salchenberger ve diğ. (1992:899) tarafından da tekrarlanmakta ve sonuç değişmemektedir. Coats ve Fant (1993:142) da makalelerinde Sinir Ağları'nın ÇAA'nden daha iyi performans gösterdiğini savunmaktadır. Zaman içinde aynı konu üzerine yapılan diğer çalışmalarda da sonuçlar benzerlik göstermektedir (Wilson ve Sharda, 1994, Rudorfer, 1995, Serrano-Cinca, 1996, Jo ve diğ., 1997, Yang ve diğ., 1999, Shah ve Murtaza, 2000). ÇAA'nin performansının Sinir Ağları ile denk olduğunu savunan nadir makalelerden biri bizzat ÇAA'nin yaratıcısı olan Altman ve diğ.'ne (1994:505) aittir. Araştırmacılar bu çalışmada, Sinir Ağları'nın ciddi modelleme güçlüklerine dikkat çekmekte ve sonunda elde edilen modelin yorumlanmasının da hayli çaba gerektirdiğini ileri sürmektedirler.

Sinir Ağları ile Logit Model'in kıyaslanması da pek çok araştırmaya konu olmuştur. Bu bağlamda yapılan ilk çalışmalardan biri Bell ve diğ.'ne aittir (1990:29). Çalışmada Sinir Ağları'nın etkinliği Logit Model ile kıyaslanmakta ve sonuç olarak da önemli bir performans farkının olmadığı savunulmaktadır. Fletcher ve Goss'un (1993:159) yaptığı çalışmanın sonucunda ise Sinir Ağları'nın Logit Model'den bariz olarak daha iyi çalıştığı ileri sürülmektedir. Benzer şekilde Wilson ve diğ.'nin (1995:31) gerçekleştirdikleri çalışma neticesinde de süreklilik riski öngörüsünün modellenmesinde Sinir Ağları'nın Logit Model'den daha başarılı olduğu savunulmaktadır. Zhang ve diğ. (1999:16) yayınladıkları makalede de Sinir Ağları lehindeki görüşler tekrarlamakta ve hatta örnek sayısının artması durumunda Sinir Ağları'nın giderek daha iyi performans gösterdiği ileri sürülmektedir.

Sinir Ağları'nı aynı anda hem Logit Model ile hem de ÇAA ile kıyaslayan çalışmalar da yapılmıştır. Bu duruma ilk örneklerden biri Tam'ın (1991:429) yayınladığı makaledir. Tam bu çalışmasının sonucunda Sinir Ağları'nı hem Logit Model ile hem de ÇAA ile kıyaslamakta ve Sinir Ağları'nın diğer iki modelden daha etkin olduğu ileri sürmektedir. Benzer bir çalışma Tam ve Kiang (1992:926) tarafından bir sene sonra tekrarlanmakta ve sonuç değişmemektedir. Agarwal (1993:116) yaptığı çalışmanın sonucunda en başarılı model olarak Sinir Ağları'nı, en zayıf olarak ise

CAA'ni işaret etmekte, Logit Model'i ise performans olarak ortaya yerleştirmektedir.

Sinir Ağları ile oluşturulan süreklilik riski öngörüsü modelleri ülke spesifik çalışmalarda da yer bulmuştur. Tsukuda ve Baba (1994:445) yöntemi Japon verisine uygulamakta ve batıklık öngörüsü konusunda başarılı sonuçlar alındığını savunmaktadır. Kiviluoto modeli (1998:191) Finlandiya verisi üzerine uygulamakta ve sonuçların benzer şekilde başarılı olduğunu ileri sürmektedir. Alam ve diğ. (2000:185-186) modelin tek bir sektöre (bankacılık) kısıtlanması durumunda da başarılı şekilde performans gösterdiğini ileri sürmektedir.

Sinir Ağı modelinin yaygın kullanımı ona başka fonksiyonların yüklenmesinin önünü açmış ve bu fonksiyonlardan birinin de süreklilik riski öngörüsündeki önemli değişkenlerin belirlenmesi olmuş görünmektedir. Jandaghi ve diğ.'nin (2011:37-38) ve Neophytou ve diğ.'nin (2001:25-26) yaptıkları çalışmalar süreklilik riski öngörüsü modellerindeki önemli değişkenlerin Sinir Ağları ile belirlenmesine uygun örneklerdir.

Sinir Ağları'nın kullanımına yönelik eleştiriler de çeşitli araştırmalara konu olmuştur. Boritz ve Kennedy (1995:503) yaptıkları çalışmada Sinir Ağları model performansının değişken seçimine ve örnekleme çok hassas olduğunu ileri sürmektedir. Lee ve diğ. (1996:63-64) modele dahil edilecek değişkenlerin belirlenmesini önemli bir sorun olarak belirtmekte ve bir çözüm olarak seçimin Çoklu Ayraç Analizi ile yapılmasını, burada belirlenen değişkenlerin daha sonra Sinir Ağları'nın oluşumunda kullanılmasını önermektedirler. Çalışma, bu açıdan, hibrit modellere ("Hybrid Model") ilk örneklerdendir. Lesnhno ve Spector (1996:125-126) ise, Sinir Ağları'nda sıklıkla yaşanan Aşırı Uygunluk durumuna vurgu yapmakta ve model eğitimi esnasında öğrenme adımlarının küçük tutulmasını önermektedirler.

Destek Vektör Makineleri çok değişkenli doğrusal olmayan Makine Öğrenmesi modellerinin yakın tarihli örneklerindedir. DVM, daha önceden fikir olarak bilinse de, Cortes ve Vapnik'in (1995:273) yeni bir dönem açan çalışmasına kadar pratikte bir uygulama alanı bulamamıştır. Çalışma, daha önceden DVM kullanımını kısıtlayan bazı ön şartları kaldırmakta, çok daha geniş bir problem yelpazesine cevap verebilecek şekilde esneklik kazandırmakta ve bu sayede model sonuçlarının genelleme başarısını (Çapraz Doğrulama performansı) yükseltmektedir. DVM hem

örnek bazlı hem de kural bazlı sınıflandırma ihtiyaçlarına cevap verebilecek şekilde ortaya çıkmaktadır.

Cortes ve Vapnik'in (1995) makalesinin yayınlanmasından kısa bir süre sonra Burges ve Scholkopf (1997:475-480) DVM modelleme hızının ve doğruluğunun artırılması yönünde öncül bir çalışma yapmakta, bunu bir yıl sonra Burges'in (1998) DVM'nin etkin kullanımına yönelik detaylı makalesi takip etmektedir. Scholkopf ve diğ. 2000 yılında (1208-1243) ele aldıkları makalede DVM'nin uygulanması sırasında yaşanabilen Aşırı Uygunluk problemine karşı etkin bir çözüm geliştirdiklerini savunmakta ve bunu kullanıcının takip edebileceği bir prosedür haline getirmektedir. Tay ve Cao (2001:310-316) ise DVM'nin zaman serilerine de uygulanabileceğini ileri sürmekte ve bunun için bir prosedür önermektedir.

DVM'nin süreklilik riski öngörü modeli olarak kullanıldığı ilk çalışmalardan biri Fan ve Palaniswami'ye (2000:354) aittir ve önemli değişkenlerin belirlenmesi hususunda başarı ile kullanıldığı savunulmaktadır. Min ve Lee (2005:603-604) DVM kullanarak gerçekleştirdikleri batıklık öngörüsü çalışmasında model parametrelerinin optimum seçimi konusuna da eğilmekte ve doğru çekirdek (DVM model optimizasyon parametrelerinden biri) seçimine odaklanmaktadır. Modelin Türkiye'ye uygulandığı ilk çalışma Erdoğan'dan (2013) gelmekte ve bankacılık sektöründeki süreklilik risklerinin öngörülmesi amaçlanmaktadır. Erdogan, çalışmanın sonucunda, DVM'nin Gauss çekirdek (eldeki örneklere yakınlıkla sınıflandırma) seçeneğiyle kullanıldığında olumlu sonuçlar alındığını ileri sürmektedir.

DVM modelinin performans açısından Sinir Ağları ile karşılaştırılması araştırmalara konu olmuştur. Kim (2003:307-308) yaptığı çalışmada DVM'nin zaman serisi öngörüsünde Sinir Ağları'ndan daha iyi olduğunu savunmaktadır. Shin ve diğ. (2005:127-128) ise süreklilik riski öngörüsü modelleri çerçevesinde bu iki modeli kıyaslamakta ve yine DVM'nin daha başarılı sonuçlar ürettiğini ileri sürmektedir. Aynı çalışmanın sonucunda veri daraldıkça DVM'nin Sinir Ağları'na kıyasla pozitif performansının belirginleştiği savunulmaktadır.

Karar Ağaçları temelde Özyinelemeli Bölümlendirme ("Recursive Partitioning") modelleri içinde yer alır ve çok değişkenli doğrusal olmayan Makine Öğrenmesi modellerinin ana kollarından biridir. Friedman'ın (1977) çığır açan makalesi ile Karar Ağaçları'nın temelleri atılmaktadır. Bu çalışmada özyinelemeli

bölümlendirilmenin nasıl yapılacağı, bunun istatistiki dayanağı detaylı şekilde açıklanmaktadır. Her şeye rağmen modellemenin gerektirdiği işlem miktarı araştırmacıları 1980'lerin ortalarına kadar Karar Ağaçları'ndan uzak tutmuş görünmektedir. Breiman ve diğ.'nin (1984) Karar Ağaçları'nın pratikte nasıl uygulanabileceğine dair kaleme aldıkları çalışma modelin kullanımına yeni bir ivme kazandırmıştır. Bu çalışmada Karar Ağaçları'nın hem sınıflandırma amaçlı hem de regresyon amaçlı kullanılabileceği, her iki durum için de yöntemlerin ne olması gerektiği detaylı ve tekrarlanabilir bir biçimde sunulmaktadır.

Karar Ağaçları'nın süreklilik riski öngörüsü için kullanıldığı ilk çalışma Frydman ve diğ.'ne (1985:269-270) aittir. Bu çalışmada Karar Ağaçları performans açısından Çoklu Ayraç Analizi ile kıyaslanmakta ve Karar Ağaçları'nın, eldeki veri çerçevesinde, daha başarılı olduğu savunulmaktadır. Daha yakın tarihli bir çalışmada Pompe ve Feelders (1997:267-268) ÇAA, Sinir Ağları ve Karar Ağaçları'nı kıyaslamakta ve bu üç modelin performans olarak birbirine yakın olduğunu ileri sürmektedir. Karar Ağaçları'nın, Destek Vektör Makineleri'nin ve Sinir Ağları'nın kıyaslandığı daha güncel bir çalışma Kirkos ve diğ.'nden (2008:213-214) gelmekte ve sonucunda en başarılı modelin Karar Ağaçları olduğu, en zayıfın ise Sinir Ağları olduğu savunulmaktadır. Gepp ve diğ. (2010:536-537) Çoklu Ayraç Analizi'ni Karar Ağaçları'na karşı tekrar sınamakta ama sonuç değişmemekte ve Karar Ağaçları'nın daha başarılı olduğu savı tekrarlanmaktadır. Shah (2014:103-104) Avustralya madencilik sektörüne odaklandığı çalışmasında Karar Ağaçları'nı, hibrit modelleri, Sinir Ağları'nı ve Çoklu Ayraç Analizi'ni veriye uygulamakta, sonuç olarak Karar Ağaçlarının ve hibrit modellerin diğerlerine oranla daha başarılı olduğunu ileri sürmektedir.

Friedman'ın orijinal Karar Ağacı modeli eşit öneme sahip iki değişken arasından daha çok seviyeye ayrılmış olana öncelik tanıyabilmektedir. Bu durum "daha etkili" olabilecek değişkenlerin modelden dışlanması sonucunu ortaya çıkarabilmektedir. Hothorn ve diğ. (2006:651-652) tam da bu noktaya eğilmekte ve daha yansız bir Karar Ağacı modeli önermektedirler. Hothorn Ağacı modeli Breiman Ağacı'na kıyasla daha gelişmiş bir model olarak görünse de eldeki verinin yapısı nihai performansı belirleyebilmektedir. Örneklemin hangi modeli öne çıkaracağı ilk anda belli olmadığından, her iki tip Karar Ağacı modelinin de veriye uygulanması ve en iyi modelin daha sonra seçilmesi takip edilebilecek en doğru yöntem olarak

görülmektedir.

Karar Ormanı modelleri birçok Karar Ağacı modelinin bir araya gelmesiyle oluşmaktadır (Breiman, 2001). Bu modelde örnekleme oluşturan veri ve değişkenler rastgele bölümlendirilmekte ve eldeki her bölüm ile ayrı bir Karar Ağacı modeli oluşturulmaktadır. Sınıflandırma kararı ormandaki tüm ağaçların eşit oyları ile belirlenmektedir. Değişkenlerin rastgele seçilmesi baskın faktör etkisini (bir değişkenin çok ön plana çıkarak diğer potansiyel değişkenleri model dışında bırakması) azaltmaktadır. Modellemenin bu şekilde yapılmasından dolayı Karar Ormanları'nın nadiren Aşırı Uygunluk sorunu yaşayacağı ileri sürülmektedir (Hothorn ve diğ., 2006:667-668).

3.2 Modelleme Esnasında Karşılaşılan Sorunlar

Süreklilik riski öngörüsü yapılmaya çalışıldığında doğru modelin belirlenmesi yanında gerek ve yeter sayıdaki değişkenin seçimi, öngörünün ne kadar önceden yapılabileceği, nihai modelin doğrulanması ve muhtemel sınıflandırmanın objektif olarak tanımlanması gibi konuların da araştırmacı tarafından ele alınması gerekmektedir.

Süreklilik riski öngörüsü üzerine yapılan çalışmalarda kullanılan değişken adedi büyük farklılıklar göstermekte (Belowary, 2007:5-7) ve tek bir finansal orana dayanan modeller yanında onlarca değişkeni bünyesinde barındıran modellere de rastlanılmaktadır. Bağımsız veri kümesi ile doğrulanan modeller baz alındığında en başarılı modellerin 2 ila 21 arasında değişkene sahip oldukları görülmekte ve bu da yüksek değişken adedinin tek başına model performansını belirleyemeyeceğine işaret etmektedir (Belowary, 2007:1).

Modele dahil edilen her gösterge probleme farklı bir bakış açısı getirebilmektedir, ancak gösterge sayısındaki artışın ciddi yan etkileri (Aşırı Uygunluk gibi) de karşımıza çıkabilmektedir. Platt ve Platt (1990:31-32) yayınladıkları makalede, Aşırı Uygunlu'ğa karşı en iyi çözümün değişken sayısının mümkün olduğunca azaltılması olduğunu ileri sürmektedir. Neophytou ve diğ.'nin (2001:3-5) yaptıkları araştırma sonucu "yeterli" gördükleri değişken sayısı sadece üçtür. Bu çalışmada yalnızca karlılığı, nakit akımını ve borçluluğu gösteren üç finansal oranın süreklilik riskini %83 başarı ile öngörebildiği savunulmaktadır.

Nakit ve fon akımlarının süreklilik riski öngörüsüne etkileri, diğer finansal göstergelere nispetle, daha sıklıkla araştırmalara konu olmuştur. Casey ve Bartczak 1984'te (61) yaptıkları çalışmada nakit akımlarının tek başına süreklilik riski öngörüsünde yeterli olmadığını ileri sürmekte ancak bir sene sonra yineledikleri ve derinleştirdikleri çalışmalarında (1985:384-385) ise operasyonlardan elde edilen nakit akımının süreklilik riski öngörüsüne en büyük katkıyı yaptığını savunmaktadırlar. Gentry ve diğ. (1985a:146-147) ise ele aldıkları makalede, operasyon kaynaklı nakit akımlarının gösterge niteliğinin zayıf olduğunu ancak temettü dağıtım politikasına bağlı nakit akımının değerli bir indikatör olduğunu ileri sürmektedirler. Gentry ve diğ. (1985b:47-48) aynı yıl yaptıkları benzer bir çalışmada fon girişlerinin değil, fon çıkışlarının genelde daha önemli olduğunu belirtmekte, temettü, yatırımlar ve alacaklardaki artışların önemli göstergeler olduğunu savunmaktadırlar. Gombola ve diğ. (1987:55-56) ise yaptıkları çalışmanın sonucunda operasyon kaynaklı nakit akımlarının süreklilik riski öngörüsü çerçevesinde önemli göstergeler olmadığını ileri sürmektedirler.

Nakit akımlarının tek başlarına kullanımı yerine diğer bilanço kalemleri ile birlikte modellere dahil edilmeleri araştırmacıların ilgisini çekmiştir. Gilbert ve diğ. (1990:161-162) yaptıkları çalışmada, süreklilik riski öngörüsü için nakit akımı kalemlerinin tahakkuk esaslı kalemlerle birlikte kullanılmasının gerektiğini ve ancak bu durumda olumlu sonuçlar alınabileceğini savunmaktadırlar. Rujoub ve diğ. (1995:75-76) de yayınladıkları makalelerinde nakit akımlarının önemli göstergeler olduğunu, ancak tahakkuk esaslı kalemlerle birlikte kullanılmaları halinde süreklilik riski öngörü performansını yükseltebildiklerini ileri sürmektedirler.

Nakit akımları yerine nakdin nicel büyüklüğünün önemini irdeleyen Laitinen ve Laitinen (1998:893-894) yaptıkları çalışmada, nakdin ciroya oranının önemli bir gösterge olduğunu ve süreklilik riski öngörüsü modeline dahil edildiğinde olumlu katkı yaptığını savunmaktadırlar.

Ortaya çıkan modellerin geçerliliğinin sağlıklı bir şekilde belirlenmesi de en az modelin kendisi kadar önemlidir. Jones (1987:131-132) bu noktayı görece erken bir tarihte vurgulamakta ve süreklilik riski öngörüsünde kullanılan modeller için bağımsız örneklerin kullanılmasını önermektedir. Jones'un bu uyarısına rağmen yapılan araştırmaların pek çoğunda sağlıklı bir doğrulamaya yer verilmemektedir (Belowary ve diğ. 2007:8). Pek çok modelin performansı üzerine inşa edildikleri veri

kümesi ile test edilmekte ve bu da modellerin pratikte olabileceğinden daha başarılı görünmesine neden olmaktadır.

Süreklilik riski öngörüsü yapmaya yönelik kurulan modellerin ne kadar önceden bu uyarıyı yapabildikleri ayrı bir farklılık unsurudur (Belowary ve diğ., 2007:10). Konu üzerine yapılan çalışmaların geneli ele alındığında ortaya çıkan modellerin bir yıl öncesinden yaptıkları tahminlerin en başarılı sonuçları ürettiği ve zaman uzadıkça tahmin gücünün zayıfladığı görülmektedir, ancak beş yıl öncesinden başarılı tahminler yaptığını ileri süren çalışmalara da rastlanılmaktadır (El Hennawy ve Morris, 1983:209-210).

Süreklilik riski öngörüsü çalışmalarında riskin gerçekleşmesi durumu için yapılan tanımlar da oldukça farklılıklar göstermektedir. Çalışmaların ekseriyeti bunu doğrudan resmi iflas başvurusu olarak, diğerleri ise finansal stres veya borçların ödenmesinde yaşanan güçlükler olarak kabul etmektedir (Dimitras ve diğ., 1996:487-488). Bunun pratik nedenleri vardır, örneğin banka iflasları üzerine Logit Model kurmaya çalışan Martin (1977:249-250) bankaların genelde resmi olarak iflas etmediğini ileri sürmekte ve tanımlamayı “ciddi finansal zorluk” olarak yapmayı tercih etmektedir. Gruszczynski (2015:97-101) ise yakın tarihli çalışmasında süreklilik riski öngörüsü modellemelerinde yaşanan en büyük üç sorundan birinin bizzat “Süreklilik Riski” tanımını olduğunu savunmaktadır.



4. ÇALIŞMADA KULLANILAN MAKİNE ÖĞRENMESİ MODELLERİ

Bu bölümde çalışmada kullanılan çok değişkenli Makine Öğrenmesi modellerinin teorik altyapıları, temel kullanım şekilleri ve dikkat edilmesi gereken yönleri sunulmaktadır. Araştırmanın bundan sonraki bölümüne tek değişkenli Makine Öğrenmesi modelleri dahil edilmemektedir ve bunun nedenleri tezin üçüncü bölümünde (3.1.1) aktarılmaktadır.

4.1 Çoklu Ayraç Analizi

ÇAA bir sınıflandırma aracı olarak kullanılmaktadır. Çoklu Ayraç Analizi'nde amaç eldeki değişken boyutu sayısını azaltmak, örneklemdaki sınıfların tek bir lineer ayraç kullanılarak birbirinden ayrılmasını sağlamak ve bunu yaparken de muhtemel ayraçlar arasından en doğrusunu seçmektir. Model en az sayıda değişkeni kullanarak en doğru sınıflandırmayı gerçekleştirmeyi amaçlamaktadır. Ayrıca dahil olan değişken sayısı azaldıkça modelin anlaşılması ve pratikte de faydalı olması mümkün olmaktadır.

Modelin dayandığı veri kurulumunda kullanılan örneklemin değişkenleri ve hedef gruplamalardan oluşmaktadır. Her değişken örnekleme ile ilgili farklı bir özelliği yansıtmaktadır. Bu değişkenlere, sınıflandırma gücüne göre, farklı ağırlıklar verilerek tek bir ayracın ortaya çıkarılması sağlanmaktadır. Her örnek için sınıflandırma kararı bu ayracın aldığı değere göre verilmektedir.

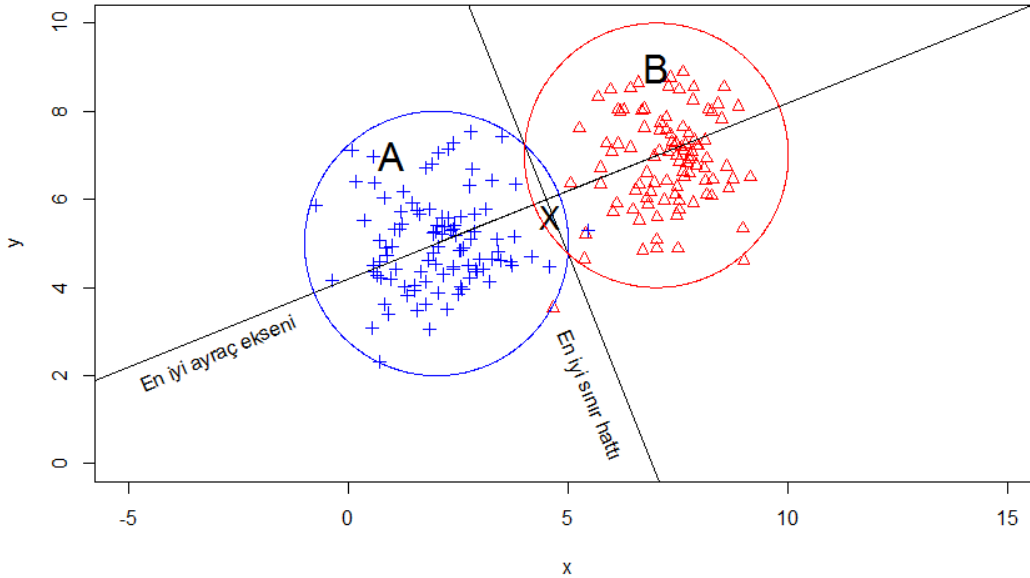
Başarılı bir sınıflandırma için ayracın hedef gruplar arasında oldukça farklı değerler, grupların kendi içlerinde ise birbirine yakın değerler üretmesi amaçlanmaktadır. ÇAA, bu ayracı oluşturan değişkenlerin optimal şekilde seçilmesini ve ağırlıklandırılmasını sağlamaktadır.

Modelin ortaya çıkardığı ayracın hangi aralığının hangi gruba işaret edeceği ayrı bir optimizasyon konusunu oluşturmaktadır. Ayraç üzerindeki bu eşik değerler “Kesme Noktaları” olarak isimlendirilmektedir. Kesme Noktaları sınıflandırma hatalarını en aza indirecek şekilde belirlenmekte, ancak bu durum arada kalan örneklerin hiç sınıflandırılmaması ile sonuçlanabilmektedir.

Sınıflandırma, önceden belirlenmiş sınıf adedi ve bu sınıfların değişken uzayındaki yerleşimine göre, bir veya birden çok ayraç ile gerçekleştirilebilmektedir. Model optimizasyonu ayraç sayısının artışından etkilenmemektedir ve hedef yine en az sayıda ayraç ile en doğru sınıflandırmanın yapılması olarak kalmaktadır.

ÇAA modeli bazı ön şartların sağlanmasını gerekli kılmaktadır. Model oluşumunda kullanılan değişkenlerin dağılımlarının normal dağılıma yakınsamaları, farklı sınıfların kovaryans matrislerinin birbirine benzemeleri ve değişkenlerle sonuç arasında lineer bir ilişkinin olması gerekmektedir.

ÇAA modelinin nasıl çalıştığına dair farazi bir örnek faydalı olabilecektir. Şekil 4.1’de A ve B sınıfına dahil örneklerinin x ve y değişkenleri cinsinden yayılımı sergilenmiştir. Şekilden de görülebileceği gibi A ve B gruplarını başarılı bir şekilde, düz bir hat vasıtası ile ayırmak mümkündür (En iyi sınır hattı). Bu hattı tanımlamanın bir yolu da sınır hattını dik olarak kesen ayraçtır (En iyi ayraç eksenini). Bu sayede A ve B gruplarını birbirinden ayırmak için x ve y eksenleri yerine sadece tek bir ayraç kullanılabilir.



Şekil 4.1: Çoklu Ayraç Analizi'ne Örnek

Gruplara ayırma işlemi tüm örneklemin bu ayraç üzerine yansıtılması ile başlamaktadır. Herhangi bir örneğin ayraç üzerine yansıtılması, ayraçın örneğe en yakın noktasının tespiti ile gerçekleştirilmektedir. Örneklem ayraça yansıtıldıktan sonra, herhangi bir örneğin model açısından değişken uzayındaki lokasyonu ayraç

üzerindeki bu noktadır. Şekil 4.1 incelendiğinde, A grubuna ait örneklemin ayracın soluna doğru olan bölgeye, B grubuna ait örneklemin ise ayracın sağına doğru olan bölgeye yansiyacağı görülmektedir.

Örnek kümesi ayraca yansıtıldığında elimizdeki problemin karmaşıklığı azalmakta, boyut sayısı düşmektedir. Optimal ayraç, grup içi farklılıkları azaltırken gruplar arası farkı bariz hale getirmektedir. Şekil 4.1’den de anlaşılabilceği gibi, gruplamaları iki boyutlu çember grup tanımlamaları ile yapmak yerine tek bir ayraç üzerinden sınıflandırma yapmak ve boyut sayısını azaltmak mümkün olmaktadır.

Şekil 4.1’de ortaya çıkan model, iki çemberin kesişim bölgesi hariç (Kesme Noktaları arasında kalan bölge), sınıflandırmayı şu kurala göre öngörmektedir: bahse konu örneğin ayraç üzerine yansımaları “X” noktasının solunda ise A gurubuna, değilse B gurubuna aittir. Model, iki çemberin kesiştiği alanda kalan örnekler için ise herhangi bir sınıflama öngörüsü yapmaktan kaçınmaktadır.

4.2 Lojistik Model

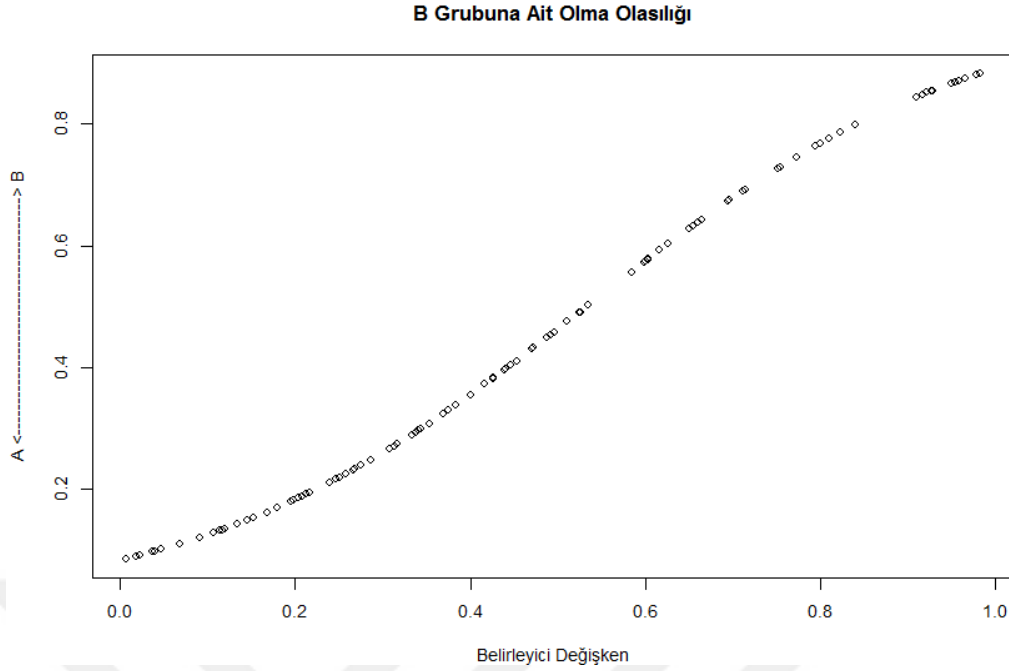
Lojistik Model de Çoklu Ayraç Analizi gibi bir sınıflandırma aracıdır ve ÇAA’ni iki yönden geliştirmeyi amaçlamaktadır. Bunlardan ilki Çoklu Ayraç Analizi’nin ürettiği ayraç değerinin “olasılık” eksenine yansıtılmasıdır. ÇAA eldeki örneğin grup tayinini gerçekleştirmekte ama bu tayinin ne derece güvenilir olduğunu veya kararın doğruluk ihtimalini kullanıcıya sunmamaktadır. Lojistik Model ise ÇAA tarafından üretilen ayraç değerini olasılık aralığına, yani (0,1) aralığına yansıtmakta, sınıflandırmayı eldeki örneğin ayracın hangi kısmına yansıdığına bakarak tayin etmektedir. Model herhangi bir örnek için %50’nin üzerinde bir değer üretiyorsa bu örneğin hedef gruba, değil ise alternatif gruba tayini öngörülmektedir. Bu durumda sınıflandırma yanında yapılan grup tayininin güvenilirliği hakkında da bir veri üretilmiş olmaktadır.

Lojistik Model Çoklu Ayraç Analizi’ne kıyasla daha az ön şartın sağlanmasını gerektirmektedir ve bu durum ÇAA’ne karşı ikinci bir uygulama avantajı sağlamaktadır. Model, ÇAA’nin modellenmesi için şart koşulan sınırlamalardan biri olan hedef grupların benzer kovaryans matrislerine sahip olmasını gerektirmemektedir.

Lojistik Model’de kullanılan ve ayraç deęerinin olasılık aralıęına yansıtılmasını saęlayan “yansıtma” (link) fonksiyonu deęişik Őekillerde belirlenebilmektedir. Lojistik Model’de, literatüre yansıdığı gibi ekseriyetle, “Logit” fonksiyonu kullanılmaktadır. Yansıtma fonksiyonunu olarak kümülatif normal daęılımının tersi kullanıldığında ise “Probit” modele ulaşılmaktadır. Eldeki örneklemin ve hedef grupların örnek uzayındaki daęılımları farklı yansıtma fonksiyonlarının kullanımını optimal hale getirebilmektedir.

Lojistik Model’in klasik regresyondan farkı model optimizasyonunda kullanılan uyumsuzluk maliyet fonksiyonunun “farkın karesi” yerine “Logit” ile deęiştirilmiŐ olmasıdır. Logit Model’de yapılan hatanın maliyeti öngörü ile gerçek durum arasındaki farkın karesi yerine kısmi logaritması ile hesaplanmaktadır. Model bu sayede, temelde hala bir regresyon fonksiyonundan faydalanılmasına raęmen, yansıtma fonksiyonları aracılığıyla, bir sınıflandırma aracı olarak kullanılabilir.

Logit fonksiyonu, tüm bir ayraç eksenini (0, 1) aralıęına yansıtılmaktadır (Őekil 4.2). Bu yansıtma sonucunda elde edilen deęerler bir tür “olasılık” verisi olarak kabul edilmektedir. Őekil 4.2’de sunulan yansıtmayı temel alırsak; eldeki örnek için model 0.6 deęeri üretiyor ise bu örneğin grubu %60 ihtimal ile B olarak öngörülecektir. Benzer Őekilde, eęer model bu örnek için 0.2’lik bir deęer üretiyorsa bu sefer eldeki örneğin grubu %80 ihtimal ile A olarak öngörülecektir. Yansıtmanın sonucu 1.0’a yaklaştıkça eldeki örneğin B grubuna ait olma olasılığı, 0.0’a yaklaştıkça ise A grubuna ait olma olasılığı yükselmektedir.



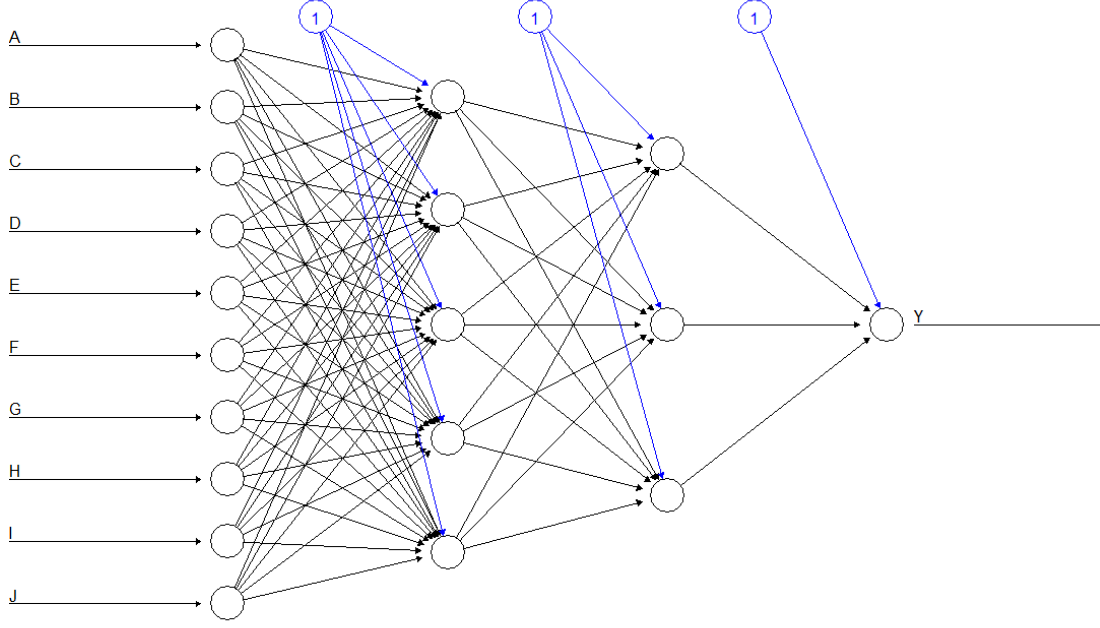
Şekil 4.2: Lojistik Model

Lojistik Model’de, optimize edilmeye çalışılan parametre büyüklüklerini sınırlamak (“regularization”), Aşırı Uygunluk sorununa karşı uygulanan yöntemlerden biridir. Bu çalışmada iki tip sınırlandırma yönteminden faydalanılmaktadır, bunlar; “Ridge” ve “Lasso” yöntemleridir. Birinde model parametrelerinin kareleri sınırlandırılırken diğesinde mutlak değerleri sınırlandırılmaktadır. Lasso yönteminde yüksek korelasyona sahip değişkenlerden sadece biri modelde kalırken diğeleri elenmekte, Ridge yönteminde ise yüksek korelasyona sahip değişkenlerin ağırlıkları birbirine yakınlaştırılmaktadır. Değişkenlerin dağılımları ve hedef grupların değişken uzayındaki lokasyonları hangi yöntemin daha başarılı sonuçlar üreteceğini belirleyebilmektedir.

4.3 Sinir Ağları

Sinir Ağları bir sınıflandırma aracıdır ve temelde birden çok Lojistik Model’in paralel ve seri bağlanması sonucunda ortaya çıkarılabilmektedir (Şekil 4.3). Model, bu yoğun bağlantı ağı sayesinde, değişkenler arası yüksek dereceli ilişkileri, karmaşık yapıları keşfedebilmektedir.

Sinir Ağları üç ana katmandan (soldan sağa) oluşmaktadır. Bunlar “girdi katmanı” (en soldaki düğüm noktaları), “saklı katmanlar” (ara kolonlardaki düğüm noktaları) ve “çıktı katmanı” (en sağdaki düğüm noktaları) olarak anılmaktadır.



Şekil 4.3: Sinir Ağı

Şekil 4.3'te sergilenen Sinir Ağı modelinde girdi adedi 10 (giriş katmanındaki A, B, C, ..., J), saklı katman adedi 2, ilk saklı katmandaki düğüm adedi 5, ikinci saklı katmandaki düğüm adedi 3 ve çıktı katmanındaki düğüm adedi 1'dir (Y). Şekil 4.3'te en üst satırda yer alan ve içlerinde "1" rakamı olan düğüm noktaları katman ortalama referans değerlerini düzenlemekte, yanlılığı engellemektedirler.

Sinir Ağları modellerinde, giriş katmanı dışındaki tüm düğüm noktaları, farklı bir Lojistik Model'i temsil etmektedir. Şekil 4.3'teki modelde ilk saklı katmanda 5, ikincide 3 ve çıkış katmanında da 1 olmak üzere toplam 9 adet farklı Lojistik Model kendine yer bulmaktadır. Giriş katmanındaki düğüm noktaları değişime uğramamış orijinal veriden direkt olarak beslenmektedir. Ara katmandaki düğüm noktaları ise bir önceki katman tarafından üretilmiş olan farklı Lojistik Modeller'in çıktılarını girdi olarak kabul etmektedir.

Sade bir Lojistik Model'e kıyasla ortalama bir Sinir Ağı'nın karmaşıklığı katlamalı olarak artmakta ve bu durum modelin sağlıklı bir şekilde eğitilmesini zorlaştırmaktadır. Sinir Ağları'nı oluşturan ilişkiler ağı modelin ortaya çıkarılmasını zorlaştırırsa da doğrusal olmayan ilişkilerin keşfi ancak bu sayede mümkün olabilmektedir. Özellikle de derin (birden fazla saklı katmana sahip) ağlarda bu durum uç noktalara ulaşabilmektedir.

Sinir Ağı genişledikçe (girdi adedi arttıkça) ve derinleştikçe (saklı katmak adedi arttıkça) eğitilmesi daha fazla veri gerektirecektir. Katmanlardaki düğüm noktaları bir sonraki katmanda bulunan tüm diğer düğüm noktalarını beslemektedir. Her bağ için optimal bir ağırlık hesaplanmasının yapılması gerekmektedir. Bu da belirlenmesi gereken pek çok parametre ve her parametrenin sağlıklı tespiti için gereken çok sayıda örnek demektir. Şekil 4.3'teki modelde optimal olarak belirlenmesi gereken ağırlık sayısı; 11 girdi ve ilk saklı katman arası 5 Lojistik Model için 55, birinci ve ikinci saklı katmanlar arası 6 girdi ve 3 Lojistik Model için 18 ve ikinci saklı katman ile çıktı arası 4 girdi ve 1 Lojistik Model için de 4 olmak üzere toplam 77'dir.

Makine Öğrenmesi modellerinde optimize edilmesi gereken her model ağırlığı için idealde ortalama 20 civarı, ama kısıtlı hallerde ise en az 10 örnek ve daha fazlası sağlıklı sonuçlar verebilmektedir (Goldberger, 1991:249-250). O halde, böyle bir modeli eğitebilmek için en az $77 * 10 = 770$, idealde ise yaklaşık olarak 1540 adet örneğe ihtiyaç duyulmaktadır. Elimizdeki toplam verinin sadece 106 örnekten oluştuğu düşünüldüğünde bu çalışma çerçevesinde optimize edilmesi planlanan sinir ağı modelinin Şekil 4.3'teki modele kıyasla çok daha sade olacağını öngörmek yanlış olmayacaktır.

Sinir Ağları model tarafından gerçekleştirilen öngörülerin gerçek durum ile kıyaslanması ve farkın (hatanın) belirlenmesi ile eğitilmektedir. Bu fark çıktı katmanından başlayarak geriye doğru sürüklenmekte ve Sinir Ağı bağlantı parametreleri bu farkı azaltacak yönde kısmi olarak düzeltilmektedir. Model eldeki örneklemin ağ boyunca sonuç katmanına doğru ilerletilmesi ile öngörü üretmekte, tespit edilen hataların sonuç katmanından geriye doğru çekilmesi ile de eğitilmektedir. Ne şiddetle bu yanlışların düzeltileceği seçimlik bir parametredir (modelin öğrenme hızı). Bu parametre eldeki veri ve modele göre değişebilmekte ve ne çok büyük ne de çok küçük olması önerilmektedir (Leshno M. ve Spector, 1996:125). Öğrenme hızı çok yüksek olduğunda optimal bir modele yakınsamak mümkün olamamakta, çok düşük olduğunda ise optimal modele yakınsamak çok uzun süre alabilmektedir.

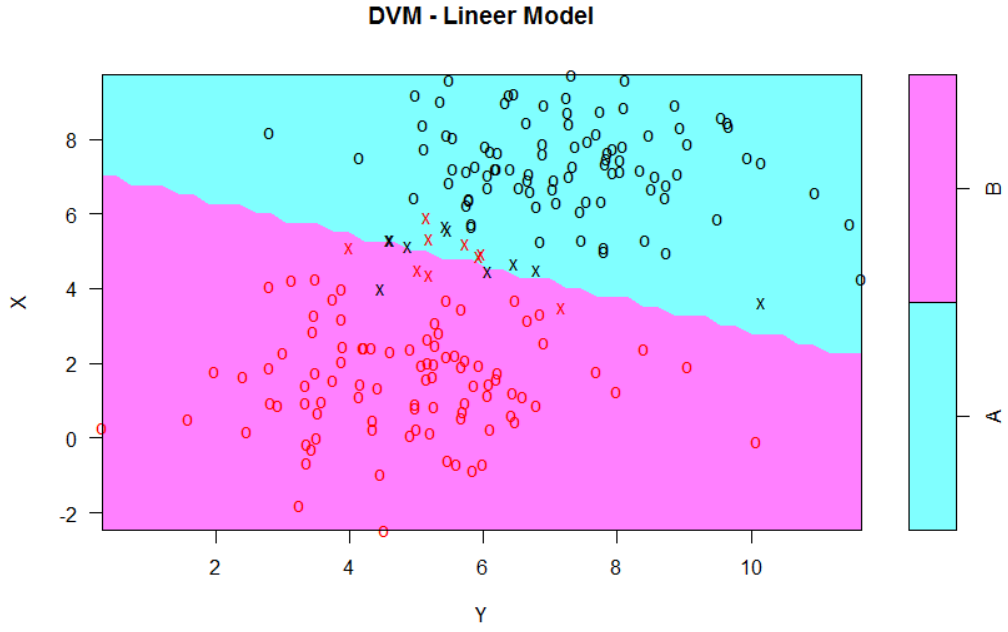
Sinir Ağı modelinin optimizasyonu esnasında çok sayıda deneme yapılmakta, bu denemeler esnasında düğüm noktaları arasındaki bağların ağırlıkları farklı değerler alabilmektedir. Düğüm noktaları arasındaki bir bağın bu denemeler boyunca aldığı değerlerin ortalaması o bağın muhtemel önemi hakkında bilgi verebilmekte,

değişkenler arası “önemli” ilişkilerin ortaya çıkarılmasına yardımcı olabilmektedir. Sinir Ağları’nın, model optimizasyonu rutini gereği, öngörü üretmesi yanında “ilişki keşfedici” bir rol üstlenmesi de bu şekilde mümkün olabilmektedir. Keşfedilen bu ilişkiler diğer Makine Öğrenmesi modellerinde kullanılabilmekte, bu da modellerin öngörü performansını olumlu yönde etkileyebilmektedir.

4.4 Destek Vektör Makineleri

Destek Vektör Makineleri (DVM) bir sınıflandırma aracıdır ve Cortes ve Vapnik (1995) tarafından Makine Öğrenmesi modellerine dahil edilmiştir. Ağırlıklı olarak Lojistik Model altyapısını kullanmakta, ancak Sinir Ağları kadar kompleks modeller de üretmemektedir. DVM’nin diğer Makine Öğrenmesi modellerinden en önemli farkı, sadece değişken dağılımları yerine, seçilmiş bir kaç örneğin değişken uzayındaki lokasyonlarını da baz alabilmesidir. Model az sayıdaki örneği yapacağı öngörülere destek olarak seçebilmekte ve sınıf ayraçlarını bu örnekler aracılığı ile tanımlayabilmektedir. Bu modelde değişkenler değil örnekler (“kural bazlı” yerine “örnek bazlı” modelleme) ön plana çıkabilmektedir.

DVM’nin iki temel kullanım şekli bulunmaktadır. İlk ve basit kullanımı (Lineer Modelleme) Çoklu Ayraç Analizi’nin gelişmiş bir versiyonuna benzemektedir. Bu yöntemde, eldeki örneklemin iki grubu ayrılmasına yardım edebilecek yeterli sayıda örnek belirlenmekte (Şekil 4.4, “X” ile işaretlenen örnekler) ve bu örnekler aracılığı ile oluşturulan doğrusal bir hat ile grupların mümkün olduğunca başarılı şekilde ayrıştırılması amaçlanmaktadır.



Şekil 4.4: DVM Lineer Model

Şekil 4.4'te görülebileceği gibi DVM, A ve B gruplarına ait veriyi optimal şekilde ayırmaya çalışmakta, “X” işaretli olan örnekler “ayırma” işinde “destek” alınan örnekleri (vektörleri) işaret etmektedir. “O” işaretli olan örnekler destek vektörü olarak kullanılmayan veriyi, farklı renkler ise farklı gruplara ait örnekleri göstermektedir.

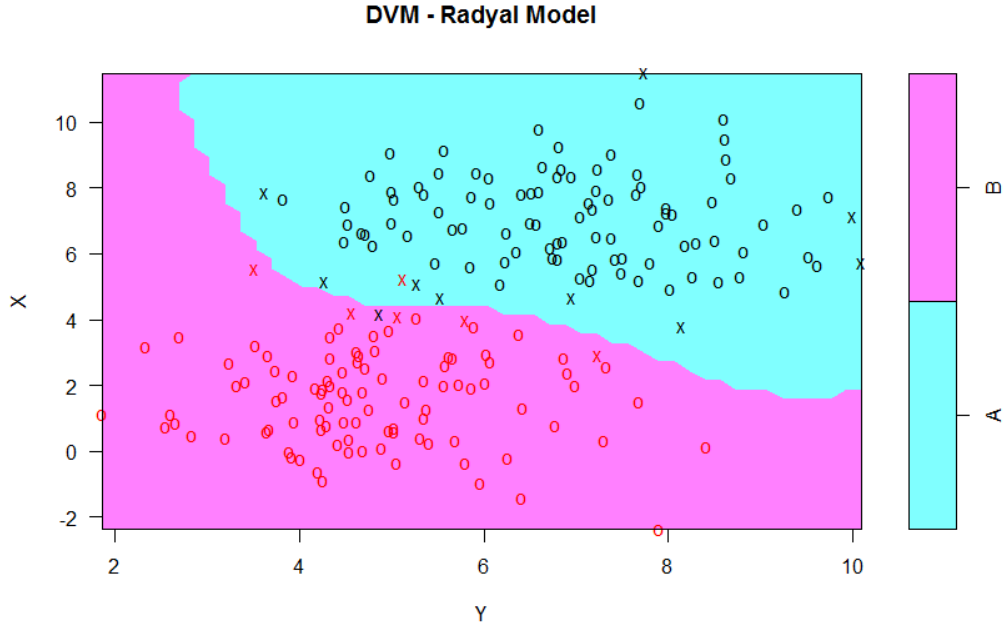
Model, destek vektörlerini kullanarak sınıf ayırımına karar vermekte ve bu ayırımı lineer bir hat aracılığı ile gerçekleştirmektedir. DVM, bu versiyonu ile Çoklu Ayraç Analizi'ne bir alternatif oluşturmaktadır.

DVM, lineer modelleme yanında, “mevcut örneklere yakınlık” yöntemiyle de sınıflandırma problemlerine çözüm üretebilmektedir. Bu seçenekte eldeki örneğin hangi grubun üyelerine daha yakın olduğu belirlenmekte ve öngörü buna göre gerçekleştirilmektedir.

Model, bu ikinci yöntemde, mevcut örneklem arasından bazı örnekleri (vektörleri) sınıflandırma gücünü artıracı “destekler” olarak belirlemekte ve bu örnekler için optimal bir etki alanı (çapı) tayin etmektedir. DVM, yeni bir örneğin sınıflandırılması gerektiğinde, önce bu örneğin tüm destek vektörlerine tek tek ne kadar yakın olduğunu (hangisinin etki alanına ne kadar girdiği) belirlemekte, daha sonra da bu hesaplamalar üzerinden ilgili örneğin sınıflandırılmasına yönelik bir öngörü üretmektedir. Seçilmiş örnekler (vektörler), karar “makinesini”

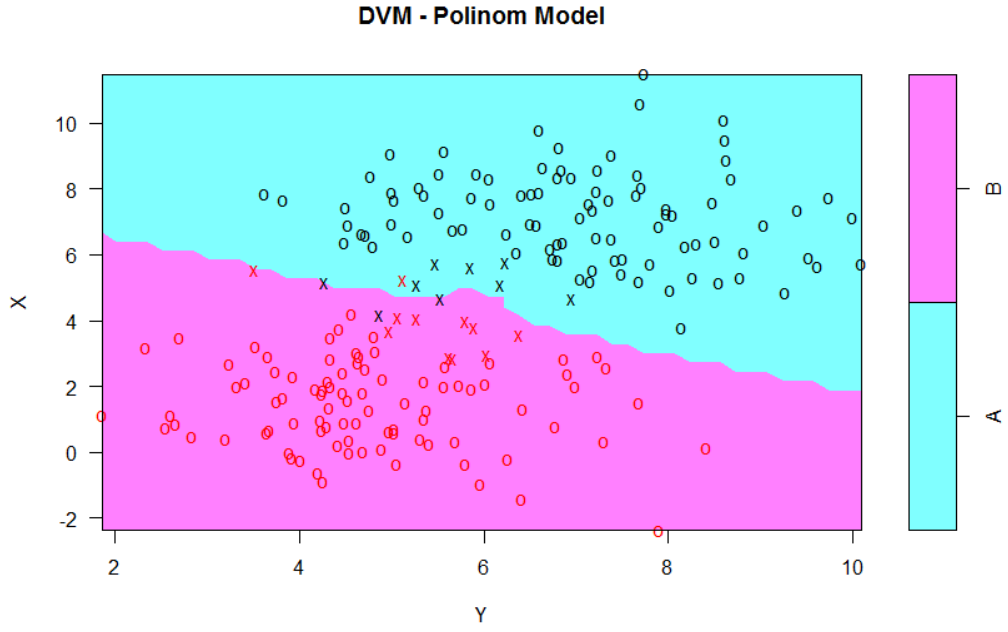
“desteklemek” için kullanılmaktadır (DVM).

Modelin belirlediği vektörlerin etki alanlarının tayin edilmesi esnasında farklı fonksiyonlar kullanılabilenkte, DVM bu fonksiyonların parametrelerini eldeki veri ile optimize edebilmektedir.



Şekil 4.5: DVM Radyal Model

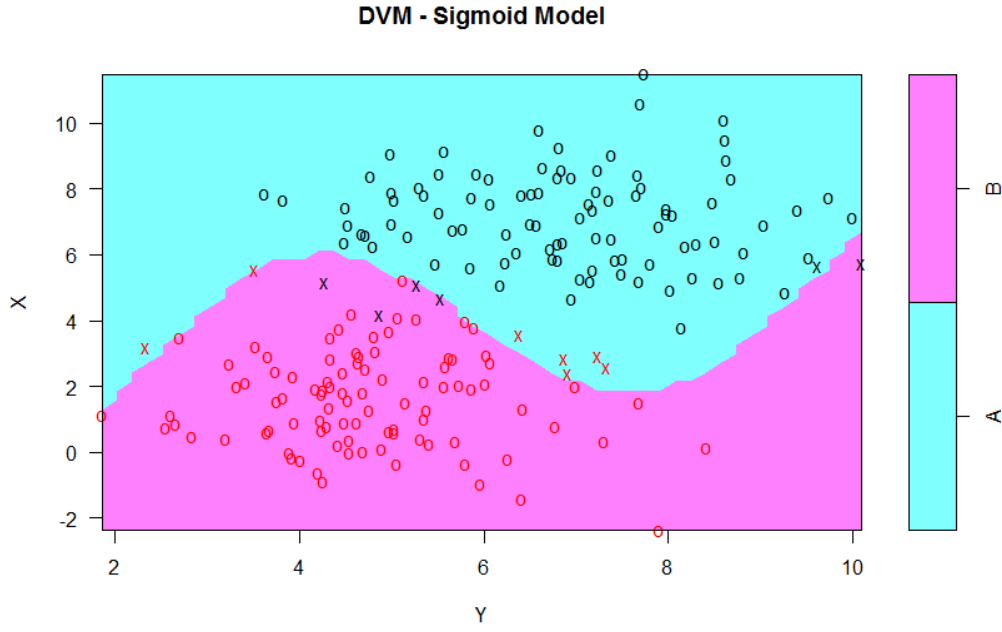
Şekil 4.5'te DVM Radyal modele bir örnek verilmekte ve model tarafından belirlenen destek vektörleri "X" ile işaretlenmektedir. Destek vektörlerinin her iki gruptan da belirlenebileceği görülmektedir (farklı renklerdeki X'ler). Şekil 4.5'ten de anlaşılacağı gibi, çok az sayıda destek vektörü aracılığı ile örneklemin başarılı bir şekilde doğru sınıflara ayrılması mümkün olmaktadır.



Şekil 4.6: DVM Polinom Model

Şekil 4.6’da DVM Polinom modele bir örnek verilmektedir. DVM Polinom model ile çok esnek etki alanı tanımları yapılabilmektedir. DVM Radyal modelde etki alanı sadece iki parametre ile tanımlanırken, DVM Polinom modelde parametre sayısı için bir sınırlama bulunmamaktadır. Şekil 4.6’daki örnekte başarılı bir polinom model uygulaması sergilenmektedir. Bununla beraber, etki alanı tayininde klasik uzaklık yerine başka bir ölçütün kullanılması modelin kavranmasını güçleştirebilmektedir.

Şekil 4.7’de DVM Sigmoid modele bir örnek verilmektedir. Bu modelde seçili vektörlerin etki alanlarının tayininde Logit fonksiyonundan yararlanılması amaçlanmaktadır. Şekil 4.7’den de anlaşılacağı gibi model başarısının önceki alternatifler kadar yüksek olmadığı anlaşılmaktadır. Farklı veri kümeleri için farklı modellerin daha başarılı olması doğal karşılanmalıdır. Eldeki örneklem çerçevesinde doğrusala yakın modeller daha başarılı sonuçlar üretmiş görünmektedir.



Şekil 4.7: DVM Sigmoid Model

DVM modellerin belki de en büyük gücü kendi değişkenlerini yaratabilmesidir. Sıradan bir Makine Öğrenmesi modeli (Sinir Ağları hariç) kullanıcının belirlediği değişkenler dışında yeni bir değişken ortaya koyamamaktadır. DVM bunu eldeki örneklerin bizzat kendilerini kullanarak gerçekleştirmektedir. Herhangi bir örneğe yakınlık yeni bir değişken olarak ortaya çıkabilmektedir.

Yukarıdaki DVM modellerinde örneklemin sadece X ve Y eksen bilgileri mevcuttur. DVM burada devreye girerek bazı veri noktalarını yeni eksenler olarak tayin etmektedir. Örneğin, Radyal yöntemde, seçili “vektörlere yakınlık” bilgisi yeni eksenleri/değişkenleri oluşturmaktadır. Bu sayede modelleme 2 yerine, örneğin, 11 destek vektörü vasıtası ile yapılmaktadır. Ayırma işlemi 2 boyuttan 11’e çıkarılmakta ve örneklemin daha başarılı bir şekilde sınıflandırılması mümkün olmaktadır.

4.5 Karar Ağaçları

Karar Ağaçları, aslen sınıflandırma araçları olmakla beraber, regresyon problemlerinde de kullanılabilir. Karar Ağaçları, yapıları gereği, hem kolay anlaşılabilir hem de kolay uygulanabilir modeller ortaya çıkarabilmektedir.

Karar Ağaçları, Özyinelemeli Bölümlendirme (Recursive Partitioning) modelleri ana başlığı altında yer almakta, budaklar (dal ayrımları) ve sınıflandırmanın belirlendiği yapraklardan (nihai duraklardan) oluşmaktadır.

Karar Ağacı modelinin oluşturulmasındaki ana amaç eldeki örnek kümesinin, herhangi bir değişken aracılığı ile ve hedef kriterlere uygun olarak, daha saf (tek bir gruba ait örneklerden oluşan) alt kümelere ayrılmasıdır. Her ayırım sonunda iki yeni alt küme oluşmaktadır. Her alt küme tekrar alt kümelere ayrılabilir. Proses yeni alt kümeler için de aynı şekilde yinelenmektedir. “Özyinelemeli Bölümlendirme” ismi de buradan kaynaklanmaktadır. Önceden belirlenmiş bir kriter gere göre alt kümelere ayırma döngüsü sonlandırılabilir. En küçük alt küme veya en fazla yaprak sayısı bu kriterlerden ikisidir.

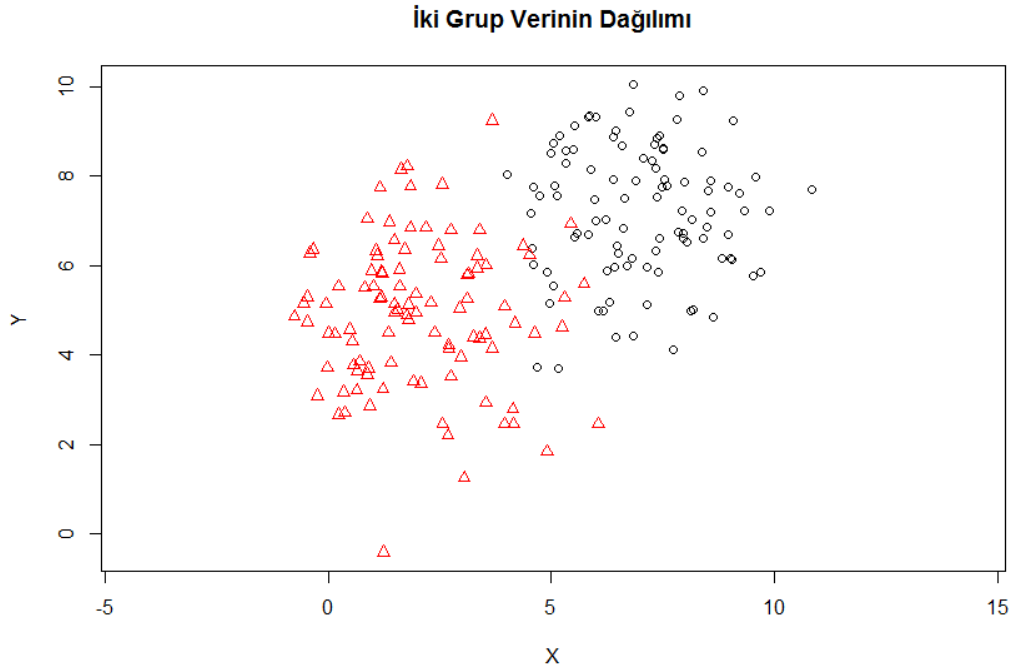
Budaklarda hangi değişkenin alt kümelere ayırma işleminden sorumlu olacağı farklı kriterlerle göre tespit edilebilir. Yaprak saflığı veya modelin öngörü başarısı bu kriterlerden bazılarıdır. Alt kümelere ayırma işini gerçekleştirecek değişkenin hangi seviyesinin kritik ayraç değeri olarak kullanılacağı da belirlenmesi gereken ayrı bir unsurdur. Bu seviyenin tespiti esnasında değişkenin aldığı tüm değerler, önceden belirlenen kriterler çerçevesinde, tek tek sorgulanmakta ve eşik değeri buna göre tespit edilmektedir. Örneğin; ana kriter yaprak saflığı olarak belirlenmiş ise eşik değerin belirlenmesinde yapraklardaki örneklerin saflığı ön plana çıkacaktır. Yaprak saflığı, yapraklara ulaşan örneklerle yönelik olarak gerçekleştirilen sınıflandırma öngörüsüne olan güveni belirleyecek, buradaki saflık azaldıkça modelin gerçekleştirdiği öngörülerin güvenilirliği azalacak, zıddı durumda ise artacaktır.

Karar Ağaçları'nın diğer Makine Öğrenmesi modellerine göre istatistiksel üstünlükleri bulunmaktadır. Örneğin; Karar Ağaçları'nın doğrusal bir model gereksinimi yoktur ve değişkenlerin normal bir dağılıma sahip olması da gerekmemektedir. Karar Ağaçları dışında, Makine Öğrenmesi modellerinin neredeyse tamamında, verinin önceden düzenlenmesi ve dağılımların, eğer gerekli ön şartlara uygunluk sağlanmamış ise, dönüştürülmesi gerekmektedir. Örnekleme oluşturan veriye uygulanan bu tip dönüşümler modellerin kavranmasını güçleştirebilir. Örnekleme dönüşüme uğramış modellerle öngöründe bulunabilmek için önce eldeki yeni örneklerin de aynı dönüşümlerden geçirilmesi, modelin öngörülerde bulunmasından sonra ise bu dönüşümlerin geriye çalıştırılması gerekmektedir. Bu durumdaki modellerin kullanımı daha zahmetli hale gelmektedir.

Karar Ağaçları'nda yaprakların tam olarak saflaştırılması her zaman mümkün olmamakta, modeli bu yönde zorlamak ise çok büyümesine ve aşırı uygun hale gelmesine sebep olabilmektedir. Bir Karar Ağacı'nın büyüklüğü bağımsız test

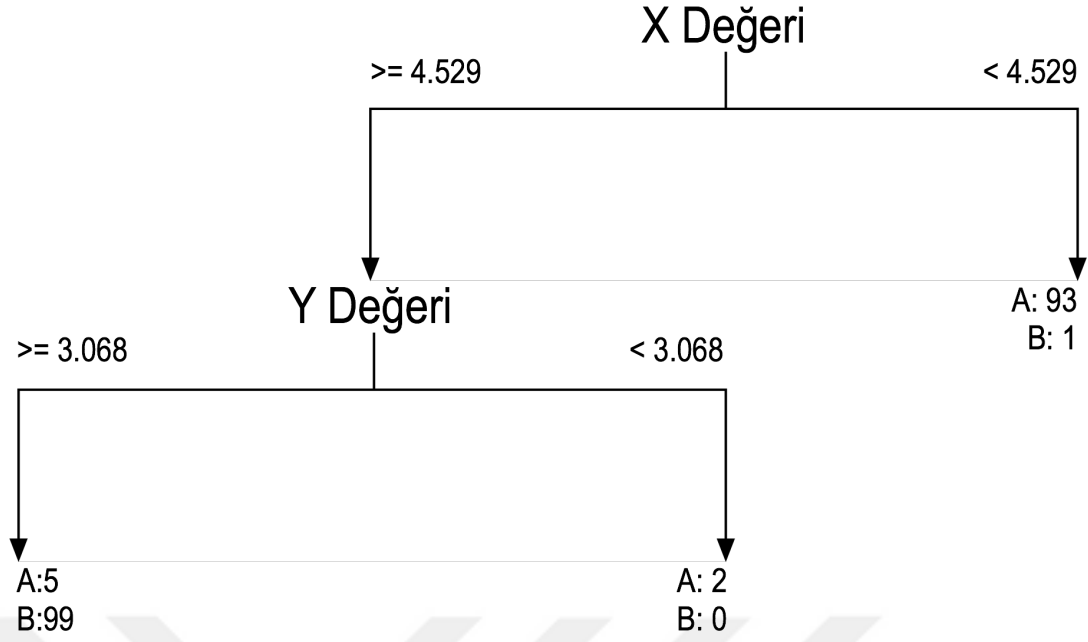
kümesindeki başarısına (Çapraz Doğrulama performansına) göre optimize edilebilmekte ve sonuca katkısı düşük olan dalları budanabilmektedir (modelden çıkarılma). Budama sonucunda elde kalan modelin pratikte (bağımsız veri üzerinde) daha başarılı sonuçlar üretmesi beklenmektedir.

Karar Ağacı modelinin çalışma prensibinin görsel bir sunumu faydalı olacaktır. Şekil 4.8’de “A (Üçgenler)” ve “B (Çemberler)” gruplarına ait örneklerin “X” ve “Y” düzlemindeki dağılımı sergilenmektedir.



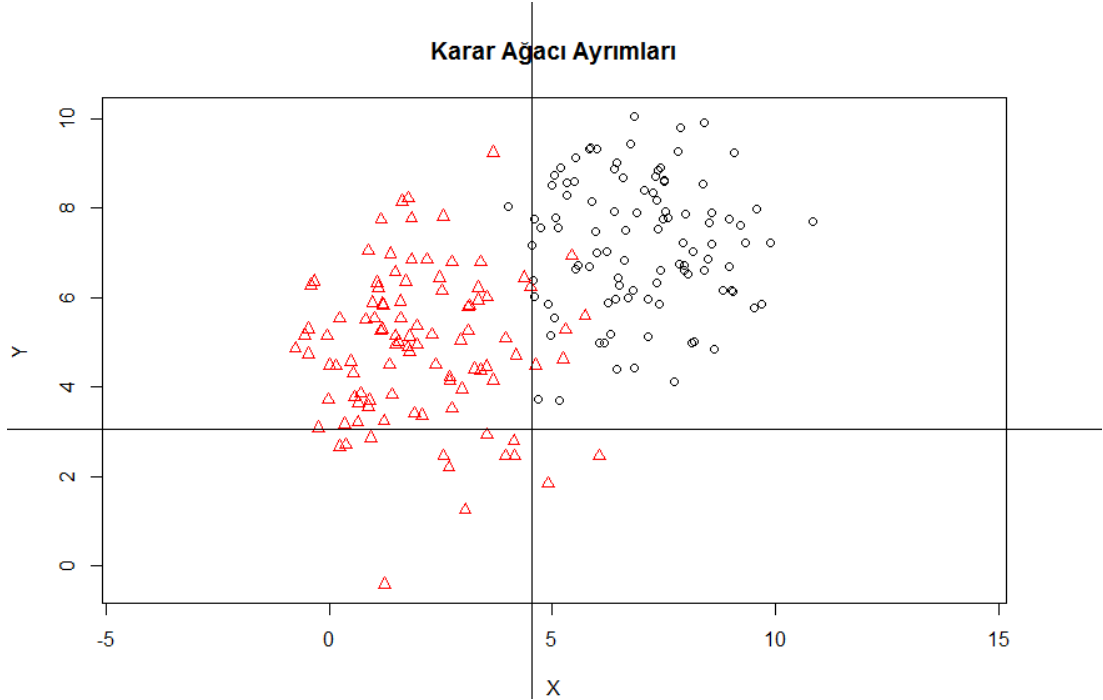
Şekil 4.8: İki Grup Verinin Dağılımı

Örnekleme, Karar Ağacı modeli ile gruplara ayrılmaya çalışıldığında iki temel ayracın ortaya çıktığı görülmektedir (Şekil 4.9). İlk ayraç X değişkeni üzerinedir ve herhangi bir örneğin X değeri 4.529’dan küçük ise A grubuna ait olduğu öngörülmektedir. Bu ayraç tek başına 94 örneği sınıflarken sadece 1 örnekte hata yapmaktadır (en sağdaki yaprak).



Şekil 4.9: Karar Ağacı

Karar Ağacı (Şekil 4.9) ilk ayracın belirleyemediği örnekler için (X değeri 4.529'a eşit ve daha büyük olanlar için) ikinci bir ayraç daha tayin etmekte ve bu da Y eksenine dayanmaktadır. İlk ayracın sınıflandıramadığı örnekler arasından Y değeri 3.068'den küçük olanların da A grubuna ait olduğu (model bu durumdaki her iki örneği de doğru sınıflandırmaktadır), geri kalanların ise B grubuna ait olduğu öngörülmektedir (104 örnekte 5 hata).



Şekil 4.10: Karar Ağacı Ayrımları

Karar Ağacı tarafından tespit edilen iki ayraç Şekil 4.10’da örnek dağılımı üzerine yerleştirilmiştir. X eksenini dik kesen çizgi birinci, Y eksenini dik kesen çizgi ise ikinci ayracı temsil etmektedir.

Çizelge 4.1’de, eldeki Karar Ağacı modelinin örnekler için öngördüğü sınıflandırma ve gerçek durum karşılaştırılmaktadır. Çizelgedeki satırlar model öngörüsünü, sütunlar ise gerçek durumu yansıtmaktadır. Örtüşen alanlar (A sütunu ve A satırı, B sütunu ve B satırı) modelin öngörü kabiliyetini, diğer alanlar ise modelin yanlış sınıflandırdığı örnekleri işaret etmektedir. Bu çizelge modelin eğitilmesinde kullanılan veri ile oluşturulmakta ve testlerde bağımsız veri kullanılmamaktadır. Eldeki çizelge, daha önce tanımlandığı üzere, bir Model İçi Performans göstergesidir. Model İçi Performans çerçevesinde bakıldığında, bu çok basit iki ayraç, %97’lik bir ekseriyetle (Çizelge 4.1: %49 + %48) iki grubu doğru şekilde ayırmayı başarabilmektedir.

Çizelge 4.1: Karar Ağacı Modeli – Model İçi Performans

		Gerçek	
		A	B
Öngörü	A	%49	%2
	B	%1	%48

4.5.1 Breiman Ağacı

Breiman Ağacı, Breiman ve diğ.’nin (1984) Makine Öğrenmesi’ne kazandırdığı bir Karar Ağacı türüdür. Model doğrusal olmayan regresyon problemlerinde de kullanılabilir, ancak asıl kullanım alanı sınıflandırma problemleridir.

Breiman Ağacı’nın oluşumu esnasında budaklarda kullanılacak değişkenlerin tespiti için eldeki tüm değişkenler ve bu değişkenlerin her seviyesi, belirlenmiş hedef kriter çerçevesinde, test edilmektedir (“greedy search”). Bu testlerde en sık başvurulan kriterlere örnek olarak yaprak saflığı ve doğru sınıflandırma başarısı gösterilebilir. Yapılan testler sonucunda hem seçilen kriteri en çok geliştiren değişken tayin edilmekte, hem de bu değişkenin ayraç olarak kullanılacak kritik seviyesi belirlenmiş olmaktadır. Dallara ayırımı belirleyen bu yöntem, bazı durumlarda, yanlışlığa sebep olmakta, daha fazla seviyeye sahip olan değişkenler, diğerlerine nazaran, daha sık seçilebilmektedirler. Bu durum, sonuca nihai etkisi daha yüksek olabilecek

değişkenlerin, sadece daha az seviyeye bölünmüş olmalarından ötürü, modelden dışlanmaları ile sonuçlanabilmektedir.

Breiman Ağaçları Aşırı Uygunluk problemiyle iki şekilde mücadele edebilmektedir. Bunlardan ilkinde “budama” yöntemi kullanılmaktadır. Bu yöntemde ilk önce eldeki örnekleme “model” ve “test” kümeleri olarak ikiye ayrılmaktadır. Daha sonra model kümesi kullanılarak aşırı büyük bir Breiman Ağacı (tamamen saf yapraklara sahip) oluşturulmaktadır. En son aşamada ise test kümesi kullanılarak Çapraz Doğrulama yoluna gidilmekte ve görece sonuca katkısı düşük olan dallar tek tek budanmaktadır (budakların ve bağlı yaprakların modelden çıkarılması). Bu sayede aşırı büyük Breiman Ağaçları’ı “makul boya” indirgenmekte ve Aşırı Uygunluk sorunu giderilebilmektedir.

Aşırı Uygunluk problemine çözüm olarak önerilen ikinci yöntemde ise Breiman Ağaçları’nın budanmaksızın Çapraz Doğrulama performanslarının yükseltilmesi amaçlanmaktadır. Bu yöntemde, daha Breiman Ağacı dallara ayrılırken, ne kadar karmaşıklaştığı ve bu karmaşıklığa karşın Çapraz Doğrulama performansına ne kadar fayda sağlandığı ölçülebilmektedir. Karar Ağaçları’nın büyüklüğü, ağaç oluşumu sırasında belirlenen Karmaşıklık Parametresi (cp - complexity parameter) ile sınırlandırılabilir ve bunun sonucunda Karar Ağaçları’nın, daha oluşturulurken, aşırı uygun hale gelmeleri engellenebilmektedir.

4.5.2 Hothorn Ağacı

Hothorn Ağacı, Hothorn ve diğ.’nin (2006) görece yakın bir dönemde Makine Öğrenmesi’ne kazandırdığı, Breiman Ağacı’na benzeyen ama teorik alt yapısı gereği daha nadir Aşırı Uygunluk sorunu yaşaması öngörülen bir Karar Ağacı alternatifidir.

Breiman Ağaçları’nda değişken seçiminde yaşanan yanlılık burada büyük ölçüde ortadan kaldırılmaktadır. Değişken seçimi önceden belirlenen hedef gruplar ile arasındaki korelasyon seviyesine göre yapılmakta, değişkenin kaç farklı seviyesi olduğu bu karara etki etmemektedir. Bu durum, Breiman Ağaçları ile en büyük farkı oluşturmaktadır.

Hothorn Ağaçları’nda dallara ayırma kararı, muhtemel ayırmadan sonraki yaprak saflığının istatistiki olarak “kayda değer” miktarda artırılabilmesine bağlı olarak verilmektedir. Model pek çok istatistiki parametreyi kendi içinde belirlemekte ve kullanıcı müdahalesine nadiren ihtiyaç duymaktadır. Bu açıdan da Hothorn Ağaçları

Breiman Ağaçları'na kıyasla pratikte uygulanması daha kolay modeller olarak ortaya çıkabilmektedir. Tüm bu artılarına rağmen, hangi Karar Ağacı modelinin daha başarılı sonuçlar üretebileceği örnekleme sıkı sıkıya bağlı kalmakta, bu konu üzerine önceden bir fikir yürütülmesi mümkün olamamaktadır.

4.6 Karar Ormanları

Karar Ormanları Breiman'ın (2001) Makine Öğrenmesi'ne kazandırdığı bir sınıflandırma aracıdır ve Karar Ağaçları'nın bir araya getirilmesinden oluşmaktadır. Karar Ormanları'ndaki her Karar Ağacı, eldeki örneklemin ve değişkenlerin rastgele seçilmiş alt bir kümesi ile modellenmekte ("bagging") ve bunun "baskın değişken" etkisini azaltabileceği ileri sürülmektedir (Breiman, 2001:5). Rastgele üretilmiş Karar Ağaçları'ndan oluşan bu orman ("Random Forest") ile yapılan sınıflandırma, ormandaki Karar Ağaçları'nın tamamının oyları ile gerçekleştirilmektedir. Karar Ormanı'ndaki Karar Ağaçları, sorgulanan her örnek için, kendilerine en uygun sınıflandırma öngörülerini üretmekte, bu öngörüler tek elde derlenmekte ve sınıflandırma kararı oy çokluğu ile alınmaktadır.

Karar Ağaçları oluşturulurken bir kaç baskın değişkenin ortaya çıkması doğal olarak kabul edilebilse de bu durum modele giremeyen diğer değişkenlerin tamamen faydasız olduğu anlamına da gelmemektedir. Modele dahil olamayan değişkenlerin farklı bir örneklem çerçevesinde veya farklı bir değişken kümesinde ön plana çıkabilmeleri mümkündür. Karar Ormanları tam da bu noktaya parmak basmakta, eldeki veriyi ve değişkenleri rastgele örnekleyerek, diğer değişkenlerin de muhtemel "önemini" ortaya çıkartmayı amaçlamaktadır. Karar Ormanları oluşturulurken Karar Ağacı modellemesinde olmayan iki şey yapılmakta; model oluşumu sırasında eldeki verinin sadece bir kısmı kullanılmakta ve dal ayırımına karar verilirken değişkenlerin sadece bir alt kümesinden faydalanılmaktadır. Bu prosedür hem örnek kümesi, hem de baskın değişken yanlılığını hafifletebilmektedir (Breiman, 2001).

Karar Ormanları'nın çok önemli iki yan ürünü bulunmaktadır. Bunlar; örneklerin birbiriyle yakınlığının (benzerliğinin) keşfi ve değişkenlerin sonuca katkısının (değişken öneminin) ortaya çıkarılmasıdır (Breiman, 2001:5). Eldeki örneklerin birbirine yakınlığı, Karar Ormanı'ndaki kaç ağaçta aynı yaprakta buluşmaları ile ölçülebilmektedir. İki örnek sıklıkla aynı yaprakta buluşuyorsa, bu örnekler birbirine "yakın", eğer bunun zıddı ise birbirinden "uzak" olduğu kabul edilmektedir.

Yakınlık/uzaklık analizinin bir sonucu da “aykırı” örneklerin (“outliers”) keşfidir. Eğer bir örnek başka hiçbir örneğe yeteri kadar yakın değil ise, “aykırı” bir örnek olduğu kabul edilebilmektedir. Aynı analizinin diğer bir faydası ise örneklemin gruplara ayrılmasında ortaya çıkmaktadır. Analiz sonucunda, yakın olarak tespit edilen örneklerin benzer özellikler taşıması ve aynı büyük gruba ait olması öngörülmektedir.

Karar Ormanları'nın ikinci önemli yan ürünü değişkenlerin sonuca katkısının ortaya çıkarılmasıdır. Bu analiz için öncelikle ilgili değişken doğal hali ile modele dahil edilmekte ve modelin genel başarısı kaydedilmektedir. Daha sonra bu değişken yerine, aynı değişkenin rastgele sıralanmış verisi örnekleme atanmakta ve modelleme tekrarlanmaktadır. Son aşamada ise elde edilen iki modelin sonuçları kıyaslanmakta ve etkisi sorgulanan değişkenin model performansına ne kadar katkıda bulunduğu ortaya çıkarılmaktadır. Değişken öneminin belirlenmesi, daha önce de vurgulandığı gibi, bazı değişkenlerin modelden güvenle çıkarılabilesine yardım etmekte ve bu sayede Aşırı Uygunluk problemine de çözüm üretilebilmektedir. Bazı durumlarda, birbiriyle yüksek korelasyona sahip birden çok “önemli” değişkene rastlanması mümkündür ve idealde yapılması gereken bunlardan sadece birinin modelde tutulması ve diğerlerinin dışlanmasıdır. Yukarıda detayları sunulan yöntem bu soruna cevap verememektedir. Çözüm Hothorn Ormanları'ndan gelmekte, önemlilik analizi değişkenler arası korelasyonların da dikkate alınması ile gerçekleştirilmekte ve Breiman Ormanları'na göre daha faydalı sonuçlar elde edilebilmektedir (Strobl ve diğ., 2009:1).

Karar Ormanları çok sayıda Karar Ağacı'nın bir araya gelmesinden oluştuğu için kullanıcı tarafından takibi zorluklar içermekte, modelin sınıflandırma öngörüsüne nasıl ulaştığını kavramak kolay olmamaktadır. Modelin pratik uygulamalarında da zorluklar bulunmakta, modele dahil yüzlerce (bazı durumlarda binlerce) Karar Ağacı'nın hatasız olarak sistemleştirilmesi büyük dikkat gerektirmektedir.

4.6.1 Breiman Ormanı

Breiman Ormanları, adından da anlaşılacağı gibi, Breiman Ağaçları'ndan oluşmaktadır. Model, yan ürünler olarak, değişken önemi, örneklerin birbirine benzerliği, ayırık örnekler gibi çok değerli çıktıları da kullanıcıya sağlamaktadır. Doğrusal olmayan, son derece karmaşık ilişkileri modelleyebilmektedir (Breiman, 2001:5). Modelin sağlıklı eğitimi için veri kümesinin mümkün olduğunca derin

olması gerekmektedir.

4.6.2 Hothorn Ormanı

Hothorn Ormanları, Hothorn Ağaçları'ndan oluşmaktadır ve Breiman Ormanları'ndan en önemli farkı da budur. Bir diğer önemli fark ise, Breiman Ormanları'nda mevcut olmayan, değişken önemi analizinin değişkenler arası korelasyonların da dikkate alınması ile gerçekleştirilebilmesidir. Bu da, yüksek korelasyona sahip değişkenlerin mevcudiyeti halinde, daha zayıf değişkenlerin tespitini mümkün kılmakta, modeldeki değişken sayısı azaltılabilmekte ve Çapraz Doğrulama performansı bundan olumlu etkilenebilmektedir.



5. HALKA AÇIK FİNANS DIŐI ŐİRKETLERDE SÜREKLİLİK RİSKİNİN MAKİNE ÖĞRENMESİ İLE ÖNGÖRÜLMESİ

Bu bölümde araştırmanın amacı, önemi, yöntemi, kısıtları, ana kütesi, seçilen örnekleme, bulguları ve sonuçları aktarılmaktadır.

5.1 Araştırmanın Amacı

Bu araştırmanın birincil amacı halka açık finans dışı şirketlerin süreklilik risklerinin çok deęişkenli Makine Öğrenmesi modelleri ile öngörülmesidir. Araştırmaya konu Makine Öğrenmesi modellerinin performans, ön şartlar ve uygulanabilirlik açısından kıyaslanması ve ilgili modellerden benzer araştırmalarda hangi nispette faydalanılabileceğinin öngörülmesi ise ikincil amacı oluşturmaktadır.

Teknik gelişmelere paralel olarak süreklilik riski öngörüsü modellerinin de güncellenmesi gereklilięi açıktır. Son on yıl içinde öne çıkan bazı istatistiki modeller (Karar Ağaçları örneğinde olduęu gibi) bu konuda oldukça ümit vericidir ve doğrusal olmayan ilişkilerin kolaylıkla modellenebilmesi yanında, verinin istatistiki dağılımı üzerindeki kısıtlamalar da hafifletebilmektedir. Bu modellerin ekseriyeti Makine Öğrenmesi (Machine Learning) sınıfına girmekte ve eldeki veriyi kullanarak, hem modeli hem de modelin parametrelerini kendi içinde belirleyebilmektedir.

Araştırmanın ikincil amacına yönelik olarak Makine Öğrenmesi başlığı altında yer alan ve son on yıllık dönemde popülerlik kazanan başlıca modeller ve görece eski ancak oldukça yaygın kullanıma sahip iki model bu çalışmaya dâhil edilmektedir. Çalışma boyunca bu modellerin kullanımına yönelik olarak dikkat edilmesi gereken noktalar yeri geldikçe vurgulanmakta ve sadece model performanslarının sunulması ile yetinilmemektedir.

Araştırmada kullanılan modellerin kendilerine has kuvvetli ve zayıf yönleri bulunmaktadır. İdeal modeller kolay anlaşılır, anlatılır, eğitilebilir ve sağlanması gereken ön şartlar bakımından kısıtlayıcı olmayan seçeneklerdir. Karar Ağaçları bu tip modellere iyi örneklerdendir, ancak her model ideal özelliklerin tamamını barındırmasa da sahip oldukları farklı avantajlar onları yine de ön plana çıkarabilmektedir. Örneğın, Sinir Ağları verinin içindeki bileşik ilişkileri ve yapıları

keşfedebilirken, Destek Vektör Makineleri kısıtlı veri ile dahi başarılı sonuçlar üretebilmektedir. Modellerin birbirine karşı farklı üstünlükleri olabilmekte, bu nedenle de herhangi bir modelin tamamen göz ardı edilmesi mümkün olmamaktadır.

5.2 Araştırmanın Önemi

Süreklilik riskinin öngörülmesi; fonlama maliyetlerinin belirlenmesi, muhtemel batıklar için karşılık ayrılması, kaynakların doğru yatırımlara yönlendirilmesi ve kurumun bekası açısından hayati bir süreci temsil etmektedir. Bu alana yapılabilecek her türlü katkı değerlidir. Kaynakların doğru şekilde kullanılması mikro seviyede ilgili şirketleri finansal açıdan sağlamlaştırırken, makro açıdan tüm endüstrinin daha verimli çalışmasını ve genel anlamda ekonomik büyümeyi destekleyecektir.

Sağlıklı bir süreklilik riski öngörüsü modeli muhasebe bilgi kullanıcıları (yöneticiler, ortaklar, kredi kuruluşları, çalışanlar, potansiyel ortaklar, devlet ve toplum) açısından önem taşımaktadır. Bu model, bağımsız denetimlerde süreklilik riskinin ortaya çıkarılmasında, yatırım kararlarında finansal riskin ve elde edilmesi gereken minimum verimin tespitinde, işletmede yüksek finansal risk oluştuğunda kurumsal tavrın belirlenmesinde, finansal riskin azaltılması gerektiğinde en doğru yolun seçilmesinde, portföy yönetiminde varlıkların risk açısından sınıflandırılmasında ve alacak portföylerinin kredi riski takibinde kullanılabilir.

5.3 Araştırmanın Yöntemi

Araştırmaya konu olan finans dışı halka açık şirketlerin süreklilik risklerinin tespitinde, tezin dördüncü bölümünde detayları verilmiş olan çok değişkenli Makine Öğrenmesi modellerinden ÇAA, Lojistik Model, Sinir Ağları, DVM, Breiman Ağacı, Hothorn Ağacı, Breiman Ormanı ve Hothorn Ormanı olmak üzere sekiz farklı model eldeki veriye uygulanmaktadır. Söz konusu modellerin veriye uygulanması R istatistik yazılımı ve bu modellere has R kütüphaneleri aracılığı ile gerçekleştirilmektedir.

5.3.1 Verinin hazırlanması

Çalışmaya dahil modellerin nihai performanslarının adil bir şekilde kıyaslanabilmesi, bu modellerin şart koştuğu tüm sınırlamalara uyulması ile mümkündür. Çok değişkenli doğrusal modeller ve Sinir Ağı modelleri bu duruma iyi örneklerdir. Bu modellerin sağlaması gereken bir kaç ön koşul birden vardır. Bu ön koşullardan biri,

belki de en önemlisi, modele dahil değişkenlerin normal (Gauss dağılımı) veya log-normal (logaritması Gauss dağılımı sergileyen) dağılım sergilemesidir. Eldeki veri eğer doğal hali ile bu koşulu sağlamıyor ise, dönüşümler aracılığı ile yeniden düzenlenmesi gerekebilmektedir. Bu dönüşümlerin yapılmaması nihai model başarısını olumsuz etkileyebilecektir.

Modellemede kullanılan verinin sadece dağılım ön koşullarını sağlaması her zaman yeterli olmamaktadır. Model optimizasyon prosedürlerinin pek çoğu değişken yayılımlarının (değişkenin aldığı en büyük ve en küçük değerler arası uzaklık) da birbirine yakınsamasını gerektirmektedir. Veriye dahil değişkenlerin yayılımları birbirine yakınsamaz ise en doğru model parametrelerinin keşfi her zaman mümkün olmayabilmekte ve nihai model performansı da bundan olumsuz etkilenebilmektedir. Sinir Ağı modeli bu duruma iyi bir örnektir. Bu modelin sağlıklı optimizasyonu için hem değişken dağılımlarının normal, hem de yayılımlarının benzer olması gerekmektedir.

Karar Ağacı ve Karar Ormanı modellerinin kullanacağı veri için ise neredeyse hiçbir ön hazırlık yapmak gerekmemektedir. Bu da toplam modelleme süresini kısaltabilmektedir. Model verisinin herhangi bir dönüşümden geçirilmemiş olması modelleme esnasında olup bitenin daha kolay kavranmasına yardımcı olabilecektir. Bunun zıddı durumlarda ise, örneğin Sinir Ağı modellerinde, bir değişkenin model tarafından belirlenmiş kritik seviyesini kavramak için öncelikle veriye uygulanmış olan tüm dönüşümlerin geri alınması ve tekrar anlaşılır hale getirilmesi gerekmektedir. Bu da modelin standart kullanıcılar tarafından kavranmasını zorlaştıran bir unsur olarak ortaya çıkabilmektedir.

Model verisine uygulanan muhtemel dönüşümlere bir örnek vermek durumun kavranmasına yardımcı olabilecektir. Normal dağılıma sahip olmayan ve de hem pozitif hem de negatif değerler alan bir değişken için yapılması gereken dönüşümler şu şekilde sıralanabilir: değişkeni tamamen pozitif değerler alacak şekilde kaydır, logaritmasını al, ortalamayı tüm değerlerden çıkar ve standart sapmasını kullanarak dağılımı standart normal dağılıma (ortalaması sıfır ve standart sapması bir olan Gauss dağılımı) yakınlaştır. Üstelik bu dönüşüm işlerin yolunda gittiği bir duruma örnektir ve pratikte çok daha fazla basamaklı dönüşümlere de ihtiyaç duyulabilmektedir.

5.3.2 Değişken sayısının azaltılması

Modellerde kullanılan değişken sayısının artırılması eldeki probleme yeni açılardan bakılabilmesini sağlarken, eklenen her değişken için en az bir parametrenin daha sağlıklı bir şekilde hesaplanmasını zorunlu kılmaktadır. Modelin belirlenmesi gereken parametre sayısı arttıkça ihtiyaç duyduğu veri sayısı da artmaktadır (Goldberger, 1991:249-250) ve başarılı bir model için bu sayının parametre başına yirmi civarında olması gerektiği öngörülmektedir. Modelleme esnasında eldeki veri tarafından sağlıklı bir şekilde belirlenemeyen parametrelerin bağımsız veri üzerindeki test sonuçlarını olumsuz yönde etkileyeceği ve bu durumda takip edilmesi gereken en doğru yolun parametre sayısının azaltılması olduğu savunulmaktadır (Platt ve Platt, 1990:31-32). Üstelik, model performansının kullanılan değişken adedine paralel olmadığı, başarılı süreklilik riski öngörüsü modellerinin iki ila yirmi bir arasında pek çok farklı sayıda değişkeni kullanabildiği de makalelerde yer almaktadır (Belowary ve diğ., 2007:1).

Mevcut çalışmada model eğitiminde kullanılacak örnek sayısının 106 olduğu göz önüne alındığında sağlıklı bir modelin en fazla altı civarında parametreye sahip olması öngörülmektedir. Modellerde kullanılacak parametre sayısının sınırlı tutulması modele dahil edilebilecek değişken adedini de kısıtlamaktadır. Hangi değişkenlerin modelde tutulacağı, hangilerinin dışlanacağı ise bu çalışma kapsamında iki yöntemle belirlenmeye çalışılmaktadır. Bunlardan ilki eldeki değişkenlerin sınıflandırma başarısına yaptığı genel katkının ortaya çıkarılması, ikincisi ise “önemli değişkenlerin” bizzat model tarafından seçilmesine izin verilmesidir.

Değişkenlerin sınıflandırma başarısına yapabileceği genel katkı, bölüm 4.6’da detaylı bir şekilde açıklandığı gibi, Karar Ormanları tarafından gerçekleştirilebilmektedir. Çalışmada bu yöntem ile toplam değişken sayısı, ilk aşamada, on ikiden sekize indirilmekte ve en zayıf etkiye sahip dört değişken modellerde kullanılacak veriden çıkarılmaktadır. Eldeki toplam veri sayısı sekiz değişken için de yeterli görünmemekte, ancak bilahare gerçekleştirilebilecek elemeler modellerin kendi iç süreçlerine terk edilmektedir. Takip edilen bu uygulama bazı açılardan hibrit modelleri anımsatmaktadır. Hibrit modellerde analize dahil olabilecek değişkenler başka bir model ile belirlenmekte ancak nihai modelleme tamamen farklı bir Makine Öğrenmesi modeli ile gerçekleştirilmektedir. Bu tip hibrit modellerin başarı ile

uygulandığı yakın tarihli çalışmalarda savunulmaktadır (Lee ve diğ., 1996, Jandaghi ve diğ., 2011, Shah, 2014).

Modellerde kullanılabilirlik değişken sayısına yönelik yapılan tüm bu ön hazırlığın, modellerin Çapraz Doğrulama performanslarına olumlu yönde katkı yapması öngörülmektedir.

Çalışmanın ilk aşamasında kullanılması öngörülen finansal oranlar Çizelge 5.1’de listelenmiştir:

Çizelge 5.1: Çalışmada Kullanılması Planlanan Finansal Oranlar

Finansal Oranlar	Finansal Oranlar
Borçlar/Aktifler	Nakit Akımı/Aktifler
Cari Oran	Nakit Akımı/Borçlar
Nakit/Aktifler	FAVÖK/Aktifler
Aktif Karlılığı	FAVÖK/Borçlar
Çalışma Sermayesi/Aktifler	Nakit/Borçlar
Öz Kaynak/Borçlar	Nakit/Kısa Vadeli (KV) Borçlar

Bu listenin oluşturulmasında benzer çalışmalarda sıklıkla kullanılan göstergeler (Dimitras ve diğ., 1996, Aziz ve Dar, 2006, Belowary ve diğ., 2007) dikkate alınmakta, ortaya çıkan listenin on ikiye indirilmesi esnasında ise tüm kaldıraç etkilerinin izlenebilmesi ve ilgili finansal oranların “kararlı” olmasına dikkat edilmektedir. Bu çalışma çerçevesinde, bir finansal oranın böleni eğer çok fazla salınım göstermiyor ise bu oran “kararlı”, aksi durumda ise “kararsız” bir gösterge olarak sınıflandırılmaktadır. Örneğin, Fiyat/Kazanç oranı oldukça kararsız bir finansal göstergedir. Bu oran, kazanç eğer sıfır civarında salınıyor ise çok büyük ve pozitif değerlerden, çok büyük ve negatif değerlere doğru hızlı geçişler yapabilmektedir. Bu da modellerin eğitilmesini zorlaştıran bir unsur olarak ortaya çıkabilmektedir.

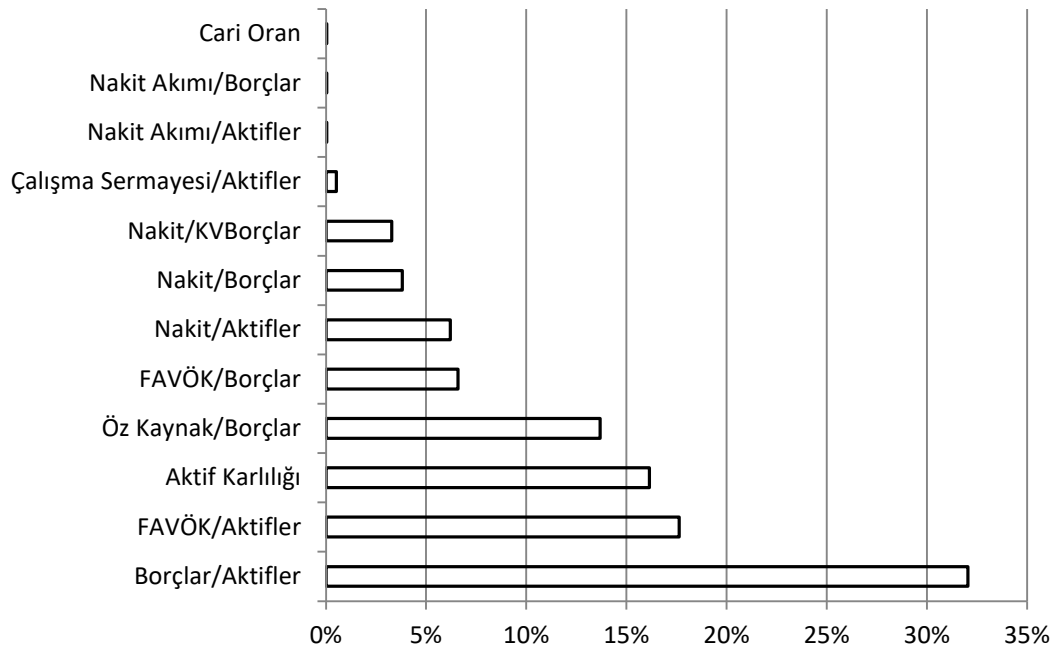
Sektörler arasında ciddi farklılıklar gösteren veriye dayalı finansal oranlar da bu listenin dışında tutulmaktadır. Örneğin; satışların büyüklüğü (toplam varlıklara nispetle) sektörler arasında ciddi farklılıklar gösterebilmektedir, bu durumda satışlara dayalı bir finansal oranın kullanılabilmesi ancak sektör odaklı bir modelde mümkün olabilecektir. Elimizdeki örneklemin sektörlere pay edilmesi durumunda ise hiçbir sektör için on adetten fazla örneğin bulunmadığı görülmektedir.

Makroekonomik deęişkenler de sektör duyarlılığı yüksek deęişkenlere benzer şekilde hareket etmekte ve bu nedenle çalışmanın dışında bırakılmaktadırlar. Her sektör makroekonomik deęişkenlerden aynı oranlarda etkilenmemektedir. Örneğin; perakende sektörü makroekonomik gelişmelere daha duyarsızken otomotiv sektörü son derece duyarlı olabilmektedir. Reel faiz oranlarının modellere dahil edilmemesi ise en az bir yıllık gecikme ile gerçek seviyesinin belirlenebilmesinden kaynaklanmaktadır. Bu çalışma çerçevesinde süreklilik riski öngörüsünün bir yıl öncesinden yapılması amaçlanmakta, bu da bir yıl gecikmeli bir deęişkenin kullanımını boşa çıkarmaktadır.

Çalışmada kullanılması planlanan öncül deęişken listesi (Çizelge 5.1) belirlendikten sonra, bu bölümün girişinde belirtilen sebeplerden ötürü, toplam deęişken sayısının azaltılması ve bunlar arasından sonuca katkısı düşük olanların ayıklanması yönüne gidilmektedir. Hangi deęişkenlerin sonuca katkısı yüksek, hangileri görece önemsiz sorgusu Karar Ormanları tarafından gerçekleştirilebilmektedir (bölüm 4.6). Hothorn Ormanları, deęişken önemi tayininde korelasyonları da dikkate alabildiği için (Şartlı Önemlilik) bir adım öne çıkabilmektedir (Strobl ve dię., 2009:1-2).

Şekil 5.1’de, eldeki sınıflandırma problemi çerçevesinde, deęişkenlerin sonuca korelasyon etkilerinden arındırılmış katkıları sunulmaktadır.

Deęişkenlerin Sonuca "Bağımsız" Katkısı



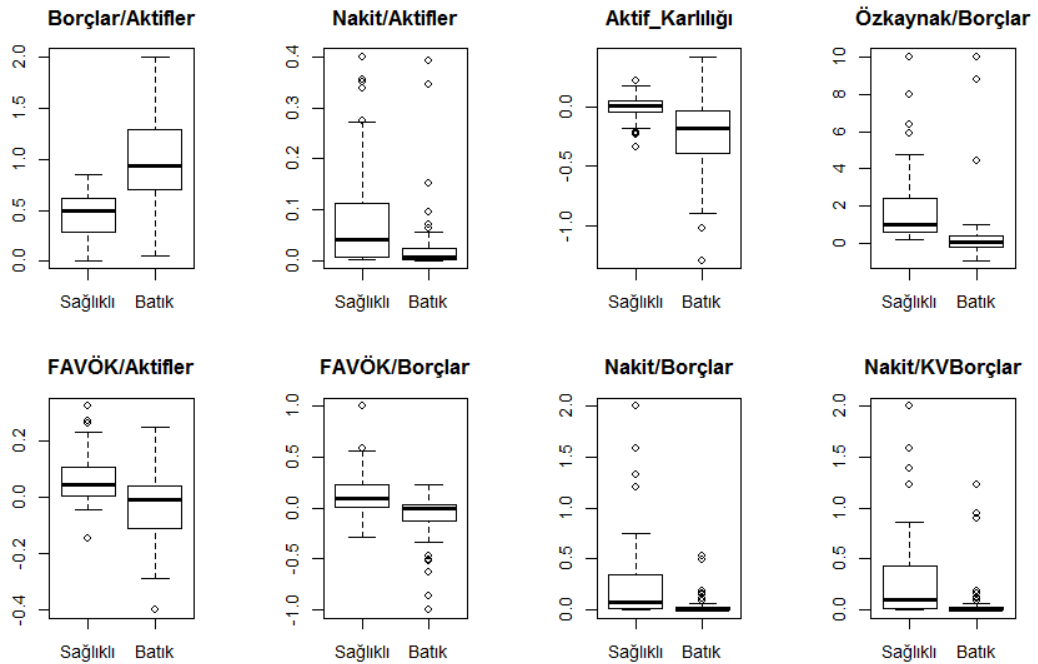
Şekil 5.1: Deęişkenlerin Sonuca "Bağımsız" Katkısı

Şekil 5.1 incelendiğinde, en tepedeki dört değişkenin sonuca toplam katkısının %2’den daha az olduğu anlaşılmakta ve bundan dolayı modellerde kullanılabilen değişken sayısının 12’den sekize güvenli bir şekilde indirilebileceği öngörülmektedir. Çalışmanın bundan sonraki bölümünde “Cari Oran”, “Nakit Akımı/Borçlar”, “Nakit Akımı/Aktifler” ve “Çalışma Sermayesi/Aktifler” modellerde faydalanılabilecek değişkenlerin dışında bırakılmaktadır. Geri kalan değişkenlerin (Çizelge 5.2) sonuca toplam katkısı %98’in üzerindedir.

Çizelge 5.2: Model Eğitimlerinde Kullanılacak “Özet” Finansal Oranlar

Finansal Oran	Finansal Oran
Borçlar/Aktifler	FAVÖK/Aktifler
Nakit/Aktifler	FAVÖK/Borçlar
Aktif Karlılığı	Nakit/Borçlar
Öz Kaynak/Borçlar	Nakit/KV Borçlar

Çizelge 5.2’de listelendiği şekliyle çalışmanın bundan sonraki kısmında kullanılabilen değişkenler belirlendikten sonra, bu değişkenlerin eldeki örneklem çerçevesinde batık ve sağlıklı şirketler arasında nasıl dağıldığı Şekil 5.2’de sergilenmektedir. Dağılımlar arası farkların niceliği ilgili finansal oranların sonuca muhtemel katkısı hakkında ipucu verebilmektedir.



Şekil 5.2: Değişkenlerin Hedef Gruplar Bazında Dağılımı

5.3.3 Model oluřturma ve seme

alıřma erevesinde veriye uygulanan ok deęiřkenli Makine ęrenmesi modellerinden bazıları kendi ilerinde de ok sayıda alternatifte sahip olabilmektedir. Bu durumda ilgili modellerin farklı parametreler kullanılarak tekrar oluřturulmaları ve bunların arasından en iyisinin seilmesi gerekebilmektedir. Buna en iyi rnek Sinir Aęları'dır. alıřmada pek ok farklı Sinir Aęı modelinin denenmesi bunun neticesindedir.

Lojistik Modeller de bu alıřmada iki farklı alternatifini ile ele alınmaktadır. Bařvurulan her iki Lojistik Model alternatifini de yapılan arařtırmaya uygundur. alıřmada kullanılan dięer modeller ise bu tip bir eřitlilięe sahip deęildir.

5.3.4 apraz Doęrulama

alıřmada, Model İi Performans Karıřıklık Matrisi ile lölürken, modelin baęımsız veri üzerindeki muhtemel performansı (nihai bařarı kriteri) apraz Doęrulama yöntemi ile tespit edilmektedir. apraz Doęrulama yöntemi, modelleme ařamasında yařanabilecek Ařırı Uygunluk sorunlarının tespitini, bu sorunların özölmesini ve farklı modellerin benzer řartlar altında performans aısından kıyaslanabilmesini mümkün kılmaktadır.

apraz Doęrulama yönteminde ilk ařamada rneklem rastgele iki alt kümeye ayrılmakta, daha sonra bu alt kümelerden biriyle (model alt kümesi) model oluřturulurken dięeriyle de (test alt kümesi) eldeki modelin baęımsız veri üzerindeki performansı lölümlenmektedir. Aynı model iin apraz Doęrulama yöntemi defalarca tekrarlanabilmekte ve modelin nihai bařarısı hakkında oldukça kesin bir kaniya varılabilmektedir.

Dięer modellerde binlerce kez "rastgele" tekrarlanabilen apraz Doęrulama testlerinin Sinir Aęları'na uygulanması mümkün olamamakta (iřlem yoğunluęu ve veri etkinlięi) ve Sinir Aęları'na özel bir apraz Doęrulama yönteminin tatbik edilmesi gerekmektedir. Sinir Aęlarına daha uygun olan bu yöntemde her test iin sadece iki rnek dıřarıda bırakılmakta ve geri kalan veri ile model oluřturulmaktadır. Yöntemin bir sonraki basamaęında ise dıřarıda kalan iki rnek ile modelin bařarısı test edilmektedir. Bu yöntem Sinir Aęları iin 5565 defa (eldeki rneklerin ikili kombinasyonları kadar) tekrarlanmakta ve ulařılan sonuçların istatistikini geerlilięi muhafaza edilebilmektedir.

5.3.5 Model eğitimi ve optimizasyonu rutini

Bu çalışma çerçevesinde model eğitimi ve optimizasyonunda aşağıdaki rutin takip edilmektedir:

- Değişkenlerin model ön koşullarına uygun hale getirilmesi,
- Eldeki tüm değişkenlerin kullanılması ve “önemi düşük” değişkenlerin model tarafından belirlenerek dışlanması,
- Geri kalan değişkenler ile farklı modellerin denenmesi,
- Model performanslarının Çapraz Doğrulama yöntemi ile öngörülmesi.

5.4 Araştırmanın Ana Kütlesi ve Örneklemi

Mevcut çalışmaya 1999-2016 döneminde Türkiye’de yerleşik, halka açık, Borsa İstanbul A.Ş.’de payları işlem gören, finans sektörü dışındaki şirketler dahil edilmektedir.

Ciddi süreklilik riski yaşayan şirketlerin tespitinde borsa kotundan çıkarılma asıl kriter olarak kabul edilmekte, ancak borsa kotundan çıkarılan her şirketin de süreklilik riski yaşamadığı anlaşılmaktadır. Bu tip şirketlerin önemli bir kısmı kendi isteği ile veya şirket birleşmeleri neticesinde kottan çıkarılmaktadır. Bazı şirketlerin ise devlet tarafından el konulan bankalar ile ortaklık ilişkilerinden dolayı kottan çıkarıldıkları anlaşılmaktadır. Çalışmanın bu aşaması kottan çıkarılma haberlerinin tek tek incelenmesi ile tamamlanmakta ve toplam 53 adet şirketin süreklilik risklerinden dolayı borsa kotundan çıkarıldıkları tespit edilmektedir.

İdeal bir süreklilik riski öngörüsü modeli bahse konu şirketler henüz aleni finansal zorluklar yaşamazken potansiyel sorunları keşfedebilmelidir. Finansal zorluklar yaşanmaya başladıktan sonra belirlenen riskler, iş işten geçtiği için, pratikte ciddi faydalar sağlamayacaktır. Bu nedenle, finansal zorluklar sebebi ile kottan çıkarılan şirketlerin ne kadar önceden bu sorunları yaşamaya başladığı tespit edilmeli ve son sağlıklı (aleni finansal sorun yaşanmayan) bilanço dönemi belirlenmelidir.

Finansal zorluklar sebebiyle kottan çıkarılan şirketlerin son sağlıklı bilanço döneminin belirlenmesinde artçı göstergelerden faydalanılmaktadır. Cirodaki hızlı dönemsel düşüşler, piyasa fiyatlarındaki keskin kırılmalar ve döneme ait finansal sorun haberleri artçı göstergeleri oluşturmaktadır. Bu göstergeler şirketin finansal sorunlarının aleni hale gelmesini takiben ortaya çıkmaktadır. Artçı göstergeler

sayesinde şirketin finansal zorluk yaşamadan hemen önceki bilanço döneminin tespiti ve modellerin bu döneme ait finansal veri ile oluşturulması mümkün olmaktadır.

Modellemelerin sağlıklı yapılabilmesi için her batık şirket bir adet sağlıklı şirket ile eşleştirilmektedir. Seçilen sağlıklı şirketlerin batık şirketler ile aynı sektörden olmasına ve bilanço dönemlerinin mümkün olduğunca yakın belirlenmesine dikkat edilmektedir. Bu şekilde 53 adet sağlıklı şirket ve bilanço dönemi tespit edilmekte ve modellerin zaman ve sektör yanlılığı en aza indirilmektedir. Toplam örnek sayısı 106'ya ulaşmakta ve sağlıklı/batık dağılımı dengelenmektedir.

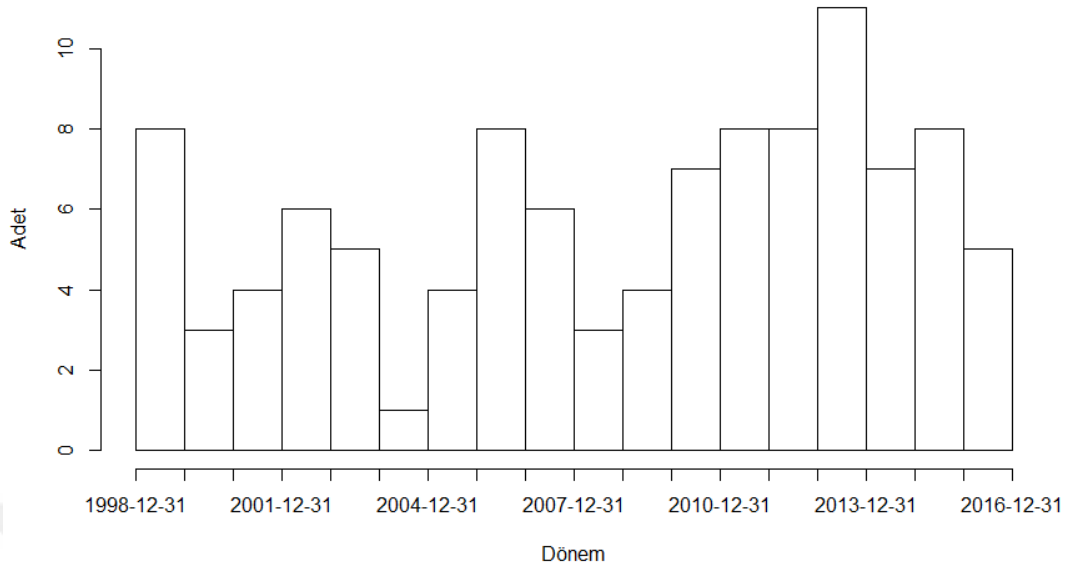
Çizelge 5.3'te çalışmaya konu şirketlerin dahil olduğu sektörler listelenmektedir.

Çizelge 5.3: Çalışmaya Konu Şirketlerin Dahil Olduğu Sektörler

Sektör	Sektör
Tarım Makineleri	Mühendislik
Alüminyum	Yedek Parça Üreticisi
Giyim ve Ayakkabı	Denizcilik ve Lojistik
Otomotiv Perakendecisi	Dinlenme Tesisleri
Teme ve Çeşitli Kimyasallar	Medya
Aracı Kurumlar	Hastane
Holdingle (finans ağırlıklı değil)	Paketlenmiş Gıda
İnşaat Malzemeleri	Özellikli Kimyasallar
Tüketici Elektronikleri	Özellikli Perakende
Çiftlik Ürünleri	Demir-Çelik
Süpermarketler	Tekstil

Şekil 5.3'te çalışmada kullanılan örneklerin tarihsel dağılımı sergilenmektedir. Veri kümesindeki tarihi yayılım borsa kotundaki şirket sayısına paralellik sergilemektedir ve son dönemlerdeki veri miktarının artışı bunun neticesindedir. Yakın dönemde, borsada işlem gören şirket sayısının artışına paralel olarak, kottan çıkarılma sayısının da artmış olması beklenen bir durumdur.

Veri Kümesinin Dönemsel Dağılımı



Şekil 5.3: Veri Kümesinin Dönemsel Dağılımı

5.5 Araştırmanın Kısıtları

Bu çalışma çerçevesinde karşılaşılan kısıtlar iki başlık altında toplanabilir: verinin hacmi ve modelleme araçlarının teorik alt yapısı.

5.5.1 Verinin hacmi

Çalışmada karşılaşılan en önemli kısıt veri hacmidir. Modellemeler için kullanılacak örnek sayısı toplam 106 ile sınırlı kalmakta ve sağlıklı bir şekilde belirlenebilecek parametre sayısı altıyı geçememektedir. Modellerde kullanılacak sağlıklı parametre sayısının son derece kısıtlı olması özellikle Sinir Ağı modellerinin performansını olumsuz yönde etkilemektedir. Bu modeller, ağ yapılarından dolayı, eğitimleri esnasında yüksek miktarda veriye ihtiyaç duymaktadır. ÇAA ve Lojistik Model ise, sadece doğrusal ilişkileri ortaya çıkarabildikleri için, yetersiz veri sorununu en az hisseden modeller olarak ortaya çıkmaktadır.

Çalışmada faydalanılan Karar Ağacı modellerinin sadece üç dal ayrımına sahip olması yine sınırlı verinin bir sonucudur. Her dal ayrımı, değişkenin ve kritik seviyenin seçimi olmak üzere, iki parametrenin belirlenmesini gerektirmekte ve eldeki örneklem sadece üç dal ayrımının sağlıklı bir şekilde tespitini mümkün kılmaktadır. Mevcut veri çerçevesinde, Karar Ağaçları'na bir dal ayrımı daha eklemek mümkün olamamakta ve bu modellerin nihai performansını yapıları değil,

verinin hacmi belirlemektedir.

5.5.2 Modelleme araçlarının teorik altyapısı

Makine Öğrenmesi modelleri teorik altyapıları çerçevesinde ciddi farklılıklar göstermekte ve bu ayrılıklar model ön koşullarının örtüşmemesi ile sonuçlanmaktadır.

Değişken dağılımına ilişkin bazı modellere ait ön koşullar son derece kısıtlayıcı olabilmekte, bu durum kimi değişkenlerin veriye hiç girememesiyle sonuçlanabilmektedir. Bölüm 5.3.3'te tanımlanan “kararsız” değişkenler bu duruma örneklerdir. Makine Öğrenmesi modellerinin diğer bir kısmı ise veriye dahil değişken dağılımlarının normal (Gauss) dağılıma yakınsamasını şart koşmaktadır. Değişken dağılımı normal değil ise dönüşümler aracılığı ile bunun düzeltilmesi yönünde çaba harcanmakta, ancak kimi zaman bu dönüşümler yeterli olmamakta ve dağılım tam olarak normalleşmemektedir.

Kararsız verinin dışlanması ve tam olarak normal dağılıma sahip olmayan değişkenlerin mevcudiyeti bir takım modellerin performansını ciddi ölçüde, diğerlerinininkini ise kısmen, olumsuz yönde etkileyebilmektedir.

5.6 Araştırmanın Bulguları

Bu bölümde araştırmada kullanılan Makine Öğrenmesi modelleri tek tek veriye uygulanmakta ve bulgular derlenmektedir.

5.6.1 Çoklu Ayraç Analizi

Analizin ön şartları sağlanmakta, veri dağılımları dönüşümler sonucunda uygun şekle sokulmaktadır. Analizde R istatistik dili ve “MASS” kütüphanesi kullanılmaktadır.

İlk aşamada tüm değişkenler kullanılarak modelleme yapılmaktadır. Modelde kullanılan değişkenlerin sonuca katkısı Çizelge 5.4'te sunulmaktadır. Bu çizelgede sonuca en yüksek katkıyı sağlayan değişkenin (Özkaynak/Borçlar) ağırlığı %100'e normalize edilmektedir. Modele dahil edilen diğer değişkenlerin sonuca katkısı ise bu değişkene oranla listelenmektedir. Örneğin; Nakit/Borçlar oranının sonuca katkısı Özkaynak/Borçlar oranının %67'si kadardır. Listenin sonuna doğru sonuca katkısı en düşük değişkenler yer almaktadır.

Çizelge 5.4: Değişkenlerin Sonuca Katkısı

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Özkaynak / Borçlar	100
Nakit / Borçlar	67
Borçlar / Aktifler	62
FAVÖK / Aktifler	61
Aktif Karlılığı	59
FAVÖK / Borçlar	27
Nakit / KV Borçlar	17
Nakit / Aktifler	5

Çizelge 5.5'te satırlar model öngörüsünü, sütunlar ise gerçek durumu temsil etmektedir. Öngörü ve gerçek durumun kesiştiği hücreler (%42 + %42 = %84) Model İçi Performans'a işaret ederken, bunların dışında kalan hücreler (%8 + %8 = %16) model öngörüsünün başarısız olduğu örnekleri temsil etmektedir.

Çizelge 5.5: ÇAA Modeli, Tüm Değişkenler Kullanılmış

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%42	%8
	Batık	%8	%42

Çapraz Doğrulama testlerinde rastgele 3 örnek bağımsız test kümesi olarak seçilmekte, geri kalan örnekler ile model oluşturulmakta ve modelin öngörü performansı bu test kümesi aracılığı ile tespit edilmektedir. Testler bu şekilde 5,000 defa tekrarlanmaktadır (bundan sonraki ÇAA modeli alternatiflerinde de Çapraz Doğrulama testleri aynı şekilde yapılmaktadır ve sonuçlar %95 güven aralığında %0.6 hata payına sahiptir). Model gerçekleştirdiği 15,000 öngörüden (5,000 * 3 = 15,000) 11,980'sinde (sonuçlar en yakın 10'a yuvarlanmaktadır) başarılı olmakta, bu da yaklaşık %80'lik bir Çapraz Doğrulama performansına işaret etmektedir.

Model İçi Performans ile Çapraz Doğrulama performansı arasındaki önemli fark (%84 - %80 = %4) Aşırı Uygunluk işareti olduğu için, değişken sayısının azaltılması yönüne gidilmektedir. En düşük katkıyı yapan değişken Nakit/Aktifler oranı olarak

tespit edilmektedir (Çizelge 5.4). Elde kalan yedi değişkenle tekrar bir model kurulmakta ve bunların sonuca katkısı Çizelge 5.6’da sunulmaktadır.

Çizelge 5.6: Değişkenlerin Sonuca Katkısı

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Özkaynak / Borçlar	100
Nakit / Borçlar	74
Borçlar / Aktifler	64
FAVÖK / Aktifler	61
Aktif Karlılığı	59
FAVÖK / Borçlar	26
Nakit / KVBorçlar	16

Değişken sayısı yediye indiğinde Model İçi Performans düşmemektedir (Çizelge 5.7). Bir değişken azaltılmış olmasına rağmen Çapraz Doğrulama performansında (15,000’de 12,010 başarılı öngörü) da ciddi bir ilerleme olmamakta ve %80’de kalmaktadır.

Çizelge 5.7: ÇAA Modeli, Değişken Sayısı 7

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%42	%8
	Batık	%8	%42

Yeni durumda en düşük katkıyı yapan değişken Nakit/KVBorçlar oranı olarak tespit edilmektedir (Çizelge 5.6). Elde kalan altı değişkenle tekrar bir model kurulmaktadır. Bu modelde kullanılan değişkenlerin sonuca katkısı Çizelge 5.8’de sunulmaktadır. Değişkenlerin sonuca katkı sıralamasında farklılıklar gözlenmektedir. Bazı değişkenler modelin dışında tutulduğunda, bu değişkenler ile en yüksek korelasyona sahip olan diğer değişkenlerin model içi ağırlığı artmaktadır. FAVÖK/Aktifler oranı buna en iyi örnektir. FAVÖK/Borçlar oranı ise en alttaki yerini korumaktadır ve diğer değişkenlerden önem açısından ayrılmış durumdadır.

Çizelge 5.8: Değişkenlerin Sonuca Katkısı

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Özkaynak / Borçlar	100
Nakit / Borçlar	94
FAVÖK / Aktifler	64
Borçlar / Aktifler	62
Aktif Karlılığı	62
FAVÖK / Borçlar	26

Modeldeki değişken sayısı bir azaldığı halde, model içi öngörü başarısı %84'te tutunmaktadır (Çizelge 5.9). Bu açıdan modelin sadeleştirilmesi başarılı olarak devam etmekte, ancak Çapraz Doğrulama performansı %80'de sabit kalmakta (15,000'de 12,040 başarılı öngörü), henüz kayda değer bir gelişme olmamaktadır.

Çizelge 5.9: ÇAA Modeli, Değişken Sayısı 6

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%42	%8
	Batık	%8	%42

Değişken sayısının tekrar azaltılması elimizdeki tek yol olduğundan, bir önceki modelde sonuca en az katkıyı yapan değişken (FAVÖK/Borçlar oranı) dışarıda bırakılarak yeni bir model oluşturulmaktadır. Bu modeldeki değişkenlerin sonuca katkısı Çizelge 5.10'da sunulmaktadır.

Çizelge 5.10: Değişkenlerin Sonuca Katkısı

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Nakit / Borçlar	100
Özkaynak / Borçlar	98
FAVÖK / Aktifler	77
Aktif Karlılığı	62
Borçlar / Aktifler	61

Nakit/Borçlar ve Özkaynak/Borçlar oranlarının sonuca katkısı neredeyse eşitlenmekte ve modelde “önemsiz” denebilecek herhangi bir değişken kalmamaktadır.

Değişken sayısı en etkin beşe indiğinde Model İçi Performans %83’e düşse de (Çizelge 5.11), Çapraz Doğrulama performansı %81’e (15,000’de 12,110 başarılı öngörü) yükselmektedir.

Çizelge 5.11: ÇAA Modeli, 5 Değişkenli

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%42	%9
	Batık	%8	%41

Değişken sayısını daha da azaltmak hem Model İçi Performans’ı hem de Çapraz Doğrulama performansını aşağı çekmektedir. En başarılı ÇAA modeli 5 değişkenli seçenek olarak ortaya çıkmaktadır. Bu modelde Aşırı Uygunluk %2’ye (%83 - %81 = %2) kadar gerilemektedir.

Çizelge 5.12’de araştırmada kullanılan farklı ÇAA modelleri performans açısından topluca kıyaslanmaktadır.

Çizelge 5.12: ÇAA Modellerinin Performans Karşılaştırması

Değişken Sayısı	Model İçi Performans (%)	Çapraz Doğrulama Performansı (%)
8	84	80
7	84	80
6	84	80
5	83	81

Farklı sayıda değişken kullanan ÇAA modellerinin performansları arasında çok büyük farkların oluşmadığı gözlenmektedir. Genel anlamda ÇAA modeli değişken sayısına duyarlı görünmemekte, bu da model kırılğanlığının (model performansının değişken sayısına hassasiyeti) düşük olduğunu göstermektedir.

5.6.2 Lojistik M

Lojistik Model eğitimi için öncelikle gerekli ön şartlar yerine getirilmekte, veri dağılımları dönüşümler neticesinde uygun şekle sokulmaktadır. Çalışmada R istatistik dili ve “GLMNET” kütüphanesi kullanılmaktadır. Bu kütüphane Lojistik Model parametrelerinin sınırlandırılabilmesini (regularization) mümkün kılmaktadır. Bu sınırlandırmalar Aşırı Uygunluk problemini hafifletebilmekte ve Çapraz Doğrulama performansını olumlu yönde etkileyebilmektedir.

GLMNET kütüphanesi parametre sınırlandırmalarını iki yöntemle yapabilmektedir. Bunlardan ilkinde model içi değişken ağırlıklarının kareleri (Ridge yöntemi), ikincisinde ise mutlak değerleri (Lasso yöntemi) sınırlandırılmaktadır. Sınırlandırmaların kullanıldığı model optimizasyonları sonucunda ilk yöntemde (Ridge) korelasyonu yüksek değişkenlerin model içi ağırlıkları birbirine yaklaşırken, ikincisinde (Lasso) değişkenlerden biri modelde tutulmakta, diğeri ise tamamen dışlanmaktadır.

Lasso yönteminin, korelasyonu yüksek değişkenlerden sadece birini modelde barındırdığı için, Aşırı Uygunluk problemi ile daha etkili mücadele edebileceği öngörülmekte, ancak mevcut çalışmada her iki yöntemde de başvurulmaktadır.

Lasso yöntemi ile yapılan çalışmanın ilk aşamasında tüm değişkenler kullanılmakta ve uygun parametre değerlerinin model tarafından tespitine izin verilmektedir. Sonuçlar incelendiğinde yöntemin beş değişken hariç diğer tüm değişkenleri modelden dışladığı görülmektedir (Çizelge 5.13). Dışlanan değişkenlerin geride kalanlarca ikame edilebildiği anlaşılmaktadır.

Çizelge 5.13: Değişkenlerin Sonuca Katkısı

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
FAVÖK / Borçlar	100
Özkaynak / Borçlar	72
Nakit / Borçlar	37
Aktif Karlılığı	30
FAVÖK / Aktifler	9
Borçlar / Aktifler	0
Nakit / KVBorçlar	0
Nakit / Aktifler	0

Çizelge 5.13'te sonuca en yüksek katkıyı sağlayan değişkenin (FAVÖK/Borçlar) ağırlığı %100'e normalize edilerek sunulmakta ve modele girebilen diğer değişkenlerin katkısı buna oranla sıralanmaktadır. Listenin sonundaki üç değişken modelden Lasso yöntemi tarafından dışlanmış görünmektedir.

CAA modellerinde aşırı uygunluğa karşı takip edilen “değişken sayısının azaltılması” süreci Lojistik Model'in Lasso yöntemi ile oluşturulması esnasında otomatik olarak gerçekleşmektedir. Yöntem değişkenlerin önemini belirlediği anda görece “önemsiz” olanları modelden derhal dışlamaktadır.

Çizelge 5.14'te satırlar model öngörüsünü, sütunlar ise gerçek durumu temsil etmektedir. Öngörü ve gerçek durumun kesiştiği hücreler (%41 + %44 = %85) Model İçi Performans'a işaret ederken, bunların dışında kalan hücreler (%9 + %6 = %15) model öngörüsünün başarısız olduğu örnekleri temsil etmektedir.

Çizelge 5.14: Lojistik Model – “Lasso” yöntemi

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%41	%6
	Batık	%9	%44

Çapraz Doğrulama testleri 3 örneğin rastgele seçilmesi ve bağımsız test kümesi olarak belirlenmesi ile gerçekleştirilmektedir. Geri kalan örnekler ile model oluşturulmakta ve modelin bağımsız test kümesi üzerindeki öngörü performansı tespit edilmektedir. Testler bu şekilde 5,000 defa tekrarlanmaktadır (Lojistik Model'in Ridge yönteminde de Çapraz Doğrulama testleri aynı şekilde tekrarlanmakta ve sonuçlar %95 güven aralığında %0.6 hata payına sahip olmaktadır). Model gerçekleştirdiği 15,000 öngörüden ($5,000 * 3 = 15,000$) 12,180'inde (sonuçlar en yakın 10'a yuvarlanmaktadır) başarılı olmakta, bu da yaklaşık %81'lik bir Çapraz Doğrulama performansına işaret etmektedir.

Model İçi Performans ile Çapraz Doğrulama performansı arasındaki fark %4 ($\%85 - \%81 = \%4$) civarındadır. Model Aşırı Uygunluk işaretleri verse de kullanılan değişken sayısının azaltıldığı durumlarda hem Model İçi Performans hem de Çapraz Doğrulama performansı gerilemektedir. Bu durumda değişken sayısının azaltılması yönüne gitmek faydalı görünmemektedir. Elde kalan tek alternatif, parametre

kısıtlaması yöntemini değiştirmektedir.

Kullanılan Lojistik Model kütüphanesinde Lasso yönteminin alternatifi Ridge yöntemidir. Ridge yönteminde model parametreleri mutlak değerleri yerine kareleri ile sınırlandırılmaktadır. Bu modellemede hiçbir değişken dışlanmamakta, ancak ciddi bir Aşırı Uygunluk durumu da gözlenmemektedir.

Ridge yöntemi kullanılarak yapılan modelleme sonucunda değişkenlerin sonuca katkısı Çizelge 5.15'te sunulmaktadır.

Çizelge 5.15: Değişkenlerin Sonuca Katkısı

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
FAVÖK / Borçlar	100
Aktif Karlılığı	67
FAVÖK / Aktifler	65
Özkaynak / Borçlar	59
Borçlar / Aktifler	57
Nakit / Borçlar	47
Nakit / KV Borçlar	46
Nakit / Aktifler	34

Çizelge 5.15'ten de anlaşılacağı gibi modelin değişkenler için belirlediği ağırlıklar arasında en düşük olan %34'tür ve en önemli değişkenle fark sadece üç kat civarındadır. Lasso yönteminde ise bu fark on bir kata kadar ulaşmaktadır.

Ridge yöntemi ile ulaşılan Model İçi Performans %86 olurken (Çizelge 5.16) Çapraz Doğrulama performansı yine %81'de (15,000 öngöründe 12,110) kalmaktadır. Modelin yaşadığı Aşırı Uygunluk sorunu %5 seviyesindedir (%86 - %81 = %5).

Çizelge 5.16: Lojistik Model – “Ridge” yöntemi

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%42	%6
	Batık	%8	%44

Çizelge 5.17’de iki Lojistik Model performans açısından kıyaslanmaktadır.

Çizelge 5.17: Lojistik Modeller’in Performans Karşılaştırması

Parametre Sınırlama Tipi	Model İçi Performans (%)	Çapraz Doğrulama Performansı (%)
Lasso	85	81
Ridge	86	81

Parametre sınırlamaları açısından Lasso ve Ridge yöntemleri arasında kayda değer bir fark görülmemekte, ancak Aşırı Uygunluk sorunu en iyi alternatifte dahi %4 civarında ve görece yüksek seyretmektedir.

Lasso yöntemi bazı değişkenleri tamamen modelden dışlamakta, bu da daha düşük Aşırı Uygunluk sorunu yaşanmasını sağlamaktadır. Ridge yöntemi ise bütün değişkenleri modelde tutmakta ve Aşırı Uygunluk problemiyle yeteri kadar mücadelede edememektedir.

Sonuç olarak, Lojistik Model alternatifleri ile yapılan çalışma, Lasso ve Ridge yöntemleri arasında ciddi bir performans farkına işaret etmemekle birlikte, Lasso yönteminin Aşırı Uygunluk açısından hafif bir üstünlüğünden bahsedilebilmektedir.

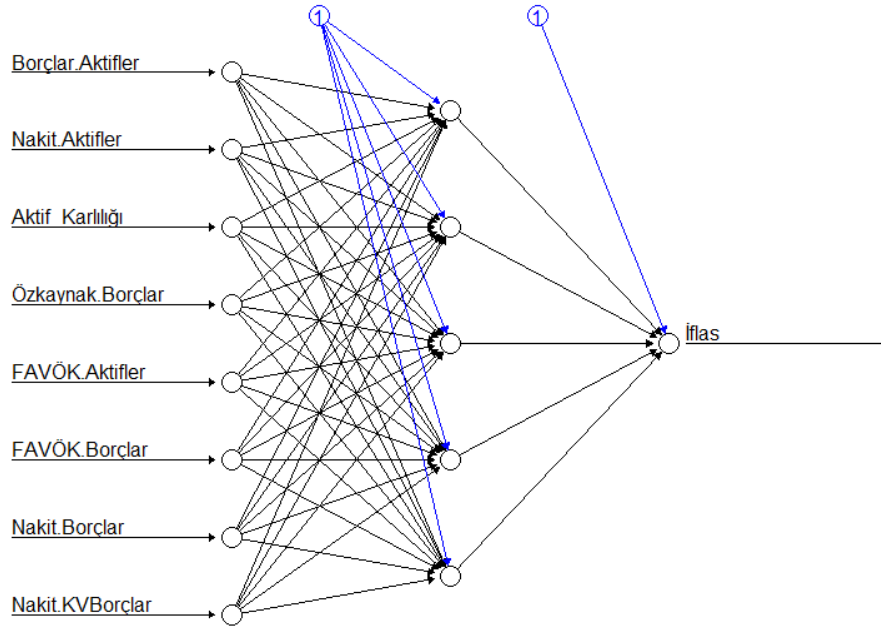
5.6.3 Sinir Ağları

Sinir Ağı modellerinin eğitilmesi ilk olarak gerekli ön şartların yerine getirilmesi ve değişken dağılımlarının dönüşümler sonucunda uygun şekle sokulması ile başlamaktadır. Çalışmada R istatistik dili ve “neuralnet” kütüphanesi kullanılmaktadır.

Sinir Ağları’nın model eğitiminin son derece işlem yoğun olmasının yanında her zaman başarılı bir şekilde sonuçlanması mümkün de olmayabilmektedir. Bunun nedeni, eldeki optimizasyon probleminin pek çok yerel optimal sonuç içermesi, “düzenli” olmamasıdır. Model eğitiminin ilk adımında rastgele tayin edilen parametre ağırlıkları çözümü başka noktalara sürükleyebilmektedir. Optimizasyonun düzenli olmaması, aynı modelin farklı başlangıç değerleri ile defalarca baştan çözülmesini gerektirmektedir. Sonuç olarak aynı model için pek çok parametre kümesi elde edilmektedir. Bu kümeler arasından en iyi Model İçi Performans’ı sergileyen parametreler nihai modeli belirlemektedir, ancak geri kalan çözüm kümeleri de tamamen işlevsiz değildir. Bu çözüm kümelerindeki parametre

değerlerinin ortalamaları, eldeki değişkenlerin ne sıklıkta ve ağırlıkta kullanıldıklarının, yani bu değişkenlerin öneminin bir göstergesidir. Sinir Ağı modelleri çerçevesinde değişkenlerin sonuca katkısı bu şekilde hesaplanmaktadır.

Sinir Ağı modelleri, esnek yapıları gereği, çok farklı mimarilerde oluşturulabilirler. Çalışmada denenen ilk Sinir Ağı modelinde (Şekil 5.4) tüm değişkenler ve saklı katmanda (girişte veya sonuçta olmayan, arada kalan düğüm noktaları) 5 düğüm noktası kullanılmaktadır.



Şekil 5.4: 1. Model - 8 Değişken, 1 Saklı Katman, 5 Düğüm Noktası

Bu model çerçevesinde değişkenlerin sonuca katkısı Çizelge 5.18’de sunulmaktadır. Bu çizelgede sonuca en büyük katkıyı sağlayan değişkenin (FAVÖK/Borçlar) ağırlığı %100’e normalize edilmekte ve diğer değişkenlerin sonuca katkısı bu değere oranlanarak sıralanmaktadır.

Çizelge 5.18: Değişkenlerin Sonuca Katkısı

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
FAVÖK / Borçlar	100
Özkaynak / Borçlar	36
Nakit / KVBorçlar	25
FAVÖK / Aktifler	14
Nakit / Borçlar	8
Borçlar / Aktifler	7
Nakit / Aktifler	5
Aktif Karlılığı	1

Çizelge 5.19’da satırlar model öngörüsünü, sütunlar ise gerçek durumu temsil etmektedir. Öngörü ile gerçek durumun kesiştiği hücreler (%50 + %50 = %100) Model İçi Performans’a işaret ederken, bunların dışında kalan hücreler (%0 + %0 = %0) model öngörüsünün başarısız olduğu örnekleri temsil etmektedir.

Çizelge 5.19: Sinir Ağı 1 – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%50	%0
	Batık	%0	%50

Sinir Ağları modelleri için yapılan Çapraz Doğrulama testleri çalışmada kullanılan diğer modeller için yapılanlarla farklılık göstermektedir. Normal şartlar altında binlerce kez rastgele tekrarlanabilen Çapraz Doğrulama testlerinin Sinir Ağları’na uygulanması pratikte mümkün olmamakta ve bu modele has bir Çapraz Doğrulama yönteminin belirlenmesi gerekmektedir.

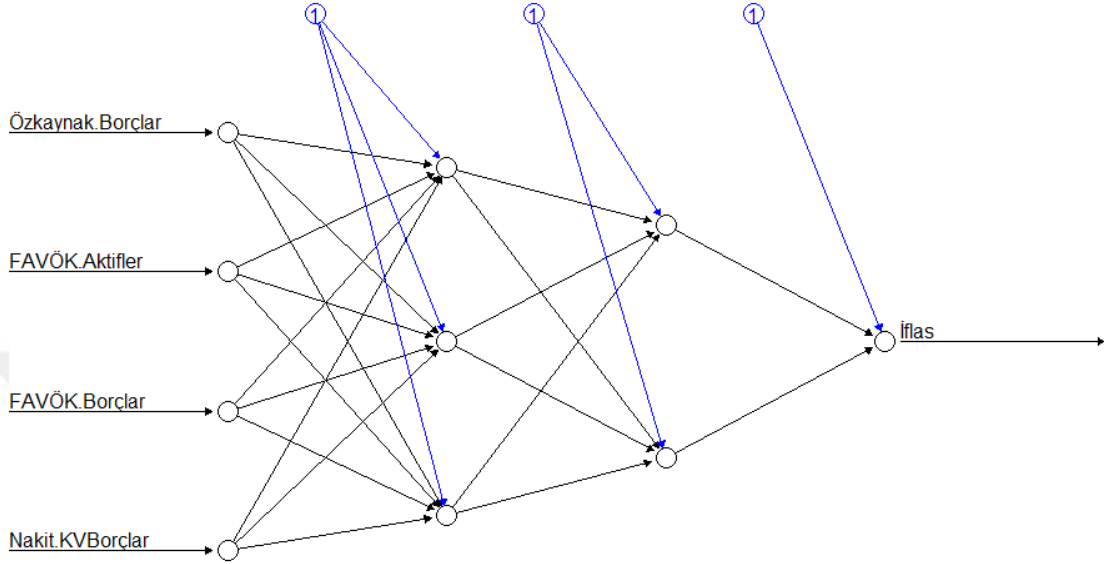
Sinir Ağlarına yönelik olarak tasarlanan Çapraz Doğrulama yönteminde her test için sadece iki örnek dışarıda bırakılmakta ve geri kalan veri ile model oluşturulmaktadır. Yöntemin bir sonraki basamağında ise dışarıda kalan örnekler ile modelin performansı tespit edilmektedir. Bu yöntem Sinir Ağları için 5,565 defa (eldeki örneklemin ikili kombinasyonları kadar) tekrarlanmakta ve bu yüzden ulaşılan sonuçların istatistiki geçerliliği (diğer Sinir Ağı modellerinde de Çapraz Doğrulama testleri aynı şekilde tekrarlanmakta ve sonuçlar %95 güven aralığında %1 hata payına sahip olmaktadır) korunabilmektedir.

Denenen ilk Sinir Ağı modeli (Şekil 5.4), Çapraz Doğrulama testleri neticesinde gerçekleştirdiği 5,565 öngörüden 4,460’ında (sonuçlar en yakın 10’a yuvarlanmaktadır) başarılı olmakta, bu da yaklaşık %80’lik bir Çapraz Doğrulama performansına denk gelmektedir. Model İçi Performans’ın %100 olduğu (Çizelge 5.19) göz önüne alındığında ciddi ölçüde bir Aşırı Uygunluk sorununun yaşandığı (%100 - %80 = %20) anlaşılmaktadır.

Değişken sayısının azaltıldığı bir sonraki modelin girdileri, bir önceki modelin ürettiği değişken önemi analizlerine göre belirlenmektedir. Sonuca katkı açısından en zayıf dört değişken (Çizelge 5.18) bir sonraki modelin dışında bırakılmaktadır.

Bunun Aşırı Uygunluk problemini hafifletmesi öngörülmektedir.

Denenen ikinci modelde (Şekil 5.5) seçilmiş 4 değişken ve 2 saklı katmanda toplam 5 düğüm noktası (3 ve 2) kullanılmaktadır. Saklı katmanlardaki toplam düğüm sayısı sabit tutulmaktadır.



Şekil 5.5: 2. Model - 4 Değişken, 2 Saklı Katman, 5 Düğüm Noktası

Değişkenlerin sonuca katkısı Çizelge 5.20’de listelenmiştir.

Çizelge 5.20: Değişkenlerin Sonuca Katkısı

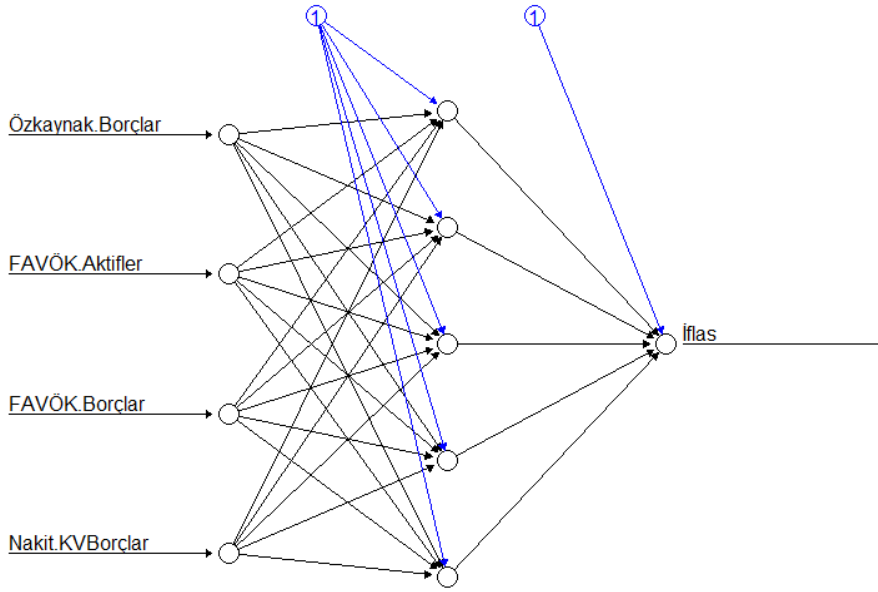
Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
FAVÖK / Borçlar	100
Özkaynak / Borçlar	52
Nakit / KVBorçlar	18
FAVÖK / Aktifler	15

Özkaynak/Borçlar oranının ağırlığı artarken, Nakit/KVBorçlar oranının ağırlığı azalmaktadır. Model İçi Performans yine %100’e ulaşmaktadır (Çizelge 5.21). Çapraz Doğrulama testlerinde performans %81’e yükselmekte (5,565 öngöründen 4,510’unda başarı), ancak bariz şekilde Aşırı Uygunluk sorunu yaşanmaya devam etmektedir.

Çizelge 5.21: Sinir Ağı 2 – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%50	%0
	Batık	%0	%50

Denenen üçüncü modelde (Şekil 5.6) seçilmiş 4 değişken ve 1 gizli katmanda yine toplam 5 düğüm noktası kullanılmaktadır. Toplam düğüm sayısı yine sabit tutulmakta ve sadece ağın mimarisi değiştirilmektedir.



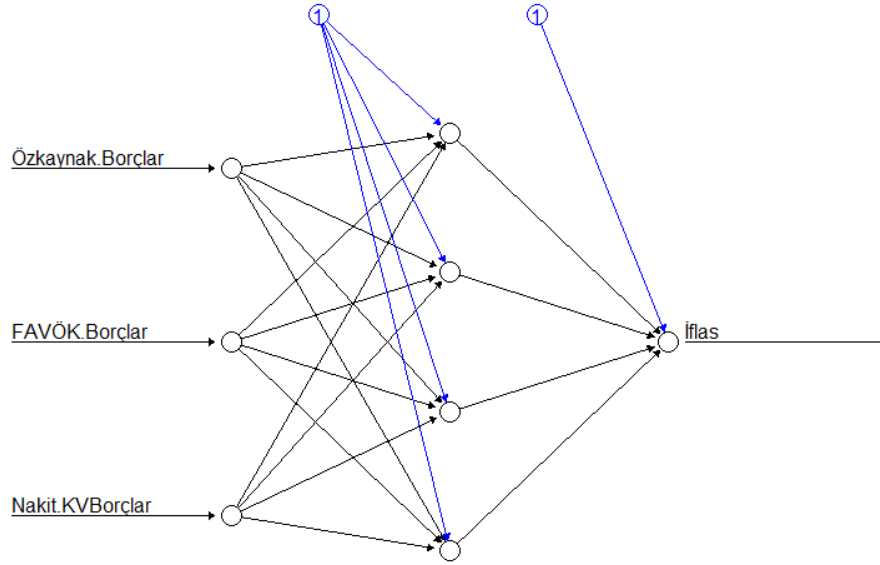
Şekil 5.6: 3. Model - 4 Değişken, 1 Saklı Katman, 5 Düğüm Noktası

Model İçi Performans yine %100 olarak ortaya çıkmakta (Çizelge 5.22) ve Çapraz Doğrulama performansı da %81'i (5,565 öngöründen 4,520'sinde başarı) aşamamaktadır. Gizli katmandaki düğüm sayısı aynı kaldıkça model performansı gelişmemektedir. Eğitilmesi gereken ilişki sayısı, yani ağırlık adedi, büyük değişim göstermemektedir.

Çizelge 5.22: Sinir Ağı 3 – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%50	%0
	Batık	%0	%50

Aşırı Uygunluk probleminin şiddeti ($\%100 - \%81 = \%19$) uç tedbirler almaya neden olmaktadır. İlk modelden sonraki modeller incelendiğinde bir değişkenin daha azaltılabileceğine kanaat getirilmektedir. Modelde kullanılan değişken sayısı 3'e indirilmektedir. Eldeki örnek veri adedinin sadece 106 olduğu düşünülerek saklı katmandaki düğüm adedinin de 4'e indirilmesinin uygun olacağı öngörülmektedir (Şekil 5.7). Değişkenlerin sonuca katkısı Çizelge 5.23'de sunulmaktadır.



Şekil 5.7: 4. Model - 3 Değişken, 1 Saklı Katman, 4 Düğüm Noktası

Çizelge 5.23: Değişkenlerin Sonuca Katkısı

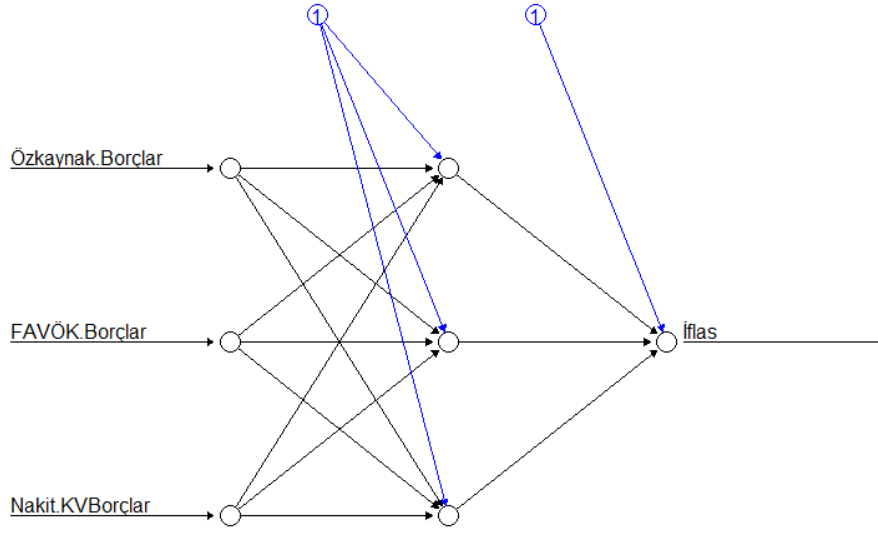
Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Özkaynak / Borçlar	100
FAVÖK / Borçlar	63
Nakit / KVBorçlar	25

Özkaynak/Borçlar oranı sonuca katkı açısından ilk sıraya yerleşmektedir. Model İçi Performans $\%95$ 'e gerilemekte (Çizelge 5.24), Çapraz Doğrulama performansı ise $\%82$ 'ye (5,565 öngöründen 4,570'inde başarı) yükselmektedir. Aşırı Uygunluk sorununun bir miktar giderilebildiği ($\%95 - \%82 = \%13$) anlaşılmaktadır.

Çizelge 5.24: Sinir Ağı 4 – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%46	%1
	Batık	%4	%49

Kullanılan değişken adedi daha da azaltıldığında hem Model İçi Performans'ın hem de Çapraz Doğrulama performansının düştüğü gözlenmektedir. Bir sonraki alternatif olarak sadece saklı katmandaki düğüm sayısında bir azaltmaya gidilmektedir (Şekil 5.8).



Şekil 5.8: 5. Model - 3 Değişken, 1 Saklı Katman, 3 Düğüm Noktası

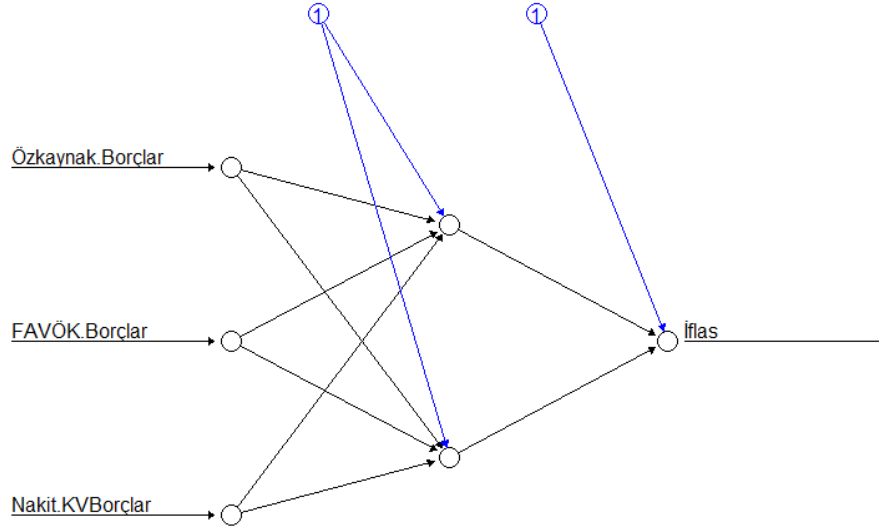
Model İçi Performans %93'e gerilerken (Çizelge 5.25) Çapraz Doğrulama performansı %84'e (5,565 öngöründen 4,670'inde başarı) yükselmektedir.

Çizelge 5.25: Sinir Ağı 5 – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%47	%4
	Batık	%3	%46

Model çok daha sağlıklı çalışmakta, Aşırı Uygunluk sorunu bir miktar daha giderilmekte (%93 - %84 = %9), ancak hala ciddi oranda yüksek seyretmektedir.

Bu aşamadan sonra denenebilecek tek model kalmaktadır, o da üç değişkenli ve tek saklı katmanda iki düğümlü modeldir (Şekil 5.9).



Şekil 5.9: 6. Model - 3 Değişken, 1 Saklı Katman, 2 Düğüm Noktası

Model İçi Performans %91'e gerilerken (Çizelge 5.26), Çapraz Doğrulama performansı %84'te (5,565 öngöründen 4,680'inde başarı) sabit kalmaktadır. Aşırı Uygunluk sorunu bu modelde bir miktar daha hafiflemekte (%91 - %84 = %7) ancak hala kabul edilebilir düzeylere inememektedir.

Çizelge 5.26: Sınır Ağı 6 – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%47	%6
	Batık	%3	%44

Eldeki en başarılı model aynı zamanda en basit model olarak karşımıza çıkmaktadır.

Bu Sınır Ağı modeli, oldukça sade bir yapıya sahip görünmekle birlikte, bünyesinde üç adet Lojistik Model barındırmaktadır (saklı katmanda iki düğüm noktası + sonuç katmanında bir düğüm noktası). Sınır Ağları kolayca çok karmaşık hale gelebilmekte ve modellenmesi için gereken veri adedi hızla yükselebilmektedir.

Çalışmada başvurulan Sınır Ağı modellerinin performansları Çizelge 5.27'de topluca sergilenmektedir.

Çizelge 5.27: Sinir Ağları'nın Performans Karşılaştırması

Değişken Sayısı	Saklı Katman Sayısı	Karar Dügümü Sayısı	Model İçi Performans (%)	Çapraz Doğrulama Performansı (%)
8	1	5	100	80
4	2	5	100	81
4	1	5	100	81
3	1	4	95	82
3	1	3	93	84
3	1	2	91	84

Yapılan analizler 3 deęişkenli ve 2 karar düęümlü mimarinin Çapraz Doğrulama performansı (%84) ve Aşırı Uygunluk (%7) açısından en başarılı Sinir Aęı modeli olduğunu ortaya koymakta, ancak Aşırı Uygunluk probleminin yeteri kadar giderilemedięi de anlaşılmaktadır. Bundan daha basit modeller ise ancak deęişken eksiltilmesi ile mümkün olmakta ve o durumda ise hem Model İçi Performans hem de Çapraz Doğrulama performansları gerilemektedir.

Çalışma kapsamındaki Sinir Aęı modellerinden çıkarılabilecek belki de en önemli sonuç eldeki verinin yetersizlięi, örnek azlıęıdır. Verinin hacmi potansiyel model seçeneklerini çok dar bir alana kısıtlamaktadır. Gerçekte çok esnek olan Sinir Aęı modeli son derece basit bir alternatifi ile süreklilik riski öngörüsüne katkı sağlayabilmektedir.

5.6.4 Destek Vektör Makineleri

DVM modellemesinin ilk aşamasında gerekli ön şartlar yerine getirilmekte, veri dağılımları dönüşümler aracılıęı ile uygun şekle sokulmaktadır. Çalışmada R istatistik dili ve “e1071” kütüphanesi kullanılmaktadır.

DVM modelleri arasından Radyal modelin (Gauss çekirdekli model) eldeki probleme veri hacmi ve kavranabilirlik açılarından daha uygun olduğu görülmektedir. Verinin dağılımı ve amaç, eldeki model ile örtüşmektedir. Radyal DVM modeli kural bazlı deęil örnek bazlı bir modelleme aracıdır. Bu açıdan mevcut çalışmada tektir. Sınıflandırma eldeki dięer örneklere benzerlikle (deęişken uzayındaki yakınlık ile) tayin edilmektedir.

DVM, tabiatı gereği, kolaylıkla Aşırı Uygunluk problemi yaşayabilecek bir modeldir. İzin verildiğinde tüm örneklerin destek vektörü olarak seçilmesi, bunun da Aşırı Uygunluk yaratması mümkündür. Eldeki tüm örnekler destek vektörleri olarak kullanıldığında optimize edilmesi gereken parametre sayısı mevcut örnek sayısının üstüne çıkabilmektedir. Bu nedenle DVM modellerinde de parametre sınırlamaları (Lojistik Model’de olduğu gibi) kritik öneme sahiptir. Mevcut çalışmada mümkün olduğunca az sayıda örneğin destek vektörü olarak belirlenmesine çalışılmaktadır. Destek vektörlerinin etki alanlarının ne olması gerektiği de ayrı bir optimizasyon problemidir. Mevcut R kütüphanesinde her iki konuda da (optimal destek vektörü adedi ve destek vektörlerinin etki alanı) kullanılabilir araçlar mevcuttur. İlk analizler optimal destek vektör adedi ve etki alanları üzerine yapılmaktadır. DVM modellemesinde destek vektör adedini ve vektörlerin etki alanlarını belirleyen “cost” ve “gamma” isimli parametreler mevcuttur.

Çizelge 5.28’de Çapraz Doğrulama yöntemi ile tespit edilen optimal cost ve gamma parametre değerleri ve bunlardan uzaklaştıkça gerileyen Model İçi Performans örneklerle sunulmaktadır.

Çizelge 5.28: Parametre Optimizasyonu

cost	gamma	Model İçi Performans (%)
8 (optimal)	0.015625 (optimal)	89
16	0.015625	88
8	0.03125	88
4	0.015625	87
8	0.0078125	85

Çizelge 5.28’in ilk satırında optimal parametre değerleri sunulurken diğer satırlarda bu değerlerden sapmalar olması halinde Model İçi Performans’ın nasıl etkilendiği sergilenmektedir.

Radyal DVM modeline has bu parametreler (cost ve gamma) eldeki problem ve örneklem çerçevesinde çok farklı değerler alabilmektedir. Model optimizasyonunun bu ilk basamağı atlanır ve analize e1071 kütüphanesinin ön değerleri ile devam edilirse başarısız sonuçlar alınması ihtimali yükselmektedir.

İlk analizler sonucu belirlenen sınırlandırma parametreleri ile ulaşılan Model İçi Performans Çizelge 5.29'da sergilenmekte, burada satırlar model öngörüsünü, sütunlar ise gerçek durumu temsil etmektedir. Öngörü ile gerçek durumun kesiştiği hücreler (%45 + %43 = %88) Model İçi Performans'a işaret ederken, bunların dışında kalan hücreler (%5 + %7 = % 12) model öngörüsünün başarısız olduğu örnekleri temsil etmektedir.

Çizelge 5.29: DVM Modeli – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%45	%7
	Batık	%5	%43

DVM modelinin Çapraz Doğrulama testleri 3 örneğin rastgele seçilmesi ve bağımsız test kümesi olarak belirlenmesi ile gerçekleştirilmektedir. Geri kalan örnekler ile model oluşturulmakta ve modelin bağımsız test kümesi üzerindeki öngörü performansı belirlenmektedir. Testler bu şekilde 5,000 defa tekrarlanmaktadır. Model gerçekleştirdiği 15,000 öngörüden ($5,000 * 3 = 15,000$) 13,030'unda (sonuçlar en yakın 10'a yuvarlanmaktadır) başarılı olmaktadır ve bu da yaklaşık %87'lik (%95 güven aralığında %0.6 hata payı) bir Çapraz Doğrulama performansına denk gelmektedir.

İlk bakışta yapılabilecek en önemli çıkarım neredeyse hiç Aşırı Uygunluk probleminin yaşanmadığıdır. Model İçi Performans %88 iken Çapraz Doğrulama performansı %87 olarak gerçekleşmekte, Aşırı Uygunluk sadece %1'de ($\%88 - \%87 = \%1$) kalmaktadır. Bu durum, analizden hemen önce gerçekleştirilen optimal parametre keşfinin başarılı olduğunu göstermektedir. Modellemenin e1071 kütüphanesi ön değerleri ile yapılması halinde ulaşılan Çapraz Doğrulama performansının %74'ü aşamadığı not edilmelidir.

İkinci bir çıkarım ise, modelin yüksek bir başarıya sahip olduğudur. Model %87 gibi oldukça iyi bir Çapraz Doğrulama performansı sergilemektedir. Bu başarı oranı pratikte kullanılabilirlik açısından da değerli görünmektedir.

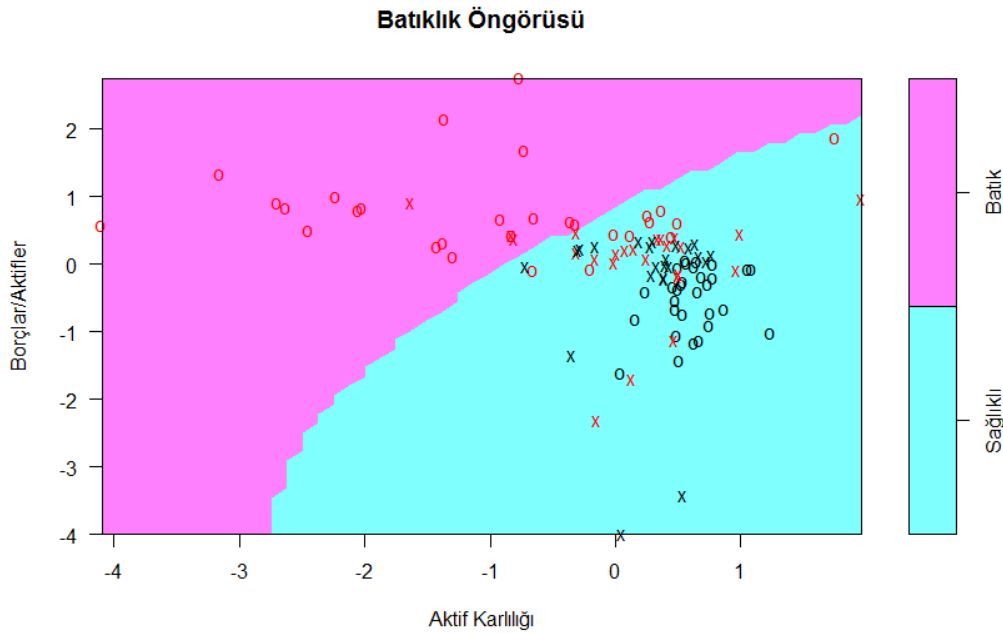
Modelin zayıf yanı ise ulaşılan performanstaki değişken katkısının kolayca anlaşılabilmesidir, çünkü asıl belirleyici olan değişkenlerden ziyade model

tarafından belirlenen birkaç destek vektörü, yani örnek veridir. Yine de değişkenlerin oynadığı rol hakkında görsel birkaç ipucu elde etmek mümkündür.

Ortaya çıkan model çok eksenlidir (sekiz değişken eksen). Sonuçların görselleştirilmesi, hangi değişkenlerin bu görsellerde kullanılacağına dair subjektif tercihleri zorunlu kılmaktadır. Çalışmanın bu aşmasında, iki boyutlu gösterimlerin kavranması daha kolay olacağından bazı ikili kombinasyonlar seçilerek grafikler hazırlanması yönüne gidilmektedir. Bu amaca yönelik hazırlanan iki görsel Şekil 5.10'da ve Şekil 5.11'de sunulmaktadır.

Şekiller incelenirken değişkenlerin dönüşümlerden geçirildiği ve dağılımların standart normal dağılıma (ortalaması sıfır ve standart sapması 1.0 olan Gauss dağılımı) yakınlştırıldığı unutulmamalıdır. Şekil 5.10 ve Şekil 5.11'de "X" simgeleri "Sağlıklı" şirketleri, "O" simgeleri ise "Batık" şirketleri göstermektedir.

Şekil 5.10'da açıkça görüldüğü gibi, batık olma ve değişkenler arasındaki ilişki doğrusal değildir. Aktif Karlılığı düşükçe daha az borçlu olan şirketler için dahi süreklilik riski artmaktadır.



Şekil 5.10: Borçlar/Aktifler – Aktif Karlılığı

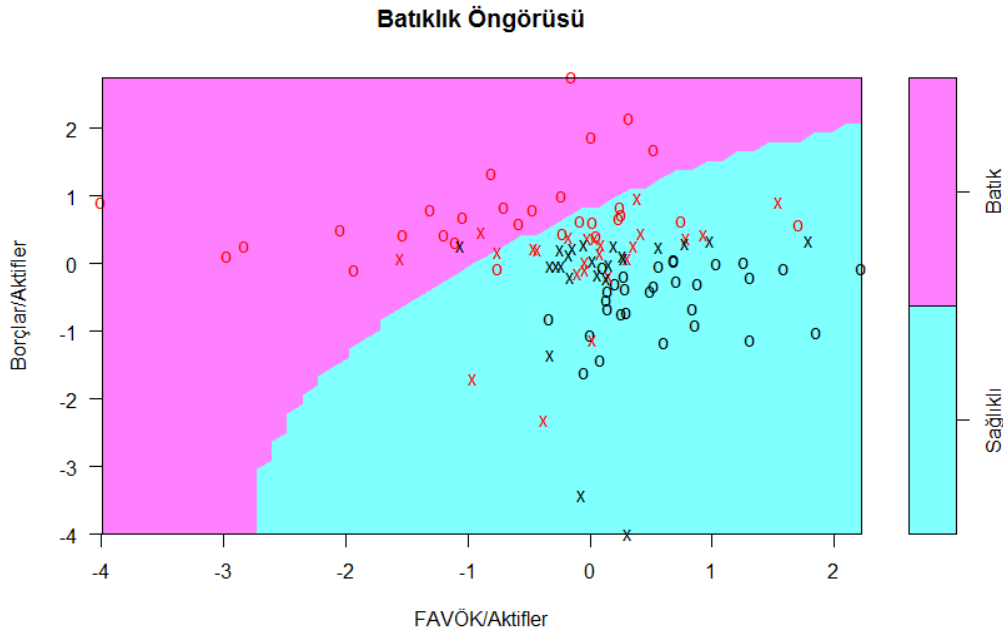
Seçilebilecek diğer ikili kombinasyonlar (toplam 56 adet farklı kombinasyon) için de grafikler hazırlanıp aralarındaki ilişkinin analiz edilmesi mümkündür ancak tüm bu kombinasyonların çalışmada yer alması mümkün değildir. Bu noktada yine subjektif

bir kararla yüksek korelasyona sahip iki değişkenin etkisine örnek verilmesinin daha değerli olacağı öngörülmektedir (Şekil 5.11).

Aktif Karlılığı ve FAVÖK/Aktifler oranı yüksek korelasyona sahip iki değişken olarak tespit edilmekte ve sonuca katkıları da oldukça benzer görünmektedir (Şekil 5.11). Bunda, analizden önce yapılan dönüşümlerin etkisi büyüktür. Örneklerin ağırlık merkezi her iki şekilde de (0,0) noktasıdır.

Şekiller yakından incelendiğinde ise örnek dağılımlarının önemli farklılıklar gösterdiği anlaşılmakta, bu da aralarında “tam” korelasyon olmayan iki değişkenin probleme farklı açılardan bakabildiğini ortaya koymaktadır.

İki değişken arasında tam korelasyon olmaması, farklı ve yeni bilginin mevcudiyetine işaret etmektedir ve şekilde gözlenen tam da budur. Probleme kısmen farklı açılardan bakılmakta ve farklı gözlemler, sonuçlar üretilebilmektedir.



Şekil 5.11: Borçlar/Aktifler – FAVÖK/Aktifler

İki şeklin ortak yönleri irdelendiğinde, borçluluk arttıkça riskin daha yumuşak bir şekilde yükseldiği (şekillerin sağ üst bölümü) ama zarardaki artışın çok daha şiddetli etki gösterdiği (şekillerin sol alt bölümü) ortaya çıkmaktadır.

Zararın aktiflere oranı belirli bir seviyeye ulaştığında işletmenin sürekliliği kabulü ortadan kalkmakta, zararın daha fazla artmasına izin verilmemektedir. Borçluluk ise bu kadar keskin bir etki göstermemekte ama düzenli bir şekilde süreklilik riskini

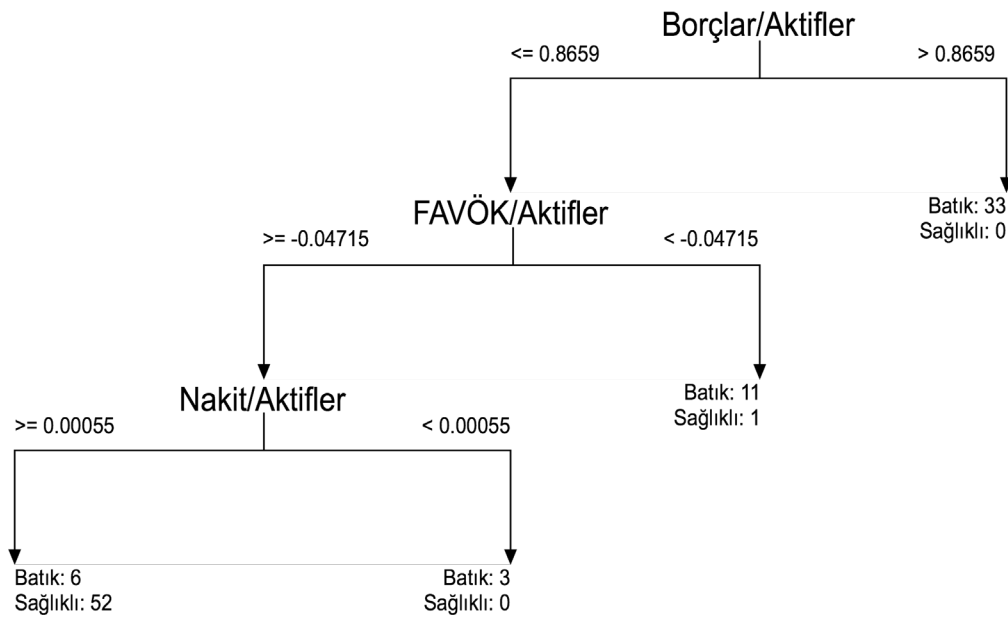
yükseltmektedir. Eldeki şekiller, doğrusal olmayan ilişkilerin DVM modeli ile ortaya çıkarılabileceğine görsel kanıtlardır.

Ulaşılan DVM modelinin %87 başarı ile çalıştığı, ancak eldeki örneklerin %13'ünde de hata yaptığı anlaşılmaktadır. Şekillerden de takip edilebileceği gibi, batık olduğu bilinen bazı örneklerin sağlıklı grubun çok içlerine girebildiği gözlenmektedir. Modelin bu örnekleri doğru sınıflandırmada yetersiz kaldığı anlaşılmaktadır.

5.6.5 Breiman Ağacı

Breiman Ağacı modeli eldeki veriyi orijinal hali ile kullanabilmekte ve örneklemin dönüştürülmesine ihtiyaç duyulmamaktadır. Modelleme esnasında ortaya çıkan tüm değerler pratikte anlamı olan seviyelerdir, bundan önceki modellerde olduğu gibi bu değerlerin kavranması için ters dönüşümler yapılması gerekmemektedir. Çalışmada R istatistik dili, “rpart” ve “partykit” kütüphanelerinden faydalanılmaktadır.

Breiman Ağacı modeli veriye uygulandığında, Çapraz Doğrulama performansı açısından, optimal Karmaşıklık Parametresi değerinin yaklaşık olarak %2 seviyesinde belirlenmesi gerektiği anlaşılmaktadır. Bu cp değeri potansiyel ağaçları 3 budağa, yani sadece dört yaprağa sınırlamaktadır. cp değerinin %2 olarak alınması ile oluşturulan Breiman Ağacı modeli Şekil 5.12'de sergilenmektedir. Çalışmanın bundan önceki kısımlarında olduğu gibi, işletmenin sürekliliği kabulünün ortadan kalktığı örnekler “batık”, süreklilik riskinin yaşanmadığı örnekler ise “sağlıklı” olarak işaretlenmektedir.



Şekil 5.12: Breiman Ağacı Batıklık Öngörüsü

Şekil 5.12'deki Breiman Ağacı'nın Model İçi Performans'ı Çizelge 5.30'da sergilenmekte, burada satırlar model öngörüsünü, sütunlar ise gerçek durumu temsil etmektedir. Öngörü ile gerçek durumun kesiştiği hücreler (%49 + %44 = %93) Model İçi Performans'a işaret ederken, bunların dışında kalan hücreler (%1 + %6 = %7) model öngörüsünün başarısız olduğu örnekleri temsil etmektedir.

Çizelge 5.30: Breiman Ağacı – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%49	%6
	Batık	%1	%44

Bu model çerçevesinde sınıflandırma öncelikle eldeki örneğin Borçlar/Aktifler oranı üzerinden gerçekleşmektedir. Bu oran 0.8659'a eşit veya daha yüksek ise örnek muhtemel batık aday olarak sınıflandırılmaktadır (başarı oranı %100). İlk kuralın dışındaki örnekler için FAVÖK/Aktifler oranına bakılmakta, bu oran -0.04715'ten daha düşük ise eldeki örnek yine muhtemel bir batık aday olarak işaretlenmektedir (başarı oranı %92). İlk iki kuralın dışında kalanlar için ise Nakit/Aktifler oranı takip edilmekte ve bu değer 0.00055'ten daha düşük ise eldeki örneğin batık (başarı oranı %100), aksi halde ise sağlıklı olduğu öngörülmektedir (başarı oranı %90).

Karar Ağacı modellerinde, modele dahil değişkenlerin sonuca katkısı kaç örneğin sınıflandırmasında rol oynadığı ile tespit edilmektedir. Bu şekilde bir analiz yapıldığında Çizelge 5.31'e ulaşılmaktadır. Çizelgede sonuca en yüksek katkıyı yapan Borçlar/Aktifler oranının ağırlığı normalize edilerek (%100'e eşitlenerek) gösterilmekte, diğer değişkenlerin katkısı buna oranlanarak sıralanmaktadır.

Çizelge 5.31: Değişkenlerin Sonuca Katkısı

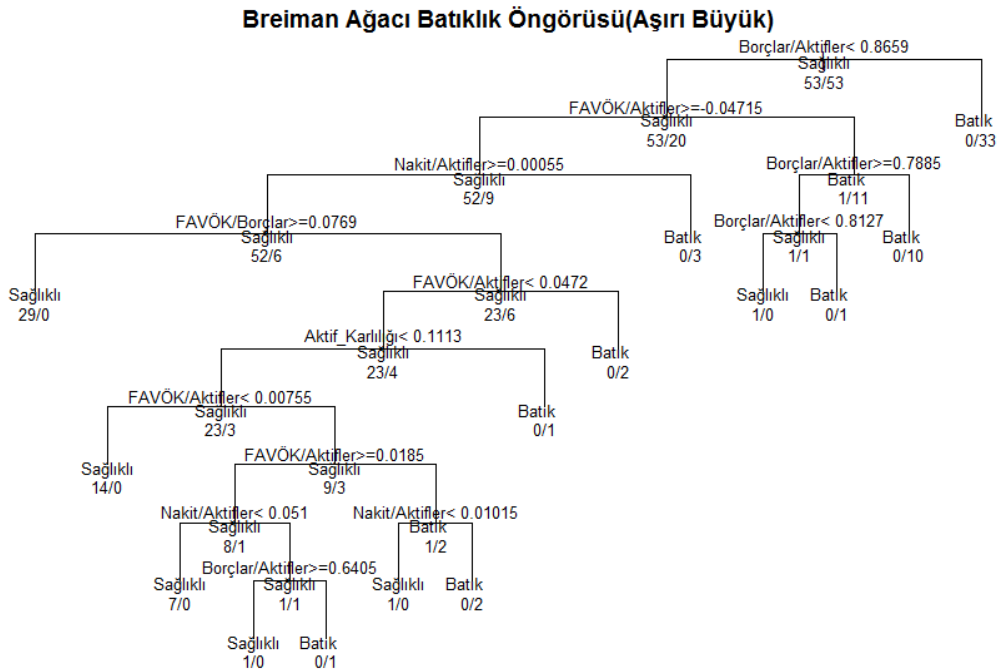
Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Borçlar / Aktifler	100
FAVÖK / Aktifler	44
Nakit / Aktifler	34

Breiman Ağacı modeli için yapılan her bir Çapraz Doğrulama testinde rastgele seçilen 3 örnek dışarıda bırakılmakta, gerisi ile model oluşturulmaktadır. Daha sonra dışarıda bırakılan 3 örnek ile öngörü performansı belirlenmektedir. Testler bu şekilde 10,000 kere tekrarlanmakta ($10,000 * 3 = 30,000$ öngörü) ve modelin 27,260 (sonuçlar en yakın 10'a yuvarlanmaktadır) öngöründe başarılı olduğu (%95 güven aralığında %0.3 hata payı ile %91) anlaşılmaktadır. Breiman Ağacı modelinin Aşırı Uygunluk sorununun %2 ($\%93 - \%91 = \%2$) seviyesinde olduğu görülmektedir.

5.6.6 Breiman Ağacı modelinde “Karmaşıklık Parametresi” etkisi

Breiman Ağacı modeli oluşturulurken optimal Karmaşıklık Parametresi değerinin belirlenmediği durumlarda model kütüphanesi bu parametre için 0.0001’i (on binde bir) ön değer olarak kabul etmekte, bu değer de aşırı büyük Breiman Ağacı modellerinin oluşmasına neden olabilmektedir (Örnek: Şekil 5.13).

Optimal cp tayininin ne derece önemli olduğunu kavrayabilmek için Şekil 5.13’teki Breiman Ağacı modelinin performansını irdelemek faydalı olacaktır.



Şekil 5.13: Breiman Ağacı Batıklık Öngörüsü (Aşırı Büyük)

Çizelge 5.32’de son Breiman Ağacı modelinin Model İçi Performans’ı sergilenmekte ve bu değer %100 olduğu anlaşılmaktadır. Model, kendi iç verisi ile hiçbir sınıflandırma hatası yapmamaktadır, ancak bağımsız veri ile gerçekleştirilen performans testleri (Çapraz Doğrulama) bundan çok farklı neticelenmektedir.

Çizelge 5.32: Aşırı Büyük Breiman Ağacı – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%50	%0
	Batık	%0	%50

Şekil 5.12’de sergilenen Breiman Ağacı modeli için kullanılan Çapraz Doğrulama testleri bu model (Şekil 5.13) için de tekrarlanmaktadır. İlk model için %91 seviyesinde gerçekleşen Çapraz Doğrulama performansının bu modelde %84’e (30,000 denemede 25,180 başarılı öngörü) gerilediği tespit edilmektedir. Model açık şekilde Aşırı Uygunluk ($\%100 - \%84 = \%16$) sergilemekte ve oldukça kompleks bir yapıya sahip olmasına rağmen pratikteki performansının ciddi oranda (%84) gerilediği (%95 güven aralığında %0.4 hata payı) gözlenmektedir.

Çalışmada başvuru iki Breiman Ağacı modelinin performansları Çizelge 5.33’te karşılaştırılmaktadır. 14 yapraklı Breiman Ağacı açıkça Aşırı Uygunluk sorunu ($\%100 - \%84 = \%16$) sergilerken, 4 yapraklı modelde bu problem sadece %2 ($\%93 - \%91 = \%2$) seviyesindedir.

Çizelge 5.33: Breiman Ağaçları Performans Karşılaştırması

Yaprak Sayısı	Model İçi Performans (%)	Çapraz Doğrulama Performansı (%)
14	100	84
4	93	91

Bu çizelgeden çıkarılacak belki de en önemli sonuç, bir modelin basit olmasının performansı her zaman kötü yönde etkilemediği ve model karmaşıklığının da performansın garantisi olmadığıdır.

Şekil 5.12’deki Breiman Ağacı modeli bu aşamaya kadar incelenen en başarılı model olarak ortaya çıkmaktadır. Model İçi Performans ve Çapraz Doğrulama performansı birbirine çok yakın gerçekleşmekte ve Aşırı Uygunluk sorunu sadece %2 nispetinde yaşanmaktadır. Modelin %91’lik Çapraz Doğrulama performansı ve kolay anlaşılır olması pratikte de faydalı olabileceğine işaret etmektedir. Eldeki model, süreklilik riski öngörüsünde hangi değişkenlerin önemli ve bu değişkenler için hangi

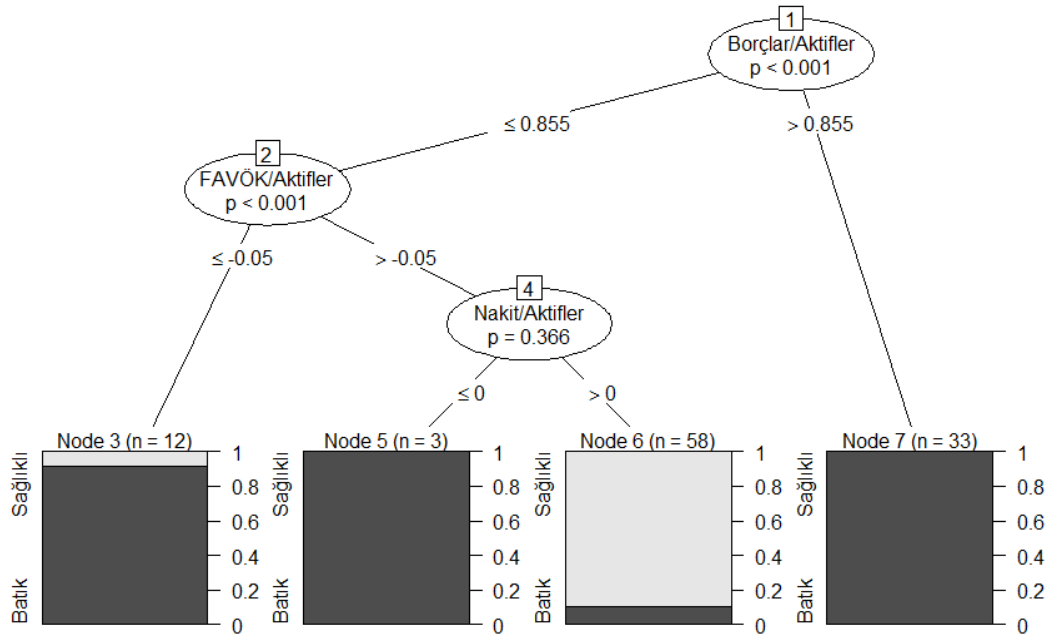
seviyelerin kritik olabileceği konusunda da bir fikir verebilmektedir.

5.6.7 Hothorn Ağacı

Hothorn Ağacı modelinin oluşturulabilmesi için eldeki verinin dönüştürülmesi gerekmemektedir. Örnekleme ait değişkenler orijinal hali ile kullanılabilir. Modelleme sırasında ortaya çıkan tüm değerler pratikte anlamı olan seviyelerdir. Çalışmada R istatistik dili ve “party” kütüphanesi kullanılmaktadır.

Hothorn Ağacı modelinin veriye uygulanması neredeyse otomatiktir. Modelleme esnasında kritik parametre düzenlemelerine ihtiyaç yoktur ve fabrika ayarları genelde makul sonuçlar üretebilmektedir. Modelin veriye uygulanması neticesinde elde edilen Hothorn Ağacı modeli Şekil 5.14’te sergilenmektedir. Çalışmanın bundan önceki kısımlarında olduğu gibi, işletmenin sürekliliği kabulünün ortadan kalktığı örnekler “batık”, süreklilik riskinin yaşanmadığı örnekler ise “sağlıklı” olarak işaretlenmektedir.

Hothorn Ağacı Batıklık Öngörüsü



Şekil 5.14: Hothorn Ağacı Batıklık Öngörüsü

Bu model çerçevesinde sınıflandırma ilk olarak eldeki örneğin Borçlar/Aktifler oranı üzerinden gerçekleştirilmektedir. Bu oran 0.855’ten yüksek ise örnek muhtemel bir batık olarak öngörülmektedir (başarı oranı %100). İlk kuralın dışında kalan örnekler için FAVÖK/Aktifler oranı kontrol edilmekte, bu oran -0.05’e eşit veya daha düşük

ise eldeki örnek yine muhtemel bir batık olarak işaretlenmektedir (başarı oranı %92) İlk iki kuralın dışında kalanlar için ise Nakit/Aktifler oranına bakılmakta ve bu değer 0.0'dan büyük ise eldeki örneğin sağlıklı (başarı oranı %90), aksi halde ise batık olduğu öngörülmektedir (başarı oranı %100). Bu kurallar Breiman Ağacı'nın ürettiği kurallara son derece yakındır.

Modele dahil değişkenlerin sonuca katkısı (Çizelge 5.34), kaç örneğin sınıflandırmasında rol oynadıkları ile tespit edilmektedir. Çizelge 5.34'te sonuca en yüksek katkıyı sağlayan Borçlar/Aktifler oranının ağırlığı normalize edilmekte (%100'e eşitlenmekte) ve diğer değişkenlerin katkısı buna oranla sıralanmaktadır. Bu çizelge Breiman Ağacı modeline ait olan ile tamamen örtüşmektedir.

Çizelge 5.34: Değişkenlerin Sonuca Katkısı

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Borçlar / Aktifler	100
FAVÖK / Aktifler	44
Nakit / Aktifler	34

Şekil 5.14'teki Hothorn Ağacı'nın Model İçi Performans'ı Çizelge 5.35'te sergilenmekte, burada satırlar model öngörüsünü, sütunlar ise gerçek durumu temsil etmektedir. Öngörü ile gerçek durumun kesiştiği hücreler (%49 + %44 = %93) Model İçi Performans'a işaret ederken, bunların dışında kalan hücreler (%1 + %6 = %7) model öngörüsünün başarısız olduğu örnekleri temsil etmektedir.

Çizelge 5.35: Hothorn Ağacı – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%49	%6
	Batık	%1	%44

Hothorn Ağacı modeli için yapılan her bir Çapraz Doğrulama testinde rastgele seçilen 3 örnek dışarıda bırakılmakta, gerisi ile model oluşturulmaktadır. Daha sonra dışarıda bırakılan 3 örnek ile öngörü performansı belirlenmektedir. Testler bu şekilde 10,000 kere tekrarlanmakta (10,000 * 3 = 30,000 öngörü) ve 26,710 öngörünün

(sonuçlar en yakın 10'a yuvarlanmaktadır) başarı ile gerçekleştiği (%95 güven aralığında %0.3 hata payı ile %89 olarak) anlaşılmaktadır. Hothorn Ağacı modelinin Aşırı Uygunluk sorununun %4 (%93 - %89 = %4) seviyesinde olduğu görülmektedir.

Hothorn Ağacı modelinin Çapraz Doğrulama performansı görece yüksek bir başarı seviyesine işaret etse de, Breiman Ağacı performansının (%91) gerisinde kalmaktadır. Modelin bu başarıya kullanılan kütüphanenin ön değerleri ile ulaştığı, yani fabrika ayarları ile bu performansın yakalandığı not edilmelidir.

5.6.8 Breiman Ormanı

Bir Breiman Ormanı modeli (orman) çok sayıda Breiman Ağacı modelinden (ağaç) oluşmaktadır. Ormanlardaki ağaç sayısı seçimlidir. Çok fazla ağaç olması her zaman sonuçları iyileştirmemektedir, gereğinden az olması ise sonuçlarda kayda değer salınımlara sebep olabilmektedir. Doğru ağaç sayısının seçimi özellikle Çapraz Doğrulama testleri için önemlidir. Bu testlerde yüzlerce kez modelin baştan kurulması gerekmektedir. Ormanın çok büyük belirlenmesi Çapraz Doğrulama testlerinin pratikte yapılamaması ile sonuçlanabilmektedir.

Bir ormanda kaç ağacın tutarlı sonuçlar üreteceği genelde değişken ve örnek adedine bağlı olmaktadır. Bunlar arttıkça ormanın daha fazla ağaca sahip olması gerekebilmektedir. Mevcut çalışma çerçevesinde ve denemeler sonucunda ormanda kullanılması gereken optimal ağaç sayısının yaklaşık olarak 2000 olduğu sonucuna varılmaktadır. Orman 2000 ağaç civarında tutarlı sonuçlar üretebilmekte, bunun altındaki adetlerde performansında dalgalanmalar görülmektedir. Model için gerçekleştirilen Çapraz Doğrulamalar 2000 ağaçlı ormanlar ile yapılmaktadır.

Breiman Orman modeli değişken önemi analizini iki farklı kıstasa göre gerçekleştirebilmektedir. Bunlar “Doğru Sınıflama” ve “Yaprak Saflığı” olarak adlandırılmaktadır. Yaprak Saflığı (Node Purity) Gini Endeksi vasıtası ile hesaplanmakta ve bir yapraktaki örneklerin hangi oranda aynı sınıfa ait oldukları ile ölçülmektedir. Her iki analiz de aşağıdaki çizelgelerde sunulmaktadır. Bu çizelgelerde en kuvvetli değişkenin sonuca katkısı normalize edilmekte (%100'e eşitlenmekte) ve diğer değişkenlerin katkısı buna oranla sıralanmaktadır.

Çizelge 5.36: Değişkenlerin Sonuca Katkısı “Doğru Sınıflama”

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Özkaynak / Borçlar	100
Borçlar / Aktifler	98
FAVÖK / Aktifler	90
Nakit / KV Borçlar	59
Nakit / Borçlar	54
FAVÖK / Borçlar	51
Aktif Karlılığı	50
Nakit / Aktifler	46

Çizelge 5.37: Değişkenlerin Sonuca Katkısı “Yaprak Saflığı”

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Özkaynak / Borçlar	100
Borçlar / Aktifler	96
FAVÖK / Aktifler	44
FAVÖK / Borçlar	36
Aktif Karlılığı	35
Nakit / KV Borçlar	26
Nakit / Borçlar	19
Nakit / Aktifler	17

2000 ağaçlı Breiman Ormanı'nın Model İçi Performans'ı Çizelge 5.38'de sergilenmekte, burada satırlar model öngörüsünü, sütunlar ise gerçek durumu temsil etmektedir. Öngörü ile gerçek durumun kesiştiği hücreler (%43 + %43 = %86) Model İçi Performans'a işaret ederken, bunların dışında kalan hücreler (%7 + %7 = %14) model öngörüsünün başarısız olduğu örnekleri temsil etmektedir. Diğer modellerden farklı olarak bu çizelge, model oluşumu için faydalanılan kütüphane tarafından bizzat üretilmekte ve ayrı bir analizin yapılmasına ihtiyaç duyulmamaktadır.

Çizelge 5.38: Breiman Ormanı – Model İçi Performans

	Gerçek		
		Sağlıklı	Batık
Öngörü	Sağlıklı	%43	%7
	Batık	%7	%43

Modelin bağımsız veri ile Çapraz Doğrulaması yine rastgele seçilmiş 3 örneğin dışarıda tutulması ile gerçekleştirilmektedir. Geri kalan veri ile yeni bir orman oluşturulmakta ve model bu üç örnek üzerinde test edilmektedir. Testler bu şekilde 3000 defa tekrarlanmaktadır. Çapraz Doğrulama performansı karışıklık matrisi sonuçlarından daha iyi neticelenmekte ve modelin başarısı %86 yerine %88 (9,000’de 7,960 başarılı öngörü) olarak tespit edilmektedir (%95 güven aralığında %0.6 sapma).

Araştırmada ilk defa bir model için Çapraz Doğrulama testleri Model İçi Performans’tan daha yüksek sonuçlar üretmektedir. Breiman Ormanları’nın oluşumuna ve kendi içinde yaptığı doğrulama testlerine bakıldığında bunun nedeni hemen ortaya çıkmaktadır. Tek tek ağaçlar oluşturulurken verinin sadece bir kısmı kullanılmakta, bu da mevcut çalışma çerçevesinde modelin yeteri kadar eğitilememesi ile sonuçlanmaktadır. Dışarıda bırakılan veri ile yapılan testler ise gerçek durumdan daha tutucu sonuçlar üretmektedir.

Model için gerçekleştirilen analizler iki önemli noktayı ortaya çıkarmaktadır. Bunlardan ilki örnek sayısının yetersizliğidir. Orman oluşumu sırasında kullanılan veri miktarının ağaçların optimal eğitimi için yeterli olmadığı anlaşılmaktadır. Analizler vasıtası ile ulaşılan önemli bir diğer sonuç ise elde edilen öngörü performansının simetrik olmasıdır. Model performansının hem batık, hem de sağlıklı örnekler için denk olduğu görülmektedir (Çizelge 5.38).

5.6.9 Hothorn Ormanı

Bir Hothorn Ormanı modeli (orman) çok sayıda Hothorn Ağacı modelinden (ağaç) oluşmaktadır. Hothorn Ormanı modelinin optimizasyonu esnasında karar verilmesi gereken tek parametre ormandaki ağaç sayısıdır ve bunun dışındaki tüm parametreler için model kütüphanesinin ön değerleri kullanılabilir. Ormanda bulunması gereken ağaç sayısı ise sadece bir kaç simülasyon sonucunda belirlenebilmektedir.

Yapılan denemeler sonucunda ormanın 500 ağaç civarında tutarlı sonuçlar ürettiği görülmektedir. Ormanın performansı 500 ağacın altında dalgalanmalar sergilemektedir. Bu model için gerçekleştirilen Çapraz Doğrulamalar 500 ağaçlı ormanlar ile yapılmaktadır.

Hothorn Ormanları'nda değişken önemi analizi değişkenler arasındaki kısmi korelasyonların da göz önüne alınması ile yapılabilmektedir. Bu sayede yüksek korelasyona sahip değişkenler arasından sonuca en yüksek katkıyı yapanları belirleyebilmek ve diğerlerini model dışına çıkarabilmek mümkün olabilmektedir. Bu analize diğer çok değişkenli Makine Öğrenmesi modellerinin kullanabileceği değişken sayısının azaltılmasında da başvurulabilmektedir.

Değişkenlerin sonuca katkısının sunulduğu Çizelge 5.39'da modelin gerçekleştirdiği sınıflandırma öngörüsüne en kuvvetli katkıyı yapan Borçlar/Aktifler oranının ağırlığı normalize edilmekte (%100 olarak belirlenmekte), diğer değişkenlerin sonuca katkısı buna oranla sıralanmaktadır. Aynı çizelgeden takip edilebileceği gibi, değişkenlerin sonuca katkısı analizi listenin sonundaki iki değişkenin model performansı açısından neredeyse “önemsiz” olduğunu ortaya koymaktadır. Breiman Ormanı değişken önemi analizleri ise bu durumu keşfedememiş görünmektedir. Çalışmanın başında bu araştırmada kullanılacak değişken sayısının Hothorn Ormanı değişken önemi analizi vasıtası ile belirlenmesinin de ana nedeni budur.

Çizelge 5.39: Değişkenlerin Sonuca Katkısı

Değişken Tanımı	En Kuvvetli Değişkene Kıyasla Sonuca Katkısı (%)
Borçlar / Aktifler	100
Aktif Karlılığı	45
FAVÖK / Aktifler	39
Nakit / Aktifler	19
Özkaynak / Borçlar	13
FAVÖK / Borçlar	10
Nakit / Borçlar	4
Nakit / KVBorçlar	1

500 ağaçlı Breiman Ormanı'nın Model İçi Performans'ı Çizelge 5.40'ta sergilenmekte, burada satırlar model öngörüsünü, sütunlar ise gerçek durumu temsil

etmektedir. Öngörü ile gerçek durumun kesiştiği hücreler (%48 + %43 = %91) Model İçi Performans'a işaret ederken, bunların dışında kalan hücreler (%2 + %7 = %9) model öngörüsünün başarısız olduğu örnekleri temsil etmektedir. Breiman Ormanı modelinde olduğu gibi bu çizelge, model oluşumu için faydalanılan kütüphane tarafından bizzat üretilmekte ve ayrı bir analizin yapılmasına ihtiyaç duyulmamaktadır.

Çizelge 5.40: Hothorn Ormanı – Model İçi Performans

		Gerçek	
		Sağlıklı	Batık
Öngörü	Sağlıklı	%48	%7
	Batık	%2	%43

Hothorn Ormanı modelinin bağımsız veri ile Çapraz Doğrulaması, rastgele seçilmiş 3 örneğin dışarıda bırakılması ve geri kalan örnekler ile bir modelin oluşturulması ile başlamaktadır. Daha sonra, modelin öngörü performansı dışarıda bırakılan bu üç örnek ile test edilmektedir. Çapraz Doğrulama prosedürü aynı şekilde 3000 defa tekrarlanmaktadır (3000 * 3 = 9000 öngörü testi). Modelin Çapraz Doğrulama performansı %90 (9,000'de 8090 başarılı öngörü) seviyesinde tespit edilmekte ve bunun Model İçi Performans seviyesine (%91) oldukça yakın olduğu (%95 güven aralığında %0.6 sapma) görülmektedir. Hothorn Ormanı modelinin kendi Model İçi Performans analizi Breiman Ormanı modeline kıyasla daha yansız sonuç üretmekte ve bağımsız Çapraz Doğrulama performansı ile uyum üst seviyede gerçekleşmektedir.

Hothorn Ormanı modeli en başarılı Karar Ormanı modeli olarak ortaya çıkmaktadır. Model İçi Performans Çapraz Doğrulama performansına çok yakın gerçekleşmekte ve Aşırı Uygunluk sorunu (%91 - %90 = %1) bulunmamaktadır. Modelin %90'lık Çapraz Doğrulama performansı pratikte de faydalı olabileceğine işaret etmektedir.

5.7 Araştırmanın Sonuçları

Bu bölümde araştırmada kullanılan modeller performans ve pratikte uygulanabilirlik açısından birbiri ile kıyaslanmakta ve ortaya çıkan en iyi model teorik açıdan irdelenmektedir.

5.7.1 Model performanslarının karşılaştırılması

Çizelge 5.41’de çalışmada kullanılan çok değişkenli Makine Öğrenmesi modellerinin en başarılı alternatifleri bir arada sergilenmektedir.

Çizelge 5.41: Modellerin En Başarılı Alternatiflerinin Karşılaştırılması

Model	Model İçi Performans (%)	Çapraz Doğrulama Performansı (%)
Çoklu Ayraç Analizi	83	81
Lojistik Model	85	81
Sinir Ağları	91	84
Destek Vektör Makineleri	88	87
Breiman Ağacı	93	91
Hothorn Ağacı	93	89
Breiman Ormanı	86	88
Hothorn Ormanı	91	90

Çizelge 5.41 incelendiğinde modellerin Çapraz Doğrulama performansının yukarıdan aşağı doğru genel olarak yükseldiği gözlenmektedir. Bu aynı zamanda tarihsel gelişime de paralellik sergilemektedir. Yeni modellerin zaman içinde giderek daha başarılı sonuçlar üretebildiği gözlenmektedir.

Tarihi gelişimin etkisi Çapraz Doğrulama performansında net olarak görülse de Model İçi Performans sonuçlarından aynı çıkarımların yapılması mümkün olmamaktadır. Bu da Çapraz Doğrulama yönteminin model seçiminde ne kadar önemli olduğunu tekrar gözler önüne sermektedir. Model İçi Performans ölçümleri, Çapraz Doğrulama performansına kıyasla, kötü bir sınav vermektedir. Sadece Model İçi Performans’a dayalı bir çalışmanın çok anlamlı olmayacağı ortaya çıkmaktadır. Örneğin; Lojistik Model ve Breiman Ormanı neredeyse aynı Model İçi Performans’a sahip görünmekte, ancak Çapraz Doğrulama performansı çok farklılaşmaktadır (%7 fark).

Tarihsel gelişim açısından bakıldığında, doğrusal modeller (CAA ve Lojistik Model) %81’lik bir başarı sergilerken, doğrusal olmayan modellere geçişin öncüsü Sinir Ağı modeli %84’lük bir başarı yakalamakta, daha güncel modeller (DVM ve Karar Ağacı bazlı modeller) ise ortalama olarak %89’luk bir başarı sergilemektedir. Bu çıkarımın eldeki veri çerçevesinde değerlendirilmesi gerektiği not edilmelidir.

5.7.2 Modellerin kuvvetli ve zayıf yönlerinin karşılaştırılması

Eldeki çalışma ve ulaşılan sonuçlar çerçevesinde başvuru Makine Öğrenmesi modelleri için tespit edilen kuvvetli ve zayıf yönler Çizelge 5.42’de özetlenmektedir.

Çoklu Ayraç Analizi’nin en kuvvetli yönü kolay anlaşılabilir olmasıdır, ancak modelin uygulanacağı örneklemin değişken dağılımı çok sayıda ön şartı tatmin etmek zorundadır. Modelin optimizasyonundan önce değişkenlerin önemli kısmı dönüşümlerden geçirilmek zorunda kalmakta, bu da model eğitimi esnasında takip edilebilirliği kısıtlamaktadır. Modelin dayandığı teorik altyapı sadece doğrusal ilişkilerin keşfine yöneliktir ve doğrusal olmayanları keşfetmeye uygun değildir.

Lojistik Model, Çoklu Ayraç Analizi ile hemen hemen aynı kuvvetli ve zayıf yönleri paylaşmaktadır. Modelin sınıflandırma öngörüsü yanında, yaptığı öngörünün güvenilirliğini de kullanıcıya sunabilmesi onu bir adım öne çıkartmaktadır.

Sinir Ağları, örneklemin sağlanması gereken dağılım ön şartları bakımından Lojistik Model ile örtüşmektedir. Bunun nedeni Sinir Ağları’nın pek çok Lojistik Model’in paralel ve seri olarak bağlanmasından oluşmasıdır, ancak modelin kuvvetli yönü de yine bu paralel ve seri bağlardan gelmektedir. Model bu sayede doğrusal olmayan, ilk bakışta keşfedilemeyen, ilişkileri de ortaya çıkarabilmektedir. Modele dahil edilen çok sayıda bağ, hem model eğitim süresini hem de ihtiyaç duyulan veri adedini oldukça artırmaktadır. Mevcut araştırma çerçevesinde karşılaşılan en hesap yoğun model Sinir Ağlarıdır. Optimal modele ulaşıldığı halde en yüksek aşırı uygunluğun (%7) da bu modelde yaşanmış olması (Çizelge 5.41), modelin yüksek veri ihtiyacının bir göstergesidir.

Çizelge 5.42: Modellerin Kuvvetli ve Zayıf Yönleri

Model	Kuvvetli Yönleri	Zayıf Yönleri
Çoklu Ayraç Analizi	Kolay anlaşılır	Değişken dağılımlarına yönelik ağır ön şartlar, doğrusal olmayan ilişkiler
Lojistik Model	Kolay anlaşılır	Değişken dağılımlarına yönelik ön şartlar, doğrusal olmayan ilişkiler
Sinir Ağları	Değişkenler arası örtülü ilişkilerin ortaya çıkarılması, doğrusal olmayan ilişkiler	Değişken dağılımlarına yönelik ön şartlar, yüksek sayıda örnek ihtiyacı, işlem yoğunluk, kapalı sistem

Çizelge 5.42 (devamı): Modellerin Kuvvetli ve Zayıf Yönleri

Model	Kuvvetli Yönleri	Zayıf Yönleri
Destek Vektör Makineleri	Doğrusal olmayan ilişkiler, örneğe dayalı sınıflandırma	Değişken dağılımlarına yönelik ön şartlar, parametre sınırlandırma hassasiyeti
Breiman Ağacı	Doğrusal olmayan ilişkiler, ön şartsız değişken dağılımı, kolay anlaşılır	Baskın değişken etkisi, daha çok seviyesi olan değişkenin avantajı
Hothorn Ağacı	Doğrusal olmayan ilişkiler, ön şartsız değişken dağılımı, kolay anlaşılır	Baskın değişken etkisi, yüksek sayıda örnek ihtiyacı
Breiman Ormanı	Doğrusal olmayan ilişkiler, ön şartsız değişken dağılımı, “değişken öneminin” belirlenebilmesi	Daha çok seviyesi olan değişkenin avantajı, kapalı sistem
Hothorn Ormanı	Doğrusal olmayan ilişkiler, ön şartsız değişken dağılımı, “değişken öneminin” belirlenebilmesi	Kapalı sistem

Destek Vektör Makineleri, modellemede kullanılacak verinin sağlaması gereken ön şartlar bakımından, Lojistik Model ile örtüşmektedir. Bu bakımdan Lojistik Model’de yaşanan benzer zorluklar DVM için de geçerlidir. DVM’nin diğer çok değişkenli Makine Öğrenmesi modellerinden en önemli farkı hem değişken bazlı (diğer modellerin tamamı sadece değişken bazlıdır) hem de örnek bazlı olarak eğitilebilmesidir. Model, bu araştırmada olduğu gibi, örnek bazlı olarak eğitildiğinde doğrusal olmayan ilişkileri de keşfedebilmekte ve bu da en önemli avantajlarından biri olarak ortaya çıkmaktadır. Modelin örnek bazlı eğitilmesi iki parametrenin (cost ve gamma) eldeki veriye uygun olarak optimize edilmesini gerektirmektedir. Bu parametreler yerine model kütüphanesinin ön değerleri kullanıldığında Çapraz Doğrulama performansı gerilemektedir. Modelin cost ve gamma parametreleri eldeki problem ve örneklem çerçevesinde ciddi oranda farklılaşabilmekte, bu da modelin zayıf bir yönü olarak ortaya çıkabilmektedir.

Breiman Ağacı modeli örneklem üzerinde hiçbir ön şart koşmamakta, doğrusal olmayan ilişkileri keşfedebilmekte ve sonuçta anlaşılması kolay bir modele ulaşılabilir. Bunlar modelin en kuvvetli yönleri olarak ortaya çıkmaktadır. Model, kullandığı özyinelemeli bölümlendirme algoritması nedeniyle baskın değişkeni her zaman ilk ayraç olarak seçmekte ve bu seçim nihai model performansı açısından optimal olmayabilmektedir. Model, yine aynı algorithmadan dolayı, daha çok seviyeye ayrılmış değişkenlere de öncelik tanıyabilmektedir.

Hothorn Ağacı modeli, kuvvetli ve zayıf yönleri açısından, Breiman Ağacı modeli ile neredeyse tamamen örtüşmektedir. Modelin daha çok seviyeye ayrılmış değişkenlere avantaj sağlamaması onu teoride bir adım öne çıkarabilmektedir. Model bu avantaja rağmen, mevcut çalışma çerçevesinde, Breiman Ağacı modelinden daha başarılı sonuçlar üretememektedir. Bu durum modelin ihtiyaç duyduğu örneklem miktarı ile açıklanabilmektedir. Hothorn Ağacı modeli için ortaya çıkan Aşırı Uygunluk seviyesinin %4 (Çizelge 5.41) gibi görece yüksek bir seviyede kalması örneklemin yeterli olmadığına açıkça işaret etmektedir.

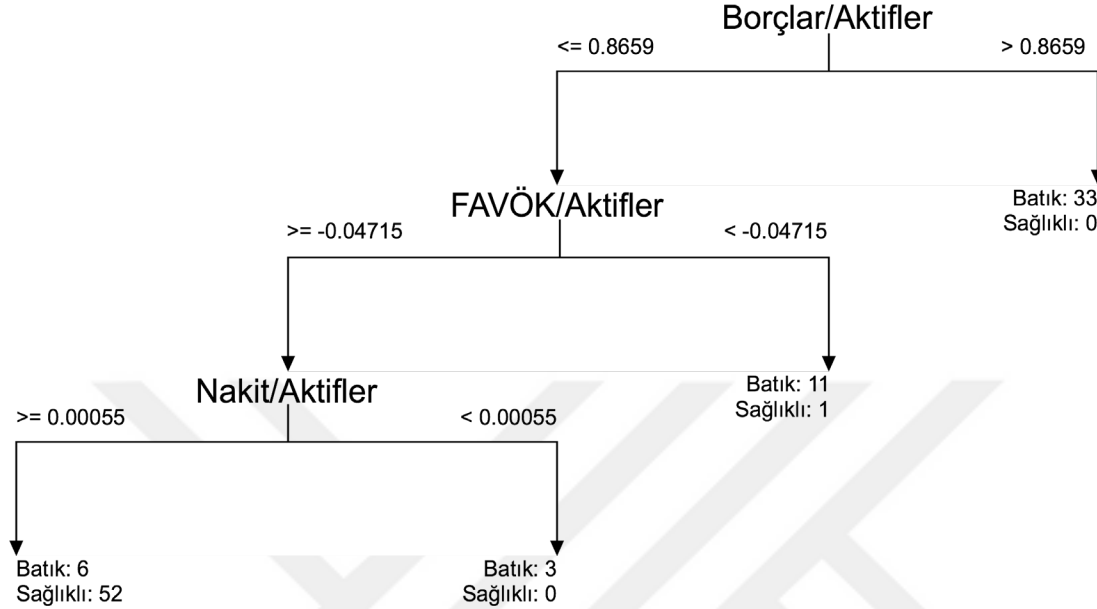
Breiman Ormanı modeli veri üzerinde hiçbir ön şart koşmamakta, doğrusal olmayan ilişkileri keşfedebilmekte ve baskın değişkene avantaj sağlamamaktadır. Bunlara ek olarak değişkenlerin sonuca hangi nispette katkı yaptığı hakkında bir sıralama da sunabilmektedir. Modelin zayıf yönlerinden biri, Breiman Ağacı modelinde olduğu gibi, daha çok seviyeye ayrılmış değişkenlere öncelik tanıyabilmesidir. Breiman Ormanı'nın diğer bir zayıf yönü ise oldukça kapalı ve takip edilmesi zor bir model üretmesidir.

Hothorn Ormanı modeli, kuvvetli ve zayıf yönleri açısından, Breiman Ormanı modeliyle neredeyse bire bir örtüşmektedir. Tek fark, Breiman Ağacı bazlı modellere hakim olan “daha çok seviyeye ayrılmış değişkenin avantajı” sorununun bu modelde bulunmayışıdır. Bu durum modele hem teoride, hem de pratikte (Breiman Ormanı modeline kıyasla) avantaj sağlamaktadır.

Çizelge 5.42 tarihsel perspektifte incelendiğinde genel olarak zaman içinde değişkenler ile ilgili ön koşulların azaldığı, modellerin kullanıcıya daha fazla bilgi sunduğu ve modelleme başarısının yükseldiği gözlenmektedir.

5.7.3 En başarılı model ve bu modelin süreklilik riski öngörüsü

Araştırmada başvurulan çok değişkenli Makine Öğrenmesi modelleri arasından, mevcut çalışma çerçevesinde, en başarılı olan bir Breiman Ağacı'dır (Şekil 5.15). Bu model son derece basit ve anlaşılır ayraç kuralları kullanmaktadır.



Şekil 5.15: En Başarılı Model

Şekil 5.15'te sunulan modelin ortaya koyduğu kurallar çerçevesinde sınıflandırma öncelikle eldeki örneğin Borçlar/Aktifler oranı üzerinden gerçekleşmektedir. Bu oran 0.8659'a eşit veya daha yüksek ise örnek muhtemel batık adayı olarak sınıflandırılmaktadır (başarı oranı %100). İlk kuralın dışındaki örnekler için FAVÖK/Aktifler oranına bakılmakta, bu oran -0.04715'ten daha düşük ise eldeki örnek yine muhtemel bir batık adayı olarak işaretlenmektedir (başarı oranı %92). İlk iki kuralın dışında kalanlar için ise Nakit/Aktifler oranı takip edilmekte ve bu değer 0.00055'ten daha düşük ise eldeki örneğin batık (başarı %100), aksi halde ise sağlıklı olduğu öngörülmektedir (başarı oranı %90).

En önemli değişken Borçlar/Aktifler oranı olarak ortaya çıkmaktadır. Eğer Borçlar/Aktifler oranı %87 ve üzerinde ise ciddi bir süreklilik riski (batık olma ihtimali) ortaya çıkmakta, ancak şirketin Borçlar/Aktifler oranı %87'den az ise hemen "sağlıklı" kanaati de oluşmamaktadır. FAVÖK/Aktifler oranı bu aşamadaki en önemli ikinci finansal oran olarak ortaya çıkmaktadır. Bu oranın -%5'ten düşük olması süreklilik riskinin devam ettiğini göstermektedir. Herhangi bir şirket ilk iki budakta pozitif tarafta kalsa bile (henüz batık olarak sınıflandırılmamakta), şirkette

bir miktar nakdin tutulması nihai önemli gösterge olarak ortaya çıkmaktadır.

Bulguları genel anlamda derlemek gerekir ise, eldeki model çerçevesinde sağlıklı bir şirketin özellikleri şöylece sıralanabilecektir: şirketin yaptığı yatırımlara oranla yeterli öz sermayesi olmalı, mali performansı “çok kötü” olmamalı ve her zaman bir miktar nakit taşıyabilmeli. Ortaya çıkan model çerçevesinde, eğer şirketin aktiflerine oranla yeterli sermayesi yoksa (borçluluk yüksek ise) veya ciddi miktarda zarar ediyorsa veya elinde neredeyse hiç nakit bulundurmuyor ise, süreklilik riskleri ciddi seviyede yükselmiş olarak kabul edilmektedir. Modelin önerdiği birinci kıstas girişimin aşırı borçlu olup olmamasına, ikincisi şirket yönetiminin performansına ve son olarak üçüncüsü de kısa vadeli taahhütlerin karşılanabilir olup olmamasına odaklanmaktadır.

Bir yönü ile 2001’de Neophytou ve diğ.’nin savundukları nokta teyit edilmektedir: karlılığı, nakdi ve kaldıracı gösteren üç gösterge gerçekten de batıkların önemli kısmını açıklayabilmektedir. Mevcut çalışmanın bu çalışma ile en önemli farkı ise performans kaynaklıdır. Neophytou ve diğ. çalışmalarında sadece Lojistik Model ve Sinir Ağı modellerini kullanmakta ve model öngörü performansları %83 ile sınırlı kalmaktadır, mevcut çalışmada ise bu modellerin yanına Karar Ağaçları’na dayalı modeller de eklenmekte ve öngörü performansı %91’e kadar yükselmektedir. Hatırlatmak gerekir ki, mevcut çalışmada da Sinir Ağları’nın performansı %84’te kalmakta ve Neophytou ve diğ.’nin 2001’de ulaştıkları başarıya çok benzer bir performans sergilenmektedir.

Breiman Ağacı gibi sade bir modelin %91’lik Çapraz Doğrulama performansına erişebilmesi gelecekte yapılabilecek benzer analizler için de ümit vericidir. Model, kolay eğitilebilir, kolay anlaşılabilir ve kolay uygulanabilir özellikleriyle ön plana çıkmaktadır.



6. SONUÇ VE ÖNERİLER

Tezin amacı halka açık finans dışı şirketlerde süreklilik riskinin “Makine Öğrenmesi” modelleri ile öngörülmesidir. Bu temel amaç yanında, araştırmada faydalanılan Makine Öğrenmesi modellerinin pratikte uygulanabilirliklerinin irdelenmesi bir diğer amacı oluşturmaktadır.

Bu tez, giriş ve sonuç bölümleri hariç, dört bölümden oluşmaktadır. Bunlar; Kavramsal Çerçeve, Literatür İncelemesi, Çalışmada Kullanılan Makine Öğrenmesi Modelleri ve Halka Açık Finans Dışı Şirketlerde Süreklilik Riskinin Makine Öğrenmesi ile Öngörülmesi’dir.

Tezin giriş bölümünde bağımsız denetim raporlarının süreklilik risklerini ne ölçüde öngörebildiğine dair son kırk yıl içinde yapılan çalışmaların bir özeti sunulmakta, mesleki yargı yanında Makine Öğrenmesi modellerinin bu konuya yapabileceği katkıya değinilmekte ve bu katkının elle tutulur nispette değerli olabileceği ileri sürülmektedir.

Tezin ikinci bölümünde bu araştırmada sıkça başvurulan temel kavramlar detaylı olarak açıklanmaktadır (Bağımsız Denetim, Bağımsız Denetçi, Denetim Raporu, Halka Açık Finans Dışı Şirketler, İşletmenin Sürekliliği, Süreklilik Riski, Finansal Risk, Finansal Süreklilik Riski, Makine Öğrenmesi, Değişken Önemi, Şartlı Önemlilik, Karışıklık Matrisi, Model İçi Performans, Çapraz Doğrulama, Aşırı Uygunluk, Logit Fonksiyonu, Probit Fonksiyonu, Özyinelemeli Bölümlendirme, Sınıflandırma Aracı, Çoklu Ayraç Analizi, Lojistik Model, Sinir Ağları, Destek Vektör Makineleri, Karar Ağacı ve Karar Ormanı). Araştırmanın devamında yapılan analizler bu kavramlar üzerine bina edilmektedir.

Tezin üçüncü bölümünde süreklilik risklerinin öngörülmesinde istatistiki modellerin tarihsel perspektifte ne seviyede kullanılabildiği ve bu modellerin gelişmesine paralel olarak öngörü performansının bundan nasıl etkilendiği özetlenmektedir. Bu bölüm kendi içinde ikiye ayrılmaktadır ve ilk bölümde Makine Öğrenmesi modellerinin tarihi gelişimi takip edilirken, ikinci bölümde bu modellerin eldeki veriye uygulanması esnasında karşılaşılan problemlere değinilmektedir. Makine Öğrenmesi

modellerinin tarihçesi de kendi içinde, değişken adedi ve doğrusallık açısından, üç bölüme ayrılmaktadır.

Tezin dördüncü bölümünde çalışmada kullanılan Makine Öğrenmesi modellerine (Çoklu Ayraç Analizi, Lojistik Model, Sinir Ağları, Destek Vektör Makineleri, Breiman Ağacı, Hothorn Ağacı, Breiman Ormanı, Hothorn Ormanı) detaylı açıklamalar getirilmekte, uygulamalarına kantitatif ve görsel örnekler verilmekte, muhtemel kuvvetli ve zayıf yönleri hakkında öncü görüşler sunulmaktadır.

Tezin beşinci bölümünde halka açık finans dışı şirketlerin süreklilik risklerinin Makine Öğrenmesi modelleri ile öngörülmesi amaçlanmaktadır. Araştırmanın bu bölümünde faydalanılan modeller sırası ile Çoklu Ayraç Analizi, Lojistik Model, Sinir Ağları, Destek Vektör Makineleri, Karar Ağaçları (Breiman ve Hothorn) ve Karar Ormanlarıdır (Breiman ve Hothorn). Çalışma kapsamında kurulan modellere 1999-2016 döneminde, Türkiye’de yerleşik, halka açık ve finans sektörü dışındaki şirketler dahil edilmekte ve toplam 106 örnekten oluşmaktadır (53 batık ve 53 sağlıklı şirket). Her örnek için faaliyet performansını, kaldıracı ve nakit dengesini işaret eden 12 finansal oran (Borçlar/Aktifler, Cari Oran, Nakit/Aktifler, Aktif Karlılığı, Çalışma Sermayesi/Aktifler, Öz Kaynak/Borçlar, Nakit Akımı/Aktifler, Nakit Akımı/Borçlar, FAVÖK/Aktifler, FAVÖK/Borçlar, Nakit/Borçlar, Nakit/ KV Borçlar) analizlere dahil edilmektedir. Çalışma verisi ağırlıklı olarak Rasyonet A.Ş. şirketi bünyesindeki EquityRT veri tabanından sağlanmaktadır. Modellemeler, analizler ve testler R istatistik dili ve ilgili R Makine Öğrenmesi model kütüphaneleri kullanılarak gerçekleştirilmektedir. Modellerin performans testleri eldeki örneklerin bağımsız “model” ve “test” gruplarına ayrılması yöntemi (Çapraz Doğrulama) ile yapılmakta ve bu testler yeteri kadar tekrarlanarak elde edilen sonuçların %95 güven aralığında %1 hassasiyete ulaşması sağlanmaktadır.

Beşinci bölümde gerçekleştirilen analizler sonucunda, tarihi perspektif içinde, Makine Öğrenmesi modellerinin giderek pratiğe daha kolay uygulanabildiği ve daha başarılı sonuçlar üretebildiği ileri sürülmektedir. Aynı analizler en başarılı modelin bir Breiman Ağacı olduğunu da ortaya çıkmaktadır. Breiman Ağacı modelinin Çapraz Doğrulama performansına göre pratikte ulaşabileceği başarı %91 olarak belirlenmektedir. Bu oran, geri kalan modellerin ortalama Çapraz Doğrulama performansının (%84) %7 üzerindedir. Breiman Ağacı modelinin %91’lik Çapraz Doğrulama performansı modelin pratikte de faydalı olabileceğine işaret etmektedir.

Model bu başarıya ulaşırken şirketin yalnızca halka açık finansal bilgilerinden faydalanmaktadır. Breiman Ağacı kolay eğitilebilir, kolay anlaşılabilir ve kolay uygulanabilir bir model olarak ortaya çıkmaktadır. Modelleme yapmak için eldeki verinin dönüşümlerden geçirilmesine de ihtiyaç duyulmamaktadır. Modelin bu başarısını benzer sınıflandırma problemlerinde de tekrarlayabileceği öngörülmektedir. Modelin pratikte uygulanabilmesi karşısında büyük bir zorluk öngörülmemektedir. Model nezdinde ortaya çıkan kurallar kolay anlaşılabilir, kolay anlatılabilir ve güncel risk takibi sistemlerine kolayca eklenebilir yapıdadır.

Moses ve Liao'nun 1987'de vurguladığı gibi, süreklilik riski öngörüsü modellerine gözü kapalı güvenmek son derece hatalı olabilir ve sonuçların bir mantık süzgecinden geçirilmesi gerektiği açıktır. Mevcut çalışma neticesinde ortaya çıkan modelin kapalı bir kutu olmaması, süreklilik riski öngörüsünü kabul edilebilir bir mantık silsilesi içinde gerçekleştirebilmesi, ona olan güveni artırabilecek unsurlardan bir kaçıdır.

Araştırmanın sonucunda belirlenen Breiman Ağacı modeli muhasebe bilgi kullanıcılarına (yöneticiler, ortaklar, kredi kuruluşları, çalışanlar, potansiyel ortaklar, devlet ve toplum) süreklilik riski öngörüsü açısından değerli bir analiz imkanı sunabilecektir. Bu modelden bağımsız denetimlerde süreklilik riskinin tespitinde, yatırım kararı alırken finansal riskin ve ilgili yatırımdan elde edilmesi gereken minimum verimin tayininde, işletmede yüksek finansal risk oluştuğunda kurumsal tavrın belirlenmesinde ve finansal riskin azaltılması gerektiğinde en doğru yolun seçilmesinde, portföy yönetiminde varlıkların risk açısından sınıflandırılmasında ve alacak portföylerinin kredi risklerinin belirlenmesinde faydalanılabilecektir.

Bağımsız denetimlerde süreklilik risklerinin belirlenmesi ağırlıklı olarak mesleki yargı ile gerçekleştirilmektedir. Mesleki yargının bağımsız denetçiler arasında farklılıklar sergilemesi ve bu durumun eldeki bağımsız denetim raporlarının süreklilik öngörülerinde tutarsızlıklara neden olması doğaldır. Bu tez sonucunda elde edilen modelin objektif kriterler çerçevesinde ikinci bir göz vazifesi gösterebilmesi, farklı denetçilere ait raporların tutarlılığına olumlu yönde etki etmesi beklenmektedir.

Bağımsız denetim raporu olmayan şirketler için bu modelin kullanılması, elde başka hiç bir süreklilik riski öngörüsü bulunmaması halinde, önemli bir gösterge olarak

muhasebe bilgi kullanıcılarının karar mekanizmalarında yer alabilecektir. Yüksek bir öngörü performansına (%91) sahip olan model bu alanda da faydalı olabilecektir.

Tez sonucunda elde edilen Breiman Ağacı modeli yatırım alternatifleri arasında görece riski düşük olanların belirlenmesinde de faydalı olabilecektir. Riskli olarak sınıflandırılan örneklerin dışarıda bırakılması elde kalan yatırım alternatiflerinden daha yüksek verimin elde edilmesine neden olacaktır.

Şirketler bu modelden kendi finansal yapılarının risklilik düzeyini belirlemede de faydalanabileceklerdir. Model hem risk düzeyini, hem de riski yaratan asıl unsuru belirlemede fayda sağlayabilecektir. Finansal riskin ana kaynağı tespit edildiğinde buna karşı kurumsal tavrın belirlenmesi ve bu yönde hareket edilmesi kolaylaşacaktır.

Alacak portföylerinin kredi riski açısından sınıflandırılması günümüzün önemli meselelerinden biridir ve genelde bu portföye dahil şirketler için bağımsız denetim raporlarına ulaşmak mümkün değildir. Eldeki Makine Öğrenmesi modeli bu soruna oldukça pratik bir çözüm sunmakta ve sadece üç finansal oran ile temel bir risk sınıflandırması gerçekleştirebilmektedir. Yapılan bu sınıflandırma neticesinde fon akımlarının daha tutarlı bir şekilde kestirilebilmesi mümkün olacak, bu da şirketin finans yönetimine ışık tutabilecektir.

Tez sonucunda elde edilen model farklı çalışmaların yapılabilmesini de mümkün kılmaktadır. Bağımsız denetim raporlarının süreklilik risklerini ne ölçüde öngörebildiği ve finansal veriyi hangi nispette kullanabildiği eldeki modelden faydalanılması ile mümkün olabilecektir. Model sadece halka açık veriyi kullanmakta ve bağımsız bir denetçiye kıyasla çok daha dar bir pencereden problemi inceleyebilmektedir. Bu şartlar altında, modelin performansı, bağımsız denetim raporlarının süreklilik riski öngörü başarısı için bir alt sınır olarak kabul edilebilecek ve bu seviyenin altında kalınması halinde bağımsız denetim yaklaşımının güncellenmesi gerekliliğini ortaya çıkarabilecektir.

Bu teze benzer bir çalışmanın uluslararası veri ile tekrarlanması, süreklilik riski öngörüsü modellerinin ülkeler ve sektörler arasında ne gibi farklılıklar gösterebileceğini de ortaya çıkarabilecektir. Verinin genişlemesi ülke veya sektör odaklı araştırmaları mümkün kılacak ve bu durum her bir modelin daha başarılı sonuçlar üretmesini sağlayabilecektir. Mevcut çalışmada faydalanılan veri derinliği

sektör ayrımlı bir arařtırmaya izin vermemektedir, ancak uluslararası bir alıřma bu sorunun üstesinden gelebilecektir.





KAYNAKLAR

- Acar D. ve Senal S.** (2011). Bağımsız Denetim Kalitesinin Arttırılmasında Kamu Gözetim Kurulunun Rolü: Bağımsız Denetim Firmaları Üzerine Bir Araştırma. *Muhasebe ve Denetime Bakış*, 35, 33-51.
- Acar D., Senal S. ve Usul H.** (2011). Bağımsız Denetim Kalitesi: Denetim Firmaları Üzerine Bir Araştırma. *Sosyal ve Ekonomik Araştırmalar Dergisi*, 11 (22), 273-306.
- Agarwal A.** (1993). *Neural Networks and Their Extensions for Business Decision Making*. Yayımlanmamış doktora tezi, Ohio State University, Ohio.
- Akdemir Y.** (2012). Bağımsız Denetimde Kalite. *Dayanışma Dergisi*, 115, 57-58.
- Akgüç Ö.** (2001). *Mali Tablolar Analizi*. İstanbul: Avcıol Basım Yayın.
- Alam P., Booth D., Lee K. ve Thordarson T.** (2000). The Use of Fuzzy Clustering Algorithm and Self-organizing Neural Networks for Identifying Potentially Failing Banks: An Experimental Study. *Expert Systems with Applications*, 18, 185-199.
- Altman E. I.** (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23 (4), 589-609.
- Altman E. I.** (1973). Predicting Railroad Bankruptcies in America. *Bell Journal of Economics and Management Science*, 4 (1), 184-211.
- Altman E. I. ve Loris B.** (1976). A Financial Early Warning System for Over-the-counter Broker Dealers. *Journal of Finance*, 31 (4), 1201-1217.
- Altman E. I., Haldeman R. ve Narayanan P.** (1977). Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance*, 1 (1), 29-51.
- Altman E. I. ve Levallee M.** (1980). Business Failure Classification in Canada. *Journal of Business Administration*, 12 (1), 147-164.
- Altman E. I., Marco G. ve Varetto F.** (1994). Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience). *Journal of Banking and Finance*, 18, 505-29.
- Aslan Ç.** (2007). Uluslararası Denetim Standartlarında Etik. *I. Uluslararası Türkiye Muhasebe Denetimi Sempozyumu, VII. Türkiye Muhasebe Denetimi Sempozyumu bildiriler kitabı* içinde (ss. 325-339). İstanbul: İSMMMÖ.
- Aziz M. A. ve Dar H. A.** (2006). Predicting Corporate Bankruptcy: Where do We Stand? *Corporate Governance: The International Journal of Business in Society*, 6 (1), 18-33.
- Beaver, W.** (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 5, 71-111.
- Bell, T., Ribar G. ve Verchio J.** (1990). Neural Nets Versus Logistic Regression: A Comparison of Each Model's Ability to Predict Commercial Bank Failures. *1990 D&T, University of Kansas Symposium on Auditing Problems bildiriler kitabı* içinde (ss. 29-53). Kansas: University of Kansas.

- Belowary J., Giacomina D. ve Akers M.** (2007). A Review of Bankruptcy Prediction Studies: 1930-Present. *Journal of Financial Education*, 33 (Güz sayısı), 1-42.
- Berk N.** (2010). *Finansal Yönetim*. İstanbul: Türkmen Kitabevi.
- Berkson J.**, (1944). Application of the Logistic Function to the Bio-Assay. *Journal of the American Statistical Association*, 39 (227), 357-365.
- Bliss, C. I.** (1934). The Method of Probits. *Science*, 79, 38-39.
- Blum, M.** (1974). Failing Company Discriminant Analysis. *Journal of Accounting Research*, 12 (1), 1-25.
- Bolak M.** (2001). *Sermaye Piyasası Menkul Kıymetler ve Portföy Analizi*. İstanbul: Beta Yayınevi.
- Bolak M.** (2004). *Risk ve Yönetimi*. İstanbul: Birsen Yayınevi.
- Boritz J. ve Kennedy D.** (1995). Effectiveness of Neural Network Types for Prediction of Business Failure. *Expert Systems with Applications*, 9 (4), 503-512.
- Bozkurt N.** (2011). *Muhasebe Denetimi* (5. bs.). İstanbul: Alfa Yayınevi.
- Breiman L.**, (2001). Random Forests. *Machine Learning*, 45 (1), 5–32
- Breiman L., Friedman J. H., Olshen R. A. ve Stone C. J.** (1984). *Classification and Regression Trees*. Monterey: Wadsworth & Brooks / Cole Advanced Books & Software.
- Bureau of Business Research** (1930). *A Test Analysis of Unsuccessful Industrial Companies*. [Bülten]. Urbana: University of Illinois Press.
- Burges C. J. C.** (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2 (2), 955-974.
- Burges C. J. C. ve Scholkopf B.** (1997). Improving the Accuracy and Speed of Support Vector Machines. Mozer M., Jordan M. ve Petsche T. (Ed.), *Advances in Neural Information Processing Systems bildiriler kitabı* içinde (ss. 475-481), Cambridge: MIT.
- Büker S., Aşıkoğlu R. ve Sevil G.** (2011). *Finansal Yönetim*. Ankara: Sözkesen Matbaacılık.
- Casey C. J.** (1980). The Usefulness of Accounting Ratios for Subjects' Predictions of Corporate Failure: Replication and Extensions. *Journal of Accounting Research*, 18 (2), 603-613.
- Casey C. J. ve Bartczak N.J.** (1984). Cash Flow – It's not the Bottom Line. *Harvard Business Review*, Cilt: Temmuz-Ağustos, 61-66.
- Casey, C. J. ve Bartczak N. J.** (1985). Using Operating Cash Flow Data to Predict Financial Distress: Some Extensions. *Journal of Accounting Research*, 23 (1), 384-401.
- Castagna A. ve Matolcsy Z.** (1981). The Prediction of Corporate Failure: Testing the Australian Experience. *Australian Journal of Management*, 6 (1), 23-50.
- Chen K. ve Shimerda T.** (1981). An Empirical Analysis of Useful Financial Ratios. *Financial Management*, 10 (1), 51-60.
- Chudson W. A.** (1945). *The Pattern of Corporate Financial Structure: A Cross-Section View of Manufacturing, Mining, Trade, and Construction*. New York: National Bureau of Economic Research.
- Coats P. K. ve Fant L.F.** (1993). Recognizing Financial Distress Patterns Using a Neural Network Tool. *Financial Management*, 22, 142-55.
- Cortes C. ve Vapnik V. N.** (1995). Support Vector Networks. *Machine Learning*, 20, 273-297.

- Cox D. R.** (1958). The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society, Seri: B*, 20 (2), 215-242.
- Çaldağ Y.** (2003). *Denetim ve Raporlama*. Ankara: Gazi Kitabevi.
- Çömlekçi F.** (2011). *Muhasebe Denetimi ve Mali Analiz*. Eskişehir: Anadolu Üniversitesi Yayınları.
- Dambolena L. G. ve Khoury S. J.** (1980). Ratio Stability and Corporate Failure. *The Journal of Finance*, 35 (4), 1017-1026.
- Dambolena L. G. ve Shulman J. M.** (1988). A Primary Rule for Detecting Bankruptcy: Watch the Cash. *Financial Analysts Journal*, Cilt: Eylül/Ekim, 74-78.
- Deakin E.** (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10 (1), 167-179.
- Dimitras A., Zanakis S. ve Zopounidis C.** (1996). A Survey of Business Failures with an Emphasis on Prediction Methods and Industrial Applications. *European Journal of Operational Research*, 90, 487-513.
- Edmister R.** (1972). An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *Journal of Financial and Quantitative Analysis*, 7 (2), 1477-1493.
- El Hennawy R. ve Morris R.** (1983). The Significance of Base Year in Developing Failure Prediction Models. *Journal of Business Finance and Accounting*, 10 (2), 209-223.
- Elgin İ.** (2010). Bağımsız Denetimde Hizmet Kalitesi. *Sermaye Piyasası Dergisi*, 2, 116-120.
- Erdogan B. E.** (2013). Prediction of Bankruptcy Using Support Vector Machines: an Application to Bank Bankruptcy. *Journal of Statistical Computation and Simulation*, 83 (8), 1543-1555.
- Erdoğan M.** (2012). *Denetim*. Eskişehir: Anadolu Üniversitesi Yayınları.
- Espahbodi P.** (1991). Identification of Problem Banks and Binary Choice Models. *Journal of Banking and Finance*, 15 (1), 53-71.
- Fan A. ve Palaniswami M.** (2000). Selecting Bankruptcy Predictors Using a Support Vector Machine Approach. *The IEEE-INNS-ENNS International Joint Conference on Neural Networks bildirileri kitabı* içinde (ss. 354-359). Como: IEEE.
- FitzPatrick, P.** (1932). A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Companies. *The Certified Public Accountant, Ekim, Kasım, Aralık* sayıları, sırasıyla 598-605, 656-662 ve 727-731.
- Flagg J., Giroux G. ve Wiggins C. Jr.** (1991). Predicting Corporate Bankruptcy Using Failing Firms. *Review of Financial Economics*, 1 (4), 67-78.
- Fletcher D. ve Goss E.** (1993). Forecasting with Neural Networks: An Application Using Bankruptcy Data. *Information & Management*, 24, 159-167.
- Foreman R. D.** (2002). A Logistic Analysis of Bankruptcy Within the US Local Telecommunications Industry. *Journal of Economics and Business*, 55 (2), 135-166.
- Friedman J. H.** (1977). A Recursive Partitioning Decision Rule for Nonparametric Classification. *IEEE Transactions on Computers*, 26 (4), 404-408.
- Frydman H., Altman E. I. ve Kao D.** (1985). Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. *The Journal of Finance*, 40 (1), 269-291.

- Fulmer J., Moon J., Gavin T. ve Erwin J.** (1984). A Bankruptcy Classification Model for Small Firms. *The Journal of Commercial Bank Lending*, 66 (11), 25-37.
- Gaeremynck A. ve Willekens M.** (2003). The Endogenous Relationship Between Audit Report Type and Business Termination: Evidence on Private Firms in a Non-litigious Environment. *Accounting and Business Research*, 33 (1), 65-79.
- Gentry J., Newbold P. ve Whitford D.** (1985a). Classifying Bankrupt Firms with Funds Flow Components. *Journal of Accounting Research*, 23 (1), 146-160.
- Gentry J., Newbold P. ve Whitford D.** (1985b). Predicting Bankruptcy: If Cash Flow's not the Bottom Line, What is? *Financial Analysts Journal*, 41 (5), 47-56.
- Gepp A., Kumar K. ve Bhattacharya S.** (2010). Business Failure Prediction Using Decision Trees. *Journal of Forecasting*, 29 (6), 536-555.
- Gilbert L., Menon K. ve Schwartz K.** (1990). Predicting Bankruptcy for Firms in Financial Distress. *Journal of Business Finance and Accounting*, 17 (1), 161-171.
- Goldberger A. S.**, (1991). *A Course in Econometrics*. Cambridge: Harvard University Press.
- Gombola M., Haskins M., Ketz J. ve Williams D.** (1987). Cash Flow in Bankruptcy Prediction. *Financial Management*, 16 (4), 55-65.
- Gruszczynski M.** (2015). Issues in Modelling the Financial Distress and Bankruptcy of Companies. *Quantitative Methods in Economics*, 16 (1), 96-107.
- Güredin E.** (2014). *Denetim ve Güvence Hizmetleri* (14. bs.). İstanbul: Türkmen Kitabevi.
- Hopwood W., McKeown J. ve Mutchler J.** (1989). A Test of the Incremental Explanatory Power of Opinions Qualified for Consistency and Uncertainty. *The Accounting Review*, 64 (1), 28-48.
- Hopwood W., McKeown J. ve Mutchler J.** (1994). A Reexamination of Auditor Versus Model Accuracy within the Context of the Going-concern Opinion Decision. *Contemporary Accounting Research*, 10 (2), 409-431.
- Hothorn T., Hornik K. ve Zeileis A.** (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15 (3), 651-674.
- Izan, H.** (1984). Corporate distress in Australia. *Journal of Banking and Finance*, 8, 303-320.
- Jandaghi G., Tehrani R., Pirani P. ve Mokhles A.** (2011). Hybrid Financial Analysis Model for Predicting Bankruptcy. *British Journal of Economics, Finance and Management Sciences*, 2 (1), 37-48.
- Jo H., Han I. ve Lee H.** (1997). Bankruptcy Prediction Using Case-based Reasoning, Neural Networks, and Discriminant Analysis. *Expert Systems with Applications*, 13 (2), 97-108.
- Johnsen T. ve Melicher R.** (1994). Predicting Corporate Bankruptcy and Financial Distress: Information Value Added by Multinomiallogit Models. *Journal of Economics and Business*, 46 (4), 269-286.
- Jones F.** (1987). Current Techniques in Bankruptcy Prediction. *Journal of Accounting Literature*, 6, 131-164.
- Jones S. ve Hensher D.** (2004). Predicting Firm Financial Distress: A Mixed Logit Model. *The Accounting Review*, 79 (4), 1011-1038.

- Joos P., Vanhoof K., Ooghe H. ve Sierens N.** (1998). Credit Classification: A Comparison of Logit Models and Decision Trees. Nedellec C., ve Rouveinol C. (Ed.), *The Workshop on Application of Machine Learning and Data Mining in Finance, 10th European Conference on Machine Learning bildiriler kitabı* içinde (ss. 59-72). Germany: Chemnitz.
- Kahya E. ve Theodossiou P.** (1999). Predicting Corporate Financial Distress: A Timeseries CUSUM Methodology. *Review of Quantitative Finance and Accounting*, 13 (4), 323-345.
- Kalmış H., Eskin İ. ve Gümüş F.** (2006). İşletmelerde Mali Bilgilerin Üretilmesinin Önemi ve Etik. *İSMMMO Mali Çözüm Dergisi*, 75, 141-160.
- Kaval H.** (2008). *Muhasebe Denetimi* (3. bs.). Ankara: Gazi Kitabevi.
- Kavut L.** (2001). Bağımsız Denetim Hizmetinden Yararlanan Firmaların Aldıkları Hizmetin Kalitesine İlişkin Görüşleri. *Yönetim Dergisi*, 40, 7-18.
- Kavut L.** (2002). Ulusal ve Uluslararası Bağımsız Denetim Firmalarında Çalışan Denetçilerin Bağımsız Denetim Hizmetinin Kalitesine İlişkin Tutumları. *Yönetim Dergisi*, 41, 7-21.
- Kavut L., Taş O. ve Şavlı T.** (2009). *Uluslararası Denetim Standartları Kapsamında Bağımsız Denetim*. İstanbul: İSMMMO Yayınları.
- Keasey K. ve McGuinness P.** (1990). The Failure of UK Industrial Firms for the Period 1976-1984: Logistic Analysis and Entropy Measures. *Journal of Business Finance and Accounting*, 17 (1), 119-35.
- Keasey K. ve Watson R.** (1986). The Prediction of Small Company Failure: Some Behavioral Evidence for the UK. *Accounting and Business Research*, 17, 49-57.
- Keasey K. ve Watson R.** (1991). Financial Distress Prediction Models: A Review of Their Usefulness. *British Journal of Management*, 2, 89-102.
- Kim K. J.** (2003). Financial Time Series Forecasting Using Support Vector Machines. *Neurocomputing*, 55 (1/2), 307-319.
- Kirkos E., Spathis C. ve Manolopoulos Y.** (2008). Support Vector Machines, Decision Trees and Neural Networks for Auditor Selection. *Journal of Computational Methods in Sciences and Engineering*, 8 (3), 213-224.
- Kiviluoto K.** (1998). Predicting Bankruptcies with the Self-Organizing Map. *Neurocomputing*, 21 (1-3), 191-201.
- Koh H.** (1987). *Prediction of Going-concern Status: A Probit Model for the Auditors*. Yayımlanmamış doktora tezi, Virginia Polytechnic Institute and State University, Blackburg.
- Koh H. ve Killough L.** (1990). The Use of Multiple Discriminant Analysis in the Assessment of the Going-concern Status of an Audit Client. *Journal of Business Finance & Accounting*, 17 (2), 179-192.
- Laitinen E.** (1991). Financial Ratios and Different Failure Processes. *Journal of Business Finance & Accounting*, 18 (5), 649-673.
- Laitinen E. K. ve Laitinen T.** (1998). Cash Management Behaviour and Failure Prediction. *Journal of Business Finance and Accounting*, 25 (7-8), 893-919.
- Lau A.** (1987). A Five-state Financial Distress Prediction Model. *Journal of Accounting Research*, 25 (1), 127-138.
- Lee K., Han I. ve Kwon Y.** (1996). Hybrid Neural Network Models for Bankruptcy Predictions. *Decision Support Systems*, 18 (1), 63-72.

- Lennox C.** (1999). The Accuracy and Incremental Information Content of Audit Reports in Predicting Bankruptcy. *Journal of Business Finance & Accounting*, 26 (5-6), 757-778.
- Leshno M. ve Spector Y.** (1996). Neural-network Prediction Analysis: The Bankruptcy Case. *Neurocomputing*, 10 (2), 125-147.
- Mahmood M. ve Lawrence E.** (1987). A Performance Analysis of Parametric and Nonparametric Discriminant Approaches to Business Decision Making. *Decision Sciences*, 18 (2), 308-326.
- Martin D.** (1977). Early Warning of Bank Failures: A Logit Regression Approach. *Journal of Banking and Finance*, 1, 249-276.
- McGurr P. T. ve DeVaney S. A.** (1998). Predicting Business Failure of Retail Firms: An Analysis Using Mixed Industry Models. *Journal of Business Research*, 43, 169-76.
- McKee T.** (1995). Predicting Bankruptcy via Induction. *Journal of Information Technology*, 10 (1), 26-36.
- McNamara R., Cocks N. ve Hamilton D.** (1988). Predicting Private Company Failure. *Accounting and Finance*, 28 (2), 53-64.
- Merwin C.** (1942). *Financing Small Corporations in Five Manufacturing Industries, 1926-1936*. New York: National Bureau of Economic Research.
- Meyer P. ve Pifer H.** (1970). Prediction of Bank Failures. *Journal of Finance*, 25 (4), 853-868.
- Min J. H. ve Lee Y. C.** (2005). Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters. *Expert Systems with Applications*, 28 (4), 603-614.
- Moses D. ve Liao S.** (1987). On Developing Models for Failure Prediction. *Journal of Commercial Bank Lending, Mart*, 27-38.
- Moyer R.** (1977). Forecasting Financial Failure: A Re-examination. *Financial Management*, 6 (1), 11-17.
- Neophytou E., Charitou A. ve Charalambous C.** (2001). Predicting Corporate Failure: Empirical Evidence for the UK. *University of Southampton, School of Management, Discussion Paper, 01 (173)*, 1-30.
- Odom M. ve Sharda R.** (1990). A Neural Network Model for Bankruptcy Prediction. *The International Joint Conference on Neural Networks bildiriler kitabı* içinde (ss. 63-68). San Diego: IEEE.
- Ohlson J.** (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18 (1), 109-131.
- Orhan S. ve Serçemeli M.** (2014). Sürekli Denetimin Uygulanma Düzeyi ile İlgili BIST 100 Şirketlerinde Bir Araştırma. *Muhasebe ve Denetime Bakış*, 43, 1-17.
- Özbirecikli M.** (2013). Türkiye’de Ticaret Kanunlarında Denetim Anlayışı Üzerine Bir İnceleme. *MUFİTAD*, 4, 152-167.
- Pearson K.** (1904). *Mathematical Contributions to the Theory of Evolution*. Londra: Dulau and Co.
- Pekdemir R. ve Selvi Y.** (2002). Teknolojik Gelişmelerin Denetimde Yarattığı Yeni Fırsatlar. 17. *Türkiye Muhasebe Kongresi bildiriler kitabı* içinde (ss. 193-200). İstanbul: TÜRMOB.
- Pettway R. ve Sinkey J. Jr.** (1980). Establishing On-site Banking Examination Priorities: An Early Warning System Using Accounting and Market Information. *Journal of Finance*, 35 (1), 137-150.

- Pinches G., Eubank A., Mingo K. ve Caruthers J.** (1975). The Hierarchical Classification of Financial Ratios. *Journal of Business Research*, 3 (4), 295-310.
- Platt H. D. ve Platt M. B.** (1990). Development of a Class of Stable Predictive Variables: The Case of Bankruptcy Prediction. *Journal of Banking, Finance and Accounting*, 17 (1), 31-51.
- Platt H. D., Platt M. B. ve Pedersen J.** (1994). Bankruptcy Discrimination With Real Variables. *Journal of Business Finance and Accounting*, 21 (4), 491-509.
- Pompe P. ve Feelders A.** (1997). Using Machine Learning, Neural Networks, and Statistics to Predict Corporate Bankruptcy. *Microcomputers in Civil Engineering*, 12, 267-76.
- Rose P. ve Kolari J.** (1985). Early Warning Systems as a Monitoring Device for Bank Condition. *Quarterly Journal of Business and Economics*, 24 (1), 43-60.
- Rudorfer G.** (1995). Early Bankruptcy Prediction Using Neural Networks. *APL Quote Quad*, 25 (4), 171-176.
- Rujoub M., Cook D. ve Hay L.** (1995). Using Cash Flow Ratios to Predict Business Failures. *Journal of Managerial Issues*, 7 (1), 75-90.
- Salchenberger L., Cinar E. ve Lash N.** (1992). Neural Networks: A New Tool for Predicting Bank Failures. *Decision Sciences*, 23, 899-916.
- Samuel A. L.** (2000). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 44 (1-2), 206-226.
- Sands E., Springate G. ve Var T.** (1983). Predicting Business Failures: A Canadian Approach. *CGA Magazine*, 17, 24-27.
- Santomero A. ve Vinso J.** (1977). Estimating the Probability of Failure for Commercial Banks and the Banking System. *Journal of Banking and Finance*, 1 (2), 185-205.
- Scaggs M. ve Crawford P.** (1986). Altman's Bankruptcy Model Revisited: Can Airline Bankruptcy be Predicted? *Review of Regional Economics and Business*, 11 (2), 11-16.
- Scholkopf B., Smola A., Williamson R. C. ve Bartlett P.** (2000). New Support Vector Algorithms. *Neural Computation*, 12, 1207-1245.
- Scott D. L.** (2003). *An A to Z Guide to Investment Terms for Today's Investor*. Boston: Houghton Mifflin Company.
- Selimoğlu S. K. ve Uzay Ş.** (2014). *Muhasebe Denetimi* (4. bs.). Ankara: Gazi Kitabevi.
- Selimoğlu S. K.** (2015). *Muhasebe Denetimi ve Finansal Analiz*. Eskişehir: Anadolu Üniversitesi Yayınları.
- Serrano-Cinca C.** (1996). Self Organizing Neural Networks for Financial Diagnosis. *Decision Support Systems*, 17 (3), 227-238.
- Shah J. ve Murtaza M.** (2000). A Neural Network Based Clustering Procedure for Bankruptcy Prediction. *American Business Review*, 18 (2), 80-86.
- Shah N.** (2014). Developing Financial Distress Prediction Models Using Cutting Edge Recursive Partitioning Techniques: A Study of Australian Mining Performance. *Review of Integrative Business and Economics Research*, 3 (2), 103-143.
- Sharma S. ve Mahajan V.** (1980). Early Warning Indicators of Business Failure. *Journal of Marketing*, 44 (4), 80-89.

- Shin K. S., Kim H. J. ve Lee T. S.** (2005). An Application of Support Vector Machines in Bankruptcy Prediction Model. *Expert Systems with Applications*, 28 (1), 127-135.
- Sinke Jr. J.** (1975). A Multivariate Statistical Analysis of the Characteristics of Problem Banks. *Journal of Finance*, 30 (1), 21-36.
- Skogsvik K.** (1990). Current Cost Accounting Ratios as Predictors of Business Failure: The Swedish Case. *Journal of Business Finance & Accounting*, 17 (1), 137-160.
- Strobl C., Hothorn T. ve Zeileis A.** (2009). Party on! A New, Conditional Variable Importance Measure for Random Forests Available in the party Package. *University of Munich, Dept. of Statistics, Technical Report*, 50, 1-4.
- Sung T., Chang N. ve Lee G.** (1999). Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction. *Journal of Management Information Systems*, 16 (1), 63-85.
- Taffler R. J.** (1982). Forecasting Company Failure in the UK Using Discriminant Analysis and Financial Ratio Data. *Journal of the Royal Statistical Society*, 145 (3), 342-58.
- Taffler R. J.** (1983). The Assessment of Company Solvency and Performance Using a Statistical Model. *Accounting and Business Research*, 13 (52), 295-308.
- Taffler R. J.** (1984). Empirical Models for the Monitoring of UK Corporations. *Journal of Banking and Finance*, 8 (2), 199-227.
- Takahashi K., Kurokawa Y. ve Watase K.** (1984). Corporate Bankruptcy Prediction in Japan. *Journal of Banking and Finance*, 8 (2), 229-247.
- Tam K.** (1991). Neural Network Models and the Prediction of Bankruptcy. *Omega*, 19 (5), 429-445.
- Tam K. ve Kiang M.** (1992). Managerial Applications of Neural Networks - The Case of Bank Failure Predictions. *Management Science*, 38 (7), 926-947.
- Tay F. E. H. ve Cao L.** (2001). Application of Support Vector Machines in Financial Time Series Forecasting. *Omega*, 29, 309-317.
- Terzi R., Atmaca M. ve Terzi S.** (2016). Denetim Açısından İşletmenin Sürekliliğinin Değerlendirilmesinde Genetik Algoritmanın Kullanımı: BIST Sınai Endeksi Örneği. *Uluslararası Yönetim İktisat ve İşletme Dergisi, ICAFR*, 16, 685-693.
- Theodossiou P.** (1991). Alternative Models for Assessing the Financial Condition of Business in Greece. *Journal of Business Finance & Accounting*, 18 (5), 697-720.
- Theodossiou P.** (1993). Predicting Shifts in the Mean of a Multivariate Time Series Process: An Application in Predicting Business Failure. *Journal of the American Statistical Association*, 88, 441-449.
- Therneau T.M.** (1983). A Short Introduction to Recursive Partitioning. *Stanford University, Department of Statistics, Orion Technical Report*, 21, 1-11.
- Tsukuda J. ve Baba S.** (1994). Predicting Japanese Corporate Bankruptcy in Terms of Financial Data Using Neural Network. *Computers and Industrial Engineering*, 27, 445-448.
- URL-1:** https://tr.wikipedia.org/wiki/Halka_ acık_ irket, (Ziyaret tarihi: 5 Haziran 2018).
- URL-2:** <https://www.ismmmo.org.tr/Mevzuat/I-Muhasebenin-Temel-Kavramlari--4003>, (Ziyaret tarihi: 5 Haziran 2018).

- Uzay Ş.** (2006). Denetçilerin Denetiminde Yeni Bir Model Olarak Kamu Gözetim Kurulu ve Türkiye’de Uygulanabilirliği. *MÖDAV Dergisi*, 8 (4), 175-210.
- Uzay, Ş ve Tañ G. Ş.** (2010). İMKB’de İşlem Gören Şirketlerin Bağımsız Denetim Raporlarında İşletmenin Sürekliliği Kavramının Analizi. *Muhasebe Bilim Dünyası Dergisi*, 12 (2), 143-179.
- 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 & Accounting* 21 (4), 547-561.
- Wertheim P. ve Lynn M.** (1993). Development of a Prediction Model for Hospital Closure Using Financial Accounting Data. *Decision Sciences*, 24 (3), 529-546.
- Westgaard S. ve Wijst N.** (2001). Default Probabilities in a Corporate Bank Portfolio: A Logistic Model Approach. *European Journal of Operational Research*, 135, 338-49.
- Widrow B. ve Hoff M. E.** (1960). Adaptive Switching Circuits. *Stanford Electronics Laboratories, Technical Report No. 1553-1*, 1-34.
- Wilcox J.** (1973). A Prediction of Business Failure Using Accounting Data. *Journal of Accounting Research*, 11, 163-179.
- Wilson N., Chong K. ve Peel M.** (1995). Neural Network Simulation and the Prediction of Corporate Outcomes: Some Empirical Findings. *International Journal of the Economics of Business*, 2 (1), 31-50.
- Wilson R. ve Sharda R.** (1994). Bankruptcy Prediction Using Neural Networks. *Decision Support Systems*, 11 (5), 545-557.
- Yang Z., Platt M. ve Platt H.** (1999). Probabilistic Neural Networks in Bankruptcy Prediction. *Journal of Business Research*, 44, 67-74.
- Yücel T., Mandacı E. ve Kurt G.** (2007). İşletmelerin Finansal Risk Yönetimi ve Türev Ürün Kullanımı: İMKB 100 Endeksinde Yer Alan İşletmelerde Bir Uygulama. *Muhasebe ve Finansman Dergisi*, 36, 106-113.
- Zavgren C. V.** (1985). Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis. *Journal of Business Finance & Accounting*, 12 (1), 19-45.
- Zhang G., Hu M.Y., Patuwo B. E. ve Indro D. C.** (1999). Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis. *European Journal of Operational Research*, 116, 16-32.
- Zimmer I.** (1980). A Lens Study of the Prediction of Corporate Failure by Bank Loan Officers. *Journal of Accounting Research*, 18 (2), 629-636.



ÖZGEÇMİŞ

Kişisel Bilgiler

Ad Soyad : GÜRKAN SAYIN
Doğum Tarihi : 21.01.1969
Doğum Yeri : ALMUS
Medeni Durumu : EVLİ



İletişim Bilgileri

Adres : Bizim Vadi No:34 Sarıyer/İstanbul
Cep : 05333387595
E-Posta : gurkansayin@yahoo.com

Eğitim Bilgileri

1991-1993 : Bilkent Üniversitesi İşletme Fakültesi
İşletme Yüksek Lisansı
1987-1991 : Bilkent Üniversitesi Mühendislik Fakültesi
Elektrik ve Elektronik Mühendisliği Lisansı
1980-1987 : Ankara Atatürk Anadolu Lisesi
Lise Diploması

İş Deneyimi

Mayıs 2017 : GO4MG Bilgisayar ve Danışmalık A.Ş.
Kantitatif Ürün ve Sistem Dizaynı
1998-2017 : Garanti Bankası A.Ş. Grubu
Kantitatif Ürün ve Sistem Dizaynı
1996-1998 : Alternatif Menkul Kıymetler A.Ş.
Kantitatif Araştırmalar
1995-1996 : Ekinciler Yatırım ve Menkul Kıymetler A.Ş.
Hisse Senedi Araştırma