

AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

Orhan ECEMİŞ

VERİ MADENCİLİĞİ YÖNTEMLERİYLE
PASLANMAZ ÇELİK SEKTÖRÜNDE SATIŞ TAHMİNİ

Ekonometri Ana Bilim Dalı
Doktora Tezi

Antalya, 2016

AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

Orhan ECEMİŞ

VERİ MADENCİLİĞİ YÖNTEMLERİYLE
PASLANMAZ ÇELİK SEKTÖRÜNDE SATIŞ TAHMİNİ

Danışman

Doç. Dr. Sezgin IRMAK

Ekonometri Ana Bilim Dalı

Doktora Tezi

Antalya, 2016

Akdeniz Üniversitesi
Sosyal Bilimler Enstitüsü Müdürlüğüne,

Orhan ECEMİŞ'in bu çalışması jürimiz tarafından Ekonometri Ana Bilim Dalı Doktora Programı tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Can Deniz KÖKSAL (İmza)

Üye (Danışmanı) : Doç. Dr. Sezgin IRMAK (İmza)

Üye : Doç. Dr. Emre İPEKÇİ ÇETİN (İmza)

Üye : Yrd. Doç. Dr. Yusuf BOZGEYİK (İmza)

Üye : Yrd. Doç. Dr. Mehmet Özer DEMİR (İmza)

Tez Başlığı : Veri Madenciliği Yöntemleriyle Paslanmaz Çelik Sektöründe Satış Tahmini

Onay : Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Tez Savunma Tarihi : 16/12/2016

Mezuniyet Tarihi : 29/12/2016

(İmza)
Prof. Dr. İhsan BULUT
Müdür

AKADEMİK BEYAN

Doktora Tezi olarak sunduđum “Veri Madenciliđi Yöntemleriyle Paslanmaz Çelik Sektöründe Satış Tahmini” adlı bu çalışmanın, akademik kural ve etik değere uygun bir biçimde tarafımda yazıldığını, yararlandığım bütün eserlerin kaynakçada gösterildiğini ve çalışma içerisinde bu eserlere atıf yapıldığını belirtir; bunu şerefimle doğrularım.

...../...../ 2016

Orhan ECEMİŞ

İmza

X X X X X

İÇİNDEKİLER

ŞEKİLLER LİSTESİ	vi
TABLolar LİSTESİ	viii
KISALTMALAR LİSTESİ	x
ÖZET	xii
SUMMARY	xiv
ÖNSÖZ	xvi
GİRİŞ.....	1

BİRİNCİ BÖLÜM TAHMİN YÖNTEMLERİ

1.1 Satış ve Talep Kavramları	6
1.2 Tahmin Yöntemleri	8
1.2.1 Tahmin Yöntemleri Seçimi	9
1.2.2 Zaman Serileri	11
1.2.2.1 Üstel Düzgünleştirme Yöntemleri	13
1.2.2.1.1 Holt Yöntemi	13
1.2.2.1.2 Winters Yöntemi	14
1.2.2.2 Otoregresif Bütünleşik Hareketli Ortalama Modelleri	14
1.2.2.2.1 Otoregresif(AR) Süreci.....	15
1.2.2.2.2 Hareketli Ortalamalar (MA Süreci).....	15
1.2.2.2.3 ARMA Modelleri	15
1.2.2.2.4 ARIMA Modelleri	16
1.2.2.2.5 Mevsimsel ARIMA Modelleri.....	16
1.2.2.2.6 Parametrelerin Tahmini	18
1.2.2.3 Durağanlık Kavramı	18
1.2.2.3.1 Katsayıların Değerlendirilmesi.....	19
1.2.2.3.2 KPSS Birim Kök Testi	19
1.2.2.3.3 Dickey Fuller Birim Kök Testi	20
1.2.3 Hiyerarşik Zaman Serileri	21
1.2.3.1 Birleşim Yöntemleri	22
1.2.3.1.1 Tüme Varım Yöntemi.....	22
1.2.3.1.2 En Uygun Tahmin Birleşim Yöntemi.....	23

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ KAVRAM VE YÖNTEMLERİ

2.1	Yapay Öğrenme.....	24
2.1.1	Denetimli Öğrenme	24
2.1.2	Denetimsiz Öğrenme	24
2.2	Veri Madenciliği.....	25
2.3	Veri Madenciliği Süreci.....	26
2.3.1	Veri Seçimi	27
2.3.2	Veri Ön İşleme	27
2.3.2.1	Veri Temizleme	28
2.3.2.2	Boyut Azaltma.....	28
2.3.3	Veri Dönüştürme	29
2.3.4	Modelleme	29
2.3.5	Yorumlama	29
2.4	Veri Madenciliği Yöntemleri.....	30
2.4.1	Karar Ağaçları	30
2.4.1.1	Karar Ağaçlarında Bölümlenme Kriteri	30
2.4.1.2	Öznitelik Seçimi	31
2.4.1.3	Kazanç Oranı	31
2.4.2	Sınıflandırma ve Regresyon Ağaçları.....	32
2.4.2.1	Twoing Algoritması.....	32
2.4.2.2	Gini Algoritması.....	33
2.4.3	Model Ağaçlar	33
2.4.3.1	Ağaç Oluşturma.....	35
2.4.3.2	Ağaç Budama	36
2.4.3.3	Nominal Değerler	36
2.4.3.4	Yerel Ağırlıklandırılmış Regresyon	36
2.4.4	Kümeleme Yöntemleri	37
2.4.4.1	Uzaklık Ölçüleri	38
2.4.4.2	Tekli Bağlantı Yöntemi	39
2.4.4.3	Tam Bağlantı En Uzak Komşu Yöntemi.....	39
2.4.4.4	Aritmetik Ortalamalı Bağlantı/Gruplar Arası.....	40
2.4.4.5	Wards Bağlantısı	40
2.4.5	Yapay Sinir Ağları.....	41

2.4.5.1	Algılayıcı	42
2.4.5.2	Ağırlıklar	42
2.4.5.3	Birleştirme Fonksiyonu	42
2.4.5.4	Aktivasyon Fonksiyonu	43
2.4.5.5	Algılayıcının Eğitilmesi	44
2.4.5.6	YSA Türleri	44
2.4.5.6.1	İleri Beslemeli YSA	44
2.4.5.6.2	Geri Beslemeli YSA	45
2.4.5.6.3	Tek Katmanlı Ağlar	46
2.4.5.6.4	Çok Katmanlı Yapay Sinir Ağları	47
2.4.5.7	Yapay Sinir Ağlarının Zaman Serilerine Uygulanması	47
2.4.6	Destek Vektör Makineleri	48
2.4.6.1	Hiper Düzlem	49
2.4.6.2	Doğrusal Olmayan Sınıflandırma	52
2.4.6.3	Çekirdek	54
2.4.7	Destek Vektör Regresyon	55
2.4.8	Gauss Süreç Regresyon	57
2.4.8.1	Tahmin Süreci	58
2.4.8.2	Parametrelerin Optimizasyonu	59
2.5	Literatürde Veri Madenciliği ile Yapılan Tahmin Çalışmaları	60

ÜÇÜNCÜ BÖLÜM

PASLANMAZ ÇELİK SEKTÖRÜNDE VERİ MADENCİLİĞİ YÖNTEMLERİYLE SATIŞ TAHMİNİ

3.1	Paslanmaz Çelik	68
3.2	Paslanmaz Çelik Üretim Aşamaları	68
3.3	Dünyada ve Türkiye’de Paslanmaz Çelik Sektörü	69
3.4	Araştırmanın Amacı ve Kapsamı	74
3.5	Uygulama Platformu	75
3.5.1	Veri Madenciliği Yazılımı	75
3.5.1.1	WEKA Yazılımı	75
3.5.1.2	R Yazılımı	76
3.6	Veri	76
3.6.1	Veri Setinin Dönüştürülmesi	77

3.6.2 Tanımlayıcı Bulgular	78
3.7 Model Ağaçlar Yöntemiyle Tahmin Modelleri	82
3.7.1 Toplam Satış Tahmini	82
3.7.2 Makine Sektörü Tahmini	83
3.7.3 Metal Sektörü Tahmini	84
3.7.4 Endüstriyel Mutfak Sektörü Tahmini	84
3.7.5 Yapı-İnşaat Sektörü Tahmini	84
3.7.6 Diğer Sektör Tahmini	85
3.7.7 Modellerin Değerlendirilmesi	85
3.8 Kümeleme Analizi	86
3.9 Zaman Serileri Satış Tahmin Modelinin Oluşturulması	86
3.9.1 Duraganlık Testleri	86
3.9.2 Hiyearşik Zaman Serisi Tahmin Modeli	88
3.9.2.1 Toplam Satış Tahmini	90
3.9.2.2 A Düğümü (Makine-Metal ve Endüstriyel Mutfak Sektörleri Birleşimli) Tahmini	90
3.9.2.3 B Düğümü (Yapı-İnşaat ve Diğer Sektörler Birleşimli) Satış Tahmini	90
3.9.2.4 Metal Sektörü Satış Tahmini	91
3.9.2.5 Makine Sektörü Satış Tahmini	91
3.9.2.6 Endüstriyel Mutfak Sektörü Satış Tahmini	91
3.9.2.7 Yapı İnşaat Sektörü Satış Tahmini	91
3.9.2.8 Diğer Sektör Satış Tahmini	92
3.10 Veri Madenciliği Yöntemleri Tahmin Modelleri	92
3.10.1 Toplam Satış İçin Tahmin Modelleri	92
3.10.1.1 Destek Vektör Regresyon Tahmini	92
3.10.1.2 Çok Katmanlı Algılayıcı Tahmini	94
3.10.1.3 Gauss Süreç Regresyon Tahmini	95
3.10.2 Makine Sektörü Satış İçin Tahmin Modelleri	97
3.10.2.1 Destek Vektör Regresyon Tahmini	97
3.10.2.2 Çok Katmanlı Algılayıcı Tahmini	98
3.10.2.3 Gauss Süreç Regresyon Tahmini	99
3.10.3 Metal Sektörü Satış İçin Tahmin Modelleri	101
3.10.3.1 Destek Vektör Regresyon Tahmini	101
3.10.3.2 Çok Katmanlı Algılayıcı Tahmini	102

3.10.3.3	Gauss Süreç Regresyon Tahmini.....	103
3.10.4	Endüstriyel Mutfak Sektörü Satış İçin Tahmin Modelleri	105
3.10.4.1	Destek Vektör Regresyon Tahmini	105
3.10.4.2	Çok Katmanlı Algılayıcı Tahmini	106
3.10.4.3	Gauss Süreç Regresyon Tahmini.....	107
3.10.5	Yapı-İnşaat Sektörü Satış İçin Tahmin Modelleri	109
3.10.5.1	Destek Vektör Regresyon Tahmini	109
3.10.5.2	Çok Katmanlı Algılayıcı Tahmini	110
3.10.5.3	Gauss Süreç Regresyon Tahmini.....	111
3.10.6	Diğer Sektör Satış İçin Tahmin Modelleri	113
3.10.6.1	Destek Vektör Regresyon Tahmini	113
3.10.6.2	Çok Katmanlı Algılayıcı Tahmini	114
3.10.6.3	Gauss Süreç Regresyon Tahmini.....	115
3.11	Modellerin Karşılaştırılması	117
3.11.1	Zaman Serisi Modellerinin Karşılaştırılması.....	117
3.11.2	Veri Madenciliği Modellerinin Karşılaştırılması	118
SONUÇ		122
KAYNAKÇA.....		125
ÖZGEÇMİŞ		135

ŞEKİLLER LİSTESİ

Şekil 1.1 Tahmin Yöntemleri Seçimi Akış Şeması	10
Şekil 1.2 Otoregresif Bütünleşik Hareketli Ortalama (ARIMA) Modeli	14
Şekil 1.3 İki Seviyeli Hiyerarşik Ağaç Diyagramı	21
Şekil 2.1 Veritabanlarında Bilgi Keşfi Süreci	26
Şekil 2.2 Model Ağaçlar İçin Çapraz Doğrulanmış RMSE değerleri	34
Şekil 2.3 Kümeleme Yöntemlerinin Sınıflandırılması	38
Şekil 2.4 Tekli Bağlantı Yöntemi Dendogram	39
Şekil 2.5 Tam Bağlantı Yöntemi Dendogram	40
Şekil 2.6 Aritmetik Ortalamalı Bağlantı Dendogram.....	40
Şekil 2.7 Ward Bağlantısı-Dendogram.....	41
Şekil 2.8 Biyolojik Sinir Hücresi ve Bileşenler.....	41
Şekil 2.9 Algılayıcı Yapısı	42
Şekil 2.10 İleri Beslemeli YSA	45
Şekil 2.11 Geri Beslemeli Yapay Sinir Ağı.....	46
Şekil 2.12 Tek Katmanlı Yapay Sinir Ağı.....	46
Şekil 2.13 Çok Katmanlı Yapay Sinir Ağı	47
Şekil 2.14 YSA'nın Zaman Serilerine Uygulanması	47
Şekil 2.15 Sınıflandırma Tipleri	49
Şekil 2.16 Destek Vektörlerinin Oluşturulması.....	51
Şekil 2.17 Destek Vektör Regresyon.....	57
Şekil 2.18 Gauss Süreç Regresyon Modeli	59
Şekil 3.1 Ülke/Bölge Bazlı Dünya Paslanmaz Çelik Üretimi	70
Şekil 3.2 Bazı Ülkeler İçin Kişi Başına Düşen Paslanmaz Çelik Kullanımı.....	72
Şekil 3.3 Dünya Geneli Sektörlere Göre Paslanmaz Çelik Talebi	72
Şekil 3.4 Mamul Türlerine Göre Paslanmaz Çelik Üretimi	73
Şekil 3.5 Sektörlere Göre Satışlar.....	79
Şekil 3.6 Sektörler ve Toplam Satış Tutarına Ait Zaman Serileri.....	80
Şekil 3.7 Aylık Emtia ve Dolar Kurunu Ait Zaman Serisi (01/2008-03/2016).....	81
Şekil 3.8 Diğer Bağımsız Değişkenlere Ait Zaman Serileri.....	81
Şekil 3.9 Kümeleme Analizine Ait Dendogram.....	86
Şekil 3.10 Satış Verilerinin Hiyerarşik Modeli	89
Şekil 3.11 Toplam Satış DVR Eğitim Aşaması.....	93
Şekil 3.12 Toplam Satış Destek Vektör Regresyon Test Aşaması.....	94

Şekil 3.13 Toplam Satış ÇKA Eğitim Aşaması.....	94
Şekil 3.14 Toplam Satış ÇKA Test Aşaması.....	95
Şekil 3.15 Toplam Satış GSR Eğitim Aşaması	96
Şekil 3.16 Toplam Satış GSR Test Aşaması	96
Şekil 3.17 Makine Sektörü DVR Eğitim Aşaması	97
Şekil 3.18 Makine Sektörü DVR Test Aşaması	98
Şekil 3.19 Makine Sektörü ÇKA Eğitim Aşaması	98
Şekil 3.20 Makine Sektörü ÇKA Test Aşaması	99
Şekil 3.21 Makine Sektörü GSR Eğitim Aşaması.....	100
Şekil 3.22 Makine Sektörü GSR Test Aşaması.....	100
Şekil 3.23 Metal Sektörü Destek Vektör Eğitim Aşaması	101
Şekil 3.24 Metal Sektörü Destek Vektör Test Aşaması	102
Şekil 3.25 Metal Sektörü ÇKA Eğitim Aşaması	102
Şekil 3.26 Metal Sektörü ÇKA Test Aşaması	103
Şekil 3.27 Metal Sektörü GSR Eğitim Aşaması.....	104
Şekil 3.28 Metal Sektörü GSR Test Aşaması.....	104
Şekil 3.29 Endüstriyel Mutfak DVR Eğitim Aşaması.....	105
Şekil 3.30 Endüstriyel Mutfak DVR Test Aşaması.....	106
Şekil 3.31 Endüstriyel Mutfak ÇKA Eğitim Aşaması.....	106
Şekil 3.32 Endüstriyel Mutfak ÇKA Test Aşaması.....	107
Şekil 3.33 Endüstriyel Mutfak GSR Eğitim Aşaması	108
Şekil 3.34 Endüstriyel Mutfak GSR Test Aşaması	108
Şekil 3.35 Yapı-İnşaat Sektörü DVR Eğitim Aşaması.....	109
Şekil 3.36 Yapı-İnşaat Sektörü DVR Test Aşaması.....	110
Şekil 3.37 Yapı-İnşaat Sektörü ÇKA Eğitim Aşaması.....	110
Şekil 3.38 Yapı-İnşaat Sektörü ÇKA Test Aşaması.....	111
Şekil 3.39 Yapı-İnşaat Sektörü GSR Eğitim Aşaması	112
Şekil 3.40 Yapı-İnşaat Sektörü GSR Test Aşaması	112
Şekil 3.41 Diğer Sektör DVR Eğitim Aşaması	113
Şekil 3.42 Diğer Sektör DVR Test Aşaması	114
Şekil 3.43 Diğer Sektör ÇKA Eğitim Aşaması	114
Şekil 3.44 Diğer Sektör ÇKA Test Aşaması	115
Şekil 3.45 Diğer Sektör GSR Eğitim Aşaması.....	116
Şekil 3.46 Diğer Sektör GSR Test Aşaması.....	116

TABLOLAR LİSTESİ

Tablo 1.1 Üstel Düzgünleştirme ve ARIMA Modellerinin Güçlü ve Zayıf Yönleri	17
Tablo 2.1 Denetimli ve Denetimsiz Öğrenme Algoritmaları	25
Tablo 2.2 Veri Kalitesini Etkileyen Faktörler	28
Tablo 2.3 Karar Ağaçları Modellerinin Güçlü ve Zayıf Yönleri.....	30
Tablo 2.4 Birleştirme Fonksiyonu Örnekleri.....	43
Tablo 2.5 Bazı Aktivasyon Fonksiyonları.....	43
Tablo 2.6 Veri Madenciliği ile Yapılan Tahmin Çalışmaları.....	61
Tablo 3.1 Paslanmaz Çelik ve Bazı Ürünlerin Talebinde Yıllık Büyüme Miktarı.....	69
Tablo 3.2 Bölgelere Göre Sıcak Haddelenmiş Paslanmaz Çelik Üretimi	71
Tablo 3.3 Paslanmaz Çelik İhracatı (İlk 5 Ülke).....	73
Tablo 3.4 Paslanmaz Çelik İthalatı (İlk 5 Ülke).....	74
Tablo 3.5 Veri Seti	77
Tablo 3.6 Sektör Kodları	77
Tablo 3.7 ARFF Formatı Dönüştürme	78
Tablo 3.8 M5p Yöntemi Model Sonuçları.....	86
Tablo 3.9 Tüm Değişkenlere Ait ADF Sonuçları.....	88
Tablo 3.10 Hiyeşik Zaman Serisi Toplam Satış Tahmini	90
Tablo 3.11 A Düğümü Tahmini (Makine, Metal ve Endüstriyel Mutfak Sektör Tahmini)	90
Tablo 3.12 B Düğümü Yapı-İnşaat ve Diğer Sektörler Birleşimli Tahminleri	90
Tablo 3.13 Metal Sektörü Birleşimli Tahmini	91
Tablo 3.14 Makine Sektörü Tahmini.....	91
Tablo 3.15 Endüstriyel Mutfak Sektörü Tahmini.....	91
Tablo 3.16 Yapı İnşaat Sektörü Tahmini.....	92
Tablo 3.17 Diğer Sektör Tahmini.....	92
Tablo 3.18 Toplam Satış DVR Eğitim Aşaması	93
Tablo 3.19 Toplam Satış DVR Test Aşaması Tahmini	93
Tablo 3.20 Toplam Satış ÇKA Eğitim Aşaması Tahmini	94
Tablo 3.21 Toplam Satış ÇKA Test Aşaması	95
Tablo 3.22 Toplam Satış GSR Eğitim Aşaması	95
Tablo 3.23 Toplam Satış GSR Test Aşaması	96
Tablo 3.24 Makine Sektörü DVR Eğitim Aşaması	97
Tablo 3.25 Makine Sektörü DVR Test Aşaması	97

Tablo3.26 Makine Sektörü ÇKA Eğitim Aşaması	98
Tablo 3.27 Makine Sektörü ÇKA Test Aşaması	99
Tablo 3.28 Makine Sektörü GSR Eğitim Aşaması.....	99
Tablo 3.29 Makine Sektörü GSR Test Aşaması.....	100
Tablo 3.30 Metal Sektörü Destek Vektör Eğitim Aşaması	101
Tablo 3.31 Metal Sektörü Destek Vektör Test Aşaması	101
Tablo 3.32 Metal Sektörü ÇKA Eğitim Aşaması	102
Tablo 3.33 Metal Sektörü ÇKA Test Aşaması	103
Tablo 3.34 Metal Sektörü GSR Eğitim Aşaması.....	103
Tablo 3.35 Metal Sektörü GSR Test Aşaması.....	104
Tablo 3.36 Endüstriyel Mutfak Sektörü DVR Eğitim Aşaması	105
Tablo 3.37 Endüstriyel Mutfak DVR Test Aşaması.....	105
Tablo 3.38 Endüstriyel Mutfak ÇKA Eğitim Aşaması.....	106
Tablo 3.39 Endüstriyel Mutfak ÇKA Test Aşaması.....	107
Tablo 3.40 Endüstriyel Mutfak GSR Eğitim Aşaması	107
Tablo 3.41 Endüstriyel Mutfak GSR Test Aşaması	108
Tablo 3.42 Yapı-İnşaat Sektörü DVR Eğitim Aşaması.....	109
Tablo 3.43 Yapı-İnşaat Sektörü DVR Test Aşaması.....	109
Tablo 3.44 Yapı-İnşaat Sektörü ÇKA Eğitim Aşaması.....	110
Tablo 3.45 Yapı-İnşaat Sektörü ÇKA Algılayıcı Test Aşaması.....	111
Tablo 3.46 Yapı-İnşaat Sektörü GSR Eğitim Aşaması	111
Tablo 3.47 Yapı-İnşaat Sektörü GSR Test Aşaması	112
Tablo 3.48 Diğer Sektör DVR Eğitim Aşaması	113
Tablo 3.49 Diğer Sektör DVR Test Aşaması	113
Tablo 3.50 Diğer Sektör ÇKA Eğitim Aşaması	114
Tablo 3.51 Diğer Sektör ÇKA Test Aşaması	115
Tablo 3.52 Diğer Sektör GSR Tahmini Eğitim Aşaması	115
Tablo 3.53 Diğer Sektör GSR Test Aşaması.....	116
Tablo 3.54 Zaman Serisi Modelleri	117
Tablo 3.55 Veri Madenciliği Tahmin Modelleriyle 1. Ay Tahmin Sonuçları.....	118
Tablo 3.56 Veri Madenciliği Tahmin Modelleriyle 2. Ay Tahmin Sonuçları.....	119
Tablo 3.57 Veri Madenciliği Tahmin Modelleriyle 3. Ay Tahmin Sonuçları.....	120

KISALTMALAR LİSTESİ

AI	: Artificial Intelligence
SQL	: Structured Query Language
ANN	: Artificial Neural Network
SQL	: Yapısal Sorgulama Dili
ARIMA	: Autoregressive Integrated Moving Averages
SVM	: Support Vector Machine
SVR	: Support Vector Regression
CART (C&RT)	: Classification and Regression Trees
TMMOB	: Türk Mühendis ve Mimar Odaları Birliği
ÇKA	: Çok Katmanlı Algılayıcı
DVM	: Destek Vektör Makine
VTBK	: Veri tabanlarında bilgi keşfi
Y.Z	: Yapay Zeka
DVR	: Destek Vektör Regresyon
YSA	: Yapay Sinir Ağları
EVDS	: Elektronik Veri Dağıtım Sistemi
GP	: Gaussian Process
GS	: Gauss Süreci
GSR	: Gauss Süreç Regresyon
ISSF	: International Stainless Steel Forum
KDD	: Knowledge Discovery in Databases
LME	: London Metal Exchange
LR	: Learning Rate
M5P	: Model Ağaçlar
MAE	: Mean Absolute Error
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
MLP	: Multi Layer Perceptron
PRE	: Pitting Resistance Equilavent
RBFN	: Radial Basis Function Network
RMSE	: Root Mean Squared Error
RRSE	: Root relative squared error

SEMMA : Sample, Explore, Modify, Model, Assess
SMO : Sequential minimal optimization
SMOreg : Sequential minimal optimization Regression



ÖZET

Ülkemiz paslanmaz çelik talebi büyük oranda ithal olarak karşılanmaktadır. Paslanmaz çeliğin ikamesi olmayan pahalı bir ürün olması, yaygın kullanımı ve ithal olarak karşılanması gibi nedenlerle paslanmaz çelik satış tahmini, gerek ülke ekonomisi gerekse sektörde faaliyet gösteren ithalatçı firmalar için önem arz etmektedir.

Bu çalışmada paslanmaz çelik sektöründe faaliyet gösteren farklı sektörlerdeki firmalara paslanmaz çelik ara mamulleri satan ithalatçı bir firmanın satış tahmini gerçekleştirilmiştir. Firma satış verileri, toplam ve sektörlerle göre düzenlenmiştir. Model Ağaç yöntemiyle Toplam ve sektörel satışların etkilendiği makro ekonomik değişkenler belirlenmiş ve geçmiş, günümüz ve geleceğe dair bilgiler veren tahmin edici modeller oluşturulmuştur. Hiyerarşik kümeleme analiziyle sektör hareketleri incelenerek ve benzer sektörlerin gruplandırılmasıyla satış hareketlerini içeren hiyerarşik zaman serisi elde edilmiştir. Hiyerarşik zaman serisi satış tahmini ARIMA ve Üstel Düzgünleştirme yöntemleriyle gerçekleştirilmiştir. Toplam ve sektörlerle göre satış tahminleri ise Destek Vektör Regresyon (DVR), Gauss Süreç Regresyon (GSR) ve Çok Katmanlı Algılayıcı (ÇKA) yöntemleri ile modellenmiştir.

Veri ön işleme aşamasında firmanın günlük satış verileri ve müşteri kodları eşleştirilerek, firmanın satış yaptığı 16 sektör belirlenmiştir. Firmanın satış stratejisine göre ağırlıklı olan Makine, Metal, Endüstriyel Mutfak ve Yapı-İnşaat sektörlerine ait satış rakamlarıyla, aylık satış bilgilerini içeren veri setleri hazırlanmıştır. Bu sektörler dışında kalan ve satıştaki payı az olan sektörler "Diğer Sektör" adıyla birleştirilmiştir. Satış veri setlerine paslanmaz çelik sektörünü etki edebilecek bağımsız değişkenler eklenerek veri hazırlama süreci tamamlanmıştır.

Model Ağaçlar yöntemiyle, Toplam ve Diğer Sektör satışlarını etkileyen en önemli değişkenin Dolar Kuru olduğu, Makine, Metal ve Yapı-İnşaat sektör satışlarını etkileyen en önemli değişkenin Üretici Fiyat Endeksi olduğu, Endüstriyel Mutfak sektör satışlarını etkileyen en önemli değişkenin Sanayi Ciro Endeksi olduğu gözlenmiştir.

Hiyerarşik Kümeleme analizi sonucunda Makine, Metal, Endüstriyel Mutfak, Yapı-İnşaat ve Diğer Sektör satış hareketleri, benzerlik özelliğine göre 2 kümede toplanmışlardır. İlk kümede (A düğümünde) Makine, Metal ve Endüstriyel Mutfak modelleri, ikinci kümede (B düğümünde) Yapı-İnşaat ve Diğer sektör satışları yer almıştır.

Hiyerarşik yapıdaki Toplam Satış, A düğümü, B düğümü ve 5 sektör olmak üzere toplam 8 düğüm için R yazılımı ile hiyerarşik zaman serisi tahmini yapılmıştır. Tümevarım

ve Optimal Birleşim olmak üzere 2 birleştirme yöntemi ile ARIMA ve Üstel düzgünleştirme yöntemleriyle toplamda 32 model oluşturulmuştur. Buna göre 32 modelin 6'sında MAPE değeri %10 değerinden küçük-eşit hesaplanmıştır. Bu durumda, çalışmada uygulanan zaman serileri yöntemlerinin kısmen başarılı olduğu söylenebilir.

DVR, ÇKA, GSR yöntemleriyle yapılan tahminlerde, MAPE değerleri, birinci ayda %5,53 ile %13,91 aralığında , ikinci ayda %5,05 ile %15,57 aralığında, üçüncü ayda %5,41 ile %16,9 aralığında hesaplanmıştır. Birinci aya ait tahminde, en başarılı tahmin ÇKA yöntemiyle (MAPE değeri %5,53) Toplam Satış için gerçekleşmiştir. Ayrıca kurulan 18 modelin 15'inde MAPE değeri %10'un altında hesaplanmıştır. İkinci ay için tahminde en başarılı tahmin, GSR yöntemiyle (MAPE değeri %5,05) Diğer Sektör Satışı için gerçekleşmiştir. Ayrıca 18 modelin 13'ünde MAPE değeri %10'un altında hesaplanmıştır. Üçüncü ay için tahminde en başarılı tahmin, GSR yöntemiyle (MAPE değeri %5,41) Diğer Sektör Satış için gerçekleşmiştir. Ayrıca 18 modelin 13'ünde MAPE değeri %10'un altında hesaplanmıştır. Tüm tahmin dönemleri incelendiğinde, 18 tahmin modelinin 16'sında DVR, 15'inde ÇKA , 10'unda GSR'nin başarılı olduğu söylenebilir.

Tez çalışmasında toplam satışı ve sektörlere göre satışları tahmin etmede Destek Vektör Regresyon yönteminin diğer yöntemlerden daha başarılı olduğu görülmüştür.

Anahtar Kelimeler: Veri Madenciliği, Model Ağaçlar, Kümeleme Analizi, Zaman Serileri, Satış Tahmini

SUMMARY

FORECASTING THE SALES VOLUMES AT STAINLESS STEEL SECTOR BY USING DATA MINING METHODS

The demand of stainless steel in our country is substantially fulfilled with importation. Sales forecasting of stainless steel is important both for the importer companies operating in the sector and for national economy because of being a widely used non-substitutional expensive and imported product.

In this study, the sales forecasting of an importer company which is active in stainless steel sector and selling midproducts of stainless steel to the firms in different sectors has been examined. The data on company sales were reorganized according to sectors and aggregate. Macroeconomic factors affecting total sales and sectoral sales by model tree method were determined and the predictive models informing about the past, present and future were formed. By hierarchical clustering analysis, sector movements were examined and hierarchical time series including sales movements were obtained by grouping similar sectors. Sales forecasting of hierarchical time series were made by ARIMA and Exponential Smoothing methods. The sales forecasting of total sales and sectoral sales were modelled by Support Vector Regression (SVR), Gaussian Process Regression (GPR) and Multi Layer Perceptron (MLP) methods.

16 sectors which buys stainless steel from the corporation were determined by matching the customer codes and corporation's daily sales data at the preprocessing phase of data. Data sets containing monthly sales information for Construction, Industrial Kitchen, Metal, Machinery sectors according to the sales strategy of the corporation were prepared. The sectors which are staying out of these sectors and those which have little shares on sales were gathered under the name of Other Sectors. Data preparing phase was completed by adding the independent variables which affect the sales of the stainless steel sector.

It was observed by using Model Trees Method that the most important factor affecting the sales of the Industrial Kitchen Sector was Industrial Endorsement Index, the most important variable affecting the sales of Machinery, Metal, and Construction Sectors was Producer Price Index, and the most important one affecting the Total Sales and the sales of Other Sectors was Dolar Exchange Rate.

Sales figures of Construction, Industrial Kitchen, Metal, Machinery and Other Sectors were gathered in 2 sets due to similarity characteristics after Hierarchical Clustering Analysis. As being the biggest cluster Machinery, Metal and Industrial Kitchen sales occurred in the first

cluster (A node), and in the second cluster (B node) includes Construction and Other sector sales.

Hierarchical time series prediction was made for 8 nodes consisting of 5 sectors and A node, B node and Aggregated Sales in hierarchical pattern by R software. Totally 32 models were formed by ARIMA and Exponential Smoothing methods and two combination methods as Bottom up and Optimal Combination methods. According to this just 6 of 32 models' MAPE values were calculated less than or equal to 10%. Therefore it can be said that the time series methods applied in this study were partially successful.

The predictions made by SVR, MLP, GPR methods, MAPE values were calculated between the intervals of 5,53%-13,91% in the first month, 5,05%-15,57% in the second month and 5,41%-16,9% in the third month. In the prediction of first month the most succesful one for Aggregated Sales actualized by MLP method (MAPE value 5,53%). In addition, 15 of 18 models' MAPE values were calculated below 10%. The most successful prediction for the second month was for the Other Sector Sales by GPR method (MAPE=5,05%). In addition, for 13 models amongst 18, MAPE values were calculated below 10%. The most succesfull prediction for the third month prediction realized for the Other Sector Sales by GPR method (MAPE value = 5,41%). In addition, MAPE values were calculated below 10% in 13 models of 18. Having analyzed all the prediction periods it can be said that SVR was succesfull in 16, MLP was succesfull in 15 and GPR was succesfull in 10 of 18 prediction models. It could be concluded that the SVR method was more succesful than the others to forecast the Aggregated Sales and the sales of each sectors.

Keywords: Data Mining, Model Trees, Clustering Analysis, Time Series, Sales Forecasting.

ÖNSÖZ

Söze eğitim hayatım boyunca ders aldığım hocalarıma teşekkür ederek başlamak istiyorum. Bu süreçte, doktora ders aşamasından başlayarak çalışmanın ortaya çıkmasını sağlayan Doç. Dr. Sezgin IRMAK hocama, öğrettikleri, destekleri ve danışmanlığı için teşekkür ediyorum.

Tez izleme sürecinde önerileri ve yönlendirmeleri için Prof. Dr. Can Deniz KÖKSAL hocama ve Doç. Dr. Emre İpekçi ÇETİN hocama teşekkürü bir borç biliyorum.

Son olarak, hayatımdaki moral kaynağım aileme , hayat arkadaşım Elif, büyük kızım İpek, süreçte ailemize katılan küçük kızım Seline daha fazla vakit ayıracağımlı -tahmin ederek- teşekkür ediyorum.

Orhan ECEMİŞ

Antalya, 2016

GİRİŞ

Hayat, öğrenme ile başlamaktadır. İnsanlık tarihi boyunca kazanımların bilgiye dönüşmesiyle ortak amacı “öğrenme” olan farklı bilim dalları gelişmiştir. Ortak noktası öğrenme olan bilim dalları doğal olarak farklılaşsa da, günümüzde bilim dallarının bir diğer ortak noktası problemlerin çözümünde ki alanlara özgü algoritmalar ve bilgisayar yazılımları daha genel bir ifadeyle bilişim teknolojileri olduğu söylenebilir. Bilgisayar donanımlarındaki gelişmeler daha fazla verinin depolanmasına aynı zamanda erişilebilir hale gelmesine imkan sağlamış, yazılım alanında ki gelişmeler ise yapay zeka kavramı ile başlamış devamında uzman sistemler ve veri madenciliği yöntemleriyle devasa veri yığınlarından nitelikli bilgilerin elde edilmesine yol açmıştır.

Tahmin kelimesi dilimize Arapçadan geçmiş olup Türk Dil Kurumu sözlüğünde “Yaklaşık olarak değerlendirme, oranlama” ve ”Akla, sezgiye veya bazı verilere dayanarak olabilecek bir şeyi, bir olayı önceden kestirme, kestirim” olarak ifade edilmektedir. Tahmin ilk çağlardan günümüze insanlık tarihinin bir parçası olmuştur. Eski çağlarda bilimsel özellikten uzak olan tahminler bazen ilahi bir işaret olarak insanları aldatmak, bazen de suç unsuru olarak haksız kazanç sağlamak amacıyla kullanılmıştır. İmparator Konstantin tarafından (M.Ö 357) matematikçilere ve kahinlere danışılması sonsuza kadar yasaklanmıştır. Benzer şekilde 1736 yılında İngilterede dolandırıcılığın artması sebebiyle “sözde tahminciler” için 3 ay kürek cezası uygulanmıştır.

Tahmin, sadece uluslararası, büyük şirketler tarafından değil, aynı zamanda küçük şirketlerin de ihtiyaç duyduğu bir olgudur; çünkü hemen hemen her organizasyon, eksik bilginin bulunduğu geleceğin koşullarını yerine getirmeyi planlamalıdır. Hükümetler ve kâr amaçlı örgütler, küçük sosyal kulüpler ve ulusal siyasi partiler, finans, pazarlama, personel ve üretim alanlarında tahmin gereklidir (Hanke vd., 2001).

Satış tahminine yönelik çalışmalar, 1 asır öncesine dayanan analitik çalışmalarla başlamış, matematik, istatistik ve ekonometri alanındaki gelişmelerle temelleri sağlamlaştırmış geçtiğimiz son on yılda bilişim sektöründeki kazanımlar, veri madenciliği yöntemleriyle ivme kazanmıştır. (Chase Jr vd., 2013: 1)

Satış tahminleri işletmelerin karar destek sistemlerinde belirgin bir rol oynamaktadır. Etkin satış tahminlerini önceden elde etmek karar vericinin üretim ve malzeme maliyetlerini hesaplamalarına, hatta satış fiyatını belirlemesine yardımcı olabilir. Bu, daha düşük bir

envanter seviyesine neden olacak ve tam zamanında hedefe ulaşmasına katkı sağlayabilir (Kuo ve Xue, 2012).

Günümüzde ekonomi, meteoroloji, tıp, mühendislik alanlarında, alanlara özgü bilimsel tahmin yöntemleri yoğun olarak kullanılmaktadır. Bu amaçla, tahmin doğruluğunu artırıcı bilimsel çalışmalar yapılmakta ve yeni bilimsel yöntemler araştırılmaktadır. İşletmeler için tahmin çalışması kuruluş aşamasıyla başlamaktadır. Kuruluş yeri seçimi, yatırım kararı alma, yeni bir ürün üretme, pazarlama stratejilerinin belirlenmesi, personel temini, satış hedeflerinin belirlenmesi gibi kritik süreçlerde uzmanlar tarafından tahmin çalışması yapılmaktadır.

Satış tahmini, işletmelerde (mal veya hizmet için) tüketici talebini anlamaya ve öngörmeye çalışan, yöneticilerin alacakları kararlara destek sağlayan analiz süreci olarak düşünülebilir. Tüketici talebi; zamana, doğal koşullara ve ekonomik faktörlere göre değişkenlik gösterebilir. Örneğin; talep yapısı ve mevsimsel dalgalanması dikkate alınmaz veya etkin bir talep tahmini yapılamazsa; satışlar, mal arzının eksikliği nedeniyle düşebilir veya mal arzının fazlası satıştan iade olması veya satılamaması nedeniyle stok maliyetini ve stok değer kaybını artırır. Oysaki etkin bir talep tahmini, bir şirket için rekabet avantajı sağlar ve pazarlama fonksiyonlarının daha işlevsel kullanılması imkânını sunar. İşletmelerin büyümelerinde/pazar payını korumasında/maliyet yönetiminde ve tüketici taleplerinin karşılanması ve müşteri tatmini için alacakları stratejik kararlarda, uygun tahmin modellerini kullanması çok önemlidir.

Birinci Dünya Savaşında başlayan, malzeme bilimindeki gelişmelerle birlikte paslanmaz çeliğin ortaya çıkışı, bir çok alanda standartları yükseltmiştir. Paslanmaz çelik her ne kadar savaş ortamında keşfedilse de günümüzde barışçıl amaçlarla pek çok alanda başta makine ve metal olmak üzere yapı-inşaat, endüstriyel mutfak, biyo-medikal, havacılık, gemicilik, gibi sektörlerde yoğun olarak kullanılmaktadır.

Paslanmaz Çelik sektöründe fiyatı etkileyen birçok faktör bulunmaktadır. Dünya Borsalarında işlem gören ham çelik, nikel, demir, krom madenleri gibi emtia fiyatları paslanmaz çelik üreticilerinin/ülkelerinin ekonomik göstergeleri ve tutumları genel olarak fiyatları etkilemektedir. Emtia fiyat dalgalanmalarına ek olarak, ithal talepten dolayı döviz kuru satış hareketlerini etkilemektedir.

Ülkemiz paslanmaz çelik talebinde büyük oranda dışa bağımlı olduğundan dolayı net ithalatçı olduğumuzu söyleyebiliriz. Firmaların satış tahminlerindeki gerçekleşmesi muhtemel büyük sapma, stokta çalışma ve bekleme maliyeti gibi ekstra maliyetler oluşturabilir. Maliyet yönetiminde problemlerin yaşanması firmanın finansal yapısını kırılgan hale getirebilir. Sektörde faaliyet gösteren ithalatçı firmaların, karşılaşılabilecekleri finansman sorununu

çözebilmeleri ve etkin kararlar alabilmeleri için talep/satış tahminlerine ihtiyaç duymaktadırlar.

Veri madenciliği, bilgi teknolojisinin doğal gelişiminin bir sonucu olarak ifade edilebilir. Veritabanı ve veri yönetimi endüstrisi, veri toplama ve veritabanı oluşturma, veri yönetimi (veri depolama ve geri alma ve veritabanı işlem işleme dahil) ve gelişmiş veri analizi (veri ambarı ve veri madenciliği dahil olmak üzere) gibi kritik işlevlerin geliştirilmesiyle gelişmiştir. Günümüzde çok sayıda veritabanı sistemi, sorgu ve işlem işlemlerini ortak uygulama olarak sunmaktadır (Han ve Kamber, 2012: 3-4). Veri madenciliği yöntemleri Tanımlayıcı ve Tahmin Edici modeller olarak ikiye ayrılabilir. Tanımlayıcı modeller veri seti hakkında açıklayıcı özet bilgiler sunmaktadır. Tahmin Edici modeller ise belirli hedef doğrultusunda değişkeninin bilinmeyen değerini tahmin etmeye çalışmaktadır (Hand, 2001: 195). Veri Madenciliği uygulamaları, ağırlıklı olarak karar alma ve planlama süreçlerinde yoğunlaşmaktadır. Buna talep, satış ve maliyet tahminlerinin yapılması, pazarlama stratejilerinin belirlenmesi gibi örnekler verilebilir.

Genel olarak bu tip çalışmalarda burada adı geçen tahmin yöntemleri kullanılmaktadır. Çalışmada veri madenciliği yöntemlerinden, Karar Ağaçları, kümeleme ve tahmin yöntemleri uygulanmıştır. Bu yöntemlerin kullanılma sebepleri şunlardır. Karar Ağaçları, sınıflandırma ve regresyon problemlerinde parametrik olmayan denetimli öğrenme yöntemleri olarak ifade edilebilir. Karar Ağaçlarının amacı, veri özellikleri dikkate alınarak basit karar kuralları öğrenilmesi, hedef değişkenlerin değerlerini tahmin edebilen modeller oluşturulmasıdır. Çalışmada kullanılan M5P model ağaçlar yöntemiyle sektörlerin hangi değişkenlerden etkilendiği ve sektöre özgü kritik kök nitelik değerlerine göre tahmin edici doğrusal modeller elde edilmiştir. Böylece geçmişte sektörlerin hangi makro ekonomik değişkenlerden ne derecede etkilendiği ve sektörlerle göre tahmin modelleri, cari değerlere göre hangi makro ekonomik değişkenden ne ölçüde etkilenebileceği ve ileriye dönük tahmin modelleri elde edilmiştir. Uygulama WEKA yazılımında gerçekleştirilmiştir.

Hiyerarşik kümeleme yöntemiyle-mesafe ölçütü varyans analizene dayanan wards yöntemi kullanılarak- sektörler hiyerarşik biçimde ifade edilmiştir. Ardından ARIMA ve Üstel Düzgünleştirme Yöntemleriyle, tahmin yönü Aşağıdan Yukarı, Optimal Birleşim, olmak üzere bütün sektörler, sektörlerin oluşturduğu düğümler ve toplam satış için tahmin, R programında gerçekleştirilmiştir.

Yapay Sinir Ağları (YSA) mühendislik, tıp ve ekonomi uygulamalarında başarılı sayılabilen uygulamalarda kullanılmaktadır. Yapay Sinir Ağları ortaya çıktığı 1943 yılından itibaren ilerleme, duraklama ve tekrar ilerleme, evrilme süreçlerini yaşamıştır. Minibüs

sürücüsünü izledikten sonra 30 km. hızla kullanan ALVINN adlı yapay sinir ağı modeli, usta düzeyde tavla oynayabilen yazılım Yapay Sinir Ağları kullanıma dair ilginç örnekler olarak ifade edilebilir (Alpaydın, 2011: 231). Çalışmamızda YSA'ile sektörlerin ileriye dönük kısa dönem tahmini gerçekleştirilmiştir. YSA performansı, zaman serilerine göre başarılı Destek Vektör ve Gauss Süreç regresyon yöntemlerine göre ise başarısız bulunmuştur.

Destek Vektör Regresyon (DVR), literatüre göre YSA'ya göre yeni sayılabilen, sınıflandırma problemlerinde başarılı sonuçlar üreten Destek Vektör Makinelerinin (DVM) regresyon problemlerine uyarlanmış hali olarak ifade edilebilir. Destek vektör makinelerinin performansı üzerinde hata terimi (ϵ), düzenleme faktörü (C) ve Kernel parametresi olmak üzere 3 hiperparametre bulunmaktadır. Model parametreleri ve Kernel tipinin seçimi destek vektör makinelerinin performansı üzerinde etkilidir. Parametrelerin doğru seçimi bazen zorluk olarak karşımıza çıkabilmektedir. Çalışmada DVR yöntemiyle başarılı tahmin değerleri elde edilmiştir.

Gauss Süreç Regresyon (GSR) yöntemi hiperparametre ayarlamalarına ihtiyaç duymadan regresyon problemlerinde kullanılabilir. Veri ön işleme aşamasında kolaylık sağlamaktadır. Gürültü seviyesi,bağımlı ve bağımsız değişkenlerin normalleştirme / standardizasyon gibi dönüşümleri, eksik değerlerin, küresel ortalama / mod ile değiştirilmesi, Nominal nitelikler binary'e dönüştürülmesine imkan sağlamaktadır. Çalışmada GSR yöntemiyle başarılı tahmin değerleri elde edilmiştir.

Çalışma kapsamında; birinci bölümde iktisat ve işletme literatürüne göre genel olarak talep, talep tahmini ve tahminde kullanılan yöntemler ele alınmıştır.

Çalışmanın ikinci bölümünde; veri madenciliği kavramı yöntem ve teknikleri (Destek Vektör Regresyon, Gauss Süreç Regresyon, Yapay Sinir Ağları, karar ağaçları, kümeleme), incelenerek, ayrıca literatürde yer alan veri madenciliği yöntemleri kullanılarak yapılan tahmin çalışmaların araştırılması ve özetlenmesi yer almaktadır.

Üçüncü bölümde ise Türkiye'de ve Dünyada paslanmaz çelik sektörüne ait bilgiler ile sektör talebine etki eden faktörler değerlendirildikten sonra karar ağacı, kümeleme ve tahmin modellerine yönelik uygulamalar ve bu uygulamalara ait sonuçların yorumlanması yer almaktadır. Paslanmaz çelik bileşimi, özellikleri, üretim aşamalarına ait bilgiler ve çeşitli istatistiksel veriler Uluslararası Paslanmaz Çelik Forumu'nun (International Stainless Steel Forum) web sitesinden elde edilmiştir. ISSF, 1996 yılında kurulmuştur, 25 ülkede 65 paslanmaz çelik üreticisinin üye olduğu, uluslararası paslanmaz çelik endüstrisi için kar amacı gütmeyen araştırma ve geliştirme organizasyonudur. Makro ekonomik değişkenler ise Türkiye Cumhuriyet Merkez Bankası elektronik veri dağıtım (TCMB-EVDS) sisteminden

elde edilmiştir. Firmanın satış stratejisine göre önemli olan Makine, Metal, Endüstriyel Mutfak, Yapı-İnşaat ve Diğer sektörlerle ait veri setleri hazırlanmıştır. Model ağacı (M5P) yöntemiyle toplam ve alt sektörlerle ait satışların hangi makro ekonomik değişkenlerden etkilendiği araştırılmıştır. Sektör satış hareketleri kümeleme yöntemiyle hiyerarşik bir yapıda ifade edilmiştir. Hiyerarşik modele ait tahmin ve sektörlerle ait kısa dönem tahmin modelleri oluşturulmuş ve yöntemlerin başarısı karşılaştırılmıştır.



BİRİNCİ BÖLÜM

TAHMİN YÖNTEMLERİ

1.1 Satış ve Talep Kavramları

Satış tahminleri, geçmiş tarihsel verilere dayalı olarak gelecekteki satışların tahmini ile ilgilidir. Rekabet ve küreselleşmeden dolayı, satış tahminleri ticari bir teşebbüsün bir karar destek sisteminde belirgin bir rol oynamaktadır. Etkin satış tahminlerini önceden elde etmek karar vericinin üretim ve malzeme maliyetlerini hesaplamalarına, hatta satış fiyatını belirlemesine yardımcı olabilir (Kuo ve Xue, 2012). Satış tahminleri, işletmenin faaliyet alanının özelliğine göre dahili ve harici ortamların etkisiyle çok karmaşık bir sorun haline gelebilir. Bu nedenle günümüzde daha doğru ve zamanında kalitatif satış tahmin yöntemlerinin uygulanması ve geliştirilmesi önemli bir araştırma konusu haline gelmektedir (Luxhøj ve Stensballe, 1996). Bazı perakendeciler piyasa bilgilerini edinerek ve tahminlerini revize ederek stoklama kararlarını geliştirirler. İyi bir tahmin yöntemi, perakendecilerin fazla stoklama ve az stoklama maliyetlerini azaltmalarına yardımcı olabilir. Böylece, satış tahminleri, belirsizlik altında tedarik zinciri yönetiminde çok önemli bir görev haline gelir ve perakendecileri ve diğer kanal üyelerini çeşitli şekillerde büyük ölçüde etkiler (Sun vd., 2008).

Satış tahminleri, işletme faaliyetlerinin çeşitli safhalarını planlamanın temeli olup dinamik piyasa talepleri doğrultusunda tedarik zinciri yönetiminde çok önemli bir görev olup perakendecileri ve diğer kanal üyelerini çeşitli şekillerde etkilemektedir (Xiao ve Yang, 2008). Etkin olmayan satış tahminleri, başarılı olmayan üretim planlanmasına dolayısıyla sipariş kaybı, yetersiz müşteri hizmetleri ve yetersiz kullanılmış kaynaklara neden olabilir. Son yıllardaki araştırmalar, etkin satış tahmininin işletme performansında iyileşme sağladığını göstermiştir.

İktisat bilimi açıdan talep tanımı, tüketicilerin satın alma niyetlerindeki miktarı işaret etsede, İşletme bilimi açısından firmanın bütün fonksiyonlarını (üretim, pazarlama, stok yönetimi vb.) harekete geçiren, bilimsel yöntemlerle doğru anlaşılması gereken bir olgudur. İşletmelerin kuruluş amacının, pazar talebini karşılayabilecek ürünlerin veya hizmetlerin üretilmesi, pazarlanması ve nihayetinde kar elde edilmesi olduğunu düşünebiliriz. Tüketici talebinin tahmini ve etik çerçevede yönlendirilmesi, işletmeyi istikrarlı bir şekilde yürütülmesinde katkı sağlayabilir. Bu amaçla işletmeler, tüketicilerin taleplerini bilimsel bir şekilde anlamalı ve buna uygun üretim sistemi kurmalı, pazarlama, reklam, Ar-Ge faaliyetleri yürütmelidir. Bu bilgiler ışığında talep tahmini, işletmelerin üreteceği/ üretmeyeceği ürünler ve hizmetlere ilişkin tüketici talebinin gelecek dönemler (kısa, orta, uzun) için tahmin edilmesi olarak tanımlanabilir.

Hangi ürünün hangi miktarda üretileceği veya üretilmeyeceği, tüketicilerin ileriye yönelik taleplerinde yaşanacak artış ve azalışların tespitinde, oluşturulacak senaryolar talep tahminleri ile yorumlanır (Çağlar, 2007).

Talep tahminlerin tekil olarak işletmeler önemli olduğu gibi genel olarak makroekonomik planda iki açıdan önemi aşağıdaki gibi ifade edilebilir.

- Mevcut talebin karşılanmaması halinde fiyatların yükselmesi ve dolayısıyla da ithal talepte artış oluşabilir. Bu durumda işletmelerin piyasa talebini karşılamak amacıyla mevcut kapasitelerinin üzerinde üretim yapmaları sonucunda maliyet artışı söz konusu olabilir.
- Talep azlığı durumunda stoklarda artış olabilir. Bu durumda işletmeler mevcut kapasitelerinin altında üretim yapmalarıyla kaynak israfı oluşabilir.

Başarılı talep tahmini ile işletmelerin karşılabileceği olumsuz durumlar önlenebilir, optimal kaynak kullanımı ve sürdürülebilir üretim verimliliği sağlanabilir (Tekin, 2009).

Talep tahmini Üretim Yönetimi, Pazarlama, Ar-Ge faaliyetlerindeki önemi aşağıdaki gibi ifade edilebilir.

Üretim Yönetimi Planlama: Talep tahmini üretim faaliyetlerinin planlamasında ki birinci basamaktır. Talep konusunda belirsizlikler giderilmeden bir planlamanın yapılması söz konusu değildir (Kobu, 1994).

Pazarlama: Beklenen karlılık düzeyine ve dönemsel hedeflere ulaşılması, pazarlama planlarında yer almaktadır. İşletmeler gerçeğe yakın olması beklenen tahmin modelleriyle kaynakları verimli şekilde kullanabilir nihayetinde optimum karlılık düzeyine ulaşabilirler. (Bolt, 1981). Pazarlama stratejilerinin belirlenmesinde önemlidir (Özdemir ve Özdemir, 2006).

Ar-Ge: Tüketicilerin ihtiyaç duyduğu yeni ürünlere olan talebin tahmini Ar-Ge faaliyetlerin başarısını firma için ise karlılığını artırmaktadır.

Satış tahmini;

Yüksek bir tahmin tutarlılığı aşağıdaki faaliyetlerin yürütülmesinde gereklidir (Tokpınar, 2014).

- Ürün/Yeni ürün konusunda müşteri eğilimlerinin tespiti,
- Üretim kapasitesi belirlenmesi
- Personel sayısının belirlenmesi
- Stok planının yapılması
- Başarılı yatırım kararların alınması
- Alternatif üretim stratejilerinin belirlenmesi
- Ürün maliyetinde azalma

- Nakit akışının düzenli ve beklenen şekilde ve düzenli bir şekilde olması
- İleriye dönük satış hedeflerinin belirlenmesi ve buna uygun planlamalar yapılması

Kullanılan yöntemlerin etkinliğine bağlı olarak talep tahminleri başarılı ya da başarısız olabilir. Talep tahminleri belirli dönemler için (kısa-orta-uzun) yapılmaktadır. Gelecekle ilgili faaliyetlerin planlanması amacıyla ürün ve hizmetlere olan talepler çeşitli yöntemlerle tahmin edilir (Tanyaş ve Baksak, 2008).

Tahmin süreci 5 aşamada ifade edilebilir (Hydman, 2014: 15)

Adım 1: Problem tanımı: Genellikle bu tahminin sürecinin en zor kısmıdır. Problemin doğru bir şekilde ele alınması, işletmenin gereksinim duyduğu tahmin dönemi, yöntemi gibi özelliklerin belirlenmesi ayrıca gelecek planlaması için işletmedeki uzmanlardan görüş alınması gibi işlemlerden oluşmaktadır.

Adım 2: Veri Toplanması: Veri toplama süreci iki farklı şekilde gerçekleşebilir. Bu süreçler:

(a) Ham verilere ulaşılması: Verileri işletme içi ve işletme dışı olarak ele alabiliriz. İşletme içi verilere, veritabanları işlevlerini kullanarak ilk elden güvenilir biçimde verilerin elde edilmesi gereklidir. İşletme dışı verilere güvenilir ve tarafsız kaynaklardan elde edilmesi gereklidir.

(b) Uzman Görüşlerinin Alınması: Ham verilere ulaşamadığı durumlarda örneğin yeni ürün çıkışında geçmişe ait veri bulunamadığından dolayı uzman görüşlerinin alınması yargısal tahmin modellerinde ve tahmin sürecinde katkı sağlamaktadır.

Adım 3: Ön Keşif analizi: Veri setinin grafiksel olarak incelenmesi başka bir ifadeyle desenlerintutarlılığı, mevsimsel etkinin olup olmadığı , önemli değişim, eğilim bulunup bulunmadığı gibi kontrollerin yapılmasıdır.

Adım 4: Model Seçimi: Veri setine uygun zaman serisi, regresyon ve veri madenciliği yöntemlerinden hangisi/hangilerinin kullanılacağına karar verilmesidir.

Adım 5: Tahmin: Tahmin yöntemiyle analizin yapıldığı ve model performansının değerlendirildiği son aşamadır. Tahmin sonuçlarının yeterli olmadığı durumlarda önceki adımlarda ifade edilen aşamalar tekrarlanmaktadır.

Kısaca tahmin sürecinde, bağımlı ve bağımsız değişkenlerin veya değişkenin kendisinin geçmiş dönem değerlerine ait verilerle, ele alınan probleme, oluşturulan verilere uygun yapay öğrenme, istatistik vb. yöntemlerle, gelecekteki değişimlerin neler olabileceğine dair ön görüde bulunmaktadır.

1.2 Tahmin Yöntemleri

Yargısal Yöntemler: Teknolojik yöntemlere kıyasla daha basit ve düşük maliyetli yöntemler olup bireysel fikir, yargı ve düşünceler tahminin oluşturulmasında kullanılır. Yargısal yöntemler geçmiş verilerin az olduğu ve/veya objektif yöntemle ilişkin uzmanlığın yetersiz olduğu durumlar için son derece uygundur. Yargısal tahmin yalnızca yargı yürütme olmayıp

ilgili tahminin tanımlanması, geçerli uygun bilginin gözden geçirilmesi, uzman görüşleri ile uyumlu olarak sorunun yapılandırılması, değerlendirme tekniklerinin kullanılması, nitel yargıların nicel tahmine dönüştürülmesi, hem sistem hem de bir süreç dahilinde olmaktadır (Tokpınar, 2014).

Teknolojik Yöntemler: Teknolojik tahmin, gelecekteki karakteristiklerin ya da yararlı makinelerin, tekniklerin veya prosedürlerin uygulanmasına yönelik yöntemlerdir. Bu yöntemler uzak dönemli tahminlerde kullanılırlar, gelecekte teknolojinin nasıl kullanılacağı, teknoloji gelişmelerin etkilerinin ve buna bağlı yaratıcılıkların nasıl olacağına yoğunlaşır. Diğer bir ifadeyle teknolojik, sosyal, ekonomik ve politik çevrenin uzun vadedeki gelişimini tahmin edilmesini amaçlar. Teknolojik yöntemler aşağıdaki gibi sıralanabilir:

- Keşifsel yöntemler
- Delphi
- Normatif yaklaşımlar
- Çapraz etki anketi
- İlişki ağacı yaklaşımı
- Senaryo analizleri
- Yapay sinir ağları

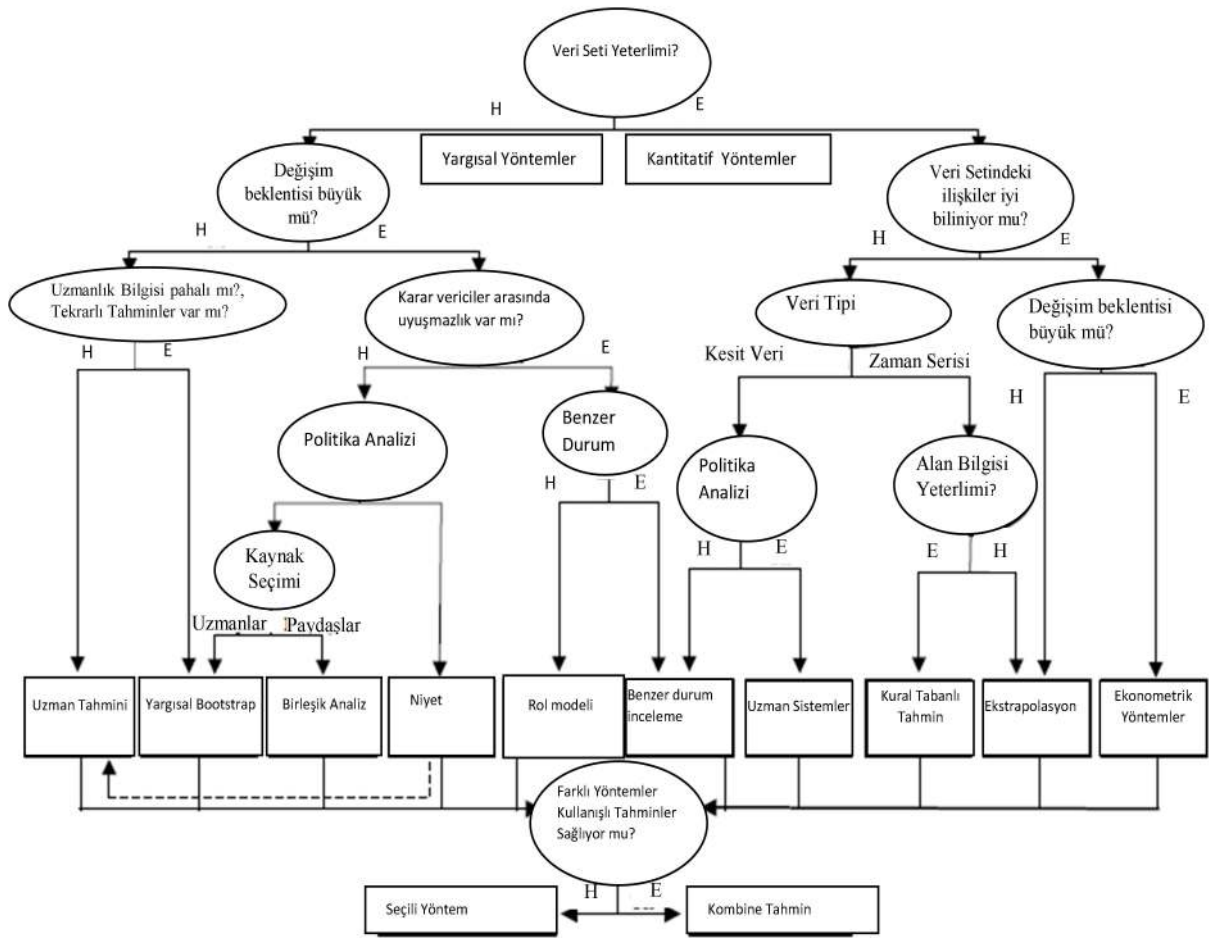
Teknolojik yöntemler “açıklayıcı” ve “normatif” teknikler şeklinde iki yaklaşım ile ele alınmaktadır.

Açıklayıcı Teknikler: Gelecekte neler olacağına ilişkin belirli tahminlerin yapıldığı teknikler olup bunlar olasılık ifade edecek şekilde de olabilir. Delphi yöntemi ve genel eğilimler açıklayıcı tekniklere örnek olarak verilebilir.

Normatif Teknikler: Gelecekteki belirli bir amaç için gelecekte günümüze gelerek hangi alanda ya da teknolojilerde neler yapılması gerektiği, bunların maliyetlerinin, sıklıklarının neler olacağı ve bu hedefe ulaşmak için potansiyel tehlike ve eksikliklerin neler olacağının belirlenmesini kapsar.

1.2.1 Tahmin Yöntemleri Seçimi

Tahmin yöntemleri seçiminde, Armstrong (2001: 376) tarafından önerilen akış şeması Şekil 1.1'de yer almaktadır. Tahmin edilecek veri setindeki sayısal bilgi yeterli olmadığı durumlarda yargısal yöntemler, yeterli olması durumunda kantitatif yöntemler seçilmektedir.



Şekil 1.1 Tahmin Yöntemleri Seçimi Akış Şeması

Kaynak: Armstrong, 2001: 376

Yargısal yöntemlerde tahmin yöntemi seçiminde beklentinin büyük veya küçük olması önemlidir. Beklenen değişiminin küçük olması durumunda tahmin yöntemlerinin doğruluğunda önemli farklar bulunmayabilir. Bu durumda uzman görüşlerinin alınması yeterli olabilir. Süreçte tekrarlı tahminlerin olmasında uzman görüşleri maliyeti göz önüne alındığında yargısal bootstrap yöntemi tercih edilebilir. Değişim beklentisinin büyük olduğu ve karar vericiler arasında uzlaşmanın olduğu durumlarda tahminler uzmanlardan veya paydaşlardan elde edebilir. Karar vericilerin farklı politikaları incelemesi için yargısal bootstrap ve birleştirici analiz yöntemleri tercih edilebilir. Uzmanların, yeteri kadar deneyime sahip olmaması durumunda, potansiyel müşterilerden görüş alınması tahmin sürecine yardımcı olabilir. Diğer bir ifadeyle birleştirici analiz, tüketicilerin alternatif tekliflere nasıl tepki verdiklerini temel alan bir tahmin modeli geliştirilmesinde kullanılabilir.

Değişim beklentisinin büyük karar vericiler arasında uzlaşmanın olmadığı durumlarda:

- Geçmişteki örnek olaylar arasında bağlantı kurulması tahmin sürecinde kullanılabilir.
- Benzerlik kurulamadığı durumlar için, rakiplerin nasıl davranacağı konusunda alternatif durumları içeren rol model tahmin sürecine katkı sağlayabilir.

Kantitatif yöntemlerde veri setindeki değişkenler arasındaki ilişkinin bilinmesi durumunda:

Değişim Beklentisinin Büyük Olması: Ekonometrik modeller uygulanmaktadır. Tahmin sürecinde değişkenler arasındaki nedensellik kavramı önem arz etmektedir.

Değişimin Beklentisinin Küçük Olması: Kısa vadeli tahmini içeren çalışmalarda ekstrapolasyon yöntemleri kullanılmaktadır. Allen ve Fildes yöntemin, ekonometrik yöntemler kadar başarılı olabileceğini belirtmişlerdir. (Allen ve Fildes, 2001)

Kantitatif yöntemlerde veri setindeki değişkenler arasındaki ilişkinin bilinmemesi durumunda yatay kesit veriler için:

- Alternatif politikaları karşılaştırmanız gerekip gerekmediği araştırılmalıdır.
- Uzmanlar tahminleri temel olarak benzetmeyi kullanabilirler.

Kantitatif yöntemlerde veri setindeki değişkenler arasındaki ilişkinin bilinmemesi durumunda zaman serisi veriler için:

- Alternatif politikaları karşılaştırmanız gerekip gerekmediği araştırılmalıdır.
- Uzmanlar tahminleri temel olarak benzetmeyi kullanabilirler.

1.2.2 Zaman Serileri

Zaman serileri, zaman ölçütüne göre değişken değerlerinin ardışık bir şekilde sıralandığı veri seti olarak düşünülebilir. Zaman serileri analizindeki temel yaklaşım, incelenen değişkenin geçmişte göstermiş olduğu seyri açıklamak ve bu bilgileri kullanmak suretiyle gelecekte beklenen davranışlarını belirlemektir (Wilson ve Keating, 2001).

Zaman serisi modelleri, ekonometrik modellerden farklıdır. Bir zaman serisi modeli ekonometrik model gibi, zaman serisi değişkeninin (Y_t), $X_{1t}, X_{2t}, \dots, X_{nt}$ gibi diğer değişkenler arasındaki davranışsal $Y_t = f(X_{1t}, X_{2t}, \dots, X_{nt})$ ilişkisini belirlelemekten ziyade Y_t gibi bir değişkendeki değişimi, kendi geçmiş değerlerindeki gelişmelerle tahmin etmeye çalışır (Sevüktekin, 2012: 156).

Zaman serisi yöntemleri veri ihtiyacının az olması, kolay model oluşturulabilmesi, kısa vadede iyi sonuç vermesi gibi avantajlara sahip olması bakımından makro ve mikro düzeyde son derece yoğun bir şekilde kullanılmaktadır. Bu yöntemler satış tahmini, hisse

senedi ve tahvil analizi, bütçe analizi, kalite kontrol ve nüfuz analizi gibi alanlarda kullanılmaktadır (Akgül, 2003a: 12).

Zaman serilerinin genel özellikleri:

1. Zaman serileri verilerinin diğer seri verilerinden farklı olmasının belirgin özelliklerinden birisi zaman süreci içerisinde serideki gözlem değerlerinin birbiri ile bağımlı olmasıdır.
2. Zaman serileri analizinin temel amacı veri setindeki değişken değerlerini kullanarak stokastik sürecin özellikleri veya temel öğeleri hakkında çıkarımlarda bulunmaktadır.
3. Zaman serilerinin geleneksel ayrışım yöntemi, serideki trend, konjonktür, mevsim etkileri ve düzensiz hareketlerin ayrıştırılmasını inceler. Bu açıdan toplamsal zaman serisi modeli, $Y_t = T_t + C_t + S_t + I_t$ biçiminde ve çarpımsal model ise $y_t = S_t \times T_t \times E_t$ biçiminde yazılabilir.
4. Ekonomik zaman serilerinde durağan- dışılığın nedeni olarak genellikle trend (deterministik veya stokastik), konjonktürel ve mevsimsel hareketlerin etkilediği ileri sürülmüştür.

Zaman serileri, verilerin sistematik ve rassal olmak üzere ikiye ayrılmasından sonra analiz edilir. Zaman serisi bileşenleri aşağıdaki gibi ifade edilebilir (Hanke ve Reitsch, 1998):

Trend: Uzun bir zaman dilimi boyunca zaman serilerindeki artış ya da azalışı gösterir.

Mevsimsellik: Zamana bağlı olarak ortaya çıkan, aynı dönemlerde görülen düzenli değişimleri ifade eder.

Çevrimsel (Konjonktürel) Hareketler: Çevrimsel Hareketler, uzun vadede (2-10 yıl), genellikle ekonomik duruma bağlı olarak serinin düzeyinde ortaya çıkan değişimlerdir.

Rassallık: Zaman serisinin trend, mevsimsellik, konjonktürel bileşenlerinin ayrıştırılmasından sonra serideki değişkenliği ifade eder. Kalitatif ya da kantitatif yöntemler ile belirlenemez.

Geleneksel zaman serisi ayrışım yöntemleri Toplamsal ve Çarpımsal Ayrıştırma Yöntemi olarak ele alınmaktadır.

Toplamsal zaman serisi modelini (y_t), mevsimsel bileşen (S_t), trend bileşen (T_t), hata terimi E_t olmak üzere aşağıdaki gibi bileşemlerin toplamı şeklinde ifade edebiliriz.

- Toplamsal modellerde mevsimsel değişim, trendden bağımsız olduğundan dalgalanma büyüklüğü zaman içinde değişmemekte, yani sabit kalmaktadır.

$$y_t = S_t + T_t + E_t$$

- Çarpımsal zaman serisi modelini (y_t), mevsimsel bileşen (S_t), trend bileşen (T_t), hata terimi E_t olmak üzere aşağıdaki gibi bileşemlerin çarpımı şeklinde ifade edebiliriz.

Çarpımsal modellerde mevsimsel değişim, trende döngüsü etrafında orantılı olarak yer almaktadır. Ekonomik zaman serisi ile çarpımsal modeller yaygındır.

$$y_t = S_t \times T_t \times E_t$$

Çarpımsal modeli eşdeğer şekilde şöyle ifade edilebilir.

$$\log y_t = \log S_t + \log T_t + \log E_t.$$

Burada amaçlanan zaman serisinin değişimi zamana göre kararlı olması içim verilerin logaritmasının alınması ve toplamsal model şeklinde ifade edilmesidir.

1.2.2.1 Üstel Düzgünleştirme Yöntemleri

Üstel Düzgünleştirme Yöntemlerinin temelleri, Holt 1957, Brown 1959 ve Winters'ın 1960 yıllarında yapmış oldukları çalışmalara dayanmaktadır. Üstel düzeltme yönteminde, daha önceki gözlem değerlerinin üstel ağırlıklı hareketli ortalaması alınmaktadır. Model trend içermeyen veri setleri için uygundur. Serinin cari düzeyi tahmin edildikten sonra, ileriye dönük tahmin amacıyla kullanılmaktadır. Üstel düzgünleştirme yöntemi, zaman serisi değerlerini eşit olmayan bir biçimde ağırlıklandırmakta yakın geçmiş gözlemlere daha fazla ağırlık vermektedir. (Hyndman ve Athanasopoulos, 2014: 165, Bowerman ve O'Connell, 1993: 379)

1.2.2.1.1 Holt Yöntemi

Holt yönteminde, zaman serisinin tahmini, trend tahmini ve ileriye yönelik tahmini hesaplanmaktadır. (Hanke ve Wichern, 2008). Ampirik kanıtlar Holt doğrusal tahmin fonksiyonunun aşırı tahmin ettiğini göstermektedir (Gardner ve McKenzie, 1985). Model kullanılan α düzgünleştirme katsayısı ve β değeri trend tahmini için düzleştirme katsayısıdır. Holt denkleminde üç denklem kullanılır (Hyndman ve Athanasopoulos, 2014: 173):

Üstel düzeltilmiş seri ya da cari
düzey tahmini

Trend tahmini

h dönem ilerisi tahmini

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad T_t = \beta (L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad \hat{Y}_{t+h} = L_t + pT_t$$

1.2.2.1.2 Winters Yöntemi

Winters Yönteminde, zaman serisinde mevsimsellik ve trend tahmin edilmektedir. Bunlardan α değeri düzeltme katsayısını, β değeri trend tahmini için düzeltme katsayısını ve γ değeri ise mevsimsellik tahmini için düzeltme katsayısını ifade etmektedir.

Winters yönteminde 4 bileşen için tahmin kullanılır (α , β , γ düzgünleştirme katsayısı)

cari düzey tahmini

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1})$$

Trend tahmini

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

Mevsimsellik tahmini

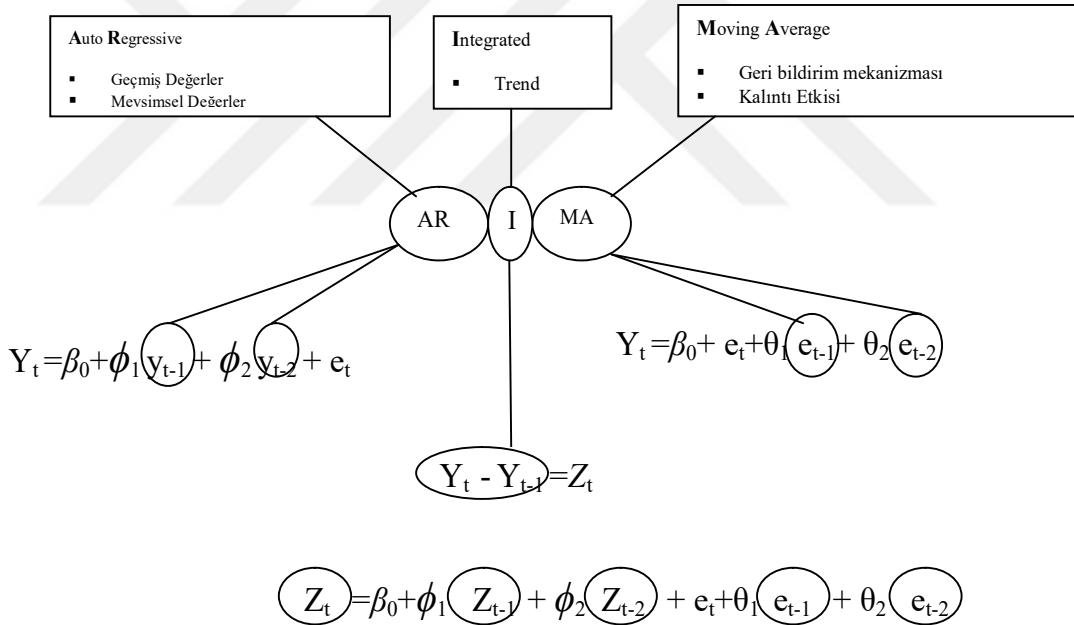
$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s}$$

h dönem tahmini

$$\hat{Y}_{t+h} = (L_t + pT_t)S_{t-s+h}$$

1.2.2.2 Otoregresif Bütünleşik Hareketli Ortalama Modelleri

Box-Jenkis tarafından ileri sürülen Otoregresif Bütünleşik Hareketli Ortalama (ARIMA) modeli yazarların soy ismiyle de Box-Jenkins veya soy isimlerinin baş harfleriyle de BJ anılmaktadır. ARIMA modeli Otoregresif Süreç (AR), Hareketli Ortalama (MA) süreçlerinin birleşimidir.



Şekil 1.2 Otoregresif Bütünleşik Hareketli Ortalama (ARIMA) Modeli

Kaynak: Chase Jr, 2013.

ARIMA modeli genel yapısı Şekil 1.2'de yer almaktadır. ARIMA modeli ile tahmin süreci dört adımda özetlenebilir (Pektaş, 2013: 164)

Model Belirleme: Zaman serisi analizinde kullanılacak B-J yöntemi belirlenir.

Parametre Seçimi: Zaman serisine ait p, d, q değerleri tespit edilir.

Tanısal Denetim: Verilerin modele yeterli derecede uygun olup olmadığı incelenir. Uygun olmadığı durumlarda (p,d,q) değerleri seçilir. (2. Adım)

Tahmin: Yeterli olduğuna karar verilen model, serinin örneklem dışı değerlerini tahmin etmek amacıyla kullanılır.

1.2.2.2.1 Otoregresif(AR) Süreci

Çoklu regresyon modelinde, bağımsız değişkenlerin doğrusal kombinasyonu kullanılarak bağımlı değişken tahmin edilmektedir. AR sürecindeyse, değişkenin kendisine ait geçmiş değerlerin doğrusal kombinasyonu kullanılmasıyla değişken tahmin edilmesidir. Başka bir ifade ile AR model değişkenin kendisinin geçmiş değerleriyle tahmin edildiği regresyon modelidir. Aşağıdaki şekilde ifade edilebilir. (Hyndman ve Athanasopoulos, 2014: 217)

$$Y_t = \beta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

$\phi_1=0$, y_t beyaz gürültü

$\phi_1=1$ ve $c=0$, y_t rassal yürüyüş

$\phi_1=1$ ve $c \neq 0$, y_t kayan rassal yürüyüş

$\phi_1 < 0$, y_t pozitif ve negatif değerleri arasında salınım eğilimi göstermektedir.

Durağan serilerde uygulanan AR sürecinde değer ve parametreler üzerindeki bazı kısıtlamalar ise şunlardır.

AR(1) model: $-1 < \phi_1 < 1$

AR(2) model: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, $\phi_2 - \phi_1 < 1$.

1.2.2.2.2 Hareketli Ortalamalar (MA Süreci)

Hareketli Ortalama Modeli, Regresyon modellerinde tahmin edilen değişkenin geçmiş değerleri kullanmak yerine, bir regresyon benzeri modelinde geçmiş tahmin hataları kullanır.

MA(q) Aşağıdaki gibi ifade edilebilir (Hyndman ve Athanasopoulos, 2014: 218).

$$Y_t = \beta_0 + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

1.2.2.2.3 ARMA Modelleri

Bir çok durağan rassal süreç pür otoregresif veya pür hareketli ortalama sürecini modellenemeyebilir. Zaman serileri, AR ve MA bileşenleri p ve q'uncu dereceden olmak üzere modellenemeyebilir. Geçmiş gözlemler ve geçmiş hata terimlerinin doğrusal bir fonksiyonu. ARMA(p,q) modeli aşağıdaki eşitlikte gösterildiği gibidir (Sevüktekin, 2010: 155):

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \delta + u_t - b_1 u_{t-1} - b_2 u_{t-2} - \dots - b_q u_{t-q}$$

1.2.2.2.4 ARIMA Modelleri

ARIMA modelleri Otoregresif (AR) ve hareketli Ortalamalar Modellerin (MA) birleşimidir. Sadece veri oluşturma sürecini diğer bir ifade ile değişkenin geçmiş ve gecikmeli değerleri ile ilgilendikleri için kuramsız modeller olarak da adlandırılmıştır. (Tarı, 2011: 444)

ARIMA(p,d,q), AR ve MA süreçlerinin birleşimidir. Parametre dereceleri: $p=AR$, $d=Fark$, $q=MA$ derecelerine göre bileşenler aşağıdaki gibi ifade edilebilir. (Hydman, 2014: 133)

$$\begin{array}{ccccc} (1-B-\dots-\phi_p B^p) & (1-B)^d y_t & = & c+(1+\theta_1 B+\dots+\theta_q B^q)e_t \\ \uparrow & \uparrow & & \uparrow \\ AR(p) & d \text{ fark} & & MA(q) \end{array}$$

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t$$

p , d ve q için uygun değerleri bazı durumlarda seçilmesi zor olabilir. Çalışmamızda kullandığımız R yazılımında `auto.arima()` fonksiyonu başarılı bir şekilde bu dereceleri hesaplamaktadır. Hiyerarşik ve gruplandırılmış zaman serileri paketinde de (`hts,gts`), `auto.arima` fonksiyonuyla dereceler hesaplamaktadır. Aşağıda ARIMA modelinin özel durumları ifade edilmiştir (Hyndman ve Athanasopoulos, 2014: 220)

Beyaz Gürültü:	ARIMA(0,0,0)
Rassal Yürüyüş:	ARIMA(0,1,0) sabit içermez
Kayan Rassal Yürüyüş:	ARIMA(0,1,0) sabit içerir
AR:	ARIMA(p,0,0)
MA:	ARIMA(0,0,q)

1.2.2.2.5 Mevsimsel ARIMA Modelleri

ARIMA modelleri, mevsimsel özellikteki veri setlerini bir modelleme yeteneğine sahiptirler. Mevsimsel ARIMA modeli, (P,D,Q) olarak ifade edilen mevsimsel şartlar eklenmesiyle aşağıdaki gibi ifade edilebilir:

$$\begin{array}{ccc} ARIMA(p,d,q) & & (P,D,Q)_m \\ \uparrow & & \uparrow \\ \text{Modelin mevsimsel olmayan bölümü} & & \text{Modelin mevsimsel bölümü} \end{array}$$

Üstel düzünleştirme ve ARIMA yöntemlerinin güçlü ve zayıf yönleri Tablo 1.1'de yer almaktadır.

Tablo 1.1 Üstel Düzünleştirme ve ARIMA Modellerinin Güçlü ve Zayıf Yönleri

Yöntem	Güçlü Yönleri	Zayıf Yönleri
Basit Üstel Düzünleştirme	- Trend/ döngüsel tahminde kullanılması - Az miktarda veriye ihtiyaç duyması - Yakın geçmiş dönemler daha fazla ağırlıklandırılması. - Hareketli ortalama yöntemine göre talep dalgalanmalarında başarılı olması	- Düzgünleştirilme katsayısının optimal değerinin bulunması - Mevsimsel verilerde başarılı olmaması - Bir dönem için tahmin üretmesi - Talep değişikliklerine geç algılaması - Açıklayıcı değişkenlerin kullanılmaması
Holt Yöntemi	- Trend/ döngüsel tahmininde kullanılması - Az miktarda veriye ihtiyaç duyması - İki farklı ağırlıklandırma yöntemi kullanması - Genellikle hareketli ortalama ve basit üstel düzeltme yöntemleri hareketli daha iyi performans göstermesi	- Düzgünleştirilme katsayısının optimal değerinin bulunması - Bir dönem için tahmin üretmesi - Ani talep değişimleri tahminindeki başarısı - Talep değişikliklerine geç algılaması - Açıklayıcı değişkenlerin kullanılmaması
Winters Yöntemi	- Trend / Döngü ve mevsimsellik tahmininde kullanılması - Az miktarda veriye ihtiyaç duyması - Üç parametre (eğilim / döngü, mevsimsellik ve düzensiz) kullanması - Yaygın olarak kullanılan matematiksel yöntem olması	- Düzgünleştirilme katsayısının optimal değerinin bulunması - Bir ve üç dönem için tahmin üretmesi - Ani talep değişimleri tahminindeki başarısı - Talep değişikliklerine geç algılaması - Açıklayıcı değişkenlerin kullanılmaması
ARIMA	- Trend/Döngüsel tahminin yanı sıra bağımsız değişken kullanarak tahmin modeli oluşturulabilir. - MAPE değerini artıran değişkenler modele dahil edilmemesi - Eğer-ise kurallarıyla talep şekillendirebilmesi - Kısa orta ve uzun vadeli tahmin modelleri oluşturulabilmesi	- Daha fazla veriye ihtiyaç duyulması - İstatistiksel bilgiye ihtiyaç duyulması

Kaynak: Chase Jr, 2013: 153

1.2.2.2.6 Parametrelerin Tahmini

En Küçük Kareler yönteminden daha güçlü kuramsal özelliklere sahip nokta tahminci yöntemi “ençok olabilirlik” (maximum likelihood), kısaca “EO” (ML) Fisher tarafından 1920’li yıllarda ileri sürülmüştür. Parametrelerin tahmini için $(c; \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ ARIMA model dereceleri $(p, d$ ve $q)$. belirlenmesi gerekmektedir. R yazılımında parametrelerin tahmininde EO yöntemi kullanılmaktadır. Gözlemlenen verileri elde etme olasılığını maksimize ederek parametre değerlerini hesaplamaktadır. Aşağıdaki gibi ifade edilebilir (Hyndman ve Athanasopoulos, 2014: 225).

$$\sum_{t=1}^T e_t^2$$

Bilgi kriterleri;

Akaike Bilgi Kriteri (Akaike’s Information Criterion, AIC), regresyon modellerinde tahmincilerin seçiminde kullanıldığı gibi ARIMA modellerinde de kullanılmaktadır. Aşağıdaki gibi ifade edilebilir (Hyndman ve Athanasopoulos, 2014: 227).

$$AIC = -2\log(L) + 2(p + q + k + 1)$$

Burada L veri olasılığı olmak üzere, $k=1$ ise $c \neq 0$ ve $k=0$ ise $c=0$ ’dır. Parantez içerisindeki son terim parametre sayısını ifade etmektedir.

ARIMA modelleri için, düzeltilmiş AIC olarak yazılabilir

$$AIC_c = \frac{AIC + 2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2}$$

Bayes Bilgi Kriteri ise aşağıdaki gibi ifade edilebilir.

$$BIC = AIC + (\log(T) - 2)(p + q + k + 1)$$

1.2.2.3 Durağanlık Kavramı

Zaman serileri analizlerindeki temel süreçlerden birisi “duragan” (stationary) olasılıksal süreçtir. Ortalaması ve varyansı zaman içerisinde değişmediği ve iki dönem arasındaki kovaryansın ise dönemlerin arasındaki uzaklığa bağlı olduğu süreçtir. Durağan serilerin özellikleri şöyle ifade edilebilir (Akgül, 2003b: 105).

$$E(Y_t) = \mu \quad (\text{tüm } t' \text{ ler için})$$

$$\text{Var}(Y_t) = E(Y_t - \mu)^2 = \sigma^2 \quad (\text{tüm } t' \text{ ler için})$$

$$\text{Cov}(Y_t, Y_{t+k}) = \gamma_k \quad \text{sabit} \quad (\text{tüm } t' \text{ ler için tüm } k \neq 0 \text{ için})$$

Zaman serilerinin durağanlık, serilerin zaman yolu grafiğinde ve onun korelogramında otokorelasyon ve kısmi otokorelasyon katsayıların değerlendirilmesi veya birim kök testleriyle incelenmektedir.

1.2.2.3.1 Katsayıların Değerlendirilmesi

Tek değişkenli zaman serisinde, değişkenin gecikmeli değerleri arasında birlikte değişimin bir ölçüsü otokovaryans ve otokorelasyon (ACF) olarak adlandırılır. Basit durağanlık sınaması, ACF'na dayanır. Gecikmesi k iken ρ_k ile gösterilen ACF şöyle tanımlanır (Sevüktekin, 2012: 106):

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad -1 < \rho_k < 1$$

$$= \frac{\text{gecikme } k \text{ iken ortak varyans}}{\text{varyans}}$$

Örneklem ACF	Örneklem Ortak Varyansı	Örneklem Varyansı
$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$	$\hat{\gamma}_k = \frac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n}$	$\hat{\gamma}_0 = \frac{\sum (Y_t - \bar{Y})^2}{n}$

Herhangi bir $\hat{\rho}_k$ 'nin istatistik bakımından anlamlılığı, standart hatasıyla belirlenmektedir. Bartlett, bir zaman serisi bütünüyle rassal ise (beyaz gürültü) örneklem otokorelasyon katsayılarının $N(0,1/n)$ sıfır ortalama ve $1/n$ varyansla yaklaşık normal dağıldığını ifade etmiştir.

1.2.2.3.2 KPSS Birim Kök Testi

KPSS testinde gözlenen serideki deterministik trendin arındırılarak serinin durağanlaştırılmaktadır. Test artıkların uzun dönem varyansının nonparametrik tahmincisine dayanmaktadır. Diğer birim kök testlerinden farklı olarak, Y_t 'nin sıfır hipotez altında durağan olduğu varsayılmaktadır. S_t kalıntıların kısmi süreç toplamı olmak üzere ($S_t = \sum_{r=1}^t \hat{u}_r$) olmak üzere KPSS test istatistiği: (Sevüktekin, 2010: 362)

$$\eta(u) = T^2 \sum S_t^2 / s^2(l),$$

$s^2(l)$ 'nin tutarlılığı için sınırlı gecikme parametresi $\ell \rightarrow \infty$ olarak belirlenmelidir. Hipotezler, hesaplanan değerle kritik değerin karşılaştırılmasıyla sınanmaktadır. KPSS testinin bir diğer özelliği sıfır hipotezinin trend durağanlığı belirtmesinden dolayı rassal yürüyüş (Random Walk) hipotezinin varyansının sıfır olmasıdır (Çiftci, 2014).

1.2.2.3.3 Dickey Fuller Birim Kök Testi

Durağanlığın belirlenmesinde birim kökün varlığını tespit eden kullanılan yöntemlerden birisi de Dickey ve Fuller yöntemidir (Sevüktekin, 2010: 106; Akgul, 2003a: 60).

$$Y_t = \rho Y_{t-1} + u_t$$

Sürecinde birim kökün varlığı araştırıldığında hipotez aşağıdaki gibi oluşturulur.

$$H_0 : \rho \geq 1 \text{ (Seri durağan değildir)}$$

$$H_1 : \rho < 1 \text{ (Seri durağandır)}$$

Eğer hesaplanan-t değeri, 0.01, 0.05 ve 0.10 kritik-t değerlerinden daha küçükse H_0 reddedilir ve serinin durağan olduğuna karar verilir. Eşitliğin her iki tarafı Y_{t-1} den çıkarılırsa

$$Y_t = \rho Y_{t-1} + u_t$$

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + u_t$$

$$\Delta Y_t = (\rho - 1)Y_{t-1} + u_t$$

$$= \delta Y_{t-1} + u_t \quad \delta = \rho - 1$$

Δ : birinci fark işlemcisi

$$\Delta Y_t = (Y_t - Y_{t-1})$$

$$H_0 : \rho \geq 1 \text{ veya } H_0 : \delta \geq 0 \text{ (Seri durağan değildir, birim kök vardır)}$$

$$H_1 : \rho < 1 \text{ veya } H_0 : \delta < 0 \text{ (Seri durağandır)}$$

Dickey-Fuller birim kök sınaması için üç model kullanılır.

Pür Rassal Yürüyüş Modeli: Bu model trendin ve sabitin yer almadığı modeldir.

$$\Delta Y_t = \delta Y_{t-1} + u_t$$

hipotezler aşağıdaki gibi oluşturulmaktadır.

$$H_0 : \rho \geq 1 \text{ veya } H_0 : \delta \geq 0 \text{ (Seri durağan değildir, birim kök vardır)}$$

$$H_1 : \rho < 1 \text{ veya } H_0 : \delta < 0 \text{ (Seri durağandır)}$$

Sabitin Yer Aldığı Rassal Yürüyüş Modeli: Modelde sabit yer almaktadır.

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t$$

hipotezler aşağıdaki gibi oluşturulmaktadır.

$$H_0 : \rho \geq 1 \text{ veya } H_0 : \delta \geq 0 \text{ (Seri durağan değildir, birim kök vardır)}$$

$$H_1 : \rho < 1 \text{ veya } H_0 : \delta < 0 \text{ (Seri durağandır)}$$

Trend ve Sabitin Yer Aldığı Rassal Yürüyüş Modeli: Model tüm deterministik bileşenleri ve stokastik kısmı içermektedir.

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t$$

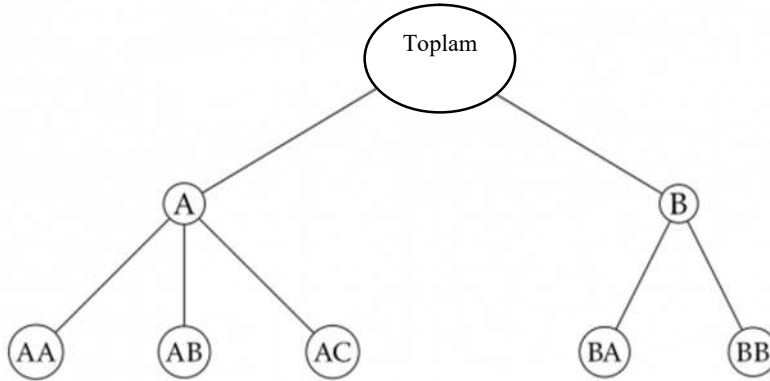
hipotezler aşağıdaki gibi oluşturulmaktadır.

$H_0 : \rho \geq 1$ veya $H_0 : \delta \geq 0$ (Seri durağan değildir, birim kök vardır)

$H_1 : \rho < 1$ veya $H_0 : \delta < 0$ (Seri durağandır)

1.2.3 Hiyerarşik Zaman Serileri

Zaman serileri, veri özelliklerine göre hiyerarşik bir yapıda oluşturulabilir. Şekil 1.3'de $K=2$ seviyede bir hiyerarşiyi göstermektedir. Hiyerarşik yapının en üstünde, 0. düzeyde verilerin toplamı olan düğüm yer almaktadır. y_t , $t=1, \dots, T$ gözleminden oluşan toplam serisi, bu seviyenin altında k_i, j düğüm sırası olmak üzere, seriler ise y_{jt} olarak ifade edilmektedir (Hyndman ve Athanasopoulos, 2014: 275).



Şekil 1.3 İki Seviyeli Hiyerarşik Ağaç Diyagramı

Hiyerarşik ağacın sonraki seviyesi olan 1. seviye verilerini j düğüm sayısı olmak üzere y_j olarak; Hiyerarşik ağacın en son düzeyi olan 2. Seviyede sırasıyla üç ve iki seri halinde ayrıştırılmış zaman serileri elde edilmektedir. Zaman serileri seviyelerine göre adlandırılmıştır. Birinci seviyede tek harfle (A,B), ikinci seviyede ise ilk harf bir üst düğümü ve ikinci harfi kendisini göstermek üzere (AA-AB) 2 harfle ifade edilmiştir (Hyndman ve Ahmed, 2011).

Bir hiyerarşik zaman serisinde, serilerin toplam sayısı (n), hiyerarşik düzeylerdeki düğüm (seri etiketi, n_i) ($n=1+n_1+\dots+n_K$) sayılarının toplamına eşittir.

Herhangi bir t zamanı için, alt düzey serilerin toplamı 1 üst düzeydeki serilerin toplamına eşit olacaktır. Bu yapı matris gösterimi kullanılarak temsil edilebilir. Şekildeki hiyerarşik yapıdan faydalanan alt düzey serileri toplayan S matrisini ($n \times n_k$):

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix}$$

Klasik zaman serilerine göre farklı bir yaklaşım, Hyndman tarafından önerilen hiyerarşik veya gruplandırılmış zaman serileri modelleridir. Önerilen model için, R yazılımında hiyerarşik seriler için hts ve gruplandırılmış seriler için gts kütüphaneleri bulunmaktadır.

Athanasopoulos ve arkadaşları yapmış oldukları çalışmada, 1998-2011 döneminde Avustralya turizm talebini hiyerarşik yapıda oluşturulan çeyrek veri setleriyle, tahmin etmiştir. Avustralya turizm talebini 2 seviyeli hiyerarşik yapıda incelemiştir. Avustralya, ekonomik ve idari özerkliğe sahip (New South Wales(NSW), Victoria(VIC), Queensland (QLD) ve diğer bölgeler coğrafi bölgelere ayrılmıştır. Bu bölgeler tanımlanmasıyla 1. Seviye düğümler oluşturulmuştur. Her düğümün altında, eyaletlerin başkentleri ve diğer turizm şehirleri olmak üzere sırasıyla Sidney, Melbourne, Brisbane olmak üzere 2. Seviye düğümler oluşturulmuştur. Tüm düğümler için tahmin değerleri üretilmiştir (Athanasopoulos ve Ahmed, 2009).

1.2.3.1 Birleşim Yöntemleri

1.2.3.1.1 Tüme Varım Yöntemi

Hiyerarşik tahmin için yaygın olarak uygulanan yöntem aşağıdan yukarıya bir yaklaşımdır. Bu yaklaşım da hiyerarşinin en alt düzeyinden başlayarak her seri için ilk üretilen bağımsız tahminleri ve daha sonra tüm hiyerarşi yukarı doğru toplayarak tahminleri hesaplanmaktadır (Hyndman ve Athanasopoulos, 2014: 277).

Daha önce Şekil 1.1'de ifade edilen alt düzey serileri için $(\hat{Y}_{AA,h}, \hat{Y}_{AB,h}, \hat{Y}_{AC,h} \text{ ve } \hat{Y}_{BA,h}, \hat{Y}_{BB,h})$ için h-adım “baz” tahminler hesaplanır. Düğümlerin bağlı olduğu A ve B düğümlerinin tahminleri ayrı ayrı hesaplanarak bir üst düğüm olan denklemin tamamının tahmini gerçekleştirilir.

$$\tilde{Y}_{A,h} = \hat{Y}_{AA,h} + \hat{Y}_{AB,h} + \hat{Y}_{AC,h}$$

$$\tilde{Y}_{B,h} = \hat{Y}_{BA,h} + \hat{Y}_{BB,h}$$

$$\tilde{Y}_h = \tilde{Y}_{A,h} + \tilde{Y}_{B,h}$$

Aşağıdan yukarıya yaklaşımda her alt düzey serisi için gözden geçirilmiş tahminlerle baz tahminler birbirine eşittir. Matris notasyonu olarak aşağıdaki şekilde ifade edilebilir.

$$\tilde{y}_h = S\hat{y}_{A,k}$$

Bu yaklaşımın en önemli avantajı toplama bağlı olarak herhangi bir bilgi kaybolmamasıdır. Fakat alt düzeyler için, veri oldukça gürültülü olabilir bu da tahmini güçleştirebilir (Hyndman, 2014).

1.2.3.1.2 En Uygun Tahmin Birleşim Yöntemi

Bu yaklaşım da, işlem adımları hiyerarşik yapıdaki her bir seri için bağımsız baz tahminlerin oluşturulmasıyla başlamaktadır. Bu baz tahminler, hiyerarşik yapıya eklenmeden, bağımsız olarak üretilir. Bu yaklaşımın amacı, en uygun bağımsız baz tahminleri birleştirerek ve tek değişkenli tahminleri mümkün olduğunca yakın, ama aynı zamanda hiyerarşik yapı ile sürekli bir araya getirerek revize tahminler kümesini oluşturmaktır (Hyndman ve Athanasopoulos, 2014: 287).

Genel lineer regresyon modeli ile hiyerarşinin tamamı için baz tahminler:

$$\hat{y}_h = S\beta_h + \varepsilon_h$$

$\hat{y}_h$: Bütün hiyerarşi için baz tahminlerin bir vektörü

β_h : Alt düzeyler için gelecek değerlerinin bilinmeyen ortalaması.

ε_h : Regresyon hatası

Genel olarak $\sum h$ bilinmemektedir. Baz tahminlerin yaklaşık (tahminlerin herhangi kümesi için gerçekleşmelidir) hiyerarşik kümeleme yapısını sağlaması şartıyla, hataların yaklaşık orijinal verilerin aynı kümeleme yapısında olduğu söylenebilir ($\varepsilon_{K,h}$ altdüzeyde tahmin hataları, $(\varepsilon_h \approx S\varepsilon_{K,h})$ olmak üzere en iyi doğrusal yansız tahmin için $\beta_h, \hat{\beta}_h$ dir. $\hat{\beta}_h = (S'S)^{-1}S'\hat{y}_n(h)$. Böylece tahmin kümesi $\tilde{y} = (S'S)^{-1}S'\hat{y}_h$ ifade edilebilir.

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ KAVRAM VE YÖNTEMLERİ

2.1 Yapay Öğrenme

Pedagojik açıdan öğrenme “Kavramsal düzenlemeler yapma süreci”, ”Alıştırma ve uygulamaların oldukça sürekli olan etkilerine verilen ad”, “Belli bilgi, beceri ve anlayışlar edinme”, “Tepki ve davranışlarda her zaman ya da kimi durumlarda yaşantıların oluşturduğu değişme” anlamlarına gelmektedir. (TDK). Yapay Öğrenme (Learning Machine) ise alanın önde gelen bilim insanlarından Ethem Alpaydın tarafından terim olarak öne sürülmüş “bilgisayarların örnek veri ya da deneyimi kullanarak başarımlarını artıracak biçimde programlaması” olarak ifade edilmiştir.

Yapay zekâ ilk olarak 1950'lerin ortalarında bir disiplin olarak tanınmayı başardığından, Yapay Öğrenme merkezi bir araştırma alanı olmuştur. Bu önemin iki nedeni olabilir. Öğrenme kabiliyeti akıllı davranışın bir işaretidir, bu nedenle zekayı bir fenomen olarak anlama girişiminde bulunmak öğrenme anlayışını içermelidir. Daha somut olarak, öğrenme, yüksek performans sistemleri oluşturmak için potansiyel bir metodoloji sağlar. (Quinlan, 1986)

Yapay öğrenme, istatistikî yöntemlerle model oluşturmakta, büyük veri üzerindeki öğrenme sürecinde ele alınan optimizasyon probleminin çözümü , öğrenilmiş modelin gösterimi ve çıkarımında bilişim teknolojilerini kullanmaktadır. Alpaydın tarafından bu süreç, yani yapay öğrenme yöntemlerinin büyük veri setlerinde uygulanması, veri madenciliği olarak ifade edilmiştir (Alpaydın, 2011: 2-3).

2.1.1 Denetimli Öğrenme

Denetimli öğrenme, modellenmesi düşünülen veri setindeki nitelikler için girdi-çıkıtı, neden-sonuç, durum-sınıf örnekleri gibi daha önceden elde edilmiş bilgilerin var olduğu durumlarda başvurulacak öğrenme yöntemidir. Başka bir ifadeyle var olan örnekleri kullanarak yeni model veya ileriye yönelik olması muhtemel durumların tahmininde kullanılan öğrenme yöntemidir.

2.1.2 Denetimsiz Öğrenme

Denetimsiz öğrenme, modellenmesi düşünülen veri setinde ki verilere ait herhangi ön bilginin (girdi-çıkıtı, neden-sonuç, durum-sınıf örnekleri) olmadığı sadece genel etiketsiz veri kümesi olduğu durumlarda başvurulacak öğrenme yöntemidir. Veri Madenciliğinde yaygın

kullanılan denetimli ve denetimsiz öğrenme algoritmaları ve görevlerine(Sınıflandırma, Sayısal tahmin, Örüntü Tanıma,Kümeleme) ilişkin bazı algoritmalar Tablo 2.1 de yer almaktadır (Lants, 2013: 22).

Tablo 2.1 Denetimli ve Denetimsiz Öğrenme Algoritmaları

Denetimli Öğrenme Algoritmaları	Görev
En Yakın Komşu	Sınıflandırma
Naif Bayes	Sınıflandırma
Karar Ağaçları	Sınıflandırma
Regresyon Ağaçlar	Sayısal Tahmin
Model Ağaçlar	Sayısal Tahmin
Yapay Sinir Ağları	Sınıflandırma ve Sayısal Tahmin
Destek Vektör Makineleri	Sınıflandırma ve Sayısal Tahmin
Denetimsiz Öğrenme Algoritmaları	Görev
Birliktelem Kuralları	Örüntü Tanıma
K-Means Kümeleme	Kümeleme

2.2 Veri Madenciliği

Veri madenciliğinin ortaya çıkmasında, gelişen veri tabanı yönetim sistemleriyle verinin ifade edilmesi ve kayıt edilmesinde önemli kolaylıklar sağlanmış, bunun sonucu olarak büyük veri yığınları oluşmuştur. Veri tabanı yönetim sistemlerinde sorgulama ve raporlama yöntemlerinde ziyade kaliteli bilgiye ulaşılması ihtiyacı yeni yöntemlerin keşfedilmesini sağlamıştır.

Veri madenciliği yöntemleri, veri setindeki anlamlı kalıpları ve kuralları keşfetmek için büyük miktarda verinin otomatik veya yarı otomatik yollarla araştırılması ve analizi için geliştirilmiştir. Yöntemler varsayımlardan ziyade keşif odaklı kullanılmaktadır (Chien ve Chen, 2008).

Veri madenciliği işletmelerde karar destek, pazar stratejisi, finansal tahminler olmak üzere pek çok alanda uygulanabilmektedir. Veri madenciliği makine öğrenme, istatistik ve veritabanları alanlarında kullanılmakta olan teknikleri birleştirmek suretiyle veritabanlarından yararlı ve değerli olan bilgilerin çıkarılmasını sağlar (Ching ve Pong, 2002).

Veri madenciliği istatistik, karar ağaçları, yapay sinir ağları, genetik algoritma ve görsel teknikleri gibi uzun yıllar boyunca geliştirilen çeşitli tekniklerle içermekte olup finans, pazarlama, bankacılık, üretim, müşteri ilişkileri yönetimi ve organizasyon öğrenme gibi birçok alanda uygulanmaktadır (Tan vd., 2006).

Veri madenciliği veritabanları ya da veri ambarlarında bulunan yığın veri içerisindeki gizli örüntüleri ve ilişkileri tespit etmek için istatistiksel algoritmaları ve yapay zeka yöntemlerini kullanan kompleks bir veri arama yöntemi olarak tanımlanabilir (Emel ve Taşkın, 2005).

Larose (2005) veri madenciliğini, istatistiksel ve matematiksel yöntemler ile örüntü tanıma teknolojilerinin kullanılarak, depolama ortamlarında mevcut olan büyük miktardaki verinin elenerek anlamlı yeni korelasyon, örüntü ve eğilimlerin keşfedilmesi olarak tanımlamaktadır.

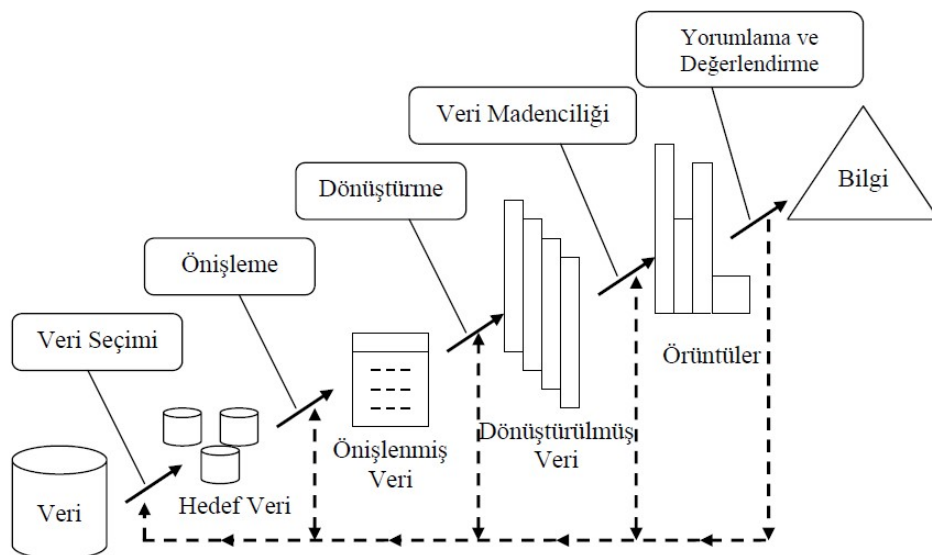
Linoff ve Berry'ye (2011) göre veri madenciliği, büyük miktardaki veriden anlamlı örüntüler ve kurallar keşfetme sürecidir.

Veri madenciliği genel olarak büyük oradaki veri setlerindeki çeşitli bilinmeyen ya da gizli kalmış kuralların keşfedilmesine ve analizine imkan tanıyan yöntem ve teknikler bütünüdür. Diğer bir ifadeyle veriden bilgi çıkarma sanatıdır (Tuffery, 2011).

Veri madenciliği, sınıflandırma, kümeleme, tahmin, ilişki analizi, örüntü tanıma gibi problemlerin çözümünde varolan/geliştirilen algoritmaların, büyük veri setlerinde ön işleme-dönüştürme-modelleme aşamalarından geçerek hızlı ve güvenilir şekilde bilgiye ulaşılmasını sağlayan süreç olarak ifade edilebilir.

2.3 Veri Madenciliği Süreci

Veri Madenciliği Süreci konusunda çeşitli öneriler bulunmaktadır. Fayyad (1996), CRISP-DM ve Semma süreç konusunda en çok bilinenleridir. Fayyad ve arkadaşları tarafından 1996 yılında ileri sürülen Veri Madenciliği Süreci, süreç için yapılan ilk tespit olmasından dolayı literatürde önemli bir yere sahiptir. Şekil 2.1'de görüldüğü gibi 5 aşamadan oluşmaktadır.



Şekil 2.1 Veritabanlarında Bilgi Keşfi Süreci

Kaynak: Fayyad vd., 1996: 41

SPSS, NCR ve Daimler Benz konsorsiyumu tarafından öne sürülen CRISP-DM süreci, birbirini izleme sırası net olmayan 6 aşamada ifade edilmiştir (Chapman vd., 2000). Bu

aşamalar İşin anlaşılması, Verinin Anlaşılması, Veri Hazırlama, Modelleme, Değerlendirme ve Kullanıcılarla Paylaşım aşamalarıdır. CRISP-DM modelinin, Fayyad'ın önerdiği modele göre farkı Kullanıcılarla Paylaşım aşamasıdır. Burada kullanıcıların belirlenmesi, eğitilmesi ve kullanıcılarda dönüt alınması gibi işlemlersürecin uzamasına ve maliyetlerin artmasına neden olabilir.

SEMMA süreci ise SAS firması tarafından "yazılımcı" bakış açısıyla süreçlerin baş harfleri Sample, Explorer, Modify, Model, Asses (Örnekle, Keşfet, Düzelt, Modelle, Değerlendir) aşamalarından oluşmaktadır.

CRISP-DM sürecinde aşama sıralamalarının kesin olmayışı, kullanıcılarla paylaşım gibi ucu açık bir aşamanın oluşu, SEMMA sürecinde ise Fayyad'ın önerdiği modele çok farklı bir yenilik getirmediği için çalışmamızda Fayyad ve arkadaşlarının önerdiği süreç takip edilecektir.

2.3.1 Veri Seçimi

Günümüzde "veri madenciliği" çatısı altında olan yöntemler aslında uzun yıllardır kullanılmaktadır. Bu süre zarfında, birçok farklı veri kümesine çok çeşitli teknikler uygulayarak hem açıklayıcı hem de tahmin edici modellere yönelik araştırmalar yapılmıştır. Hali hazırdaki yöntemlerin veya geliştirilen yeni yöntemlerin başarılı olması için eskimeyen kural ise "iyi verilerle" çalışmak olduğu söylenebilir (Pyle, 1999: 26).

Veri seçimi, veri madenciliği sürecinin ilk basamağıdır. Veri madenciliği analizlerinde kullanılacak verilerin probleme uygun olarak seçilmesi işlemidir. Probleme uygun olmayan veriler gereksiz zaman kaybı ve maliyete yol açabilir. Kaynak veriden hedef veriye ulaşıldığı aşama olarak ifade edilebilir.

2.3.2 Veri Ön İşleme

Veri ön işleme aşamasında veri temizleme, boyut indirgeme, sıra dışı değer analizi gibi çeşitli yöntemler uygulanmaktadır. Veri ön işleme aşağıdaki sebeplerden dolayı verilere uygulanmaktadır (Oğuzlar, 2003: 70).

- Veriler üzerinde herhangi bir analiz türünün uygulanmasını engelleyecek durumların giderilmesi
- Verilerin doğasının anlaşılması ve anlamlı veri analizinin yapılması
- Verilen bir veri kümesinden daha anlamlı bilginin çıkarılması.

Veri setinin etkileyen faktörler Tablo 2.2'de yer almaktadır (Akpınar, 2014: 86).

Tablo 2.2 Veri Kalitesini Etkileyen Faktörler

Faktör	Tanım
Anlaşılabilirlik	Veri kümesindeki alanların ve değerlerin anlaşılabilir olması.
Değer Katabilirlik	VM sürecine faydalı olması
Erişilebilirlik	Verinin ulaşılabilir ve hazır olması
Güncellik	VM sürecine uygun güncel kayıtların bulunması
İnanılrlık	Veri kümesinde yer alan kayıtların gerçek ve tüm paydaşlar tarafından kabul edilmiş olması
Nesnellik	Verilerin yansız,önyargısız ve tarafsız olması
Tamlık	Veri setinde eksik kayıt olmaması
Tutarlılık	Veri setinde tutarsız kayıtların olmaması
Uygunluk	VM Sürecine uygun olması
Veri miktarının uygunluğu	Veri setinin VM sürecine uygun büyüklükte olması
Yorumlanabilirlik	Veri setindeki kayıtların tanımlanabilmesi

Veri Ön işleme aşamasında bazı kaynaklara göre farklı adlandırılırsa genel olarak veri kaynağının güvenilirliği sağlandıktan sonra veri temizleme, veri dönüştürme ve boyut azaltma işlemleri yapılmaktadır (Silahtaroglu, 2013: 22; Akpınar, 2014: 89).

2.3.2.1 Veri Temizleme

Veri temizleme, veri setindeki gürültünün ortadan kaldırılması, mükerrer kayıtların giderilmesi, kayıp verilerin tamamlanması gibi işlemlere veri kalitesini etkileyen faktörlere karşı olumlu cevap verebilecek hale getirmektir. Başka bir ifade ile Veri Temizleme, VM sürecine uygun ve kaliteli veri setinin elde edilmesidir.

2.3.2.2 Boyut Azaltma

Sınıflandırma ve regresyon modellerinin karmaşıklığı girdi (öznitelik) sayısına bağlıdır. Girdi değerine göre modelin eğitim zaman ve uzay karmaşıklığı, eğitim örnek sayısını belirlenmektedir. Aşağıdaki nedenlerden dolayı veri kümesindeki boyut azaltma işlemi gereklidir (Alpaydın, 2011: 89).

- Öğrenme algoritmasındaki karmaşıklık genel olarak örneklem büyüklüğü (N) ve girdi boyutu (d) değerine bağlıdır. Hesaplama da kullanılacak bellek miktarını ve işlem sayısını azaltmak için d değeri optimum olmalıdır.
- Muhtemelen kullanılmayacak olan girdi değeri işlem sayısını ve bellek boyutunu artıracaktır.
- Küçük veri kümesi için girdi değeri optimum olan basit modeller daha güvenilir olmaktadır. Böylece gürültü, aykırı gözlemlerle uğraşmak daha kolay olmaktadır.
- Veri üretim sürecinin anlaşılması, tahmin ve bilgi çıkarımı kolaylaşmaktadır.

- Bilgi kaybına neden olmadan yapılan boyut azaltma sürecinde görselleştirme sonucunda veri yapısı hakkında ve aykırı değerler hakkında fikir yürütebiliriz.

2.3.3 Veri Dönüştürme

Veri setindeki değerlerin kendi aralarındaki uyumu ve farklı veri madenciliği yöntemlerinde kullanılmasını sağlamak amacıyla biçimlendirilmesidir. Literatürde normalleştirme ve standardizasyon olarak yer almaktadır. Başlıca veri dönüştürme biçimleri: (Akpınar, 2014: 115)

10 Tabanına Göre Logaritma: Veri setindeki değerlik onluk tabana göre logaritmasının alınmasıdır.

Z Skor Normalleştirme: Z Skor normalleştirme değerler aşağıdaki gibi hesaplanır.

$$x'_i = \frac{x_i - \bar{x}}{\sigma_x}$$

Min-Max Normalleştirme: Veri setinin en küçük değeri 0 en büyük değeri 1 olmak üzere [0,1] aralığında ifade edilmesidir. Değerler aşağıdaki gibi hesaplanır.

$$x'_i = \frac{(x_i - \min(x))}{(\max(x) - \min(x))}$$

[-1,1] Aralığında Normalleştirme: Veri setindeki en büyük ve en küçük değerler kullanılarak aşağıda ifade edilen dönüşüm yapılmaktadır.

$$x'_i = \frac{x_i - \left(\frac{\max(x_i) + \min(x)}{2}\right)}{\frac{\max(x_i) - \min(x)}{2}}$$

2.3.4 Modelleme

Probleme ve veri setine uygun Yapay Öğrenme algoritmaların büyük miktarlardaki veri setinde uygulama aşamasıdır. Bu aşamada yeni bilgiler elde edilmektedir.

2.3.5 Yorumlama

Veri madenciliği aşamasında elde edilen bilgilerin yorumlama ve problem çözümünde yeterli olup olmadığının tespiti aşamasıdır. Sonuçlar yeterli olmadığında önceki aşamalarda ifade edilen aşamalar tekrarlanabilir.

2.4 Veri Madenciliği Yöntemleri

2.4.1 Karar Ağaçları

Karar Ağaçları sezgisel olarak özyinelemeli bölümlenme yöntemiyle oluşturulan, sınıflandırma ve regresyon problemlerinde başvuru alan, dağılımdan bağımsız bir yöntemdir. Bu yaklaşım da, "böl ve yönet" olarak ifade edilen genellikle benzer sınıfların daha küçük alt kümelerle ayırmak için öznitelik değerleri, bir sabit veya diğer öznitelik değerleriyle karşılaştırılmaktadır. Karar ağacı, verilerin sınıf dağılımlarıyla ilgilenmeden, verileri birbirinden ayıran sınırları belirlediğinden dolayı yöntem, ayırtaç (diskriminant) tabanlıdır (Alpaydın, 2011: 153; Written ve Frank, 2005: 95; Lantz, 2013: 121).

Uygulamada karar ağacının daha karmaşık yöntemlerden önce kullanılması ve sonuçların karşılaştırılması tercih edilebilir (Alpaydın, 2011: 170). Karar ağacı modellerinin diğer yöntemlere göre uygulamada ki bazı avantajları ve dezavantajları Tablo 2.3'de yer almaktadır.

Tablo 2.3 Karar Ağaçları Modellerinin Güçlü ve Zayıf Yönleri

Güçlü Yönleri	Zayıf Yönleri
- Karar ağacı modellerinin anlaşılması ve yorumlanması kolaydır. - Veri ön işleme aşamasına gerek duyulmayabilir. - Bazı Karar Ağaçlarında sürekli ve kategorik değişkenlerin aynı anda kullanılabilir. - Büyük veri setlerinde uygulanabilir. - Ağaç yapısı sayesinde, örnek tabanlı yöntemlere göre yaprak, daha az hesap yapılarak bulunur. - Ağaç büyütüldükten sonra öğrenme kümesini saklanmadığından dolayı bellek karmaşıklığı düşüktür.	- Öznitelik seçiminde kullanılan sezgisel yöntemler de sonuçlar yerel optimumda yer alabilir. - Ezbere öğrenme sorunu yaşanabilir. - Çözümü zor olan problemlerde (XOR,parity vb) ağaç aşırı büyüyebilir.

Kaynak: Akpınar, 2014: 218; Alpaydın, 2011: 170

Literatürde söz konusu dezavantajların önlenmesi için bazı yöntemler bulunmaktadır. Karar ağacı modellerinde, sonuçları global optimuma yaklaştıran, çifte enformasyon uzaklığı, ağacın aşırı büyümesini engelleyen istatistiksel ilişkisel öğrenim, tümevarım mantıksal programlama gibi yöntemler kullanılmaktadır (Akpınar, 2014: 218).

2.4.1.1 Karar Ağaçlarında Bölümlenme Kriteri

Karar ağaçlarında temel süreç herhangi bir düğümün bölümlenmesi için kriter belirlenmesidir. Karar Ağaçları bölümlenme kriteri algoritmalarına göre gruplandırılabilir. Söz konusu algoritmalar aşağıdaki gibi gruplandırılabilir (Özkan, 2013: 55):

- Entropiye dayalı algoritmalar (ID3,C4.5)
- Sınıflandırma ve regresyon (CART,TWOING,GINI,M5P)
- Bellek tabanlı sınıflandırma algoritmaları (K-en yakın komşu)

Entropi, Shannon tarafından öne sürülmüştür, enformasyondaki belirsizlik olarak ifade edilmektedir. Bağımsız olarak n adet $\{m_1, m_2, \dots, m_n\}$ mesaj üreten S kaynağının m_i mesajını üretme olasılığı p_i 'dir. $P = \{p_1, p_2, \dots, p_n\}$ olasılık dağılımına sahip mesajları üreten S kaynağının entropisi $H(S)$:

$$H(S) = -\sum_{i=1}^n p_i \log_2(p_i)$$

2.4.1.2 Öznitelik Seçimi

Sınıf niteliği olmayan X değişkenine göre, sınıf niteliği olan T kümesi alt kümelerine ayrıldığında, T'nin bir elemanının sınıfını belirlemek için gerekli bilgi, T_i 'nin bir elemanının sınıfının belirlenmesinde gerekli olan bilginin ağırlıklı ortalamasıdır. Matematiksel olarak şöyle ifade edilebilir (Özkan, 2013: 58).

$$H(X, T) = \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i)$$

T kümesinin X değerine göre bölünmesi sonucu elde edilen bilgileri “kazanç ölçütü” ile ölçülür. Bu ölçü:

$$Kazanç(X, T) = H(T) - H(X, T)$$

Kazanç(X,T) değeri için en yüksek bilgi kazancını sağlayan nitelik seçilir.

2.4.1.3 Kazanç Oranı

Quinlan tarafından öne sürülen kazanç oranı T kümesi için X niteliğinin değerini belirlemek için gerekecek bilgi miktarı olarak ifade edilebilir.: $H(P_{x,T})$ T kümesindeki X niteliği için bilgi bölünmesidir.

Bu değer :

$$H(P_{x,T}) = H\left(\frac{T_1}{T}, \frac{T_2}{T}, \dots, \frac{T_k}{T}\right)$$

Elde edilen bilgi bölünmesi $H(P_{x,T})$ değeri ile Kazanç Oranı aşağıdaki gibi hesaplanır (Özkan, 2013: 76).

$$KazançOran(X, T) = \frac{Kazanc(X, T)}{H(P_{x,T})}$$

2.4.2 Sınıflandırma ve Regresyon Ağaçları

Sınıflandırma, yapay öğrenme ve veri madenciliği uygulamalarında önemli bir görevi temsil eder. Sınıflandırma da bilinen sınıf değerleriyle (eğitim seti) bir sınıflandırıcı oluşturulması ve ardından bilinen öznitelik değerleriyle yeni nesnelerin sınıf değerini (kategori) tahmin edilmesi amaçlanmaktadır (Jenhani vd., 2008: 1).

Breiman tarafından öne sürülen Sınıflandırma ve Regresyon Ağacı yönteminde ise (Classification and Regression Tree, CART) yöntemi, bağımlı değişkeninin kategorik olması durumunda sınıflandırma problemlerinde (Classification Tree, CT) veya sürekli olması durumunda ise regresyon problemlerinde (Regression Tree, RT) kullanılan veri madenciliği yöntemidir (Chang vd., 2006).

CART algoritmasında, bir düğümde belirli bir kriterle karşılaştırılarak bölünme işlemi gerçekleştirilir. Daha sonra bu bölünmeler üzerinde seçme işlemi Twoing veya Gini algoritmalarıyla gerçekleştirilir (Özkan, 2013: 89)

2.4.2.1 Twoing Algoritması

Twoing algoritması 4 adımda gerçekleşmektedir. İşlem adımları (Özkan, 2013: 89):

Adım 1. Eğitim kümesi sağ ve sol olmak üzere iki parçaya ayrılır: (t_{sol} , $t_{sağ}$)

Adım 2. Her bir bölüm için olasılıklar şu şekilde hesaplanır.

$$P = \frac{t_{sol} \text{ ve } t_{sağ} \text{ daki herbir nitelik degerinin ilgili nitelik sutunundaki frekansı}}{\text{Eğitim setindeki örneklerin sayısı}}$$

$$P(j | t_{sol}) = \frac{t_{sol} \text{ daki kayıtların } j \text{ sınıfları sayısı}}{t_{sol} \text{ daki herbir nitelik degerlerinin ilgili nitelik sutunundaki tekrar sayısı}}$$

$$P_{sag} = \frac{t_{sag} \text{ daki herbir nitelik degerinin ilgili nitelik sutunundaki } i \text{ tekrar sayısı}}{\text{Eğitim setindeki örneklerin sayısı}}$$

$$P(j | t_{sag}) = \frac{t_{sag} \text{ daki kayıtların } j \text{ sınıfları sayısı}}{t_{sag} \text{ daki herbir nitelik degerlerinin ilgili nitelik sutunundaki } i \text{ tekrar sayısı}}$$

Adım 3. t düğümün aday bölünmesi olan s , uygunluk ölçüsü hesaplanır.

$$\Phi(s | t) = 2P_{sol}P_{sag} \sum_{j=1}^n |P(j | t_{sol}) - P(j | t_{sag})|$$

Adım 4. Adım 1'e dönülerek ağacın alt kümesine aynı işlemler uygulanır.

2.4.2.2 Gini Algoritması

Gini yöntemi, nitelik değerlerinin sol veya sağ olmak üzere iki bölüme ayrılmasına dayalı bir sınıflandırma yöntemidir (Özkan, 2013: 106).

a) Her nitelik değerleri ikili ağaç biçiminde gruplanır. Bu şekilde sol ve sağ bölünmelere karşılık gelen sınıf değerleri hesaplanır.

b) Her bir nitelik ile ilgili sol ve sağ taraftaki için $Gini_{sol}$ ve $Gini_{sag}$ değerleri hesaplanır.

$$Gini_{sol} = 1 - \sum_{i=1}^k \left(\frac{L_i}{|T_{sol}|} \right)^2$$

k : Sınıf sayısı

T : Bir düğümdeki örnekler

$|T_{sol}|$: Sol taraftaki örneklerin sayısı

L_i : Sol taraftaki i kategorisindeki örneklerin sayısı

$$Gini_{sag} = 1 - \sum_{i=1}^k \left(\frac{R_i}{|T_{sag}|} \right)^2$$

$|T_{sag}|$: Sağ taraftaki örneklerin sayısı

R_i : Sağ taraftaki i kategorisindeki örneklerin sayısı

c) Her j niteliği için, n eğitim kümesindeki satır sayısı olmak üzere aşağıdaki bağıntının değeri hesaplanır:

$$Gini_{sag} = \frac{1}{n} \left(|T_{sol}| Gini_{sol} + |T_{sag}| Gini_{sag} \right)$$

d) Her j niteliği için hesaplanan $Gini_j$ değerleri arasından en küçük olanı seçilir ve bölünme bu nitelik üzerinden gerçekleştirilir.

e) En baştaki adıma dönülerek işlemlere devam edilir.

2.4.3 Model Ağaçlar

Model Ağaçlar (M5), sürekli değişkenlerin tahmini için ağaç algoritması olarak Quinlan tarafından 1992 yılında önerilmiştir. Model ağaçların, veri setindeki bilgi keşfini grafiksel ve if-then-else kurallarıyla sunabilen ayrıca lineer regresyon fonksiyonlarıyla sayısal değerleri tahmin edebilen bir karar ağacı türü olduğu söylenebilir (Frank, 1998; Bartolomei vd., 2010).

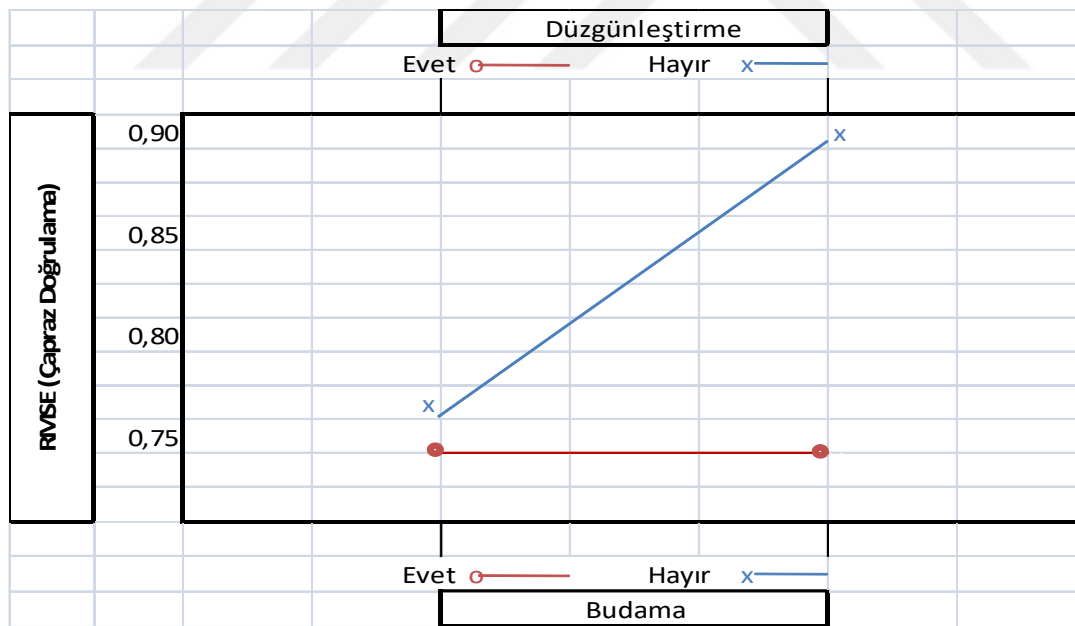
M5 ve CART algoritmaları arasındaki temel fark CART yöntemi ile oluşturulan karar ağacında yapraklar değer alabilmekteyken M5 algoritmasında çok değişkenli doğrusal modeller kurulabilmektedir. Böylece M5 modellerinde daha esnek tahminler yapılabilir. (Zhan, 2011) M5 algoritmasında Ağaç Oluşturma, Ağaç Budama ve Ağaç Düzgünleştirme süreçlerinden sonra Model Ağacı oluşturulabilmektedir.

Ağaç Oluşturma Süreci: Standart sapma azaltmasını sağlayan (SDR) adı verilen bir ölçünün en büyük değerine ulaşılması hedeflenmektedir.

Budama Süreci: Aşırı öğrenmenin engellenmesi hedeflenmektedir.

Düzgünleştirme Süreci: Ortaya çıkabilecek süresiz durumların telafi edilmesi hedeflenmektedir.

Kuhn (2013: 176), Model Ağaç yöntemi için, budama ve düzgünleştirme süreci ile ilgili deneysel çalışma yapmışlardır. Şekil 2.2'de model ağaçlara ait çapraz doğrulanmış RMSE değerleri yer almaktadır. Düzgünleştirme ve budama olmaksızın 159 adet yol içeren ağaç elde etmişlerdir. Daha sonra düzgünleştirme ve budama süreciyle yol sayısı 159'dan 18'e düşmüştür. Budanmamış ağaçta verilerin önemli ölçüde aşırı uyumlu olduğu ve düzgünleştirmenin önemli ölçüde hata oranını artırdığını tespit etmişlerdir.



Şekil 2.2 Model Ağaçlar İçin Çapraz Doğrulanmış RMSE değerleri

Kaynak: Kuhn, 2013: 176.

Tahmin sürecinde Model Ağacı, yaprağa doğru, örneğin öznelilik değerlerini kullanarak her düğüm için yönlendirme kararları oluşturmaktadır. Böylece yaprak, bazı

öznitelik değerlerine dayalı doğrusal bir modeli içermekte ve test örneği için elde edilen ham tahmin değerleri elde edilmektedir.

Eğitim örnek sayısının az olduğu modellerde, budanmış ağaç yapraklarıyla komşu doğrusal modeller arasında oluşabilecek süreksiz/kesikli durumları engellemek için ham değerlerinin doğrudan kullanılması yerine düzgünleştirme işlemi kullanılmaktadır. Düzgünleştirme, her iç düğüm için doğrusal modellerle birlikte yaprak üretildiğinde, ağaç oluşturulduğunda tamamlanır. Daha sonra, yaprak modeli bir test örneği için elde edilen ham tahmin değeri, yol boyunca geriye köke doğru filtrelenir, her düğüm için doğrusal model tarafından tahmin değeri ile birleştirilerek her düğüm noktasında düzgünleştirilir.

Düzgünleştirme işlemi aşağıda ifade edildiği gibi gerçekleşmektedir.

$$p' = \frac{np + kq}{n + k}$$

p': Üst düğüme aktarılan tahminidir.

p: Alt düğümden aktarılan tahmindir

q: Düğümden tahmin edilen değerdir.

n: Alt düğümlere ulaşmak için kullanılan eğitim örneklerinin sayısıdır.

k: Düzgünleştirme sabitidir.

2.4.3.1 Ağaç Oluşturma

Bölme kriteri belirli bir düğüm sayısındaki eğitim verilerini T oranında en iyi bölen niteliği belirlemek için kullanılmaktadır. Düğümlerde, T sınıf değerlerinin standart sapması ve düğümden her özellik için beklenen hatanın hesaplanmasına dayanmaktadır. Beklenen hatayı en fazla azaltan özellik düğümü bölmek için seçilir. Beklenen hata, Standart Hata Azaltılması SDR olarak adlandırılır (Written, 2005: 244-264).

$$SDR = sd(T) - \sum_i \frac{|T_i|}{T} sd(T_i)$$

T_1, T_2 seçilen özniteliğine göre bölme işlemi sonucu kümesidir. Sınıf değerlerinin standart sapması çok az farklılık gösterdiği orjinal örneğin standart sapmasına %5'ten daha az bir oranda yaklaştığı bir düğüme ulaşıldığı zaman bölme işlemi sona ermektedir. Deneylerde elde edilen sonuçlara göre bu eşiklerin seçiminin çok hassas olmadığını ortaya çıkmıştır.

2.4.3.2 Ağaç Budama

İç düğümler, doğrusal modeli ve ağaç yapraklarını oluşturmak ayrıca düzgünleştirme işlemi için gereklidir. Budama süreci öncesinde, model budanmamış ağacın her bir düğüm için hesaplanmaktadır. Model (Kuhn, 2013: 175; Written vd., 2005: 244-264):

$$w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k$$

a: öznitelik değerlerini

w: ağırlığı ifade etmektedir.

Budama sürecinde her düğüm için test verilerinin beklenen hatası tahmin edilmektedir. İlk olarak düğümdeki eğitim verilerinde sınıfın için tahmin edilen değeri ile gerçek değeri arasındaki mutlak farkların ortalaması hesaplanmaktadır. Hesaplanan ortalama, görünmeyen durumlar için, beklenen hatayı daha küçük değerde tahmin edecektir. Bu durumu önlemek ve modelin tahmin hatası en aza indirmek için ortalama, $\frac{(n+v)}{(n-v)}$ telafi faktörü ile çarpılmaktadır. Burada n ulaşılan düğümdeki eğitim örneklerinin sayısıdır. v ise düğümün sınıf değeri veren doğrusal modelin parametre sayısıdır.

Bütün iç düğümler için doğrusal modeller oluşturulduktan sonra beklenen tahmin hatasını azaltacak şekilde ağaç geriye doğru budanır. Düğümdeki doğrusal model için beklenen hata değeri ile alt ağaç için beklenen hata değeri karşılaştırılır. Daha sonra her dal için hata değeri, düğümdeki bütün değerlerin ağırlıkları ve her daldaki eğitim örneklerinin ağırlıkları kullanılarak oluşturulan doğrusal modelle tahmin edilen hata değeri birleştirilerek tek bir değer şeklinde elde edilir.

2.4.3.3 Nominal Değerler

Model ağaçlarda nominal öznitelikler ikili (binary) forma dönüştürülerek kullanılmaktadır. Her bir nominal değer, eğitim örneklerinin numaralandırılmasıyla mümkün olan sınıf ortalama değerlerinin hesaplanması ve bu ortalama değerlerin sıralanmaktadır. k sayıda değeri olan nominal değer için k-1 sayıda yapay binary değişkenle yer değiştirme yapılır (Written vd., 2005: 244-264).

2.4.3.4 Yerel Ağırlıklandırılmış Regresyon

Sayısal tahmin için kullanılan bir diğer yöntem yerel ağırlıklandırılmış regresyon yöntemidir. Model ağaçlarda, ağaç yapısı bölgelere bölünmektedir ve her bölge için doğrusal model oluşturulabilmektedir. Eğitim örnekler uzayının nasıl bölüneceği açıklamada katkı sağlamaktadır. Diğer yandan oluşturulan yerel model tahminleri komşu modeldeki test örneklerine yüksek ağırlık vermektedirler. Özellikle eğitim örneklerinin ağırlıkları test

örneklerine olan uzaklıklarına göre ayarlanır ve doğrusal regresyonda ağırlıklandırılmış veriler etkilidir. Eğitim örnekleri test örneklerine yakın olduklarında ağırlıkları artmakta uzak olduklarında ise azalmaktadırlar (Written vd., 2005: 244-264).

Yerel ağırlıklandırılmış regresyon yönteminde eğitim örnekleri için uzaklık, temelli şemanın hangisi olacağına karar vermek gereklidir. Yaygın olarak test örnekleri için ters Öklid Mesafesi kullanılmaktadır. Ayrıca Öklid Uzaklığı ve Gauss Çekirdek fonksiyonu birleşimi de kullanılabilir. Ağırlık fonksiyonunun seçiminin önemi konusunda yapılan çalışmalarda açık bir kanıt bulunmamaktadır. Burada önemli olan "düzgünleştirme parametresinin" seçimidir. Çünkü düzgünleştirme parametresi'nin tersi uzaklık ölçüsü ile çarpılmaktadır. Eğer düzgünleştirme parametresi küçük bir değer olursa uzaklık çok yakın olacağından dolayı test örnekleri yüksek derecede ağırlıklandırılacaktır. Daha büyük değer seçildiğinde model üzerinde önemli etkisi olacaktır. Düzgünleştirme parametresi k-yakın yöntemiyle seçilebilir. böylece değer küçük olduğunda eğitim örneklerinin önemi artacaktır. en iyi k parametresi seçimi veri setinin gürültü durumuna bağlıdır. Daha çok gürültü , daha çok komşu doğrusal model oluşturmaktadır. Düzgünleştirme parametresinde seçiminde genel yaklaşım çapraz doğrulama yöntemidir.

Model hatasını tahmin etmek için korelasyon katsayısı (R^2) hesaplanmaktadır. Bazen korelasyon katsayısı ile determinasyon katsayısı arasında bir belirsizlik vardır. Korelasyon katsayısı aşağıdaki gibi ifade edilebilir (Hoonakker vd., 2010).

$$R^2 = 1 - \frac{\sum_1^n (a_i - p_i)^2}{\sum_1^n (a_i - \bar{a})^2}$$

Burada p, tahmini değeri, a gerçek değeri ve n test grubundaki örnek toplam sayısını belirtmektedir.

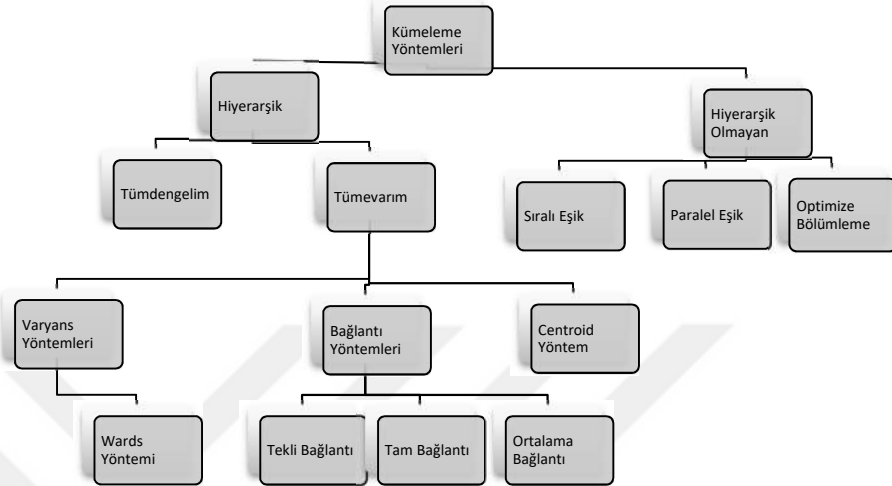
2.4.4 Kümeleme Yöntemleri

Kümeleme analizi sınıf özellikleri bilinmeyen veri setinde, seçme kriterlerine göre birbirine benzeyen verilerin aynı küme içerisinde yer almasını sağlamaktadır. Analiz sonucunda elde edilen farklı kümeler arasında heterojen yapı, kendi içlerinde homojen yapı oluşmaktadır (Akpınar, 2011: 284). Kümeler çok boyutlu uzayda gösterildiğinde, küme elemanların birbirlerine yakın, kümelerin de birbirinden uzak olması beklenmektedir (Suner ve Çelikoğlu, 2010).

Literatürde bir çok kümeleme algoritması bulunmaktadır. Bu algoritmalar temelde hiyerarşik ve hiyerarşik olmayan yapı olarak ikiye ayrılmaktadır Hiyerarşik yöntemler yığılmacı (agglomerative) ve bölücü (divisive) yaklaşımlar olarak ikiye ayrılmaktadırlar.

Hiyerarşik olmayan yöntemler ise ayırma (partitioning), yoğunluk-tabanlı (density-based), ızgaratabanlı (grid-based) ve diğer yaklaşımlar adı altında literatürde yer almaktadır (Taşkın, 2010).

Yöntemlerin özelliklerine göre Kümeleme yöntemleri Şekil 2.3'teki gibi sınıflandırılabilir.



Şekil 2.3 Kümeleme Yöntemlerinin Sınıflandırılması

Kaynak: Malhotra ve Birks, 2007: 676

Genel olarak kümeleme algoritmalarında temel ölçüt veriler arasındaki uzaklıklardır. Verilerin birbirine olan uzaklıkları yaygın olarak Öklid, Manhattan ve Minkowski uzaklık bağlantıları hesaplanmaktadır.

2.4.4.1 Uzaklık Ölçüleri

Kümeleme çözümlemelerinde birçok uzaklık bağıntısı kullanılabilir. Yaygın kullanılan uzaklık ölçütleri (Çakmak, 2005: 5):

- Öklid uzaklığı

$$d(g_1, g_2) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- Manhattan Uzaklığı :

$$d(g_1, g_2) = \sum_{i=1}^n |x_i - y_i|$$

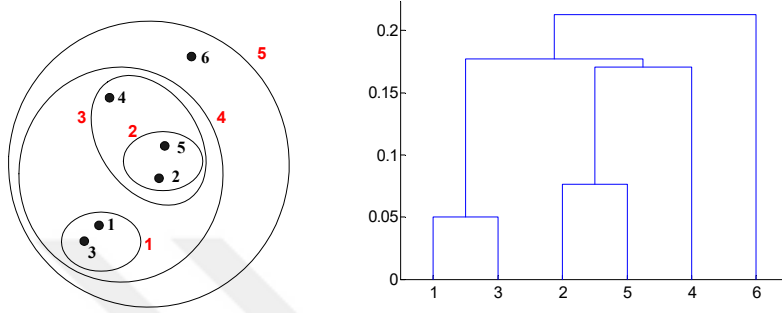
- Minkowski Uzaklığı

$$d(g_1, g_2) = \sqrt[m]{\sum_{i=1}^n (x_i - y_i)^m}$$

2.4.4.2 Tekli Bağlantı Yöntemi

Tekli Bağlantı Yöntemi, (Single linkage/SLINK/nearest neighbor method/) toplasimli (agretition) metoda göre iki veya daha fazla kümenin, birbirlerine en yakın nesnelerin uzaklığını baz alarak gerçekleştirmektedir (Akpınar, 2014: 304). Algoritmanın zaman karmaşıklığı $O(n^2)$ dir ve işlem adımları aşağıdadır (Silahtaroglu, 2013: 165).

$$D_{sl}(C_i, C_j) = \min_{x,y} \{d(x, y) | x \in C_i, y \in C_j\}$$



Şekil 2.4 Tekli Bağlantı Yöntemi Dendogram

Kaynak: Tan vd., 2006: 520

Oku

1. Veri Kümesini
2. Bağlantı eşik değeri

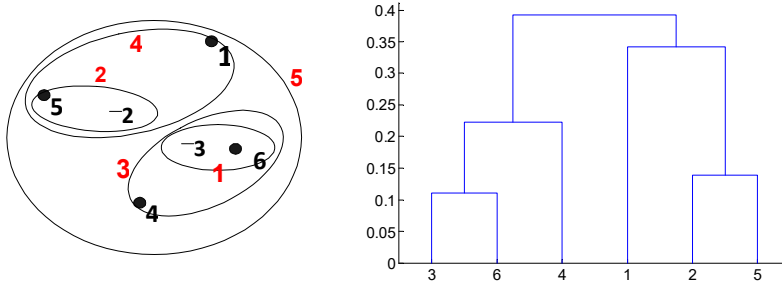
Adımlar

3. Verilerin mesafe/benzerlik matrisini oluştur.
4. Matrisi ağaç şeklinde ifade et.
5. Minimum maliyetli ağaç modelini uygula ve çıkart
6. Bağlantı eşik değerine göre budama yap ve kümeleri oluştur.
7. Dur

2.4.4.3 Tam Bağlantı En Uzak Komşu Yöntemi

Tam bağlantı yönteminde nesneler arası uzaklık en büyük uzaklık $d(u,v)w = \max \{d_{uw}, d_{vw}\}$ baz alınarak kümeleme işlemi gerçekleştirilmektedir. Şekil 2.5'te yöntemle ait kümeleme ve dendogram yer almaktadır.

$$D_{cl}(C_i, C_j) = \max_{x,y} \{d(x, y) | x \in C_i, y \in C_j\}$$

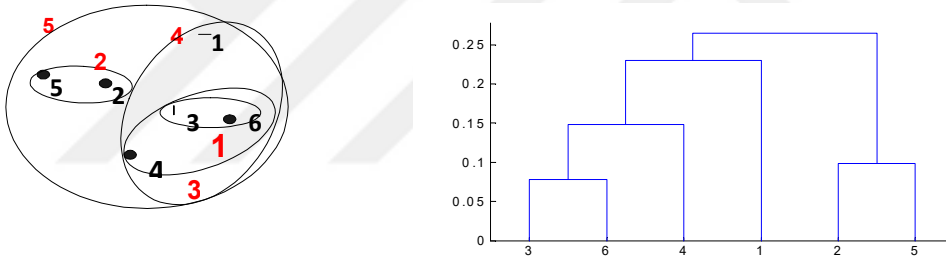


Şekil 2.5 Tam Bağlantı Yöntemi Dendogram

Kaynak: Tan vd., 2006: 521

2.4.4.4 Aritmetik Ortalamalı Bağlantı/Gruplar Arası

Aritmetik Ortalamalı Bağlantı, Tek ve Tam Bağlantı arasında uzlaşma olarak ifade edilebilir (Şekil 2.6). Gürültü ve aykırı değerlere göre daha az duyarlıdır. Küresel kümelerle karşı ön yargılı olması yöntemin bir sınırlılığı olarak ifade edilebilir. Şekil 2.6'da yönteme ait kümeleme ve dendogram yer almaktadır.



a) Aritmetik Ortalama Kümeleme

b) Aritmetik Ortalama Dendogram

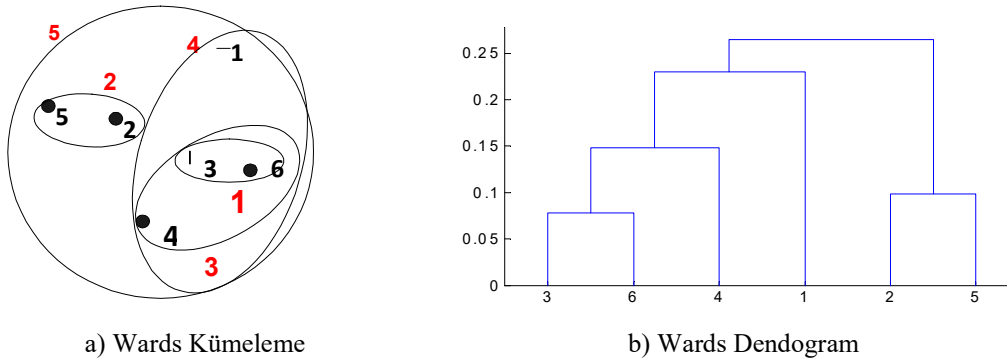
Şekil 2.6 Aritmetik Ortalamalı Bağlantı Dendogram

Kaynak: Tan vd., 2006:522

2.4.4.5 Wards Bağlantısı

Wards Bağlantısında uzaklık ölçüsü olarak varyans analizi kullanılmaktadır. Hiyerarşik yapı oluşturmak için her birinin içinde tek bir birim bulunan n tane küme ile oluşturulur. Her aşamada iki alt küme, hata kareler toplamındaki artış minimize edilerek birleştirilerek kümeleme süreci tamamlanmaktadır (Akpınar, 2014: 308). Şekil 2.7'de Wards kümeleme ve dendograma yer almaktadır.

$$D_w(C_i, C_j) = \sum_{x \in C_i} (x - r_i)^2 + \sum_{x \in C_j} (x - r_j)^2 - \sum_{x \in C_{ij}} (x - r_{ij})^2$$

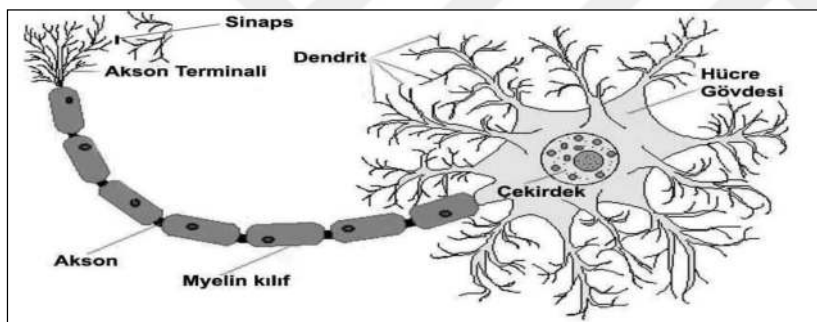


Şekil 2.7 Ward Bağlantısı-Dendrogram

Kaynak: Tan vd., 2006: 523

2.4.5 Yapay Sinir Ağları

Yapay sinir ağları sınıflandırma ve regresyon problemlerinde kullanılan parametrik olmayan tahmin yöntemidir (Alpaydın, 2011: 197). YSA başlangıcı, nörolojideki insan beyninin yapısı fonksiyonlarının anlaşılması gerçek hayat problemlerine uygulanmasına dayanmaktadır. Biyolojik sinir ağları bir çok modelde baz alınarak YSA tasarlanmaktadır.



Şekil 2.8 Biyolojik Sinir Hücresi ve Bileşenler

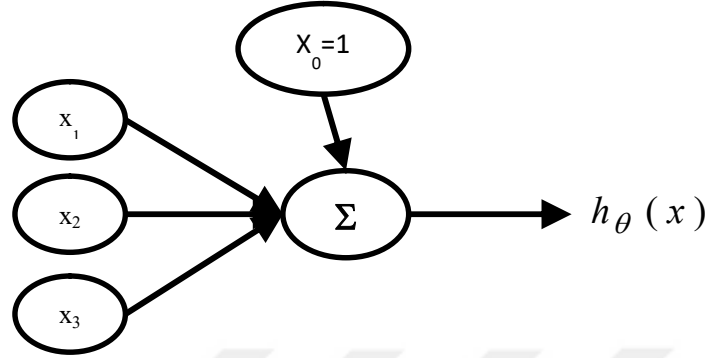
Kaynak: Akpınar, 2014: 237

Şekil 2.8'de görüldüğü üzere bir sinir hücresinin temel elemanları hücre gövdesi, dendrit ve aksondur. Nöronlar, bütün vücuttaki hisleri,duygu ve düşünceleri ileten karmaşık yapıdaki sinir ağlarını oluşturan temel birimdir. YSA'da ki temel birim olan algılayıcı ise tipik bir nöron gibi davranarak ağ yapısını oluşturmaktadır.

Nöronlarda mesajların alındığı birdenfazla giriş ve mesajların iletildiği bir tane çıkış bulunmaktadır. Nöronlar almış oldukları bir veya daha fazla mesajı toplayarak diğer nörona iletmektedirler. Yapay Sinir Ağları düğüm, nöron veya yapay sinir hücresi olarak adlandırılan çok sayıda işlem biriminin katmanlar halinde bir araya gelmesinden oluşmaktadır (Elmas, 2011: 30, 41). Bu bağlamda Yapay Sinir Ağlarının genel yapısı ve işleyişini, birden fazla girişe ve tek çıkışa sahip olan ve almış olduğu sinyali aktivasyon/eşik olarak adlandırılan fonksiyonlarla işleyen, işlem sonucunu tek bir sinyalle çıkışa gönderen hücrelerden oluşan bir ağ şeklinde ifade edebiliriz.

2.4.5.1 Algılayıcı

YSA'nın temel birimi olan nöronda $X_0=1$ bias, X_1, X_2, X_3 girişler olmak üzere girdi değişkenleri doğrusal toplanmakta, aktivasyon fonksiyonuyla çıkış değeri elde edilmektedir (Şekil 2.9).



Şekil 2.9 Algılayıcı Yapısı

YSA'daki nöronlar, ele alınan problemin özelliğine göre birbirlerine bağlı olarak büyük bir ağ yapısını oluşturabilirler. Böylece karmaşık ilişkilerin modellenmesine imkan sağlamaktadırlar. YSA yapısında giriş, gizli, çıkış katmanı olmak üzere üç katman bulunmaktadır.

2.4.5.2 Ağırlıklar

Ağırlıklar, bir yapay sinir hücresine gelen bilginin önemini ve hücre üzerindeki etkisini gösterirler. Ağırlıklar değişken veya sabit değerler olabilirler. Yapay sinir ağı içinde girdilerin nöronlar arasındaki iletimini sağlayan tüm bağlantıların farklı ağırlık değerleri bulunmaktadır. Böylelikle ağırlıklar, her nöronun her girdisi üzerinde etki yapmaktadır.

2.4.5.3 Birleştirme Fonksiyonu

Birleştirme fonksiyonu, bir hücreye gelen net girdiyi hesaplayan bir fonksiyondur. Genel olarak toplama fonksiyonu kullanılsa da bazı durumlar için toplama fonksiyonu yerine, en az (min), en çok (max), mod, çarpım, çoğunluk gibi fonksiyonlar kullanılabilir. Herhangi bir nöronunun toplam girdisi, diğer nöronlardan gelen değerlerin ağırlıklı toplamı ile eşik değerinin toplamına eşittir (Kamruzzman, 2006: 3).

Toplama fonksiyonu:

$$NET = w_{ji} x_i + \theta_j$$

NET : nöronunun net girdisini,

w_{ji} : j ve i nöronları arasındaki bağlantının ağırlığını,

x_i : i nöronunun çıktısını,
 θ_j : eşik değeri ifade etmektedir.

Tablo 2.4’de literatürde yer alan bazı birleştirme fonksiyonları yer almaktadır (Öztemel, 2006: 50).

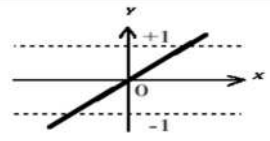
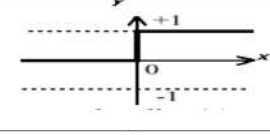
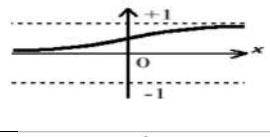
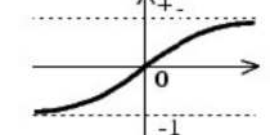
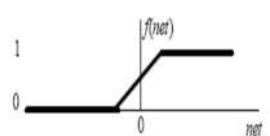
Tablo 2.4 Birleştirme Fonksiyonu Örnekleri

Toplam	$NET = \sum_i^n w_{ij} x_i + \theta_j$	Minimum	$NET = \min (w_{ij} x_i)$
Çarpım	$NET = \prod_i w_{ij} x_i$	Maksimum	$NET = \max (w_{ij} x_i)$

2.4.5.4 Aktivasyon Fonksiyonu

Algılayıcı gelen net girdiyi işlenmesinin ardından üretilecek çıktıyı belirleyen fonksiyondur. Hücrenin gerçekleştireceği işleve göre çeşitli tipte aktivasyon fonksiyonları kullanılabilir.

Tablo 2.5 Bazı Aktivasyon Fonksiyonları

Aktivasyon Fonksiyonu	Matematiksel Gösterimi	Matematiksel Gösterimi
Lineer Fonksiyon	$F(NET) = NET$	
Step Fonksiyonu	$F(NET) = \begin{cases} 1 & \text{eğer } NET \geq \theta \\ 0 & \text{eğer } NET < \theta \end{cases}$	
Sigmoid Fonksiyonu	$F(NET) = \frac{1}{1 + e^{-NET}}$	
Hiperbolik Tanjant Fonksiyonu	$F(NET) = \frac{e^{NET} - e^{-NET}}{e^{NET} + e^{-NET}}$	
Eşik Değer Fonksiyonu	$F(NET) = \begin{cases} 0 & \text{eğer } NET \leq 0 \\ NET & \text{eğer } 0 < NET < 1 \\ 1 & \text{eğer } NET \geq 1 \end{cases}$	

Kaynak: Demuth ve Beale, 2004

Aktivasyon fonksiyonun seçimi, problemde ele alınan veri setine bağlıdır. Bir problem için en uygun fonksiyon tasarımcının denemeleri sonucunda belirlenmektedir. Tablo 3.4’de

bazı aktivasyon fonksiyonları yer almaktadır. Literatürde Sigmoid Fonksiyonu, hiperbolik tanjant fonksiyonu daha sık rastlanmaktadır.

2.4.5.5 Algılayıcının Eğitilmesi

Yapay sinir ağlarını eğitiminde genellikle çevrimiçi öğrenme yöntemi kullanılmaktadır. Çevrim içi öğrenmede, öğrenme kümesinin tamamı yerine örnekler ardışık olarak kullanılmaktadır. Her örnekte, ağ parametrelerinin güncellenmekte ve zaman içinde veriye uyarlanmaktadır. Bu yöntemin avantajları;

- Öğrenme kümesinin tamamını bilmek, parametre değerlerini hesaplanırken ara sonuçları saklamak gerekmediğinden daha az bellek kullanılmaktadır.
- Öğrenilmesi gereken iş zaman içinde değişebilir ve önceden bir öğrenme kümesi oluşturmak doğru olmayabilir.
- Fiziksel ortam zamanla değişebilir. Örneğin robotik mekanik bileşenler aşınır ya da alıcıların hassasiyeti azalır.

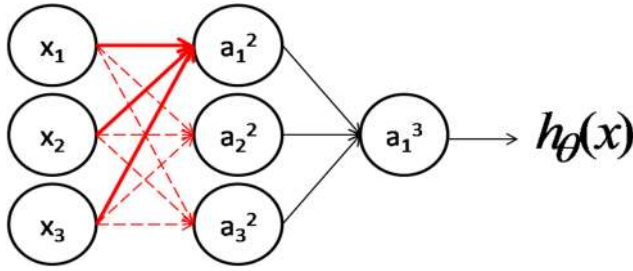
2.4.5.6 YSA Türleri

Yapay Sinir Ağlarını sahip olduğu ağ mimarisi, öğrenme şekli veya katman sayısına göre sınıflandırılmaktadır.

2.4.5.6.1 İleri Beslemeli YSA

İleri Beslemeli ağlar, sınıflandırma ve regresyon problemleri için kullanılan bir diğer makine öğrenme yaklaşımıdır. Tek gizli katmanlı ileriye beslemeli ağlar, pratik uygulamalarda önemli rol oynamaktadır. En iyi gizli düğüm sayısı probleme bağlı olduğu ve önceden bilinmediği için, kullanıcılar genellikle deneme yanılma yöntemiyle gizli düğüm sayısını seçerler. Mimari sabitlendikten sonra çıktı ve gizli tabakalardaki ağırlıkları eşzamanlı olarak ayarlamak için geri yayılım gradiyent inişi gibi yinelemeli bir öğrenme algoritması normal olarak uygulanır (Romero ve Alquézar, 2012).

Aktivasyon değerleri, nöron girdilerine kullanarak aktivasyon fonksiyonuyla hesaplanmaktadır. Aktivasyon fonksiyonundaki θ_{ij} , (j. hedef nöronun katmanı, i. aynı katmandaki hedef nöron sayısı olmak üzere)'ye nöron ağırlığını ifade etmektedir.



Şekil 2.10 İleri Beslemeli YSA

$$a_i^j = g((\theta_i^j)^T \cdot x)$$

- θ_i^j bir vektördür, θ_{ik}^j j-1. katmandaki k. nöronun j. katmandaki i. nörona etkisi demektir
- Bütün k'ler için θ_{ik}^j 'lar θ_i^j vektörü oluşturmaktadır. $(\theta_i^j)^T \cdot x = \sum \theta_{ik}^j x_k$
- Bütün i'ler için θ_i^j vektörleri bir matris oluşturur. Bu matrise j. katmanın ağırlık matrisi denir.

$$a_1^{(2)} = g(z_1^2) = g((\theta_1^2)^T \cdot x)$$

$$a_2^{(2)} = g(z_2^2) = g((\theta_2^2)^T \cdot x) \quad a_1^{(3)} = g(z_1^3) = g((\theta_1^3)^T \cdot a^{(2)})$$

$$a_3^{(2)} = g(z_3^2) = g((\theta_3^2)^T \cdot x)$$

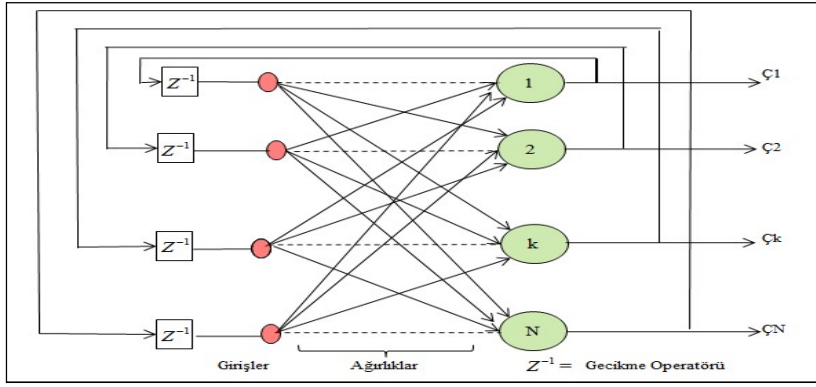
YSA modeli, karmaşık sayılabilecek bir fonksiyon şeklinde aşağıdaki gibi ifade edilebilir.

$$h_\theta(x) = a^{(3)} = g((\theta^3)^T \cdot g((\theta^2)^T \cdot x))$$

YSA başarısını artırmak için parametre seçiminde tekrar dereceli azaltma gibi optimizasyon algoritmaları kullanılmaktadır.

2.4.5.6.2 Geri Beslemeli YSA

Geri Beslemeli YSA'da temel fikir a_{ij} katmanındaki nöron aktivasyon değerlerinin olması gereken değerlerden farkı olarak ifade edilen "hata" yı minimize edilmesi amaçlanmaktadır. Shmueli, hata değerlerini kullanarak ağırlıkların düzeltilmesi sürecini öğrenme olarak ifade etmiştir. (Shmueli vd., 2007: 172). Geri Beslemeli YSA'da ağda geriye doğru yayılarak öğrenme sürecindeki nöron ağırlıkları hata değerine göre tekrar hesaplanmakta, çıktı değerinin, bağımlı değişken değerlerinden farkıyla oluşturulan hata fonksiyonu(Maliyet Fonksiyonu) üretilmektedir. Hopfield, Elman ve Jordan ağırları,geri beslemeli ağlara örnek olarak verilebilir. Şekil 2.5'te, geri beslemeli bir YSA yapısı gösterilmektedir.



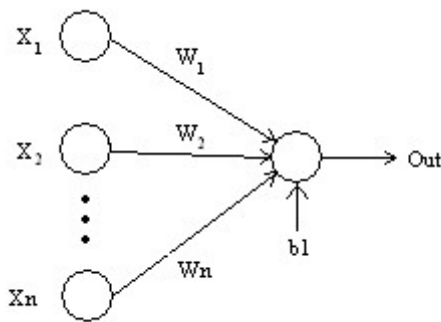
Şekil 2.11 Geri Beslemeli Yapay Sinir Ağı

Kaynak: Öztemel, 2006: 171

Yapay sinir ağının yapılandırılmasında başlangıçta nöron ağırlıkları rassal olarak atanmaktadır. Nöronlara rassal atanan ağırlıkların sonucu olarak, nöronlara farklı hataların gelmesiyle, daha başarılı öğrenme süreci gerçekleştirilmektedir (Pyle, 1999: 368).

2.4.5.6.3 Tek Katmanlı Ağlar

Tek katmanlı YSA'lar, girdi ve çıktı katmanlarından oluşmaktadır. Hesaplanan eşik değeri ile bağlantı ağırlıklarının ve ağın çıktısının sıfır olması önlenmektedir. Gizli Katman barındırmayan tek katmanlı ağlar, doğrusal problemlerin çözümünde kullanılmaktadır. Tek katmanlı ağlarda her ağın bir veya daha fazla girdisi ($x_i; i=1,2,\dots,n$) ve sadece bir çıktısı ($z_k; k=1$), her bağlantının bir ağırlığı ($w_{ik}; i=1,2,\dots,n; k=1$) ile gösterilmektedir. Şekil 2.12'deki YSA modeli tek katmanlı yapıya sahiptir.



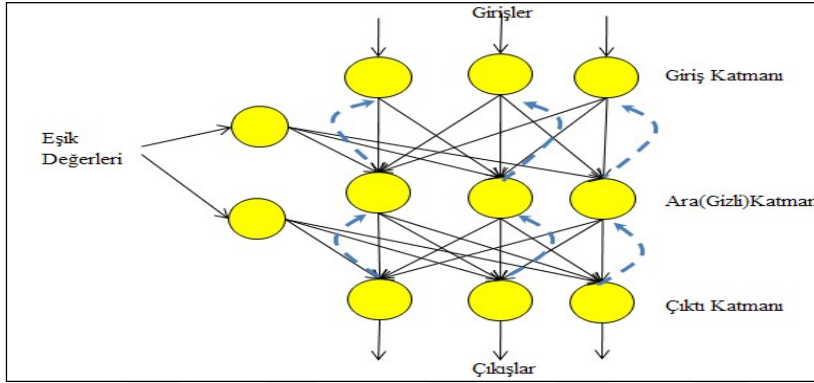
Şekil 2.12 Tek Katmanlı Yapay Sinir Ağı

Çıktı değeri, ağırlıklandırılmış girdi değerleriyle ($w_i x_i$), eşik değerin (θ) toplamına eşittir (Öztemel, 2006: 59-60):

$$\text{Çıktı} = f\left(\sum_{i=1}^m w_i x_i + \theta\right)$$

2.4.5.6.4 Çok Katmanlı Yapay Sinir Ağları

Tahmin ve karmaşık problemlerin çözümünde kullanılan Çok Katmanlı Algılayıcı Modeli (ÇKA), Geriye Yayım Modeli veya “Hata Yayma Modeli olarak da bilinmektedir. ÇKA yapısında girdi katmanı, gizli katman ve çıktı katmanı bulunmaktadır. Ele alınan probleme göre gizli katman sayısı ayarlanabilir (Zhang vd., 1998: 37, 38).



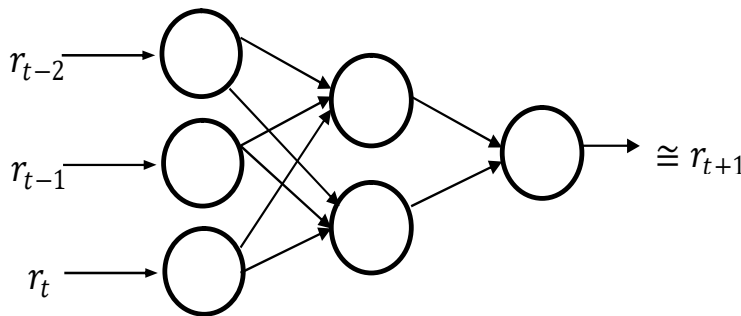
Şekil 2.13 Çok Katmanlı Yapay Sinir Ağı

Kaynak: Zhang, 2004

Şekil 2.13'de bir girdi katmanı, bir gizli katman ve bir çıktı katmanı olan ÇKA modeli yer almaktadır. Öğrenme algoritması olarak ÇKA'da, danışmanlı öğrenme modelini kullanılmaktadır. Eğitim sürecinde girdiler ve girdilere karşılık üretilmesi gereken çıktılar öğretilmektedir (Öztemel, 2006: 77).

2.4.5.7 Yapay Sinir Ağlarının Zaman Serilerine Uygulanması

Literatürde Yapay sinir ağlarının zaman serilerine başarılı bir şekilde uygulandığını gösteren çalışmalar bulunmaktadır. Veri setindeki nitelik değeri ve onun önceki dönem değerleri giriş (input) ($r_t, \dots, r_{t-1}$), oluşturmaktadır. Çıkış değeri (r_{t+1}) olarak ifade edilmektedir. Zaman serisi yöntemlerine göre tahmin başarısı, YSA'da bulunan öğrenme oranı, gizli katman sayısı, nöron sayısının ayarlanmasıyla artmaktadır (Ogasawara, 2009).



Şekil 2.14 YSA'nın Zaman Serilerine Uygulanması

Zaman serisi tahmini sinir ağıları kullanılarak tipik bir veri madenciliği süreci üç temel adımda tarif edilebilir:

Veri Setinin Hazırlanması: Probleme uygun veri kümesinin hazırlandığı aşamadır.

Yapay Sinir Modelinin Oluşturulması: YSA modeli ve parametrelerinin (giriş değerleri, öğrenme oranı) belirlenmesi aşamasıdır.

Model Tahmini: Bu aşamada elde edilen tahmin sonuçları denenmektedir. Tahmin sonuçlarının tutarlı olmaması durumunda önceki aşamalara dönülebilir.

2.4.6 Destek Vektör Makineleri

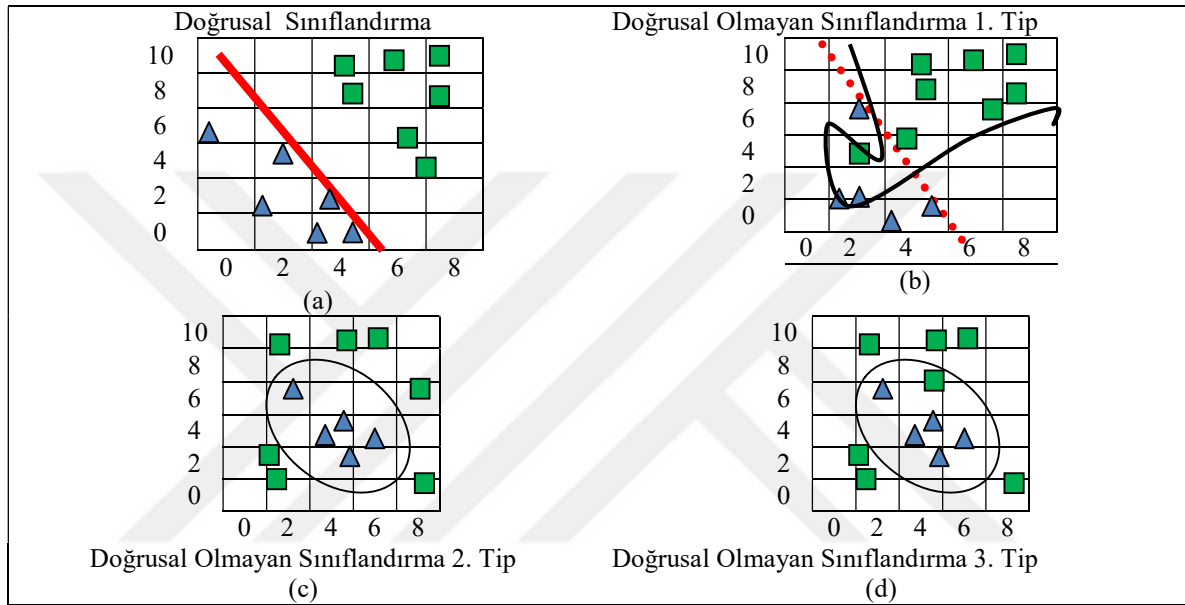
Destek vektör makineleri temel fikri, 1960'lı yıllarda, Hesaplanabilir Öğrenme Teorisinin temelini oluşturan Öğrenmenin Temel Teorisi diğer bir ifade ile Vapnik-Chervonenkis teorisine dayanmaktadır. DVM, sınıflandırma problemlerinin çözümü için Vapnik tarafından geliştirilmiştir (Cortes ve Vapnik, 1995). DVM son derece karmaşık sınıflandırma problemlerinde örneğin el yazısı tanıma, zaman serisi analizi, kanser teşhisi birçok alanda başarılı olduğu söylenebilir (Lantz, 2013: 225; Akpınar, 2014: 268). DVM, kenar payını ölçüt olarak kullanarak model parametreleri öğrenme örneklerinin bir altkümesinin etkilerinin ağırlıklı toplamı olarak hesaplanır ve bu etkiler uygulamaya özgü benzerlik çekirdeğince tanımlanır (Alpaydın, 2011: 261; Akpınar, 2014: 268; Lantz, 2013: 225).

Destek vektör makinelerinin genel özellikleri:

- DVM, örneklerin kendi özellik değerlerine göre çok boyutlu uzayda çizilen noktalar arasında homojen sınıfları oluşturan bir düzlem sınır tanımlamayı amaçlamaktadır.
- DVM, yalnızca sınıra yakın ve ayırtıcı tanımlayan örneklerin saklandığı seyrek en yakın komşu sınıflandırıcısına benzetilebilir.
- Sınıra en yakın muğlak ya da hatalı örnekler üzerinden modelin genelleme hatası kestirilebilir. Ayrıca Modelin örnekler üzerinden yazılabilmesi çekirdek kullanımına imkan tanır.
- Çıktı, çekirdek fonksiyonuyla hesaplanan destek vektörlerin etkilerinin toplamı olarak ifade edilmektedir.
- DVM'de eğer sınıflar doğrusal olarak ayrılamıyorsa doğrusal olmayan bir ayırtaç yerine doğrusal olmayan taban işlevleriyle yeni bir uzaya geçilmektedir.
- Çekirdek tabanlı algoritmalar dışbükey optimizasyon problemleri olarak tanımlanır ve en iyi olan tek bir çözüm vardır.

- DVM performansı üzerinde etkili olan 3 faktör bulunmakta olup bunlar hata terimi (ϵ), düzenleme faktörü (C) ve Kernel parametresi ve denklemi. Model parametreleri ve Kernel tipinin seçimi destek vektör makinelerinin performansı üzerinde etkilidir.

Diğer sınıflandırma yöntemlerine göre eğitim süresinin daha fazla olmasına rağmen, yüksek güvenilirlik düzeyi, aşırı öğrenmeye dayanıklı olması ve doğrusal olmayan sınıflandırma problemlerinde başarılı bir şekilde uygulanması ile öne çıkmaktadır.



Şekil 2.15 Sınıflandırma Tipleri

Kaynak: Akpınar, 2014: 269

Şekil 2.15'de sınıflandırma örnekleri yer almaktadır. Doğrusal sınıflandırma problemine uygun olarak Şekil 2.15 a'da doğrusal ayrılabilir veriler yer almaktadır. Doğrusal olmayan sınıflandırma 3 tipte yer almaktadır. Şekil 2.15 b'de veriler hatalı olsada doğrusal olarak sınıflandırılabilir. Doğrusal olmayan 2. Tip veriler elips şeklinde dağıldığından dolayı doğrusal olarak sınıflandırılmamaktadır. Doğrusal olmayan 3. Tip veriler hatalı sınıflandırılmıştır.

2.4.6.1 Hiper Düzlem

DVM'de hiper düzlem, kenar payının mümkün olan en büyük değerinin elde edilmesiyle sınıflandırma başarısını artırılması amaçlanmaktadır.

$$\mathcal{X} = \{\mathbf{x}^t, r^t\} \text{örnekleme için}$$

$\mathbf{x}^t \in C_1$ ise $r^t = +1$ ve $\mathbf{x}^t \in C_2$ $r^t = -1$ değerini alsın. $\mathbf{w}$ ve w_0 parametrelerinin sağlaması gereken koşullar:

$$\mathbf{w}^T \mathbf{x}^t + w_0 \geq +1 \text{ eğer } r^t = 1$$

$$\mathbf{w}^T \mathbf{x}^t + w_0 \leq -1 \text{ eğer } r^t = -1$$

$$r^t (\mathbf{w}^T \mathbf{x}^t + w_0) \geq 1$$

$\mathbf{x}^t$ örneğinin ayırtaca olan uzaklığı:

$$\frac{|\mathbf{w}^T \mathbf{x}^t + w_0|}{\|\mathbf{w}\|}$$

Ayırtacın her iki tarafındaki örneklerin ayırtaca olan uzaklığı p , bu değerin en az olması için:

$$\frac{r_t (\mathbf{w}^T \mathbf{x}^t + w_0)}{\|\mathbf{w}\|} \geq p, \forall t$$

p değerini olabildiğince büyütmek için $p\|\mathbf{w}\|=1$ olarak sabitleyip kenar payını olabildiğince büyütmek için $\|\mathbf{w}\|$ değerini (vektör boyunu) olabildiğince küçük olması gereklidir.

$$\min \frac{1}{2} \|\mathbf{w}\|^2$$

$$\text{kısıt: } r^t (\mathbf{w}^T \mathbf{x}^t + w_0) \geq +1, \forall t$$

Böylece girdi boyutu d değerine bağlı en iyi $\mathbf{w}$ ve w_0 değerlerini bulan ikinci dereceden optimizasyon problemi elde edilmektedir. Hiperdüzlemin etrafındaki örnekler sınıra en az $1/\|\mathbf{w}\|$ 'a ve kenar payı en az $2/\|\mathbf{w}\|$ 'a eşit olmaktadır.

En iyi hiperdüzlemi bulurken optimizasyon probleminde karmaşıklığı girdi boyutu d yerine örneklem boyu N değeri ile ifade edilerek taban işlevlerinden çekirdek işlevlerine geçişi sağlamaktadır. Bunun için:

α^t Langrange çarpanlarıyla ifade edilir.

$$L_p = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{t=1}^N \alpha^t [r^t (\mathbf{w}^T \mathbf{x}^t + w_0) - 1]$$

$$= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{t=1}^N \alpha^t r^t (\mathbf{w}^T \mathbf{x}^t + w_0) \sum_t \alpha^t$$

KKT koşullarını kullanarak dual problemi çözülebilir. Dual, L_p problemini α^t değişkenlerine göre büyütmektir. Kısıtlar ise eğimin 0 ve $\alpha^t \geq 0$ olmasıdır.

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{t=1}^N \alpha^t r^t \mathbf{x}^t$$

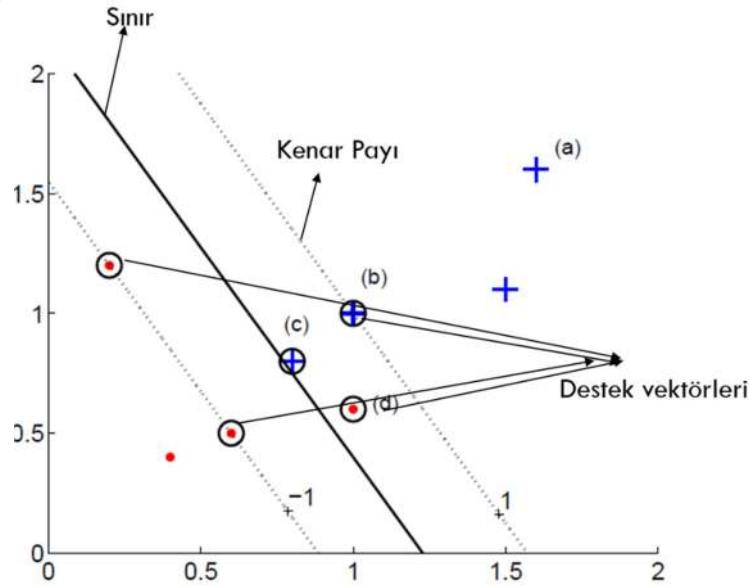
$$\frac{\partial L_p}{\partial w_0} = 0 \Rightarrow \sum_{t=1}^N \alpha^t r^t = 0$$

$$\begin{aligned} L_d &= \frac{1}{2} (\mathbf{w}^T \mathbf{w}) - \mathbf{w}^T \sum_t \alpha^t r^t \mathbf{x}^t - w_0 \sum_t \alpha^t r^t + \sum_t \alpha^t \\ &= -\frac{1}{2} (\mathbf{w}^T \mathbf{w}) + \sum_t \alpha^t \\ &= -\frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s r^t r^s (\mathbf{x}^t)^T \mathbf{x}^s + \sum_t \alpha^t \end{aligned}$$

Kısıtlar:

$$\sum_t \alpha^t r^t = 0 \text{ ve } \alpha^t \geq 0, \forall t$$

α^t değerleri hesaplanmasyla, örneklerin küçük bir kısmının $\alpha^t > 0$ olduğu, geri kalanının 0'a eşit olduğu görülmektedir. $\alpha^t > 0$ olan $\mathbf{x}^t$ örneklerine destek vektörleri denir. Şekil 2.16'da doğrusal sınıflandırma için destek vektörleri yer almaktadır.



Şekil 2.16 Destek Vektörlerinin Oluşturulması

Kaynak: Alpaydın, 2011: 267

Şekil 2.16 incelendiğinde a örneği dışında kalanlar destek vektörleridir. Diğer örnekler için:

- a) Örnek ayırtacın doğru yanında ve kenardan uzaktır. $r^t(g(\mathbf{x}^t)) > 0, \xi^t = 0$
- b) Örnek doğru yanda ve kenarın üstündedir. $\xi^t = 0$

c) Örnek doğru yanda ama kenarın içinde ayırtaca yeterince uzak değildir.

$$\xi^t = 1 - g(x^t), 0 < \xi^t < 1$$

d) Örnek ayırtacın yanlış yanında ve yanlış sınıflandırılmıştır. $\xi^t = 1 - g(x^t) > 1$

Dual değişken olarak yazılırsa

$$a) \alpha^t = 0;$$

$$b) \alpha^t < C;$$

$$c) \alpha^t = C$$

$$d) \alpha^t = C$$

Destek vektörleri olan $\alpha^t > 0$ örnekleri w ağırlık vektörünü tanımlanmasında, $\alpha^t < C$ örnekleri kenar üzerinde olduğu için w_0 değerini hesaplanmasında kullanılmaktadır. x^t örnekleri $r^t(w^T x^t + w_0) = 1$ koşulunu sağlamaktadır. Herhangi bir destek vektörüyle, w_0 değerini hesaplamakta kullanabiliriz.

$$w_0 = r^t - w^T x^t$$

Bütün destek vektörleri için hesaplanan w_0 değerlerinin ortalaması alınarak elde edilen ayırtaç, destek vektör makinesini oluşturmaktadır.

2.4.6.2 Doğrusal Olmayan Sınıflandırma

Sınıfların doğrusal olarak ayrılmadığı durumlarda, hatalı sınıflandırılan ve doğru sınıflandırılan fakat yeterince uzak olmayan örneklerin artık değişkenleri tanımlanır. Artık değişkenini eklenmesiyle;

$$r^t(w^T x^t + w_0) + \xi^t \geq 1$$

$$\xi^t \geq 0$$

KISITLAR

$\xi^t = 0$ ise x^t doğru sınıflandırılmıştır.

$0 < \xi^t < 1$ ise x^t doğru sınıflandırılmasına rağmen kenar payının içerisinde.

$\xi^t \geq 1$ ise x^t yanlış sınıflandırılmıştır. Yanlış sınıflandırılan örnek sayısı ve doğrusal ayrılamayan örnek sayısını verir.

$$\sum_t \xi^t$$

Bu eşiksiz hata ölçütünü bir ceza terimi olarak denkleme eklenir.

$$L_p = \frac{1}{2} ||w||^2 + C \sum_{i=1}^n \xi^t$$

C , düzenleme parametresidir. Karmaşıklığı burada ağırlık vektörünün L_2 normu olarak ölçülür. Hata ise doğrusal ayrılmayan örnek sayısıdır. Yeni kısıtlar eklendiğinde denklem:

$$L_p = \frac{1}{2} ||w||^2 - C \sum_t \xi^t - \sum_t \alpha^t - [r^t(w^T x^t + w_0) - 1 + \xi^t] - \sum_t \mu^t \xi^t$$

Burada μ^t gösterdiğimiz ξ^t değerlerinin negatif olmamasını sağlayan yeni Lagrange parametreleridir. Türev alıp sıfıra eşitlenirse:

$$\frac{\partial l_p}{\partial w} = w - \sum_t \alpha^t r^t x^t = 0 \rightarrow w = \sum_t \alpha^t r^t x^t$$

$$\frac{\partial l_p}{\partial w_0} = \sum_t \alpha^t r^t = 0$$

$$\frac{\partial l_p}{\partial \xi^t} = C - \alpha^t - \mu^t = 0$$

$\mu^t \geq 0$ olduğu için son satır $0 \leq \alpha^t \leq C$ koşulunu gerektirir. $L_d = \sum \alpha^t - \frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s r^t r^s (x^t)^T x^s$

Kısıtlar

$$L_d = \sum \alpha^t r^t = 0 \text{ ve } 0 \leq \alpha^t \leq C, \forall t$$

Böylece Sınırın doğru yanında ve yeterince uzak örnekler için $\alpha^t = 0$ olur.

- Destek Vektörleri $\alpha^t > 0$ olan örneklerdir. w ağırlık vektörünü tanımlar.
- $\alpha^t < C$ olanlar kenarın üzerindedir ve w_0 değerini hesaplar ve $\xi^t = 0$ olur.

$r^t(w^T x^t + w_0) = 1$ eşitliğini sağlarlar.

- Kenar payının içine düşen veya yanlış sınıflandırılan örnekler için $\alpha^t = C$ dir.

Destek vektörler öğrenme kümesinde yer alarak, yanlış sınıflandırma veya doğru dahi olsa sınıra çok yakın olabilecek örneklerin olmasını engellemektedirler. Destek vektör sayısının beklenen hata için bir üst sınır oluşturmaktadır. Beklenen hata oranı, Vapnik tarafından:

$$E_N[P(hata)] \leq \frac{E_N[\#destekvektörü]}{N}$$

$E_N[.]$: N büyüklüğündeki öğrenme kümeleri üzerinde beklenen değeri gösterir. Beklenen hata girdi boyutu d yerine destek vektör sayısına bağlıdır. Eğer örnek yanlış taraftaysa, ya da kenar payı < 1 ise

$y^T = w^T x^t + w_0$ çıktısı ve r^t istenen değer için menteşe kayıp fonksiyonu:

$$L_m(y^t, r^t) = \begin{cases} 0 & \text{eger } y^t r^t \geq 1 \\ 1 - y^t r^t & \text{degilse} \end{cases}$$

2.4.6.3 Çekirdek

Doğrusal olmayan Sınıflandırma Probleminde, doğrusal olmayan model oluşturmak yerine doğrusal olmayan taban fonksiyonlarıyla bir uzaya gidip orada doğrusal bir model oluşturabiliriz. Yeni uzaydaki doğrusal model, ilk uzaydaki doğrusal olmayan bir modele karşılık gelir. Bu yaklaşım sınıflandırma ve regresyon için kullanılabilir.

Yeni boyutlar taban işlevleriyle tanımlanır:

$$z = \varphi(x) \quad \text{öyleki } z = \varphi(x)_j, j=1, \dots, k$$

ve böylece d boyutlu x uzayından k boyutlu z uzayına geçip ayırtacı bu yeni uzay:

$$\begin{aligned} g(z) &= w^T z \\ g(x) &= w^T \varphi(x) \\ &= \sum_{j=1}^k w \varphi_j(x) \end{aligned}$$

Ayrıca bir w_0 değeri kullanmayıp $z_1 = \varphi(x) \equiv 1$ olarak ek bir yeni boyut olduğunu varsayılmaktadır.

$$\begin{aligned} L_p &= \frac{1}{2} ||w||^2 - C \sum_t \xi^t \\ r^t w^T \varphi(x^t) &\geq 1 - \xi^t \end{aligned}$$

Langrange terimleri:

$$L_p = \frac{1}{2} ||w||^2 - C \sum_t \xi^t - \sum_t \alpha^t - [r^t w^T \varphi(x^t) - 1 + \xi^t] - \sum_t \mu^t \xi^t$$

Parametrelere göre türev alınıp sıfıra eşitlenir

$$\frac{\partial L_p}{\partial w_0} = w = \sum_t \alpha^t r^t \varphi(x^t)$$

$$\frac{\partial L_p}{\partial \xi^t} = C - \alpha^t - \mu^t = 0$$

Bu durumda dual:

$$L_d = \sum_t \alpha^t - \frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s r^t r^s \varphi(x^t)^T \varphi(x^s)$$

Kısıtlar:

$$L_d = \sum \alpha^t r^t = 0 \text{ ve } 0 \leq \alpha^t \leq C, \forall t$$

Çekirdek makineleri, taban işlevlerinde $\boldsymbol{\varphi}(x^t)^T \boldsymbol{\varphi} x^s$ iç çarpımını ilk girdiler üzerinde bir $K(x^t, x^s)$ çekirdek fonksiyonu olarak yazılmasıdır

$$L_d = \sum \alpha^t - \frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s r^t r^s K(x^t, x^s)$$

Çekirdek işlevi ayrıtaça:

$$\begin{aligned} g(\mathbf{x}) &= \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) = \sum_t \alpha^t r^t \boldsymbol{\varphi}(x^t)^T \boldsymbol{\varphi}(\mathbf{x}) \\ &= \sum_t \alpha^t r^t K(x^t, \mathbf{x}) \end{aligned}$$

Çekirdek fonksiyonu kullanarak yeni uzay eşlenmesine gerek duyulmamaktadır. Her geçerli çekirdek için karşılık gelen bir taban işlevi olmasına rağmen $K(x^t, \mathbf{x})$ değerini hesaplayıp kullanmak yerine $\boldsymbol{\varphi}(x^t)^T \mathbf{v} \boldsymbol{\varphi}(\mathbf{x})$ değerlerini hesaplanması işlemleri kolaylaştırdığı söylenebilir.

2.4.7 Destek Vektör Regresyon

Destek Vektör Regresyonda, DVM'ye benzer biçimde kenar payı, artık değişkenler tanımlanmaktadır (Alpaydın,2011:279-280).

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$$

Doğrusal regresyonda hata, gözlenen değer ile beklenen değerlerin farkının karesidir.

$$e_2(r^t, f(x^t)) = [r^t - f(x^t)]^2$$

Destek Vektör Regresyonda ϵ duyarlı kayıp işlevi kullanılmaktadır.

$$e_\epsilon(r^t, f(x^t)) = \begin{cases} 0 & \text{eğer } |r^t - f(x^t)| < \epsilon \\ |r^t - f(x^t)| - \epsilon & \text{değilse} \end{cases}$$

ϵ duyarlı kayıp işleviyle, hem ϵ değerine kadar olan hatalar gözardı edilmekte hem de daha büyük hataların etkisi klasik regresyondan farklı olarak karesel değil, doğrusal olmaktadır. Böylelikle gürültüye dayanıklı, gürbüz bir regresyon modeli elde edilmektedir. Hiper düzlemde olduğu gibi ϵ alanından (hatalı) sapmalar için artık değişkenler tanımlanmaktadır.

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_t (\xi^t_+ + \xi^t_-)$$

Sapmalar artı ve eksi yönde iki yönde (artı ve eksi) tanımlanır ve kısıtlar:

$$\begin{aligned} r^t - (w^T x + w_0) &\leq \epsilon + \xi^t_+ \\ (w^T x + w_0) - r^t &\leq \epsilon + \xi^t_- \\ \xi^t_+, \xi^t_- &\geq 0 \end{aligned}$$

Bu hata işlevini sırt sırta ters oturmuş, artı ve eksi sapmalara bakan iki menteşe kayıp işlevi olarak da görebiliriz.

Langrange terimleri:

$$\begin{aligned} L_p = & \frac{1}{2} ||w||^2 + C \sum_t \xi^t_+ + \xi^t_- \\ & - \sum_t \alpha^t_+ [\epsilon + \xi^t_+ - r^t + (w^T x + w_0)] \\ & - \sum_t \alpha^t_- [\epsilon + \xi^t_- + (w^T x + w_0) - r^t] + \\ & - \sum_t \alpha^t_- [\epsilon + \xi^t_- + (w^T x + w_0) - r^t] \\ & - \sum_t (\mu^t_+ \xi^t_+, \mu^t_- \xi^t_-) \end{aligned}$$

Kısmi türevleri alınması:

$$\frac{\partial L_p}{\partial w} = w - \sum_t (\alpha^t_+ - \alpha^t_-) x^t = 0 \rightarrow w = \sum_t (\alpha^t_+ - \alpha^t_-) x^t$$

$$\frac{\partial L_p}{\partial w_0} = \sum_t (\alpha^t_+ - \alpha^t_-) x^t = 0$$

$$\frac{\partial L_p}{\partial \xi^t_+} = C - \alpha^t_+ - \mu^t_+ = 0$$

$$\frac{\partial L_p}{\partial \xi^t_-} = C - \alpha^t_- - \mu^t_- = 0$$

Duali:

$$\begin{aligned} L_d = & -\frac{1}{2} \sum_t \sum_s (\alpha^t_+ - \alpha^t_-) (\alpha^s_+ - \alpha^s_-) (x^t)^T x^s \\ & - \epsilon \sum_t (\alpha^t_+ + \alpha^t_-) - \sum_t r^t (\alpha^t_+ - \alpha^t_-) \end{aligned}$$

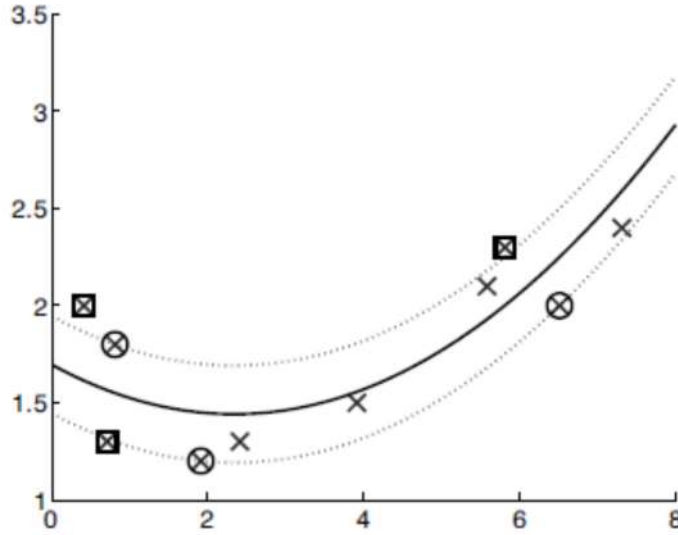
Kısıtlar:

$$0 \leq \alpha^t_+ \leq C, 0 \leq \alpha^t_- \leq C, \sum_t (\alpha^t_+ - \alpha^t_-) = 0$$

İstenilen duyarlılıkta tahmin edilen örnekler ϵ alanında yer almaktadır. İki tür destek vektörü oluşmaktadır. İlk türde ϵ alanın sınırında oluşan destek vektörleri için $\alpha^t_+ > 0$ yada

$\alpha^t > 0$ koşulları sağlanarak w_0 değeri hesaplanmaktadır. İkinci tür ϵ alanın dışında yer aldığından dolayı başarılı tahmin yapılamamıştır.

Şekil 2.16'da regresyon çizgisi ve kuadratik çekirdeği kullanan ϵ -tüpleri gösterilmiştir ($C = 10$, $\alpha = 0.25$). Çember içindeki örnekler, kenar boşluklarındaki destek vektörleridir; kare örnekler, dışlayıcı olan destek vektörleridir.



Şekil 2.17 Destek Vektör Regresyon

Kaynak: Alpaydın, 2014: 273

2.4.8 Gauss Süreç Regresyon

Gauss Süreç Regresyon Bayesçi yapay öğrenme yaklaşımıdır (Chatzis vd., 2012). Bir Gauss Süreci (GS) tesadüfî değişkenlerin, ortak bir Gauss dağılıma sahip olduğu herhangi bir sonlu sayıda bir kümesi başka bir ifadeyle Gauss olasılık dağılımının bir genellemesi olarak ifade edilebilir. Bu genelleme de, fonksiyonlar üzerinde bir dağılım yerine vektörlerin üzerinde dağılım tercih edilmektedir. Model verileri parametrik olmayan bir yöntemdir ve Gauss Süreci sonsuz boyutlara taşımaktadır. Ancak, aynı çıkarım veren sadece bir sonlu alt kümesi ile çalışır (Petelin, 2011).

Gauss Dağılımı ortalama ve kovaryans tarafından belirtilen vektörlerin üzerinde $N(\mu, \Sigma)$ dağılımı ifade etmektedir.

Gauss Süreci ise ortalama fonksiyonu ve kovaryans fonksiyonu tarafından belirlenen GS ($M(X)$, $k(X, X')$) fonksiyonları üzerinde dağılımı ifade eder.

GS'de, x giriş vektörüne y çıkış değeri olmak üzere $y = f(x)$ tanımlanan fonksiyonda $y_1, \dots, y_N \sim N(0, \Sigma)$, $\Sigma_{pq} = Cov(p, y_q) = C(x_p, x_q)$ giriş noktalarına (x_p, x_q) karşılık gelen çıkış noktaları arasındaki kovaryans hesaplanmaktadır. Kovaryansın hesaplanması, çekirdek işlevine benzetilebilir (Alpaydın, 2011: 305).

Böylece Gauss Süreci özelliği olan ortalama $\mu(x)$ ve kovaryans fonksiyonu $C(x_p, x_q)$ elde edilmektedir.

$C(x_p, x_q)$ kovaryans fonksiyonunun değeri, x_p, x_q girişlerine göre $f(x_p)$ ve $f(x_q)$ bireysel çıkışlar arasındaki korelasyondur.

Kovaryans fonksiyonu $C(\cdot, \cdot)$ bir pozitif yarı-tanımlı kovaryans matrisi üreten bir fonksiyon olarak iki parçadan oluşmaktadır.

$$C(x_p, x_q) = C_f(x_p, x_q) + C_n(x_p, x_q)$$

C_f : Fonksiyonel bölümünü temsil etmektedir.

C_n : Gürültü bölümünü temsil eder.

Gürültü için en sık kullanılan sabit kovaryans fonksiyonudur. Fonksiyonel kısımın kovaryans fonksiyon seçiminde durağan verilerde en sık kullanılan kovaryans fonksiyonu kare üstel kovaryans fonksiyonu tercih edilmektedir. Durağan olmayan verilerde ise polinom veya özel durum doğrusal kovaryans fonksiyonu kullanılmaktadır (Petelin, 2011).

$$C(x_p, x_q) = \sum_{d=1}^D w_d x_{dp} x_{dq} + \delta_{pq} v_0$$

w_d ve v_0 : Kovaryans fonksiyonunun hiper parametrelerdir.

D: Giriş boyutudur

δ_{pq} : Modelin varsayılan beyaz gürültüsüdür. ($\delta_{pq} = 1, p = q$ ise 0'dır.)

w_d : Girdilerin ağırlığıdır.

Hiperparametreler vektör olarak yazılabilir. $\Theta = [w_1, \dots, w_d, v_0]^T$. Verilen bir problem için, hiperparametre değerleri eldeki verileri kullanılarak öğrenilir. Ayrıca w_d sıfır veya sıfıra yakınsa, d boyutlu girişlerin az bilgi içerdiği ve muhtemelen ihmal edilebileceği anlamına gelmektedir.

2.4.8.1 Tahmin Süreci

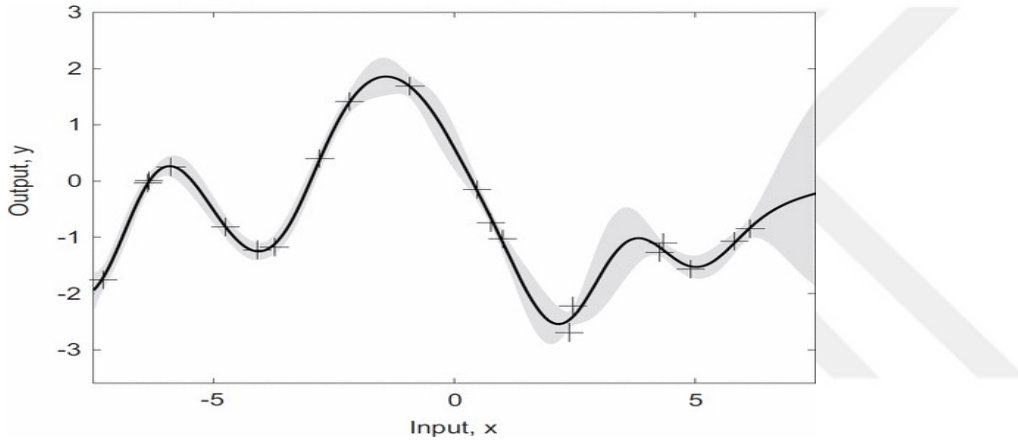
N boyutlu giriş vektörleri $X = [x_1, x_2, \dots, x_N]$ ve $y = [y_1, y_2, \dots, y_N]^T$ vektörlerin çıkış verileri olmak üzere; (X, y) verileriyle yeni giriş vektörü olan X^* tanımlayarak y^* 'ın (çıkış) tahmin dağılımını kovaryans matrisi K NxN olarak tanımlanır. N+1 veri noktası için, $x^* = [x_1(N+1), x_2(N+1), \dots, x_D(N+1)]^T$. Kovaryans matrisi $K^* = \begin{bmatrix} K & k(x^*) \\ k^t(x^*) & k(x^*) \end{bmatrix}$ ve $[y_1, \dots, y_N, y^*]^T$ eğitim verilerinin çıkış vektörü olmak üzere rastgele değişkenleri $[y, y^*]^T \sim N(0, K^*)$ gibi ifade edilebilir (Alpaydın, 2011: 305; Petelin, 2011).

Tahmin edilen çıkış için ortalama ve varyans ile normal olasılık dağılımı ise:

$$\mu(y^*) = k(x^*)^T K^{-1} y,$$

$$\sigma^2(y^*) = k(x^*) - k(x^*)^T K^{-1} k(x^*)$$

burada $k(x^*) = [C(x_1, x^*), \dots, C(x_N, x^*)]^T$, test ve eğitim durumları arasında kovaryans vektörüdür. $k(x^*) = C(x^*, x^*)$ test girişi ile kendisi arasındaki kovaryanstır. Elde edilen model ortalama değere ek olarak varyans tarafından tahmin güven aralığı bilgi vermektedir. Genellikle, tahmin güven yaklaşık % 95 güven aralığında olan 2σ aralıktadır. Bu güven bölgesi gri bir bant halinde Şekil 2.16' da, görülebilir. Bu tahmin kalitesinin düşüklüğünü ve ortalama etrafında daha geniş bir güven bant belirtilmesi nedeniyle verinin gürültülü veya eksikliğinden dolayı kötü girdi alanı alanları göstermektedir.



Şekil 2.18 Gauss Süreç Regresyon Modeli

Kaynak: Petelin, 2011a.

GSR'de diğer yöntemlerden farklı olarak, sabit bir model yapısı içinde herhangi bir model parametresi belirlenmemekte, kovaryans fonksiyonu hiperparametrelerinin optimizasyonu ile oluşmaktadır.

2.4.8.2 Parametrelerin Optimizasyonu

Eğitim verilerindeki korelasyonu doğru bir şekilde belirlemek için kovaryans fonksiyonunun hiperparametrelerinin optimize edilmesi gereklidir. Gaus Süreçlerinde model parametrelerinin ve model hatasını tanımlayan maliyet fonksiyonunu en küçükleyen optimizasyonu yerine, olasılıklı yaklaşım daha uygun olacağı belirtilmektedir (Petelin, 2011). Model hatasını en aza indirmek için, modelin olasılığı maksimize edilmektedir. Modelin olasılık maksimizasyonu genellikle En çok- olabilirlik kestirimi yöntemi ile yapılır. Bu maksimize edilecek maliyet fonksiyonu:

$$L(\theta) = -\frac{1}{2} \log(|K|) - \frac{1}{2} y^T K^{-1} y - \frac{N}{2} \log(2\pi)$$

Maliyet fonksiyonu optimize etmek için eşlenik gradyanlar yöntemi kullanılmaktadır. Zaman serisi modellerinde, t zamanında y gecikmeli çıktı değerleri, u 'ya (dışsal girdilere) bağlıdır.

$$y(t) = f(y(t-1), \dots, y(t-n), \\ u(t-1), \dots, u(t-n)) + \epsilon(t)$$

$\epsilon(t)$ beyaz gürültü ve çıkış $y(t)$ olup t zamanı için durum vektörüne $(t) = [y(t-1), y(t-2), \dots, y(t-n), u(t-1), u(t-2), \dots, u(t-n)]$ T'ye bağlıdır. t bilinen/gözlenen değerler olmak üzere h adım için yapılacak tahminde $x(t+h)$ karşılık gelen $y(t+h)$ 'dağılımını bulunması gereklidir.

Bazı kovaryans fonksiyonları şunlardır (Hong, 2012).

Kareliüstel(SE)kovaryansfonksiyonu

$$k_{SE} = \exp\left(-\frac{r^2}{2l^2}\right)$$

Matern kovaryans fonksiyonu

$$k_{Matern}(r) = \frac{2^{1-v}}{r(v)} \left(\frac{\sqrt{2\nu}r}{l}\right)^v K_\nu\left(\frac{\sqrt{2\nu}r}{l}\right)$$

2.5 Literatürde Veri Madenciliği ile Yapılan Tahmin Çalışmaları

Satış/Talep tahminine yönelik çalışmalarda zaman serileri kullanımı 1950'li yıllarda başlamıştır. Günümüzde zaman serileri analizinde, veri madenciliği yöntemleri de kullanılmaktadır.

Veri madenciliği tahmin yöntemleri, satışlar ile açıklayıcı değişkenler (sayısal veya nominal) arasındaki bağlantılar veya ilişkiler gibi veri kümelerindeki desenleri bulmak ve tanımlamak için yardımcı olur. Bu desenler yorumlanarak gelecekteki verilerin tahminlerini gerçekleştirmek üzere genelleştirilebilir. Bir veri kümesindeki desenleri (karar ağaçları, sınıflandırma kuralları, kümeleme, doğrusal modeller, bulanık sistemler veya sinir ağları) temsil etmek için birçok farklı yol vardır ve bunların her biri farklı bir bilgi kazanımı sağlamaktadır (Thomassey ve Fiordaliso, 2006).

Yapılan çalışmalarda Yapay Sinir Ağları, Destek Vektör Makine/Regresyon, Gauss Süreç Regresyon, Karar Ağaçları yöntemleri satış/talep tahmini konusunda başarılı sonuçlar

elde etmişlerdir. Tahmin çalışmalarında Veri Madenciliği yöntemleri zaman ve maliyet tasarrufu sağlamaktadır. Bu bağlamda talep/satış tahminine yönelik çalışmalarda VM'yöntemlerinin önemi artmaktadır.

Talep/satış tahminine yönelik yapılmış bazı çalışmalar Tablo 2.6'da yer almaktadır.

Tablo 2.6 Veri Madenciliği ile Yapılan Tahmin Çalışmaları

Yazar	Yöntem	Konu
Kim, 2003	DVM-YSA	Çalışmada DVM ve YSA yöntemleriyle finansal zaman serileri kullanılarak borsa fiyat indeksini tahmini gerçekleştirilmiştir. Uygulama sonucunda DVM yönteminin YSA'ya göre tahmin çalışmasında daha başarılı olduğu görülmüştür.
Emel, 2005	C&RT	Karar ağacı yöntemiyle bir işletmenin müşterilere göre kişiselleştirilmiş satış hareketlerini içeren veri tabanından yararlanarak sınıflama tipi satış tahmin modeli geliştirilmiştir. Geliştirilen model, ayrıntılı satış tutarları, müşterilerin satın alma davranış profillerini açıklamaktadır. Çalışmada geliştirilen modelin, işletme pazarlama stratejilerine olumlu katkı yaptığı sonucuna ulaşılmıştır.
Tolun, 2008	DVM-YSA	Bu çalışmada Türkiye’de 1995-2001 yılları arasında faaliyet gösteren özel sermayeli ticaret bankaların (37 banka) 64 adet değişkenle başarısızlık tahmini için oluşturulan modeller oluşturulmuştur. Destek Vektör Makineleri yönteminin banka başarısızlığının tahmin edilmesinde tutarlılık ve tahmin gücü açısından tatmin edici sonuçlar verdiği görülmüştür.
Etemad-Shahidi, 2009	M5p, YSA	Bir göldeki dalga boyunu tahmini için M5P ve yapay sinir ağlarıyla (üç katmanlı ileri beslemeli) modeller geliştirilmiş ve sonuçlar karşılaştırılmıştır. Çalışmada 2 adet veri seti kullanılmıştır. Eğitim seti olan ilk veri seti 6 Nisan-10 Kasım 2000 döneminde 4045 adet dalga ve rüzgar hızını kapsamaktadır. Test veri seti olan 2. veri seti ise 19 Nisan-6 Kasım 2001 döneminde aynı bilgileri içeren 3259 kayıt yer almaktadır. M5p yönteminin YSA'ya göre biraz daha başarılı olduğunu bulmuşlardır.
Khashei, 2010	ARIMA-YSA	ARIMA modelini kullanan yapay sinir ağır tasarlamışlardır. Yapılan çeşitli denemelerde en iyi sonuç veren yapay sinir ağının 8-3-1 yapısında olduğunu ve çıktı katmanında genellikle transfer fonksiyonu olarak lineer fonksiyon kullanılmasının tahmin başarısını artırabileceği sonucuna ulaşmışlardır.

Yazar	Yöntem	Konu
Khah, 2011	ARIMA- RBF-YSA	Elektrik fiyat tahmini için dalga dönüşümü, ARIMA, radyal temelli ANN (Yapay Sinir Ağı) içeren hibrit bir model tasarladıkları çalışmada, dalga dönüşümü ve ARIMA yöntemleri ile elektrik fiyat tahmin probleminde lineer kısımlar tespit etmişler, ANN'yi ise diğer yöntemlerden elde edilen tahminlerdeki hataları düzeltmek için kullanmışlardır.
Wang, 2011	ANFIS- YSA	Otomobil satışlarını tahmin etmek için Bulanık Mantık Sistemine Dayalı Uyarlanabilir Ağ (ANFIS) kullanmışlar ve bunun sonuçlarını ARIMA ve YSA ile kıyaslamışlardır. Çalışma neticesinde ANFIS modelinin ARIMA ve YSA göre daha iyi olduğunu bildirmişlerdir.
Jirong, 2011	GA-DVM	Konut fiyatlarını tahmin etmek için genetik algoritma ve destek vektör makineleri kullanarak yapmış oldukları çalışmada genetik algoritma ve destek vektör makinelerini birleştirerek hibrid bir model oluşturmuşlardır. Model ile yapılan deneysel çalışmalarda klasik modele göre daha doğru sonuçlar elde edildiği görülmüştür
Shuhaida, 2011	DVM	Öz düzenleyici haritalar (SOM)-En Küçük Kareler Destek Vektör Makineleri (LSSVM) ile oluşturdukları hibrid modelin esnekliğini LSSVM modeli ile karşılaştırarak incelemişlerdir. SOM-LSSVM modelinin etkinliğini değerlendirmek için Wolf Yıllık Sunspot Verileri ve Aylık genç işsiz kadın verileri kullanılmıştır. Yapılan deney, SOM-LSSVM'nin ortalama mutlak hata (MAE) ve Ortalama Karasel hatanın karakökü (RMSE) kriterlerine dayalı olarak LSSV modelinden daha üstün olduğunu göstermektedir. Yapılan çalışma neticesinde aynı zamanda SOM-LSSVM'nin zaman serileri tahmininde alternatif bir teknik olduğu görülmüştür
Petelin vd., 2011a	GSR, BVAR	Gauss Süreç Regresyon ve BVAR modelleri ile , 1990-2005 dönemi için ABD emtia piyasaları verileri (Avustralya Doları , İngiliz Poundu, Kakao , Kanada Doları,Ham Petrol, Koton, Besi Sığırtı, Altın, Kalorifer Yakıtı, Benzin , Buğday) üzerinde tahmin (h=14 günlük, kapanış verileri) tahmin modelleri geliştirmişlerdir. GSR Kalorifer yakıtı ve Benzin haricinde, diğer emtia fiyatlarında BVAR modeline göre daha iyi sonuç vermiştir.

Yazar	Yöntem	Konu
Zhan, 2011	M5p,	Otobanda meydana gelen kazalardan sonra şeritlerin trafik akışını tekrar sağlanması için M5p algoritması ile tahmin modeli geliştirilmiştir. Florida'da 2006 Nisan-2007 Mart döneminde meydana otoban olayları/kazaları veri seti hazırlanmıştır. Zaman,Arıza süresi,Çevresel,Olay tipi,Trafik işletme, Olay bilgisi faktörleriyle elde edilen model geleneksel regresyon ve karar ağacı yöntemlerine göre başarılı bulunmuştur.
Lu vd., 2012	Mars-DVR- YSA- ARIMA	Bilgisayar toptancıları için satış tahmini üzerine gerçekleştirdikleri çalışmada çok değişkenli adaptif regresyon eğrileri (MARS) ve YSA ile karşılaştırma gerçekleştirmişlerdir. Yapılan çalışma neticesinde MARS modelinin tahmin başarısı uygulamada kullandıkları diğer yöntemlere göre (YSA,DVM,ARIMA)daha başarılı olduğu görülmüştür.
Huang, 2012	GA-DVR	Huang (2012) yapmış olduğu genetik algoritmalar ve destek vektör regresyonu kullanarak bir hibrid stok modeli konulu çalışmada genetik algoritmaları (GAs) ve destek vektör regresyonu (SVR) kullanarak etkili bir stok seçim metodu geliştirmeyi amaçlamıştır. Araştırmacı öncelikli olarak gerçek stok gelirleri/dönüşleri yerine geçecek olan bir vekil oluşturmak için SVR metodunu kullanmıştır. Böylelikle bir portfolyodan en üst sıradaki stoklar seçilebilmiştir. Bu modelin üzerine model parametrelerinin optimizasyonu ve SVR modeline girdi değişkenlerinin optimal altkümelerine özellik seçimi için GA kullanılmıştır.
Lin, 2012	GREY	İnşaat projeleri için yeni bir dinamik ilerleme tahmin yaklaşımı konulu çalışmalarında “Grey Dinamik Tahmin Modeli” ve “Rezidüel Modifiye Model”i birleştirerek dinamik bir tahmin modeli oluşturmuşlardır. Yapmış oldukları deneysel çalışmalar neticesinde oluşturdukları bu yeni modelin inşaat sürecinin tahminde daha önceki klasik tahmin yöntemlerine göre daha başarılı bulunmuştur.

Yazar	Yöntem	Konu
Göndör, vd., 2012	REPTree, M5P	Romanya'da uygulanan Maliye Politikalarının ,enflasyon,faiz oranları,kurumlar vergisi,döviz kurları değişkenleriyle,Sermaye Piyasasına olan etkileri araştırılmıştır. Çalışmada 2003-2010 dönemi verileri kullanılmıştır. Maliye Politikasının sermaye piyasasına etkisinin genel kanının aksine, önemli etkisinin olduğu,maliye politikasının sermaye piyasasındaki tutarsızlıklara karşı önlem almada başarılı bir araç olabileceği ve Romanya'daki belirsizliklerin sermaye piyasasını olumsuz etkilediği sonucuna ulaşılmıştır.
Jiang, 2012	Grey-DVR	Finansal zaman serileri tahmini için lokal destek vektöre regresyonunda Grey İlişkisel Derecelendirme üzerine yapmış oldukları çalışmada Lokal Grey Destek Vektör Makinesi (LG-SVR) ile Grey ilişkisel derecelendirme lokal Destek makinelerini birleştirerek finansal zaman serileri tahmini gerçekleştirmişlerdir. Çalışma neticesinde LG-SVR'nin tahmin doğruluğunu ve hızını artıracakları görülmüştür.
Marvuglia, 2012	RBF YSA	İtalya'daki evlerde bir saat sonraki elektrik tüketimini radyal tabanlı dayalı yapay sinir ağı kullanarak tahmin etmeye çalışmışlardır. Çalışmada aynı zamanda klima kullanımının elektrik tüketimi üzerindeki etkisi de araştırılmıştır. Elektrik tüketimi ile ilgili hava durumu verilerinin yanı sıra evde klima olup olmama durumu da bir girdi değişkeni olarak modele ilave edilmiştir.
Hicham, 2012	Bulanık Küme, YSA	Bulanık kümeleme ve Uyarlanmış Öğrenme Oranı (FCBPN) ile Geri-yayılım (BP) Yapay Sinir Ağları üzerine kurulu olan hibrid satış tahmin modeli öne sürmüşlerdir. Önerilen model üç aşamadan oluşmaktadır: (1) Trend etkisini göz önüne almak için Winters Üssel Düzeltme yöntemi (2) Bulanık C-Means kümeleme yöntemi kullanılarak, her normalize edilmiş veri kaydının kümelenme üyelik seviyeleri belirlenmesi (3) Her bir küme, eğitim verileri kayıtlarının küme üyeliği seviyesine uyarlanmış bir öğrenme oranı ile paralel Geri Yayılım ağlarına beslenecektir. Bulanık kümeleme yöntemiyle kümelerin daha büyük olduğu ve dolayısıyla önerilen tahmin modelinin endüstriyel tahmin için başarılı olabileceği sonucuna ulaşmışlardır.

Yazar	Yöntem	Konu
Zhu ve Wei, 2013	ARIMA-EKDVM	Zhu ve Wei (2013), Karbon fiyat tahmini için yeni bir hibrid method geliştirmek amacıyla yapmış oldukları çalışmada ARIMA ve En Küçük Kareler Destek Vektör Makineleri (LSSVM) metodlarını kullanmışlardır. Aynı zamanda LSSVM'nin optimal parametrelerini bulmak için Parçacık Sürü Optimizasyonu (PSO) kullanılmıştır ki bu sayede tahmin doğruluğunun geliştirilmesi amaçlanmıştır.. Yapılan çalışma neticesinde oluşturulan bu yeni hibrid modelin karbon fiyat tahmini için uygun olduğu görülmüştür.
Yu vd., 2013	DVR	Yu ve ark. (2013), gazete/dergi satışlarını tahmin etmek için destek vektör regresyonunu kullandıkları çalışmalarında daha önce kullanılan tahmin yöntemleri ile karşılaştırma yapmışlardır. Çalışma neticesinde destek vektör regresyonunun gazete/dergi satış tahmininde başarılı bir şekilde kullanılabileceği görülmüştür.
Kheirkhah vd., 2013	YSA-VZA	Kheirkhah ve ark. (2013) yapmış oldukları çalışmada elektrik talebini tespit etmek için YSA -Temel Bileşen Analizi - Veri Zarflama Analizi -Varyans Analizi (ANOVA) modelini kullanmışlardır. Çalışmada PCA, VZA ve YSA modellerinde girdi değişkenlerini seçmek için PCA yöntemi kullanılmıştır.
Kheirkhah vd., 2013	YSA-VZA	Kheirkhah ve ark. (2013) yapmış oldukları çalışmada elektrik talebini tespit etmek için YSA -Temel Bileşen Analizi - Veri Zarflama Analizi -Varyans Analizi (ANOVA) modelini kullanmışlardır. Çalışmada PCA, VZA ve YSA modellerinde girdi değişkenlerini seçmek için PCA yöntemi kullanılmıştır.
Lu, 2014	DVR	Bilgisayar ürünlerinin değişken seçim düzeni ve destek vektör regresyonuna dayalı olarak satış tahmini üzerine yapmış olduğu çalışmada değişken seçim düzeni ile destek vektör regresyonu birleştirilerek hibrit bir tahmin modeli oluşturulmuştur. Yapılan deneysel çalışmalar neticesinde bu yeni oluşturulan hibrit modelin diğerlerine göre daha başarılı olduğu görülmüştür.
Çelik vd., (2014),	YSA-DVM	YSA-DVM yöntemleriyle maden ocağından elde edilen sismik darbe verilerledeprem tahminine yönelik çalışma yapılmıştır.Deprem tahmininde YSA ile %83, DVR ile %91 oranında erken tespit saptanmıştır.

Yazar	Yöntem	Konu
Wang ve Du, 2014	DE-SVR	Wang ve Du (2014), Tayvan'da anakart üreticilerinden gönderi sistemi için DVR yöntemini, diferansiyel evrim yöntemi ile birlikte kullanmışlardır. Çalışmada ayrıca yeni bir hibrid parça küme optimizasyon (HPSO) algoritması sunulmuştur. DE algoritmali DVR modelinin diğer modellere göre daha başarılı sonuçlar verdiği görülmüştür.
Kaytez, 2015	YSA-DVM	Kaytez vd. (2015) yapmış oldukları çalışmada elektrik tüketimini tahmin etmek için regresyon analizi, sinir ağları ve en küçük kareler destek vektör makinelerini karşılaştırmışlardır. En küçük kareler destek vektör makinelerinin diğerlerine göre daha doğru sonuçlar verdiği tespit edilmiştir.
Azofra, 2015	M5P	İspanyada yenilenebilir enerji kaynaklarının(rüzgar ve güneş enerjisi), elektrik fiyatlarının toptan satış fiyatı etkisini belirlemek için açıklayıcı model geliştirilmiştir. Elektrik piyasasının fiyat tespitinde işlem karmaşıklığı nedeniyle yapay zeka teknikleri kullanılmıştır. Literatürdeki var olan çalışmaların tahmin edici modeller yerine açıklayıcı model geliştirilmiştir. M5P yöntemiyle modelde yer alan her özneliliğin son fiyatı ne kadar etkilediği tespit edilmeye çalışılmıştır.
Lin, Lei, 2016	M5P-HBDM	Çalışmada Otoyol trafik kazaları süresinin tahminine yönelik önceki çalışmalardan farklı olarak "olay-süre zamanı" modeli yerine tehlike tabanlı süre modeli ("HBDM") ile birlikte M5P yöntemi kullanılmıştır. Önerilen M5P-HBDM modelinde, M5P ağacının yapraklarını, doğrusal regresyon modelleri yerine HBDM'ler oluşturmaktadır. Uygulama iki farklı, otoyol kazası veri setinde gerçekleştirilmiştir. M5P-HBDM yönteminin M5P veya HBDM'lerden daha önemli ve anlamlı değişkenleri tanımladığı ve MAPE değerinin daha düşük olduğu sonucuna ulaşılmıştır.
Mustakim, 2016	DVR-YSA	Çalışmada Endonezya da önemli bir ekonomik değer olan palmye yağının üretimine yönelik tahmin çalışması Destek Vektör Regresyon ve Yapay Sinir Ağları yöntemleriyle gerçekleştirilmiştir. Tahmin modelleri 2005-2013 dönemi verileriyle oluşturulmuştur. DVR yönteminin YSA'YA göre daha başarılı olduğu sonucuna ulaşılmıştır

Yazar	Yöntem	Konu
Azadi vd., 2016	PCA-M5P, YSA	Depolama sahalarında üretilen sızıntı suyundaki Kimyasal Oksijen Talebi (KOİ) yükünü öngörmek için Yapay Sinir Ağı (ANN) ve Ana Bileşen Analizi-M5P (PCA-M5P) modellerinden oluşan iki uzman sistem geliştirilmiştir. Tahmin sonuçlarının istatistiksel analizi, her iki modelin PCA-M5P üzerinde ANN için hafif bir üstünlükle iyi tahmin yeteneğine sahip olduğunu gösterdi. Test verilerine dayanarak, ANN ve PCA-M5P modelleri için ortalama mutlak yüzdelik hatası sırasıyla % 4 ve % 12 olarak hesaplanmıştır.
Yang, 2016	RBF	Otomobil satışlarını tahmin çalışmasında Mevsimsel Endeks ve RBF sinir ağı temelli kombinasyon tahmini modeli önerilmiştir. Önerilen modelde MAPE ve RMSE kriterlerine göre diğer yöntemlerledaha başarılı bulunmuştur.
Megahed, 2016	YSA- ARIMA	Dergi satışlarına yönelik yapılan çalışmada satışlara ait zaman serileri ve dergi içeriğinin popülerliği ile ilgili veriler kullanılmıştır. Tahmin modeli Geri Yayılım Ağlar ve ARIMA birlikte kullanılarak oluşturulmuştur. Önerilen tahmin modelinin, geleneksel satış tahmin tekniklerine göre daha başarılı olduğu sonucuna ulaşılmıştır.
Vhatkar, 2016	YSA	Kısa raf ömrüne sahip, hızlı tüketim ürünleri olan ağız bakım ürünlerinin (Jel Esaslı-380 mg. Beyazlatıcı diş macunu Jel-500gm) satış tahmini Geriye Yayılım algoritmasıyla gerçekleştirmişlerdir. Geriye Yayılım YSA'nın bu tip ürün satış tahmininde başarılı bir şekilde uygulanabilirliği sonucuna ulaşılmıştır.
Akdağ, 2016	YSA, DVR, Üstel Düzeltilme, ARIMA	Diyarbakır kent merkezi su talebini etkileyen değişkenlerle ilgili 2003-2013 yıllarına ait aylık verilerle Yapay Sinir Ağları, Destek Vektör Regresyon, Üstel Düzeltilme ve ARIMA yöntemleriyle içme suyu talep tahmini yapılmıştır. Yapay Sinir Ağları'nın tüm performans ölçütlerinde diğer yöntemlere göre daha iyi tahmin ürettiği sonucuna ulaşılmıştır.

ÜÇÜNCÜ BÖLÜM

PASLANMAZ ÇELİK SEKTÖRÜNDE VERİ MADENCİLİĞİ YÖNTEMLERİYLE SATIŞ TAHMİNİ

3.1 Paslanmaz Çelik

Paslanmaz çelik, 1913 yılında İngiliz Metalurjist Harry Brearley tarafından, %12,8 krom ve % 24 karbon içeren bir çelik geliştirilmesiyle bulunmuştur.

Paslanmaz çelikler içeriğinde bulunan Krom, çeliği paslanmaz yapmaktadır. Paslanmaz çeliken az % 10 krom içeren, çelik alaşımlarıdır (İkiz, 2012). Paslanmaz çelik içeriğinde nikel, molibden ve mangan elementleri bulunmaktadır. Paslanmaz çeliklerin korozyon dayanımı PRE (Pitting Resistance Equilavent) Oyuklanma Direnciyle belirlenmektedir (Özer, 2009). Paslanmaz çelik türleri:

Ostenitik: Diğer paslanmaz çelik türlerine göre daha yüksek Nikel içerdiğinden dolayı yaygın olarak kullanılmaktadır. Ostenitikler 200 ve 300 grup numaralı paslanmaz çeliklerdir. Yapı inşaat sektöründe yaygın olarak kullanılmaktadır.

Ferritik : İçeriğinde yüksek Krom, daha düşük Karbon içermesinden dolayı yumuşak çeliklere benzetilebilirler, Ostenitiklere göre paslanmaya karşı direnci düşüktür. Otomotiv sektöründe, mutfak gereçleri üretiminde, dekoratif uygulamalarda yaygın olarak kullanılmaktadır.

Dupleks: Östenitik ve ferritik özellikleri içermesinden dolayı dupleks paslanmaz çelik olarak anılırlar. Yüksek oranda krom % 18-28 ve orta miktarda Nikel (% 4,5 -8) içeren çeliklerdir. Dupleks çelik çoğunlukla ilaç sanayi, gemi yapımı, petrokimya sektöründe sanayide kullanılır. (Kaluç, 2007). Dupleks çelikler 4 grupta incelenmektedir (Sönmez, 2016).

- %23 Cr içerenler, Mo içermeyenler (PRE≈25)
- %22 Cr ve Mo içerenler (PRE≈30-36)
- %25 Cr ve 0-%2,5 Cu içerenler (PRE≈32-40)
- Süper dupleks paslanmaz çelikler (PRE>40)

Martensitik : Martensitikler % 0,1 karbon ve % 11-13 krom içermektedirler. Kolay işlenebilen, yüksek olmayan korozyon dayanımına sahip paslanmaz çelik türüdür. Genelde ameliyat aletleri, türbin kanatları gibi çeşitli ürünlerin üretilmesinde kullanılmaktadır.

3.2 Paslanmaz Çelik Üretim Aşamaları

Paslanmaz Çelik, hadde düzeneği ile inceltirilerek üretilmektedir. Haddeleme işlemi sıcak ve soğuk olmak üzere iki türde yapılmaktadır. Daha sonra asit banyosu ve tavlama işlemi gibi çeşitli aşamalardan sonra üretim tamamlanmaktadır (www.worldstainless.org).

Sıcak haddeleme de önce yassı kütükler 1250°C sıcaklıkta ısıtılmaktadır. Daha sonra kaba haddeleme ile malzeme kalınlığı yaklaşık 25mm'ye kadar inceltilmektedir. Sıcak haddelenmiş paslanmaz çelik görsel açıdan önem arz etmeyen tank, depolama ve makine parçası imalatında kullanılmakta ve tavlama, asit banyosu ve soğuk haddeleme yapılmadığı için düşük üretim maliyetine sahip olmaktadır. Sıcak haddelenmiş yarı mamul daha çok talep gördüğü görsel açıdan önemli alanlarda (Metal sektörü, Yapı-İnşaat vb) kullanılması için soğuk haddeleme ünitesine aktarılır.

Soğuk haddeleme sürecinde paslanmaz çelik, %80 oranında inceltilmesi sağlanmaktadır. Soğuk haddelenme ardından tavlama, asit banyosu işlemleri ve son haddeleme işlemleri yapılmaktadır.

Çeliğin yumuşatılması için tavlama ısı işlemi ardından homojenliğin sağlanması ve yüzey temizliği için asit banyosu işlemleri gerçekleştirilmektedir. Asit banyosuyla malzeme yüzeyi temizlenmektedir.

3.3 Dünyada ve Türkiye’de Paslanmaz Çelik Sektörü

İlk paslanmaz çelik 1. Dünya Savaşında kullanılmasına rağmen, diğer malzemelere göre aynı anda yüksek, yangın ve korozyon direncine sahip olması, uzun kullanım süresi ve düşük bakım maliyetine, geri dönüşüm süreci kolay bir malzemedir. Dayanıklı Tüketim ve sermaye malları, tıbbi malzemeler, ayrıca bina, demiryolları, metro, tünel ve köprüler gibi kamu yatırımlarında yoğun olarak kullanılmaktadır. Bu özellikler, Paslanmaz çelik kullanımında büyüme oranının diğer malzemelerden yüksek olmasını açıklayabilir. Uluslararası Paslanmaz Çelik Forumu (ISSF) verilerine göre paslanmaz Çelik talebi 1980-2015 döneminde ortalama %5,2 oranında artmıştır. Bu oran seçilmiş diğer ürünlere olan talep ve diğer ürünlerin talep ortalamasından oldukça yüksek çıkmıştır (Tablo 3.1).

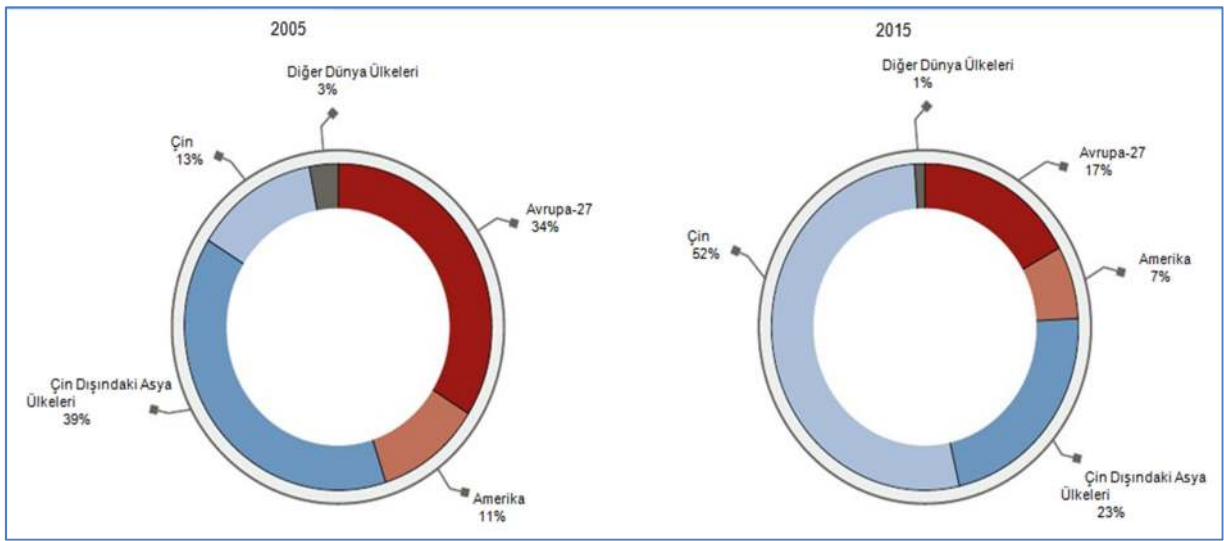
Tablo 3.1 Paslanmaz Çelik ve Bazı Ürünlerin Talebinde Yıllık Büyüme Miktarı

Ürün Adı	Büyüme oranı %
Kurşun	1,78
Bakır	2,74
Çinko	2,38
Alüminyum	3,86
Karbon Çelik	2,24
Paslanmaz Çelik	5,27
Ortalama	3,045

Kaynak: ISSF, 2015

ISSF tarafından yayımlanan verilere göre, 2012 yılında dünya paslanmaz ham çelik üretimi 2011 yılına göre %5,2 oranında artarak 35,4 milyon ton seviyesine ulaşmıştır. 2002-2012 yılları arasında toplam üretim 15 milyon ton artarak %70'lik bir artış göstermiştir.

Şekil 3.1 de yerde sol tarafta yer alan 2005 yılı verilerine göre dünya paslanmaz çelik üretimi bölgesel olarak ele alındığında Asya Ülkeleri (Çin Haricindeki) %39'luk, Avrupa Birliği ülkeleri %34'lük, Çin ise %13'lük üretim payına sahiptir. 2015 yılı verilerine göre ise, Çin %52'lik bir üretim gerçekleştirerek en büyük üreticisi konumuna geçmiştir. 2015 yılı itibariyle paslanmaz çelik üretiminin %75'lik bölümü Çin ve diğer Asya ülkelerinde gerçekleşmiştir.



Şekil 3.1 Ülke/Bölge Bazlı Dünya Paslanmaz Çelik Üretimi

Kaynak: ISSF, 2015

Tablo 3.2'de ISSF verilerine göre 2009-2015 döneminde bölgelere göre sıcak haddelenmiş paslanmaz çelik üretimi yer almaktadır. Tablo 3.2'de sırasıyla Avrupa Birliği ülkeleri ve üretim toplamı, Amerika ülkeleri ve üretim toplamı, Doğu Asya Ülkeleri ve üretim toplamı yer almaktadır.

Tablo 3.2 Bölgelere Göre Sıcak Haddelenmiş Paslanmaz Çelik Üretimi

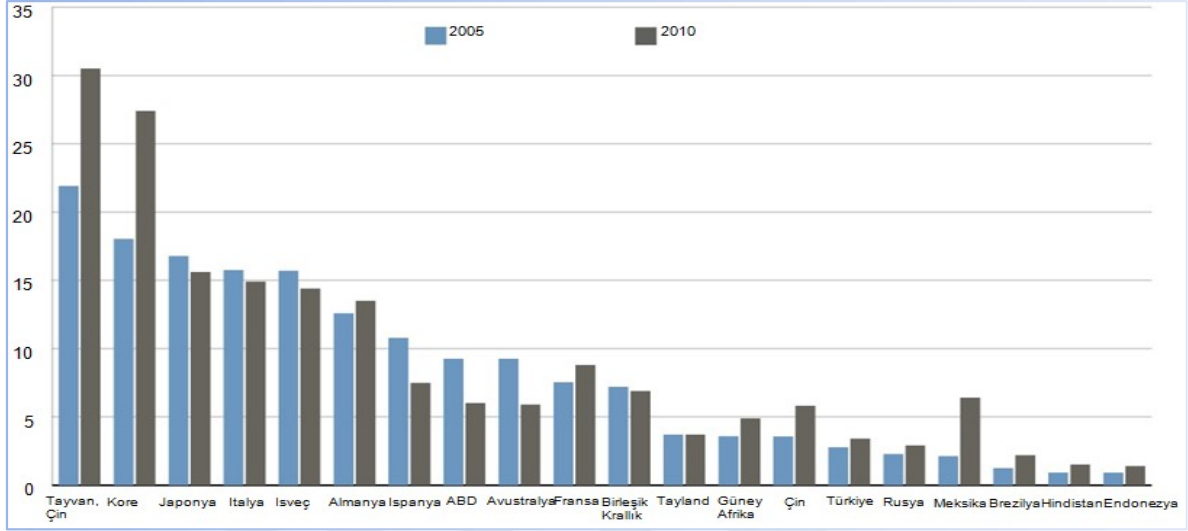
Ülke/bölge	2009	2010	2011	2012	2013	2014	2015
Belçika	1,045	1,306	1,241	1,241	1,298	1,388	
Belçika/Avusturya							1,607
Finlandiya	726	998	1,003	1,078	1,080	1,216	
İsveç	455	546	586	510	501	541	-
Birleşik krallık	224	279	330	294	257	295	-
Finlandiya/İsveç/Birleşik krallık							2,215
Fransa	202	276	300	285	300	323	291
Almanya	1.320	1.509	1.502	1.313	1.091	864	459
İtalya	1,216	1,583	1,602	1,696	1,556	1,457	1,452
İspanya	693	844	807	844	855	945	979
Diger Avrupa Ülkeleri	144	152	188	195	211	223	165
Avrupa Birliği Toplamı	5,986	7,494	7,559	7,455	7,147	7,252	7,169
ABD	1,617	2,201	2,074	1,977	2,030	2,389	2,346
Brezilya	324	409	413	391	425	424	401
Amerika Kıtası Toplamı	1,942	2,609	2,486	2,368	2,454	2,813	2,747
Japonya	2,607	3,427	3,247	3,166	3,175	3,328	3,061
Güney Kore	1,677	2,048	2,157	2,167	2,143	2,038	2,231
Tayvan , Çin	1,468	1,514	1,203	1,107	1,067	1,108	1,109
Çin	8,805	11,256	14,091	16,087	18,984	21,692	21,562
Hindistan	1,721	2,022	2,163	2,834	2,891	2858	3,060
Asya Toplamı	16,277	20,267	22,861	25,361	28,260	31,025	31,024
Güney Afrika	546	480	443	503	492	472	514
Rusya	86	122	125	112	152	123	65
Ukrayna	67	118	147	118	n/a	n/a	n/a
Dünya	24,904	31,090	33,621	35,917	38,506	41,686	41,546

Kaynak: ISSF, 2015

Tablo 3.2'de Yıllar itibari ile bazı Avrupa ülkeleri için üretim rakamları diğer Avrupa ülkeleri üretim rakamlarıyla toplanarak, birlikte gösterilmiştir. Ukrayna'daki siyasi gelişmelerden dolayı 2013 yılından beri üretim rakamı yer almamaktadır. Sıcak haddelenmiş üretimde en fazla üretim Çin başta olmak üzere Asya ülkelerinde gerçekleşmektedir. Paslanmaz çelik üretiminde, Çin 2002 yılında 1,1 milyon ton üretirken, hızlı bir büyüme göstererek 2015 yılında ise 21,56 milyon ton üretimle, paslanmaz çelik pazarının büyümesine destek olmuştur.

Sıcak haddelenmiş paslanmaz çelik üretimi bölge bazlı incelediğinde Asya'da %90, Amerika kıtasında %41, Avrupa'da %19,76 oranında artmıştır. Dünya Geneline üretim ise 24,9 milyon tondan 41,54 milyon tona çıkarak %66 oranında artmıştır. Tüm zamanlarda üretimin düzeyini en fazla artıran ülke Çin'dir. Çindeki 2009-2015 yılları üretim düzeyi %144 oranında artmıştır.

Kişi başına düşen paslanmaz çelik ve paslanmaz çelik içeren ürünleri kullanımında, en yüksek miktarda tüketim Güneydoğu Asya ülkelerinde gerçekleşmektedir. Bu ülkelerde yüksek miktarda paslanmaz çelik üretildiğinden dolayı, tüketiciler paslanmaz çelik ürünlerine ulaşmaları diğer ülkelere göre avantajlı konumdadırlar.

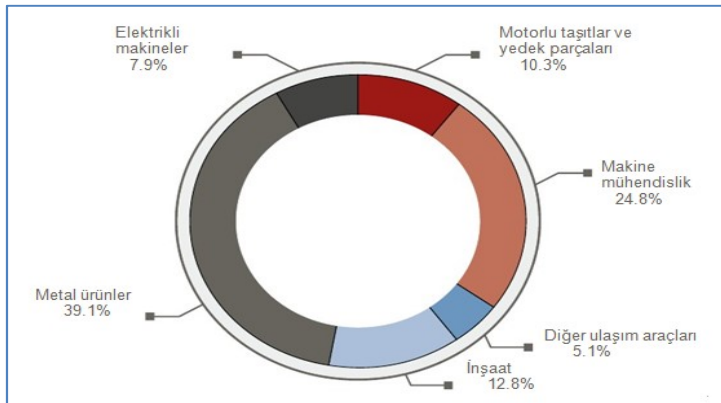


Şekil 3.2 Bazı Ülkeler İçin Kişi Başına Düşen Paslanmaz Çelik Kullanımı

Kaynak: ISSF, 2015

ISSF verilerine göre kişi başına en fazla paslanmaz çelik kullanımı Tayvan,Çin ve Güney Kore'de gerçekleşmektedir (Şekil 3.2). Bu ülkeler aynı zamanda 2006-2010 yılları döneminde paslanmaz çelik kullanımını önemli ölçüde artırmışlardır. Anılan dönemlerde ülkemizde kişi başına paslanmaz çelik tüketimi yaklaşık %10 artmıştır. Paslanmaz çelik talebi en fazla artan ülke ise Meksika olmuştur.

ISSF verilerine göre Dünya geneli, 2013 yılına ait sektörlere göre paslanmaz çelik talebine ait grafik Şekil 3.3'te yer almaktadır.

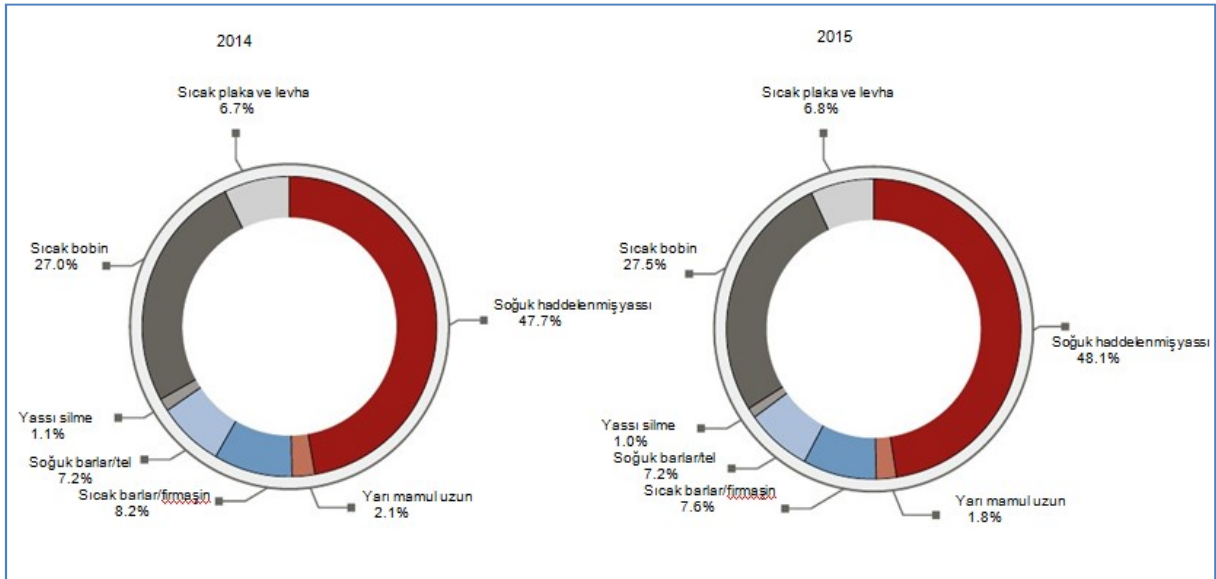


Şekil 3.3 Dünya Geneli Sektörlere Göre Paslanmaz Çelik Talebi

Kaynak: ISSF, 2015

Talep miktarındaki ilk üç sıralama, %39,1 Metal ürünler, %24 Makine, Yapı-İnşaat sektörü %12,8, sektörlerinde gerçekleşmiştir.

Şekil 3.4'te 2014-2015 yılları, Paslanmaz çelik soğuk ve sıcak (haddelenmiş, yassı, plaka, levha) türleri üretim miktarları değişimini gösteren grafik aşağıda yer almaktadır.



Şekil 3.4 Mamul Türlerine Göre Paslanmaz Çelik Üretimi

Kaynak: ISSF, 2015

2014 yılında Soğuk haddelenmiş yassı ürün, toplam üretimin %47,7 sini oluştururken 2015 yılında %48,1'e yükselmiştir. Anılan yıllarda Sıcak bobin üretimi ise %27'den, %27,5'e yükselmiştir. Soğuk haddelenmiş ve sıcak bobin dışında kalan mamüllerde üretim düzeyinde Şekil 1.3'te incelendiğinde önemli bir değişiklik olmadığı söylenebilir.

Tablo 3.3'te 2010-2012 yılları arasında paslanmaz çelik ihracatı yapan ilk 5 ülke yer almaktadır. Çin ve Güney Kore ihracatlarını iki dönemde artırmıştır. Diğer ülkelerde ihracatlarında azalma olduğu görülmektedir.

Tablo 3.3 Paslanmaz Çelik İhracatı (İlk 5 Ülke)

Ülke	2010	% Değişim	2011	% Değişim	2012
Tayvan	712,156	1,42	722,253	-9,54	653,345
Çin	462,628	28,96	596,601	4,41	622,898
Güney Kore	458,084	4,01	476,442	0,58	479,227
Avrupa	564,073	-13,31	488,992	-3,04	474,144
Japonya	416,448	-3,27	402,846	-8,88	367,082

*HR ve CR sheet >600mm büyük, <3mm kalın

Kaynak: ISSF, 2012

Tablo 3.4’de 2010-2012 yılları arasında en çok paslanmaz çelik ithalatı yapan ilk 5 ülke yer almaktadır. Türkiye’de 2010-2012 yılları arasında paslanmaz ithalatı 236,29 mt. tondan 257,77 mt. ton rakamına yükselerek %9’luk artış göstermiştir. Avrupa ülkelerinde ithalat 2011 yılında bir önceki yıla göre %13’lük artış göstermesine rağmen 2012 yılında %12,9 luk bir azalmayla 499,07 mt. ton seviyesine gerilemiştir. Çin’de ise paslanmaz çelik ithalatı aynı dönem için %24 oranında azaltarak 298,46 mt ton seviyesine gerilemiştir.

Tablo 3.4 Paslanmaz Çelik İthalatı (İlk 5 Ülke)

Ülke	2010	Değişim%	2011	Değişim%	2012
Avrupa	503,28	13,86	573,01	-12,90	499,07
Çin	391,55	-16,20	328,11	-9,04	298,46
Tayvan	250,22	10,64	276,84	1,97	282,29
USA	239,34	-1,82	234,98	10,47	259,57
Türkiye	236,29	5,31	248,83	3,59	257,77

*HR ve CR sheet >600mm büyük, <3mm kalın

Kaynak: ISSF, 2013

Ülkemizin paslanmaz çelik sektöründe en çok ithalat yapan ülkelere olması, 2016 yılı itibari ile ülkemiz paslanmaz çelik sektöründe faaliyet gösteren 1 yerli üretici firmanın bulunmasından dolayı değişmemiştir, paslanmaz çelik sektöründe dışa bağımlı olduğumuzu ifade edebiliriz (Uçkun, 2010).

3.4 Araştırmanın Amacı ve Kapsamı

Bu araştırmanın amacı, veri madenciliği yöntemleriyle, ülkemizde paslanmaz çelik sektöründe faaliyet gösteren, toptan satış yapan distribütör bir firmanın toplam satışının, döviz kuru, nikel,ham çelik, krom madeni,üretici fiyat endeksi,sanayi ciro endeksi gibi bağımsız değişkenlerle birlikte ileriye yönelik satış tahminin hiyerarşik zaman serisi ve veri madenciliği yöntemleriyle gerçekleştirilmesidir. Model ağaç yöntemiyle sektörlere etki eden makro ekonomik değişkenlerin elde edilmesi, değişkenlerin kritik değerlerine göre geçmiş, günümüz ve geleceğe ait tahmin edici modeller oluşturulmuştur. Hiyerarşik zaman serisi ve veri madenciliği yöntemleriyle tahmin edici modeller elde edilirken en iyi sonuçları üreten modelleri bulmak amacıyla modeller arasında karşılaştırma yapılmıştır. Diğer bir ifadeyle Hiyerarşik zaman serisi tahmininde, tümevarım ve optimal birleşim modelleri arasında, veri madenciliği yöntemlerinde DVR, GSR, ÇKA yöntemleri arasında karşılaştırma yapılmıştır.

Bu bağlamda firmanın veri tabanından 2008-Ocak-2016 Mart dönemindeki günlük satış verileri elde edilmiştir. Müşteri bilgilerinden yola çıkılarak firmanın satış yaptığı alt

sektörler belirlenmiştir. Müşteri bilgileri ve satış tarihi, sektör adı ve satış tutarı bilgilerini içeren yeni bir veri seti oluşturulmuştur. Daha sonra aşağıda yer alan hedeflere uygun analizler yapılmıştır.

- Model Ağaçlar Yöntemiyle, toplam satışın ve sektörel satışların etkilendiği, bağımsız değişkenin kritik değerlerine göre elde edilen tahmin edici doğrusal modellerin oluşturulması.
- Hiyerarşik Kümeleme analizi ile sektör davranışların incelenmesi ve hangi sektörlerin satış hareketlerinin benzeştiğinin tespit edilmesi.
- Kümeleme analizi sonucunda oluşturulan hiyerarşik yapıya ait ileriye dönük satış tahminleri, birleşim yöntemleri tümevarım ve optimal birleşim olmak üzere ARIMA, Üstel Düzgünleştirme yöntemleriyle gerçekleştirilmesi.
- Toplam satış ve sektörel satış verilerinin kendi geçmiş değerleri kullanılarak, veri madenciliği yöntemleriyle (Destek Vektör Regresyon, Yapay Sinir Ağları, Gauss Süreç Regresyon) ile ileriye dönük satış tahmini yapılması.

3.5 Uygulama Platformu

Uygulama çalışması, sektör bilgilerini içeren satış hareketleri ve diğer değişkenlerin elde edilmesi ve değişkenlerin özelliklerine göre oluşturulan aylık modellerle, talep tahminlerinin gerçekleştirilmesidir. Bu başlık altında tahmin modelinde yer alacak değişkenlerin belirlenmesi, elde edilen değişkenler ile yeni veritabanının oluşturulması ve kullanılan Veri Madenciliği Yazılımları ele alınmıştır.

3.5.1 Veri Madenciliği Yazılımı

Günümüzde az sayıda tahmin yazılımı sistemi, bilimsel son gelişmeleri takip edebilmektedir (Kusters vd., 2006: 612). WEKA yazılımı ve R programlama dili tahmin yazılımlarının ötesinde, yeni algoritmaların geliştirilmesine imkan sunması, Genel Kamu Lisansına sahip olması nedeniyle veri madenciliği alanında yoğun olarak kullanılmaktadır.

3.5.1.1 WEKA Yazılımı

WEKA, makine öğrenimi algoritmalarını içeren, Waikato Üniversitesindeki (Yeni Zelanda) bilim insanları tarafından geliştirilmiştir. Genel Kamu Lisansına sahip, açık kaynak kodlu bir veri madenciliği programıdır (Frank vd., 2011). WEKA'da grafik arayüzü veya Java kodları yazarak, veri madenciliği uygulamalar gerçekleştirilebilmektedir. WEKA kendine özgü ARFF (Attribute Relationship File Format) metin tipi dosya biçimini dışındaki dosya tiplerinide (csv, odt, xls) desteklemektedir. ARFF'in tanımlama satırlarında ilk satırda ilişki

(relation) bildirilmekte diğer satırlardaysa veri kümesindeki öz nitelik (attributes) adı ve türü bildirilmektedir. Veriler ise tanımlama sırasına uygun bir şekilde, her satır bir örnek (instance) olarak bildirilmektedir.

3.5.1.2 R Yazılımı

R yazılımı, 1976 yılında geliştirilen S programlama dilinin devamı niteliğinde olan R programlama dili ile uygulama geliştirilmesini sağlamaktadır. Auckland Üniversitesi bilim insanları Ross Ihaka ve Robert Gentleman tarafından geliştirilen açık kaynak kodlu yazılım, istatistik, ekonometri, yöneylem ve veri madenciliği kütüphanelerini içermektedir. Dünyanın çeşitli yerlerindeki araştırmacıların bir araya gelerek oluşturduğu “R core team” R yazılımını kısa sayılabilecek dönemlerde güncellemektedir (Özdemir, 2010: 10-12).

R yazılımı yaygın olarak kullanılan gibi istatistik paket programların ötesinde yazılım geliştirme ortamıdır. Veri madenciliği açısından incelendiğinde veri işleme ve saklama özelliklerinin yanı sıra dizi ve matris hesaplamalarında kullanılan operatörler bulunmaktadır. Veri madenciliği haricinde istatistiki ve ekonometrik analizler, yöneylem uygulamaları gerçekleştirilebilmektedir (Racineve Hyndman, 2002)

3.6 Veri

Firmaya ait veritabanından 01.01.2008-31.03.2016 yılları arasındaki Satış Hareketler tablosu ve Müşteriler tablolarından alınan veriler eşleştirilerek, satış yapılan firma düzeyinde günlük satış verileri elde edilmiştir. Paslanmaz çelik bileşenleri olan nikel, ham demir, krom madenlerine ait veriler ise infomine.com web adresindeki tarihsel veri seçeneği kullanılarak elde edilen grafiklerin sayısallaştırılması ile günlük fiyatları elde edilmiştir. Infomine Web Sitesinde, paslanmaz çelik piyasasına ait verilere, Londra Metal Borsası verilerine, diğer emtia fiyatlarına, döviz verilerine erişim imkanı bulunmaktadır. Emtia fiyatları metrik ton/USD üzerinden elde edilmiştir. Firma Amerikan Dolar Kuru üzerinden işlem yaptığı için Dolar_Kuru değişkeni olarak Satış tahminine etki edebilecek diğer değişkenler ise döviz kuru, enflasyon oranı, faiz oranı Türkiye Cumhuriyet Merkez Bankası (TCMB) Elektronik Veri Dağıtım Sistemi olan EVDS’den elde edilmiştir.

Tablo 3.5 Veri Seti

Sıra No	Kısaltma	Açıklama	Kaynak
1.	Satis_Tar	Satış Tarihi	Firma
2.	Sektor_ID	Sektör Kodu	Firma
3.	Satis_Tutar	Firmanın günlük satış tutarı	Firma
4.	Dolar_Kuru	Aylık Dolar Kuru ortalama değeri	T.C.M.B
5.	Nikel	Aylık Nikel madeni ortalama değeri	www.infomine.com
6.	Demir	Aylık ham Demir madeni ortalama değeri	www.infomine.com
7.	Krom	Aylık Krom madeni ortalama değeri	www.infomine.com
8.	KKO	Aylık Ana metal sanayi Kapasite Kullanım oranı	T.C.M.B
9.	Faiz	T.C.M.B faiz oranı	T.C.M.B
10.	Üfe	Üretici fiyat endeksi	T.C.M.B
11.	Sanurin	Sanayi üretim endeksi	T.C.M.B
12.	Doviz_Kuru	Ufe bazlı reel efektif döviz kuru	T.C.M.B
13.	San_Ciro	Sanayi ciro endeksi	T.C.M.B
14.	Elektrik	Yurt içi elektrik fiyatları	TÜİK
15.	Tuk_Guv_End.	Tüketici Güven Endeksi	TCMB

3.6.1 Veri Setinin Dönüştürülmesi

Firma veri tabanından yüzbinden fazla kayıt bulunan satış hareketleri tablosu ve müşteri tablosundan elde edilen verilerle, 2 farklı türde veri seti oluşturulmuştur. Birinci veri setinde müşteri ticari unvanı, günlük satış tutarı ve tarihi bilgileri veri seti oluşturulmuştur.

İkinci veri seti, günlük satış verilerine sektör bilgisinin eklenmesi ile oluşturulmuştur. Firma uzmanlarıyla birlikte 1'i diğer olmak üzere 5 sektör kodu belirlenmiştir. Bu sektörler Tablo 3.6'da yer almaktadır.

Tablo 3.6 Sektör Kodları

No	Sektör Adı
1.	Makine
2.	Metal
3.	Endüstriyel Mutfak
4.	Yapı-İnşaat
5.	Diğer

Firma veritabanındaki satış hareketlerindeki müşteri kodu (MusteriID) ve müşteri tablosundaki müşteri kodu (MusteriID) eşleştirilerek SQL kodu ile günlük satış tutarı, tarihi, sektör kodu (SektorID) bilgilerini içeren veri seti oluşturulmuştur. Böylece kümeleme analizi ve kümelerin satış tahminini gerçekleştirilmesi için altyapı hazırlanmıştır.

```
SELECT Ciro.ciroID, Giro.Tutar, Musteri.SektorID, Giro.Tarih
FROM Ciro
INNER JOIN Musteri
```

ON ciro.MusteriID=Musteri.MusteriID;

Aylık satış verileri toplamı için aşağıdaki SQL cümlesi yazılmıştır. (Satış verileri aylık olarak gruplandırılarak satış verilerinin toplamı ve diğer değişkenlerin ilgili aydaki ortalama fiyatı tespit edilmiştir.)

```
SELECT      ciro.ay_no,      Sum(ciro.Satis_Tutar)      AS      aylık_toplam,
Round(Avg(ciro.Dolar_Kuru),4) AS Dolar_Kur_Ort, Round(Avg(ciro.Nikel),2) AS Nikel_Ort,
Round(Avg(ciro.Demir),4) AS Demir_Ort, Round(Avg(ciro.Krom),2) AS Krom_Ort FROM
ciro GROUP BY ciro.ay_no;
```

Aylık yayınlanan makroekonomik değişkenler aylık tabloya eklenmiştir. Elde edilen veriler csv formatıyla kaydedilmiştir. Daha sonra weka yazılımına uygun dosya formatı olan “arff” biçimine, ilişki (relation) ve nitelikler (attribute) satırları eklenerek dönüştürülerek, veri ön işleme aşaması tamamlanmıştır. Veri setinde eksik bilgi bulunmamaktadır. Düzenleme ilgi satırlar Tablo 3.7’de yer almaktadır.

Tablo 3.7 ARFF Formatı Dönüştürme

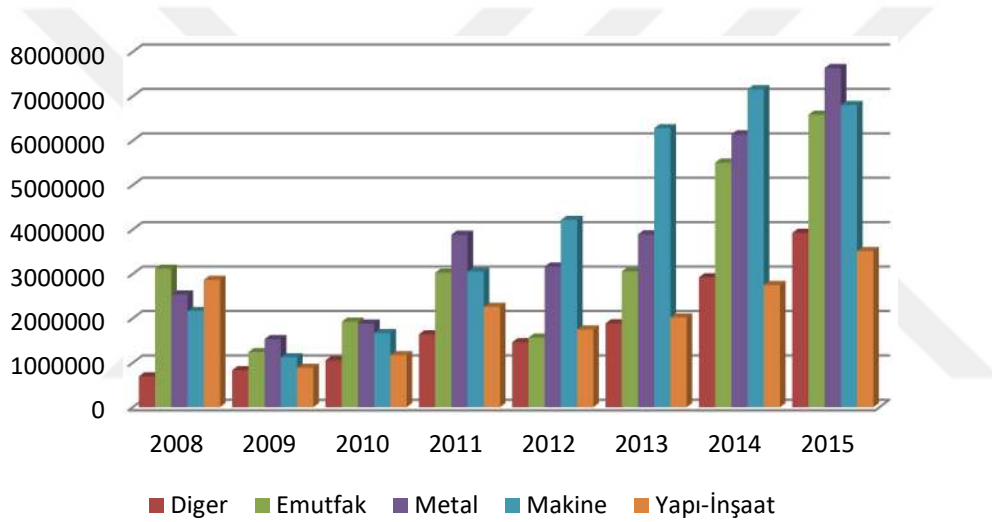
```
@relation Tutar //Tutar niteliğinin ilişkileri
@attribute Date date 'dd-MM-yyyy' // Zaman bilgisi (96 ay)
@attribute Tutar numeric // Aylık toplam satış tutarı niteliği sayısal olarak tanımlanmıştır.
@attribute Dolar_Kur_Ort numeric// Dolar Kuru aylık ortalama niteliği sayısal olarak
tanımlanmıştır.
@attribute Nikel_Ort numeric// Nikel aylık ortalama değeri, niteliği sayısal olarak
tanımlanmıştır.
@attribute Demir_Ort numeric// Demir aylık ortalama değeri, niteliği sayısal olarak
tanımlanmıştır.
@attribute Krom_Ort numeric// Krom aylık ortalama değeri, niteliği sayısal olarak
tanımlanmıştır.
@attribute KKO numeric// Kapasite Kullanım Oranı niteliği sayısal olarak tanımlanmıştır.
@attribute ufe numeric// ufe niteliği sayısal olarak tanımlanmıştır.
@attribute faiz numeric// Faiz niteliği sayısal olarak tanımlanmıştır.
@data // Veriler için bildirim satırı
"1-1-2008",1165892.25,1.1706,27633.46,60.8,4600,79.8,146.94,15.5 // aylık veriler
```

3.6.2 Tanımlayıcı Bulgular

Firma Ana Metal Sanayi kolunda faaliyet yürütmektedir. Dolayısıyla satış yapılan sektörler, ağırlıklı olarak Metal,Makine sektörleri ayrıca paslanmaz çeliğin yoğun olarak

kullanıldığı Endüstriyel Mutfak ve Yapı-İnşaat sektörlerini kapsamaktadır. Toplam satışa ve önde gelen sektörlerle göre oldukça az ağırlıklı (%1-%2) olan sektörler "diğer sektör" olarak gruplandırılmıştır.

2008-2016 döneminde firmanın sektörlerle göre satış grafiği Şekil 3.5'te yer almaktadır. 2008 yılında yaşanan ekonomik kriz 2009-2010 yılları arasında satışların önemli derecede azalmasına neden olmuştur. 2011-2012 döneminde metal, makine ve endüstriyel mutfak sektör satışları artması toplam satışları artırmıştır. 2013-2014 döneminde makine sektör satışları toplam satışın artmasında katkı yapmıştır. 2015 yılında ise bir önceki yıla göre makine sektöründe kısmi düşüş yaşanmasına rağmen özellikle metal sektörü ve diğer sektörler de 2008-2016 döneminde tüm sektörler için en yüksek değerine ulaşmıştır.



Şekil 3.5 Sektörlere Göre Satışlar

Makine sektör satışları 2008-2015 döneminde toplam satıştaki yüzdesini artırmıştır. 2008 yılında toplam satışların %19'unu oluşturmuştur. 2012 yılında %36 oranıyla en yüksek değerini elde etmiştir. 2008-2015 yılları döneminde %27 oranıyla en çok satış yapılan sektör olmuştur.

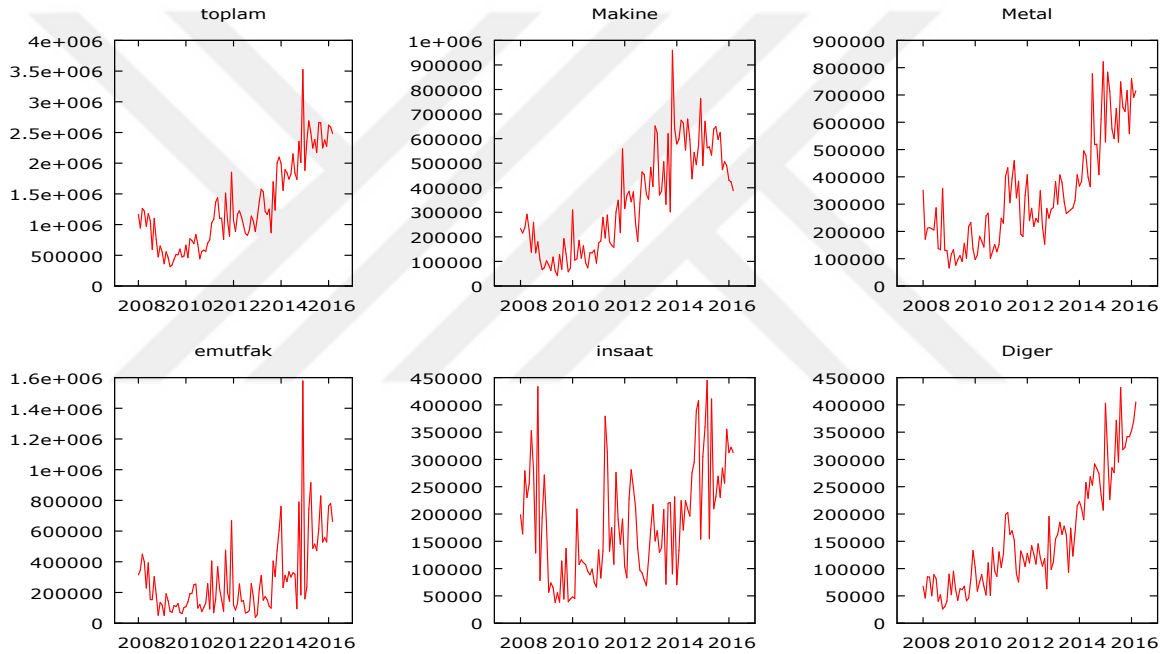
Metal sektör satışları 2008-2015 dönemi için toplam satışın %25'ini oluşturarak en fazla satış yapılan ikinci sektör olmuştur. 2009 ve 2015 yıllarında toplam satışın %27'sini oluşturarak en yüksek satış yapılan sektör olmuştur.

Endüstriyel Mutfak sektör satışları 2008-2015 dönemi için toplam satışın %22'ini oluşturarak en fazla satış yapılan üçüncü sektör olmuştur Endüstriyel Mutfak satışları 2008 yılında %28'lik satış oranıyla en çok satış yapılan sektör olmuştur. 2011 yılında en düşük satış oranı %13 ile gerçekleşmiştir.

Yapı ve inşaat sektör satışları 2008-2015 dönemi için toplam satışın %14'ünü oluşturarak en fazla satış yapılan dördüncü sektör olmuştur Yapı ve İnşaat sektör satışları 2008 yılında %25'lik satış oranıyla en çok satış yapılan ikinci sektör olmuştur. 2013-2014 yıllarında en düşük satış oranı %11 ile gerçekleşmiştir.

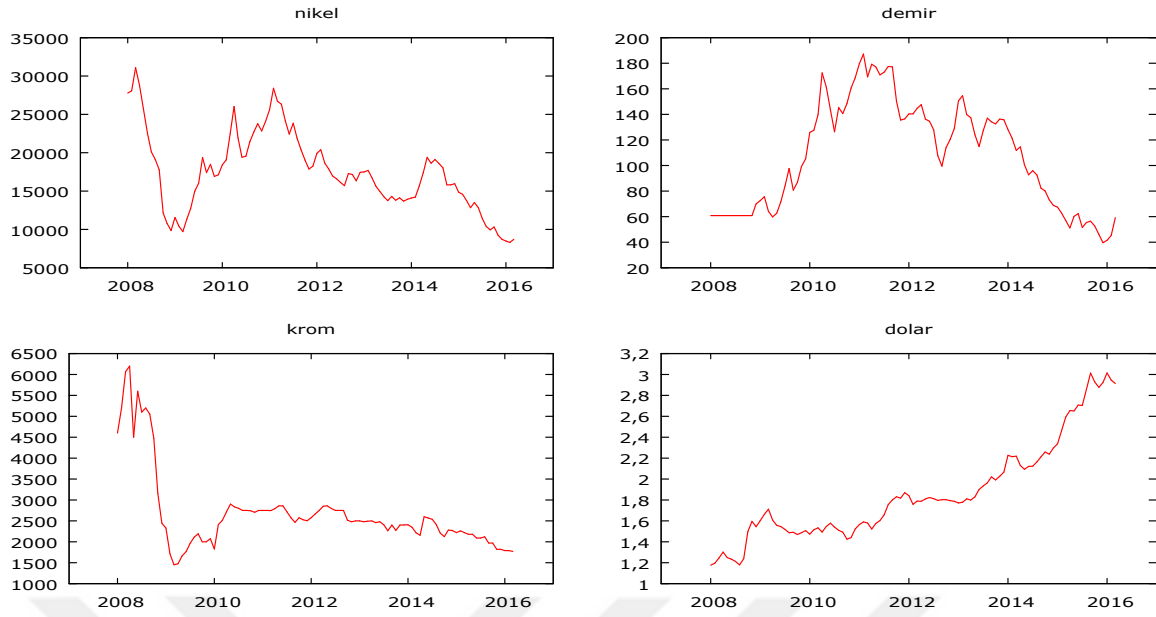
En çok satış yapılan ilk 4 sektörün ardından çeşitli sektörlerin oluşturduğu "Diğer Sektör" satışları 2008 yılında %6 oranıyla en düşük satış performansını göstermiştir. 2009-2011 döneminde satış oranını en fazla artıran sektör olmuştur. sonraki dönemlerde diğer sektör satışları %11-%14 oranında gerçekleşmiştir.

Toplam Satış ve firmanın satış yaptığı Metal, Makine, Endüstriyel Mutfak ve Yapı İnşaat sektörlerine ait zaman serileri Şekil 3.6'da yer almaktadır.



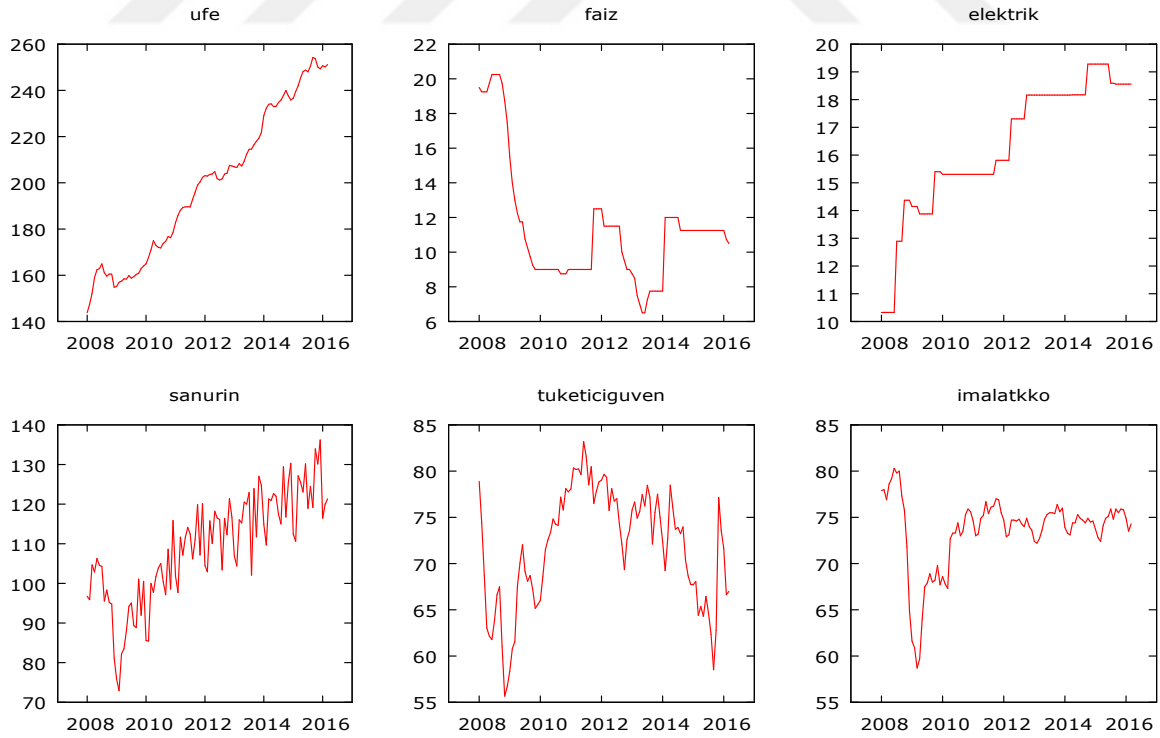
Şekil 3.6 Sektörler ve Toplam Satış Tutarına Ait Zaman Serileri

ISSF verilerine göre küresel krizin etkisiyle paslanmaz çelik üretiminde yavaşlama ve Nikel fiyatlarında 2009 yılında önemli bir düşüş yaşanmıştır. 2010-2012 döneminde artış yaşanmasına rağmen fiyatlar 2014 yılından beri düşüş eğilimindedir. Demir fiyatları 2009 yılından sonra yatay seyrini değiştirmiş ve fiyatlarda artış gözlenmiştir. 2013 yılından itibaren fiyatlar düşüş eğilimindedir. Krom fiyatlarında 2008-2009 yıllarında önemli bir düşüş görülmektedir. 2010 yılından sonra kısmi bir artış yaşanmasına rağmen 2014 yılına kadar yatay seyretilmiş daha sonra düşme görülmüştür. Paslanmaz çeliğin hammaddesi olan nikel, demir, ferrokrom maden fiyatları ve dolar kuruna ait zaman serisi Şekil 3.7'de yer almaktadır.



Şekil 3.7 Aylık Emtia ve Dolar Kurunu Ait Zaman Serisi (01/2008-03/2016)

Amerikan Doları, 2008-2009 döneminde, Türk Lirası karşısında 1.10 seviyelerinden 1.8 seviyesine ulaşmıştır. 2009 yılı son çeyreğinden 2011 yılı son çeyreğine kadar yatay seyrete de ileriki dönemlerde artmıştır. 2016 yılında 3 TL olarak en yüksek değeri elde etmiştir.



Şekil 3.8 Diğer Bağımsız Değişkenlere Ait Zaman Serileri

Paslanmaz Çelik talebini etkileyen emtia fiyatları ve döviz kurunun dışında kalan değişkenler olarak ÜFE, Faiz, Elektrik Fiyatları, Sanayi Üretim Endeksi, Tüketici Güven

Endeksi ve İmalat Sanayi Kapasite Kullanım oranları gibi belirlenmiştir (Şekil 3.8). 01.2008-03.2016 döneminde ÜFE ve elektrik fiyatları artış göstermektedir. Faiz oranlarında 2008 yılından itibaren azalma eğilimindedir. 2011-2012 yıllarında artmasına rağmen 2013 yılında en alt seviyeyi görmüştür. 2014 yılında artış yaşanmasına rağmen daha sonraki dönemlerde yatay seyretmiştir.

3.7 Model Ağaçlar Yöntemiyle Tahmin Modelleri

3.7.1 Toplam Satış Tahmini

Toplam Satış için M5P yöntemiyle 5 regresyon denklemi elde edilmiştir. SDR değeri en yüksek olan Dolar kuru en önemli değişken olarak kök düğümü oluşturmuştur. Dolar kur değeri 2,006 altında iken elde edilen 1. regresyon denklemi, Dolar Kuru 2,006 ve 2,248 değerleri arasında 2. regresyon denklemi elde edilmiştir. Dolar Kur 2. 2,248 değeri büyük reel efektif döviz kurunun 101,69 değerinden küçük aynı zamanda, ÜFE değerinin 250,14 değerinden küçük olması durumunda 3. regresyon, büyük olması durumunda 4. regresyon denklemi elde edilmiştir. Dolar Kur değerinin 2,248 değerinden büyük, reel efektif döviz kur değerinin 101,695 değerinden büyük olması durumunda 5. regresyon denklemi elde edilmiştir. Reel efektif döviz kurunun denklemde yer almasının sebebi firmanın satış yaptığı sektörlerde kullanılan dövizin kurunun etkisinin incelenmesidir. Reel Efektif döviz kurunun denklemde yer almasıyla korelasyon katsayısı iyileşmiş ve model sayısı yükselmiştir.

```
dolar <= 2,006 : LM1 (69/%29,46)
dolar > 2,006 :
| dolar <= 2,248 : LM2 (13/%22,93)
| dolar > 2,248 :
| | dovizkuru <= 101,695 :
| | | ufe <= 250,145 : LM3 (7/%13,74)
| | | ufe > 250,145 : LM4 (6/%13,94)
| | dovizkuru > 101,695 : LM5 (4/%43,011)
```

Toplam Satış tahmininde M5p ağacına ait elde edilen doğrusal modeller aşağıdadır.

1 .Doğrusal Model

toplam = -621,14 * demir - 1203733,04 * dolar + 12614,45 * ufe + 10300,43 * sanurin-
25397,94 * dovizkuru + 4194,29* imalatkkko + 1968970,27

2 . Doğrusal Model

toplam = -4599,92 * demir + 447619,13 * dolar - 7457,36 * ufe + 3377,06 * sanurin +
7829,359 * imalatkkko + 2151039,01

3 . Doğrusal Model

$$\text{toplam} = -4169,86 * \text{demir} + 410962,28 * \text{dolar} - 14019,07 * \text{ufe} + 995,61 * \text{sanurin} - 19961,19 * \text{dovizkuru} + 7829,35 * \text{imalatkko} + 6186854,25$$

4 . Doğrusal Model

$$\text{toplam} = -4169,86 * \text{demir} + 410962,28 * \text{dolar} - 13863,42 * \text{ufe} + 1929,83 * \text{sanurin} - 19961,19 * \text{dovizkuru} + 7829,35 * \text{imalatkko} + 6171666,75$$

5 . Doğrusal Model

$$\text{toplam} = -4169,86 * \text{demir} + 410962,28 * \text{dolar} - 22539,75 * \text{ufe} + 7767,54 * \text{sanurin} - 29416,50 * \text{dovizkuru} + 7829,35 * \text{imalatkko} + 8559571,97$$

3.7.2 Makine Sektörü Tahmini

Makine Sektörü için M5P yöntemiyle 5 regresyon denklemi elde edilmiştir. SDR değeri en yüksek olan UFE değişken değerine göre yapılandırılmıştır. Oluşturulan ağaç yapısı aşağıdaki gibi ifade edilebilir. Güncel değerler en yakın olan 5. modelde Üretici Fiyat ve Sanayi Üretim Endekleri, Reel Efektif Döviz Kuru yer almaktadır.

```

ufe <= 202,08 :
| sanurin <= 101,805 :
| | dovizkuru <= 107,12 : LM1 (12/%12,21)
| | dovizkuru > 107,12 : LM2 (16/%29,22)
| sanurin > 101,805 :
| | dovizkuru <= 112,27 : LM3 (17/%28,11)
| | dovizkuru > 112,27 : LM4 (5/%13,06)
ufe > 202,08 : LM5 (49/%54,99)

```

Toplam Satış tahmininde M5p ağacına ait elde edilen doğrusal modeller aşağıdadır.

1 .Doğrusal Model

$$\text{Makine} = 921,06 * \text{ufe} + 901,53 * \text{sanurin} + 2158,82 * \text{dovizkuru} + 14,08 * \text{sanayiciro} - 351701,13$$

2 . Doğrusal Model

$$\text{Makine} = 921,06 * \text{ufe} + 901,53 * \text{sanurin} + 1880,26 * \text{dovizkuru} + 82,91 * \text{sanayiciro} - 309307,13$$

3 . Doğrusal Model

$$\text{Makine} = -178,09 * \text{ufe} - 692,70 * \text{sanurin} + 2215,01 * \text{sanayiciro} + 71205,76$$

4 . Doğrusal Model

$$\text{Makine} = -837,59 * \text{ufe} - 1649,24 * \text{sanurin} + 3162,26 * \text{sanayiciro} + 165210,70$$

5. Doğrusal Model

$$\text{Makine} = 935,45 * \text{ufe} + 5861,67 * \text{sanurin} - 6046,67 * \text{dovizkuru} + 211255,95$$

3.7.3 Metal Sektörü Tahmini

Metal Sektörü Satışları, Makine Sektörüne benzer şekilde ÜFE değişkeni önem arz etmektedir. Sektör için M5P yöntemiyle 2 regresyon denklemi elde edilmiştir. İlk regresyon denklemi ÜFE değeri 233,53 değerinden küçük olduğu durumu ifade etmektedir. İkinci regresyon denklemi ÜFE değerinin 233,535 değerinden büyük olduğu durumu ifade etmektedir. Elde edilen modeller:

1. Doğrusal Model

$$\text{metal} = 54553,81 * \text{dolar} + 1951,79 * \text{ufe} + 7641,95 * \text{imalatkko} - 767304,55$$

2. Doğrusal Model

$$\text{metal} = 130642,04 * \text{dolar} + 4618,23 * \text{imalatkko} - 72806,90$$

3.7.4 Endüstriyel Mutfak Sektörü Tahmini

Endüstriyel Mutfak Sektörü Satışlarında Sanayi Ciro Endeksi değişkenine göre 2 regresyon denklemi elde edilmiştir. Birinci regresyon denklemi Sanayi Ciro Endeksinin 156,75 değerinden küçük eşit olduğu durumu, ikinci regresyon denklemi ise Sanayi Ciro Endeksinin 156,75 değerinden büyük olduğu durumu ifade etmektedir. Birinci doğrusal modelden farklı olarak ikinci doğrusal modelde demir madeni fiyatı yer almaktadır.

1. Doğrusal Model

$$\text{Emutfak} = 97165,27 * \text{dolar} - 1526,56 * \text{ufe} + 771,67 * \text{sanayiciro} + 10210,25 * \text{imalatkko} - 529037,60$$

2. Doğrusal Model

$$\text{Emutfak} = -4146,19 * \text{demir} + 550734,74 * \text{dolar} - 20340,18 * \text{ufe} + 1543,35 * \text{sanayiciro} + 5147,91 * \text{imalatkko} + 3713568,53$$

3.7.5 Yapı-İnşaat Sektörü Tahmini

İnşaat Sektörü Satışlarında karar ağacı en önemli değişken SDR değeri en yüksek olan ÜFE değişkenidir. Ağaç, ÜFE ve Dolar Kuru değişkenlerine göre yapılandırılmıştır. ÜFE değerinin 235,28 değerinden küçük olduğu ve dolar kurunun oranının 2,895 değerinden küçük eşit olması durumunda 1. regresyon denklemi , büyük olması durumunda 2. regresyon denklemi elde edilmiştir. 3. regresyon denklemi dolar kur değerinin 2,895 değerinden büyük

olduğu durumlarda geçerlidir. Günümüze en yakın olan 3. modelde, önceki modelde yer alan Demir Madeni yerine Nikel Maden fiyatı yer almış ayrıca Sanayi Ciro Endeksi ve Kapasite Kullanım oranı yer almamıştır.

Ağaç yapısı aşağıdaki gibi ifade edilebilir.

```

ufe <= 235,28 : LM1 (79/%49,879)
ufe > 235,285 :
| dolar <= 2,895 : LM2 (14/%77,09)
| dolar > 2,895 : LM3 (6/%46,39)

```

Üretilen 3 regresyon denklemi aşağıda yer almaktadır.

1. Doğrusal Model:

$$\text{insa} = 950,59 * \text{demir} + 4,4606 * \text{krom} - 267353,38 * \text{dolar} + 19625,21 * \text{faiz} + 289,38 * \text{sanurin} + 515,50 * \text{dovizkuru} + 3787,47 * \text{sanayiciro} - 9394,97 * \text{imalatkko} + 396155,35$$

2 . Doğrusal Model

$$\text{insa} = -3,69 * \text{nikel} + 11,97 * \text{krom} + 77095,80 * \text{dolar} + 4187,52 * \text{faiz} + 77,2105 * \text{sanurin} + 1384,48 * \text{dovizkuru} - 142099,97$$

3. Doğrusal Model

$$\text{insa} = -5,10 * \text{nikel} + 11,97 * \text{krom} - 264215,03 * \text{dolar} + 4187,52 * \text{faiz} + 777,21 * \text{sanurin} + 1384,48 * \text{dovizkuru} + 913705,08$$

3.7.6 Diğer Sektör Tahmini

Diğer Sektör tahmin modelinde, Dolar değişkeni değerine göre ağaç yapılandırılmıştır. Dolar değişkeninin 2,04 değerinden küçük eşit olması durumunda birinci regresyon denklemi tersi durumunda ise ikinci regresyon denklemi elde edilmiştir. Regresyon denklemlerinde Dolar ve Sanayi Üretim Endeksi değişkenleri kullanılmıştır.

1. Doğrusal Model

$$\text{Diger} = 72492,40 * \text{dolar} + 1952,47 * \text{sanurin} - 215992,77$$

2. Doğrusal Model

$$\text{Diger} = 145546,65 * \text{dolar} + 508,04 * \text{sanurin} - 138700,25$$

3.7.7 Modellerin Değerlendirilmesi

M5p yöntemiyle tüm sektörler için oluşturulan tahmin modellerine göre sonuçlar Tablo 3.8'de yer almaktadır. Korelasyon katsayısı Toplam Satış ve Diğer Sektör satış tahminlerinde 0,92, Metal Sektörü Satış tahmininde 0,89, Makine Sektörü Satış tahmininde 0,84, Yapı-İnşaat sektörü satış tahmininde 0,64, Endüstriyel Mutfak Sektörü satış tahmininde 0,54 olarak hesaplanmıştır.

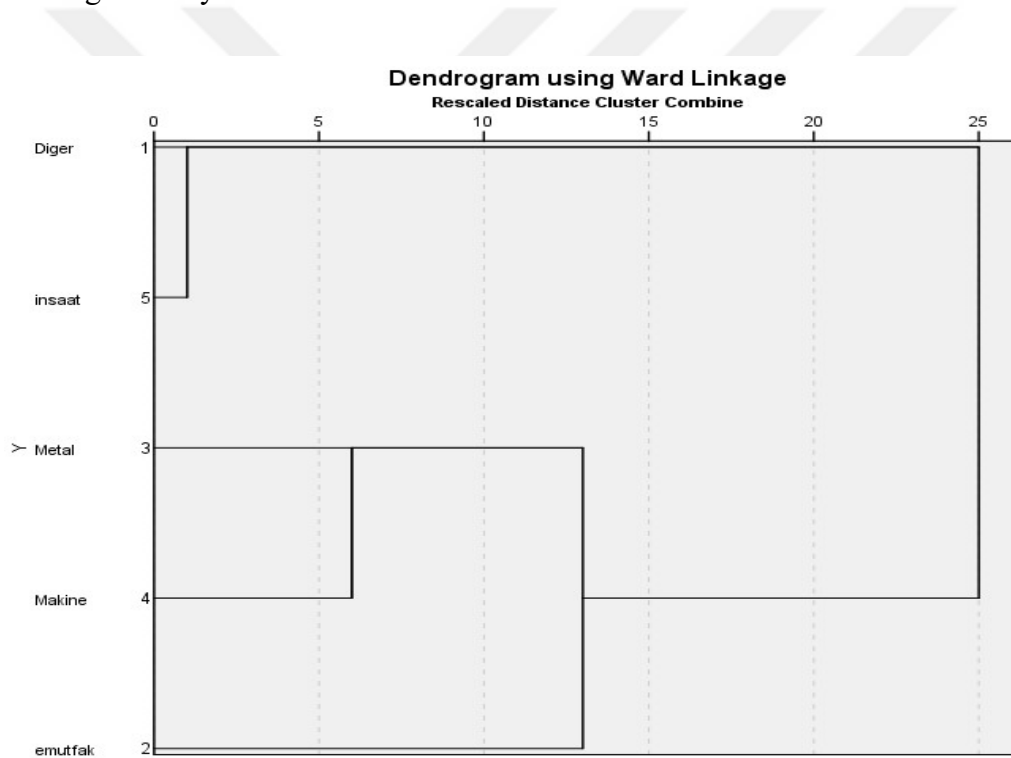
Tablo 3.8 M5p Yöntemi Model Sonuçları

Ölçüt	Toplam	Diğer Sektör	Metal	Makine	Yapı İnşaat	End. Mutfak
r	0,92	0,92	0,89	0,84	0,6843	0,54
MAE	204341,51	29638,13	65989,36	84381,66	66560,24	143970,77
RMSE	273951,70	38923,37	86922,19	111762,98	86625,12	211669,53

Korelasyon katsayısı (r) dikkate alındığında Toplam, Diğer, Metal ve Makine sektörlerindeki modellerin başarılı olduğu söylenebilir.

3.8 Kümeleme Analizi

Satış yapılan sektörler için oluşturulan zaman serileri için yapılan hiyerarşik kümeleme analizinde Wards yöntemi ve Öklid Mesafesi ölçüsü kullanılmıştır. Analiz sonucu aşağıdaki Dendrogram'da yer almaktadır.

**Şekil 3.9 Kümeleme Analizine Ait Dendrogram**

Dendrogramı incelendiğimizde sektörlerin iki kümede toplandığını görmekteyiz. İlk kümede Yapı-İnşaat ve Diğer sektör yer almaktadır. İkinci kümede Metal, Makine ve Endüstriyel Mutfak Sektörleri yer almaktadır.

3.9 Zaman Serileri Satış Tahmin Modelinin Oluşturulması

3.9.1 Durağanlık Testleri

Çalışmada kullanılan veri madenciliği yöntemlerinde, ayrıca bir analize gerek kalmadan, çekirdek yapısı sayesinde durağan olmayan verileri durağanlaştırılmaktadır.

Hiyerarşik zaman serisi tahmininde R yazılımında durağanlık testi Kwiatkowski, Phillips, Schmidt ve Shin tarafından önerilen KPSS birim kök testi ile yapılmakta, ARIMA modeline ait p ve q parametreleri ise Hyndman-Khandakar algoritması ile belirlenmektedir. Hyndman-Khandakar algoritması aşağıdaki gibi ifade edilebilir.

1- KPSS testi ile durağan olmayan serilerin durağanlaştırılması. (d parametresinin belirlenmesi)

1- p ve q değerleri, AICc'yi en aza indirgeyerek seçilir. Algoritmada p ve q parametrelerinin her olası kombinasyonunu yerine algoritma, kademeli bir arama kullanmaktadır.

a. En iyi model (en küçük AICc değerine sahip) aşağıdaki dört model arasından seçilmektedir.

ARIMA(2,d,2),

ARIMA(0,d,0),

ARIMA(1,d,0),

ARIMA(0,d,1).

Eğer $d = 0$ ise, sabit c dahil edilir; Eğer $d \geq 1$ ise, sabit c sıfır olarak ayarlanır. Böylece "geçerli model" bulunur.

b. Mevcut modeldeki değişiklikler göz önüne alınır:

Mevcut modelden p ve / veya q'yu ± 1 artırılır/azaltılır.

c'yi geçerli modele eklenir / çıkartılır.

Bu adımda en iyi model (mevcut model ya da bu varyasyonlardan biri) yeni geçerli model olarak belirlenir.

c. Daha düşük bir AICc bulunana kadar Adım 2 (b) 'yi tekrarlanır.

Durağanlık testi için kullanılan bir diğer yöntem ise Genişletilmiş Dickey Fuller (ADF) testidir. ADF testi için sıfır hipotez, verilerin durağan olmamasıdır.

Genişletilmiş Dickey Fuller testi sonucuna göre test kritik değeri tüm seviyelerdeki değerden büyük olduğu için;

- Endüstriyel Mutfak, Yapı-İnşaat, Metal sektörleri durağandır.
- Krom maden fiyatı ve Sanayi Üretim Endeksi durağandır.
- Diğer Seriler durağan değildir.

Tablo 3.9 Tm Deęiřkenlere Ait ADF Sonuları

Deęiřken	ADF Test İstatistięi	10%	5%	1%	P-deęeri	Durum
Endstriyel Mutfak	-8,5604	-4,0543	-3,4563	-3,1539	0,0000	Duraęan
Yapı-İnřaat	-6,4635	-4,0543	-3,4563	-3,1539	0,0000	Duraęan
Metal	-6,0050	-4,0543	-3,4563	-3,1539	0,0000	Duraęan
Makine	-3,4484	-4,0554	-3,4568	-3,1542	0,0510	Duraęan Deęil
Dięer	0,0539	-3,4999	-2,8918	-2,5830	0,9605	Duraęan Deęil
Toplam Satıř	-3,3887	-4,0554	-3,4568	-3,1542	0,0589	Duraęan Deęil
Nikel	-1,8923	-4,0543	-3,4563	-3,1539	0,6510	Duraęan Deęil
Demir	-1,4593	-3,4991	-2,8915	-2,5828	0,5500	Duraęan Deęil
Krom	-5,0198	-4,0575	-3,4578	-3,1548	0,0004	Duraęan
Sanayi retim	-6,7602	-4,0543	-3,4563	-3,1539	0,0000	Duraęan
Dolar	-1,0853	-4,0543	-3,4563	-3,1539	0,9257	Duraęan Deęil
Elektrik	-2,7078	-4,0543	-3,4563	-3,1539	0,2359	Duraęan Deęil
Faiz	-1,7554	-4,0543	-3,4563	-3,1539	0,7187	Duraęan Deęil
KKO	-3,2004	-4,0554	-3,4568	-3,1542	0,0905	Duraęan Deęil

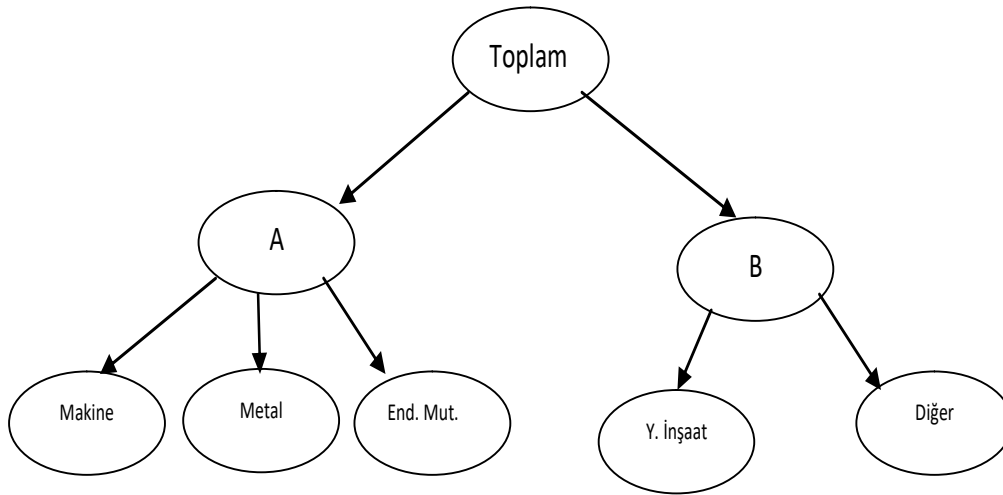
Tahmin modelleri aylık dnem řeklinde oluřturulmuřtur. Modellerin yapısı ;

- Hiyerarřık Kmeleme analizi sonucuna sektrlerin hiyerarřık biimde modellenmesi baz alınmaktadır.
- Hiyerarřık tahmin iin, Tme Varım ve En Uygun Tahmin Birleřimiyle btn dęmler iin ARIMA ve stel Dzgnleřtirme yntemleriyle tahmin yapılmıřtır.
- Firmanın toplam satıř verilerinin tek deęiřkenli ve baęımsız deęiřkenlerle ok deęiřkenli olarak modellenmesidir.

3.9.2 Hiyerarřık Zaman Serisi Tahmin Modeli

Hiyerarřık tahmin modelinde kmeleme analizi sonucu oluřan yapı kullanılmıřtır. Bu yapı ařaęıdaki řekilde aęa řeklinde ifade edilmiřtir. Analiz R programında gerekleřtirilmiřtir. Bu modelde Yntem tahmin srecinin ynn (Ařaęıdan-Yukarı, Ortadan Bařlayan, En Uygun Birleřimi) ifade etmektedir. Tahmin yntemi ise analizlerde kullanılan olan ARIMA, stel Dzgnleřtirme yntemlerini ifade etmektedir.

Hiyerarřık kmeleme analizi ile baz alınarak, satıř verileri hiyerarřık yapısı řekil 3.10'da yer almaktadır. İlk dzeyde (0. Seviye) Toplam satıř, sonraki dzeyde A ve B dęmleri yer almaktadır. A dęmn makine,metal ve endstriyel mutfak, B dęmn ise Yapı-İnřaat ve Dięer sektr satıř verileri oluřturmuřtur.



Şekil 3.10 Satış Verilerinin Hiyerarşik Modeli

R programında hts kütüphanesi hiyerarşik tahmin modelini gerçekleştirmemizi sağlamaktadır. Veri Setinin İlk 77 ayı Eğitim ve Son 12 ayı Test olmak üzere 2'ye bölünmüştür. Tahmin başarısı için accuracy.gts fonksiyonu kullanılmıştır. Bununla ilgili R kodu aşağıdadır.

```
library(hts)
```

```
data <- window(ekhts, start = c(2008,1), end = c(2015,2))
```

```
test <- window(ekhts, start = c(2015,3), end=c(2016,3))
```

Hiyerarşik tahmin süreci yöntemleri R dilinde ifadesi şunlardır.

```
method=<metod_adi>
```

```
method="bu" # Tüme Varım (Bottom-Up)
```

```
"comb" # Optimal Birleşim (Optimal combination)
```

Tüme varım ve Optimal Birleşim yöntemiyle oluşturulan hiyerarşik tahmin modellerinde ARIMA,Üstel Zaman Serileri, tahmin yöntemleri kullanılmıştır. Tahmin modeli,fmethod parametresiyle belirlenmektedir.

```
fmethod =< "ets">,<"arima">
```

Bu bilgiler ışığında tahmin modelimizi ve tahminin başarı ölçütlerini hesaplayan R kodunu aşağıdaki şekilde ifade edebiliriz. Tahmin modeli, forecast fonksiyonuyla eğitim setinde, h=12 aylık dönem için yöntem adı ve tahmin yöntemi adı yazılarak hesaplanır.

```
<tahminmodeli><- forecast(data, h=12,method = "yontem_adi",fmethod =  
"tahmin_yontemi_adi")
```

Bir önceki adımda oluşturulan <tahmin_modeli> nin, başarısı ise accuracy.gts fonksiyonuyla test verisinden hesaplanmaktadır.

```
<tahmin_basarisi><- accuracy.gts(<tahmin_modeli>, test)
```

3.9.2.1 Toplam Satış Tahmini

Toplam satışı en iyi tahmin eden model Optimal Birleşim Üstel Düzgünleştirme modelidir. MAPE değeri %6,25'tir. Diğer birleşim yöntemlerine göre sonuçlar Tablo 3.10'da yer almaktadır.

Tablo 3.10 Hiyerarşik Zaman Serisi Toplam Satış Tahmini

Yöntem	MAPE (%)	ME	RMSE	MAE
Optimal_ÜDY	6,25	-17323,7	174251,5	150350,6
Optimal_Arima	7,1	-85197,8	205003,3	168426,9
Tümevarım_ÜDY	7,5	82473,27	235822,7	186834,6
Tümevarım_Arima	10	-128874	296072,7	237089,7

3.9.2.2 A Düzümü (Makine-Metal ve Endüstriyel Mutfak Sektörleri Birleşimli) Tahmini

Hiyerarşinin birinci seviyesinin ilk düğümü olan ve Makine,Metal ve Endüstriyel Mutfak sektörlerinin toplam tahmini Tablo 3.11'de yer almaktadır.Optimal Birleşim Üstel Düzgünleştirme %8,93 MAPE değeri ile ilk sırada, Tüme Varım Arima modeli ise %14,32 MAPE değeri ile son sırada yer almaktadır.

Tablo 3.11 A Düzümü Tahmini (Makine, Metal ve Endüstriyel Mutfak Sektör Tahmini)

Yöntem	MAPE (%)	ME	RMSE	MAE
Optimal_ÜDY	8,93	-792,49	189860,3	159763,5
Optimal_Arima	10,39	-56833,5	216337,8	182224,1
Tümevarım_ÜDY	12,28	85404,07	256718,5	224041,5
Tümevarım_Arima	14,32	-96474,3	311074,9	250304,4

3.9.2.3 B Düzümü (Yapı-İnşaat ve Diğer Sektörler Birleşimli) Satış Tahmini

Hiyerarşinin 1. seviyesinin ikinci düğümü olan ve Yapı-İnşaat ve Diğer sektörlerin tahmini Tablo 3.12'de yer almaktadır. Tümevarım Arima %12,2 MAPE değeri ile ilk sırada, Optimal Birleşim Üstel Düzgünleştirme modeli ise %12,98 MAPE değeri ile son sırada yer almaktadır.

Tablo 3.12 B Düzümü Yapı-İnşaat ve Diğer Sektörler Birleşimli Tahminleri

Yöntem	MAPE (%)	ME	RMSE	MAE
Tümevarım_Arima	12,2	-32399,6	83993,79	67271,93
Tümevarım_ÜDY	12,38	-2930,8	82819,32	71576,29
Optimal_Arima	12,77	-28364,3	90129,61	70501,6
Optimal_ÜDY	12,98	-16531,2	86223,85	73701,23

3.9.2.4 Metal Sektörü Satış Tahmini

Metal Sektörünü ait tahmin sonuçları Tablo 3.13'de yer almaktadır. MAPE değerleri %11,7 ve %12,74 aralığında yer almaktadır. Optimal Birleşim ARIMA modeli en iyi MAPE değerine sahiptir.

Tablo 3.13 Metal Sektörü Birleşimli Tahmini

Yöntem	MAPE	ME	RMSE	MAE
Optimal_Arima	11,77	-30758,2	89069,61	70865,87
Tümevarım_Arima	12,1	-40426,9	88161,94	71885,14
Tümevarım_ÜDY	12,11	-40487,1	88189,57	71895,18
Optimal_ÜDY	12,74	-61491,4	96760,88	74014,57

3.9.2.5 Makine Sektörü Satış Tahmini

Makine Sektörünü ait tahmin sonuçları Tablo 3.14'de yer almaktadır. MAPE değerleri %19,14 ve %25,68 aralığında yer almaktadır. Tüme Varım Üstel Düzgünleştirme modeli en iyi MAPE değerine sahiptir.

Tablo 3.14 Makine Sektörü Tahmini

Yöntem	MAPE	ME	RMSE	MAE
Tümevarım_ÜDY	19,14	-64789	107094,2	88665,77
Optimal_ÜDY	20,75	-88780,8	117608,2	95307,07
Optimal_Arima	22,84	-98568,9	132476,7	103823,3
Tümevarım_Arima	25,68	-109467	147397,9	116696

3.9.2.6 Endüstriyel Mutfak Sektörü Satış Tahmini

Endüstriyel Mutfak Sektörüne ait tahmin sonuçları Tablo 3.15'de yer almaktadır. MAPE değerleri %22,92 ve %34,44 aralığında yer almaktadır. Optimal Birleşim ARIMA modeli en iyi MAPE değerine sahiptir.

Tablo 3.15 Endüstriyel Mutfak Sektörü Tahmini

Yöntem	MAPE	ME	RMSE	MAE
Optimal_Arima	22,92	72493,61	207392,2	156591,5
Optimal_ÜDY	26,42	149479,6	216093	181307,3
Tümevarım_Arima	27,52	53419,85	245075,1	182917,6
Tümevarım_ÜDY	34,44	190680,2	261025,3	228049,9

3.9.2.7 Yapı İnşaat Sektörü Satış Tahmini

Yapı İnşaat Sektörüne ait satış tahmin sonuçları Tablo 3.16'da yer almaktadır. MAPE değerleri %36,55 ve %40,66 aralığında yer almaktadır. Tüme Varım Üstel Düzgünleştirme modeli en iyi MAPE değerine sahiptir.

Tablo 3.16 Yapı İnşaat Sektörü Tahmini

Yöntem	MAPE	ME	RMSE	MAE
Tümevarım_ÜDY	37,32	-74624,5	99820,74	84378,88
Optimal_Arima	37,88	-73856,7	101699,7	84801,08
Tümevarım_Arima	37,93	-76621,5	101322,4	85785,29
Optimal_ÜDY	40,66	-83793,2	108084,6	92427,91

3.9.2.8 Diğer Sektör Satış Tahmini

Diğer Sektöre aitt satış tahmin sonuçları Tablo 3.17'de yer almaktadır. MAPE değerleri %12,27 ve %19,55 aralığında yer almaktadır. Tüme Varım ARIMA modeli en iyi MAPE değerine sahiptir.

Tablo 3.17 Diğer Sektör Tahmini

Yöntem	MAPE	ME	RMSE	MAE
Tümevarım_Arima	12,27	44221,99	60498,83	46059,42
Optimal_Arima	13,1	45492,32	62938,54	48859,2
Optimal_ÜDY	18,25	67261,96	80957,46	67261,96
Tümevarım_ÜDY	19,55	71693,67	84835,29	71693,67

3.10 Veri Madenciliği Yöntemleri Tahmin Modelleri

Veri madenciliği tahmin yöntemlerinden Destek Vektör Regresyon, Gauss Süreç Regresyon, Çok Katmanlı Algılayıcı yöntemleriyle için yapılan ön işlemler şunlardır:

- Ön işleme aşamasından geçmiş olan .arff biçimindeki veri seti WEKA programında açılır. Base Learner Configuration sekmesinden ilgili model algoritması seçilir.
- Periodic attributes: Aylık değer seçilir.
- Evulation : Değerlendirme sekmesinde 2 işlem yapılmaktadır.
 - İlki modelin eğitim ve test aşamasının belirlenmesidir. Uygulamamızda sırasıyla tüm modeller için eğitim aşaması %10 olarak seçilmiştir.
 - Model değerlendirme ve karşılaştırma için ölçütler seçilmiştir.
- Output: Model çıktı seçeneğinde 3 aylık dönem seçilir.

3.10.1 Toplam Satış İçin Tahmin Modelleri

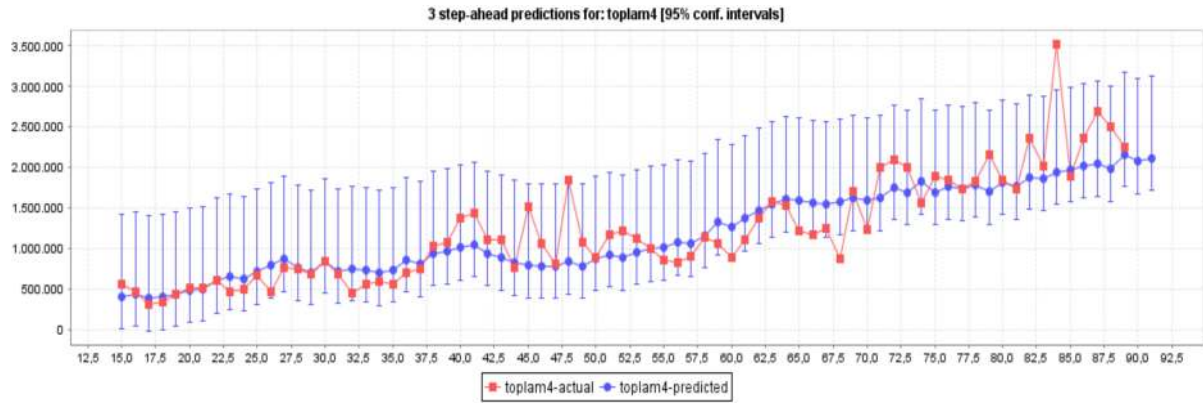
3.10.1.1 Destek Vektör Regresyon Tahmini

Destek Vektör Regresyon modeline ait eğitim aşaması sonuçları Tablo 3.18'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %16,84, %18,29 ve %18,59 olarak hesaplanmıştır. (c=1)

Tablo 3.18 Toplam Satış DVR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	205885,04	218527,74	221480,57
MAPE	16,84	18,29	18,59
RMSE	324168,73	332738,02	334408,27
MSE	105085368175,21	110714588762,01	111828893463,43

DVR eğitim aşamasına ait grafik Şekil 3.11'de yer almaktadır.

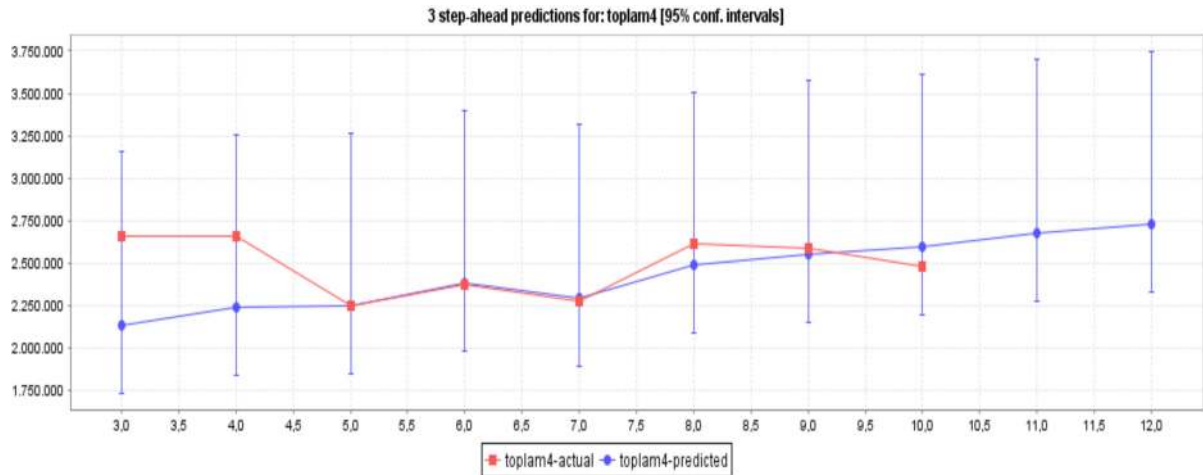
**Şekil 3.11 Toplam Satış DVR Eğitim Aşaması**

Destek Vektör Regresyon modeline ait test aşaması sonuçları Tablo 3.19'da yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %6,21, %5,58 ve %5,92 olarak hesaplanmıştır.

Tablo 3.19 Toplam Satış DVR Test Aşaması Tahmini

	1.ay	2.ay	3.ay
N	10	9	8
MAE	158845,08	145159,70	155570,18
MAPE	6,21	5,58	5,92
RMSE	237224,75	232413,61	245273,56
MSE	56275582630,25	54016086355,87	60159119749,91

Eğitim aşamasına ait grafik Şekil 3.12'de yer almaktadır. Eğitim sürecinde 5. aydan itibaren başarılı olduğu ayrıca 2. ve 3. dönem tahminlerinin başarılı olduğu söylenebilir.



Şekil 3.12 Toplam Satış Destek Vektör Regresyon Test Aşaması

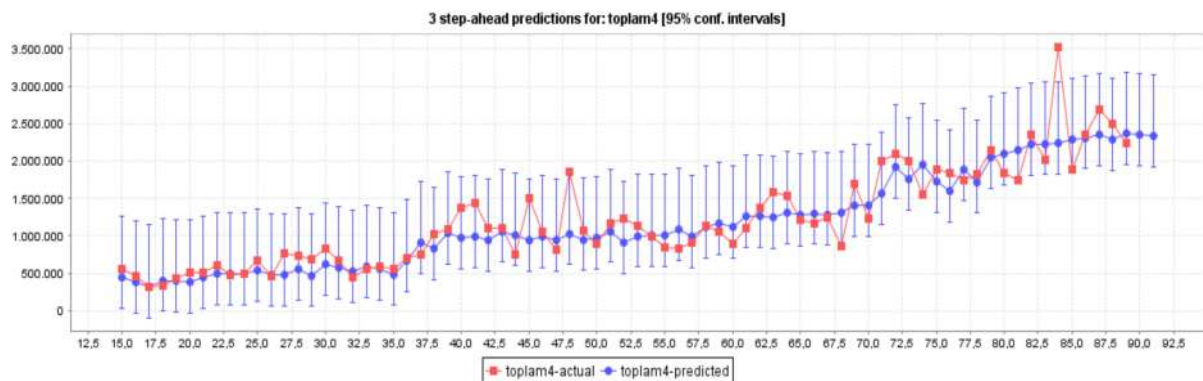
3.10.1.2 Çok Katmanlı Algılayıcı Tahmini

Çok Katmanlı Algılayıcı modeline ait eğitim aşaması sonuçları Tablo 3.20'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %14,32, %15,32 ve %15,55 olarak hesaplanmıştır. (LR=0.1,momentum=0.1)

Tablo 3.20 Toplam Satış ÇKA Eğitim Aşaması Tahmini

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	177022,62	186706,58	191244,58
MAPE	14,32	15,32	15,55
RMSE	261651,79	266967,30	271785,72
MSE	68461657378,80	71271540414,67	73867477450,48

Çok Katmanlı Algılayıcı modeline ait eğitim aşaması grafiği Şekil 3.13'de yer almaktadır. Tahmin süreci ilerledikçe MAPE değeri artmaktadır.



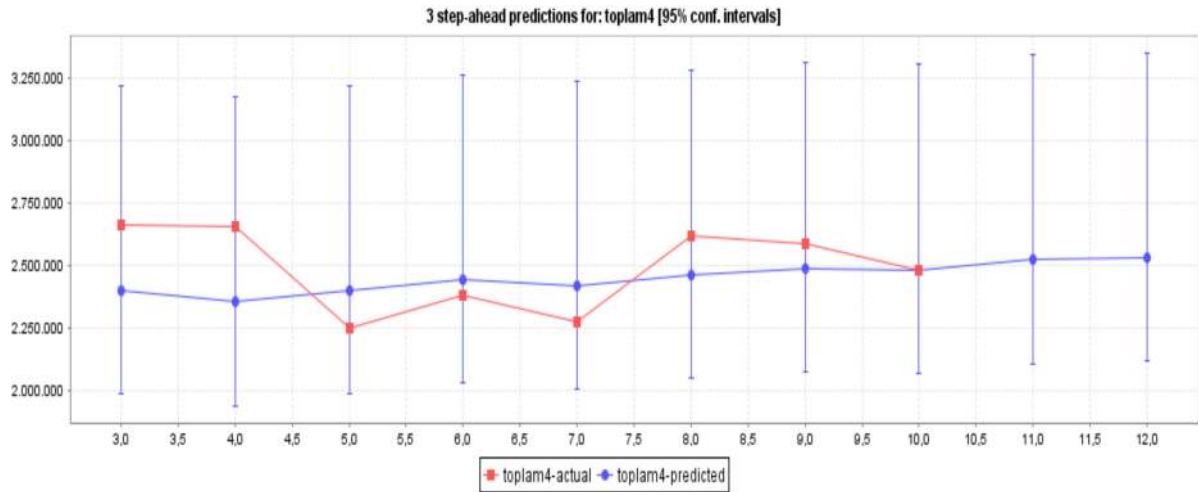
Şekil 3.13 Toplam Satış ÇKA Eğitim Aşaması

Çok Katmanlı Algılayıcı modeline ait test aşaması sonuçları Tablo 3.21'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %5,53, %5,99 ve %5,88 olarak hesaplanmıştır (LR=0.1,momentum=0.1).

Tablo 3.21 Toplam Satış ÇKA Test Aşaması

N	10,00	9,00	8,00
MAE	136785,90	148232,17	148263,15
MAPE	5,53	5,99	5,88
RMSE	163459,00	172275,69	174823,34
MSE	26718845451,94	29678914959,30	30563200711,33

Test aşamasına ait zaman grafiği Şekil 3.14 'de yer almaktadır.



Şekil 3.14 Toplam Satış ÇKA Test Aşaması

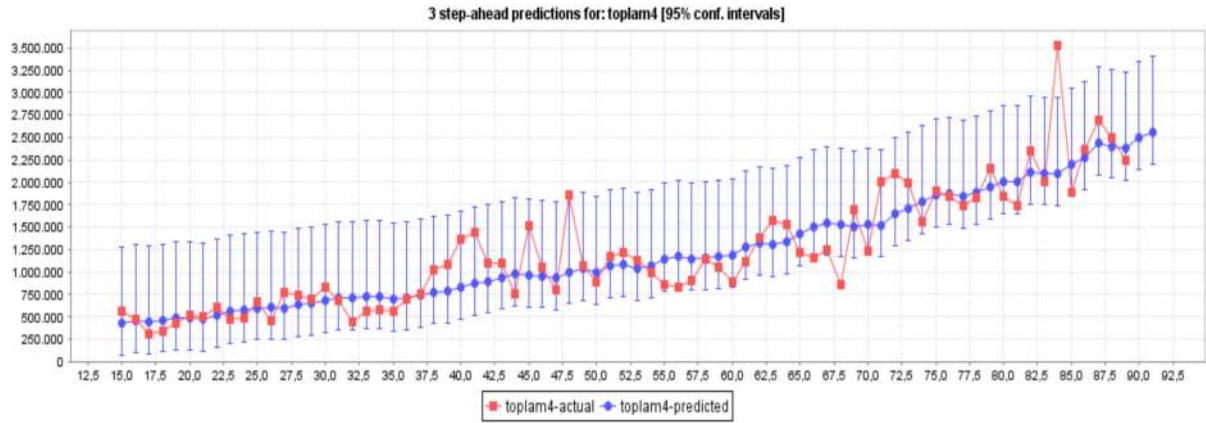
3.10.1.3 Gauss Süreç Regresyon Tahmini

GSR modelinde veriler normalize edilerek RBF çekirdeği kullanılmıştır. Eğitim aşaması sonuçları Tablo 3.22'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %20,18, %20,92 ve %20,63 olarak hesaplanmıştır.

Tablo 3.22 Toplam Satış GSR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	221356,08	228201,80	230909,04
MAPE	20,18	20,92	20,63
RMSE	320149,10	325829,23	330428,07
MSE	102495448148,12	106164684789,72	109182711615,56

Modele ait zaman serisi grafiği Şekil 3.15'de yer almaktadır.



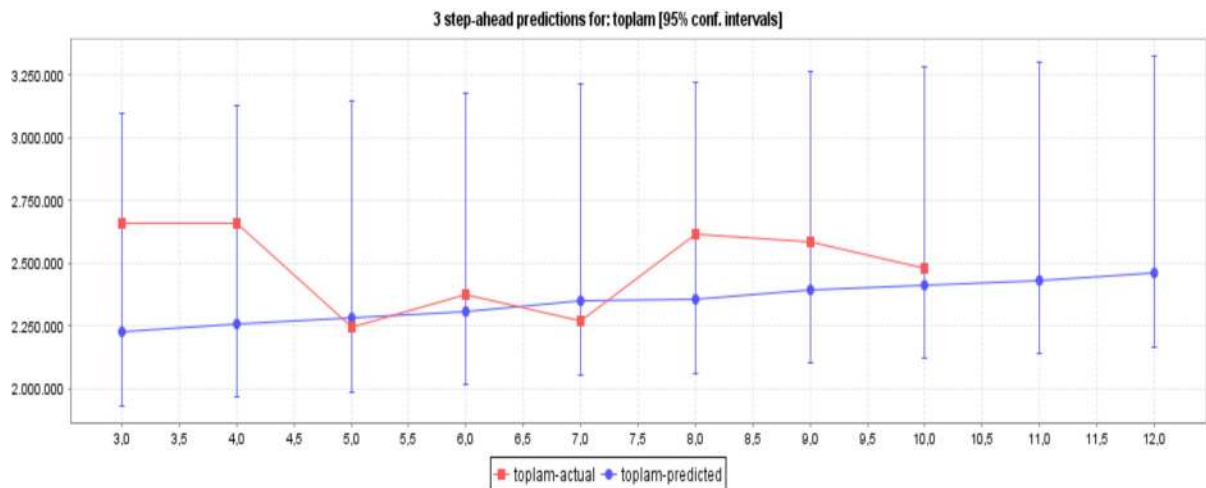
Şekil 3.15 Toplam Satış GSR Eğitim Aşaması

GSR modeline ait test aşaması sonuçları Tablo 3.23'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %6,75, %6,65 ve %7,41 olarak hesaplanmıştır.

Tablo 3.23 Toplam Satış GSR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	172318,77	171715,90	191821,01
MAPE	6,75	6,65	7,41
RMSE	219741,36	226970,71	241972,01
MSE	48286264241,07	51515704754,12	58550451837,15

Modele ait zaman serisi grafiği Şekil 3.16'da yer almaktadır. Model için 5-8 aylar aralığında oldukça başarılı olduğu söylenebilir.



Şekil 3.16 Toplam Satış GSR Test Aşaması

3.10.2 Makine Sektörü Satış İçin Tahmin Modelleri

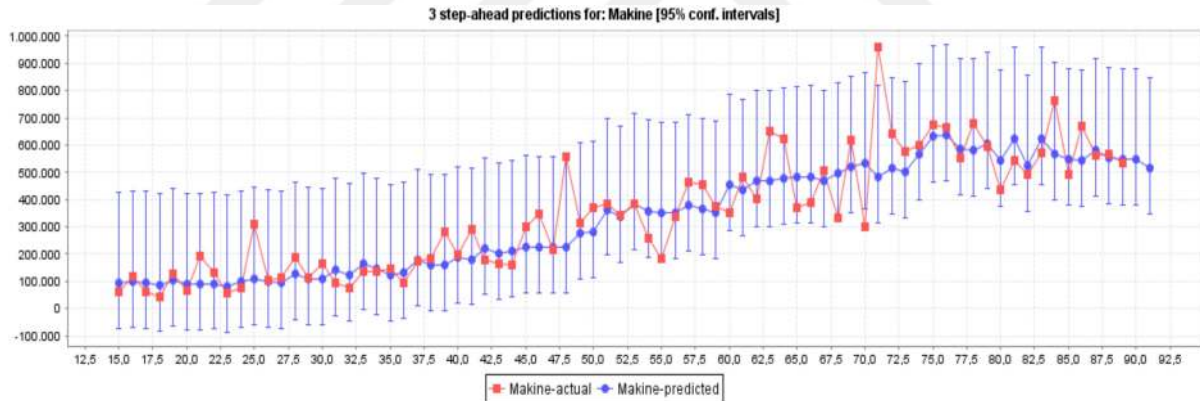
3.10.2.1 Destek Vektör Regresyon Tahmini

Destek Vektör Regresyon modeline ait eğitim aşaması sonuçları Tablo 3.24'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %16,84, %18,29 ve %18,59 olarak hesaplanmıştır. ($c=0.4$)

Tablo 3.24 Makine Sektörü DVR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	68662,57	69679,22	72071,97
MAPE	24,20	24,71	25,21
RMSE	103464,61	102975,52	104975,19
MSE	10704925210,18	10603958134,81	11019791327,05

Şekil 3.17'ye göre eğitim aşamasında MAPE değerlerinde artış görülürken, test aşamasında MAPE değerlerinde azalma yaşanmıştır. Eğitim aşamasına ait grafik aşağıda yer almaktadır.



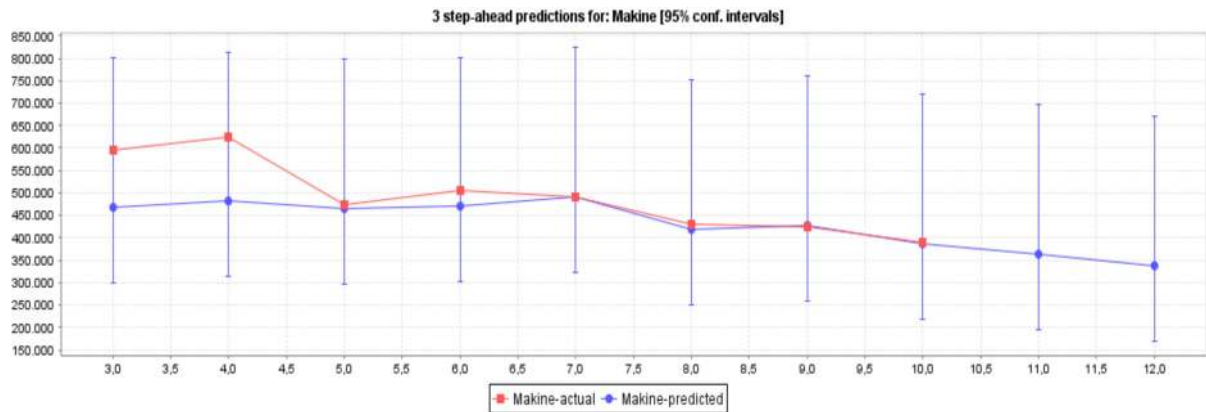
Şekil 3.17 Makine Sektörü DVR Eğitim Aşaması

Destek Vektör Regresyon modeline ait test aşaması sonuçları Tablo 3.25'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %8,25, %7,84 ve %7,21 olarak hesaplanmıştır ($c=0.4$).

Tablo 3.25 Makine Sektörü DVR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	50015,39	47490,07	41823,87
MAPE	8,25	7,84	7,21
RMSE	72092,83	75057,08	69056,08
MSE	5197375880,98	5633565290,88	4768742590,86

Makine sektörü test aşaması grafiği Şekil 3.18'de yer almaktadır. Modelin 5. aydan itibaren başarısını artırdığı söylenebilir.



Şekil 3.18 Makine Sektörü DVR Test Aşaması

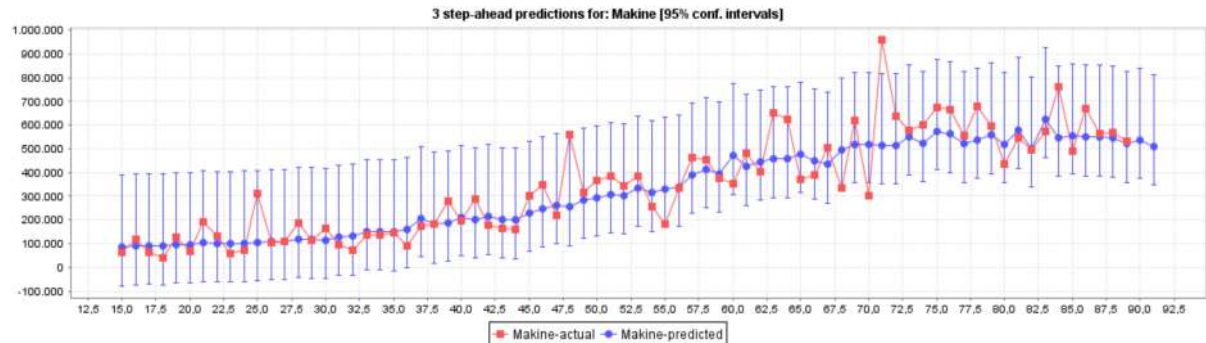
3.10.2.2 Çok Katmanlı Algılayıcı Tahmini

Çok Katmanlı Algılayıcı modeline ait eğitim aşaması sonuçları Tablo 3.26'da yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %24,85, %24,61 ve %24,95 olarak hesaplanmıştır. (LR=0.1,momentum=0.1)

Tablo3.26 Makine Sektörü ÇKA Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	72152,01	71016,29	72622,43
MAPE	24,85	24,61	24,95
RMSE	100109,46	101891,11	103103,16
MSE	10021904006,29	10381799281,21	10630262027,59

Eğitim aşamasına ait zaman grafiği Şekil 3.19'da yer almaktadır.



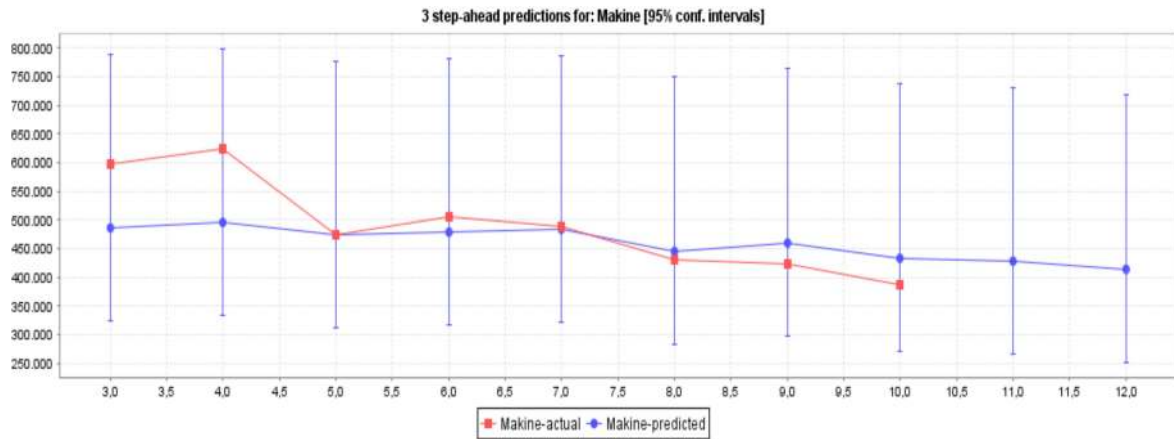
Şekil 3.19 Makine Sektörü ÇKA Eğitim Aşaması

Test aşamasına ait sonuçlar Tablo 3.27'de yer almaktadır. Tahmin süreci 1. aylık tahminde MAPE değeri %13,8 olmasına rağmen 2. ve 3. aylık tahminlerin %9,33 ve %8,97 değerleri ile daha başarılı olduğu söylenebilir.

Tablo 3.27 Makine Sektörü ÇKA Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	81088,26	55039,79	50534,90
MAPE	13,80	9,33	8,97
RMSE	105907,45	79428,97	73331,59
MSE	11216388261,00	6308961621,50	5377521630,07

Test aşamasına ait zaman grafiği Şekil 3.20'de yer almaktadır.

**Şekil 3.20 Makine Sektörü ÇKA Test Aşaması**

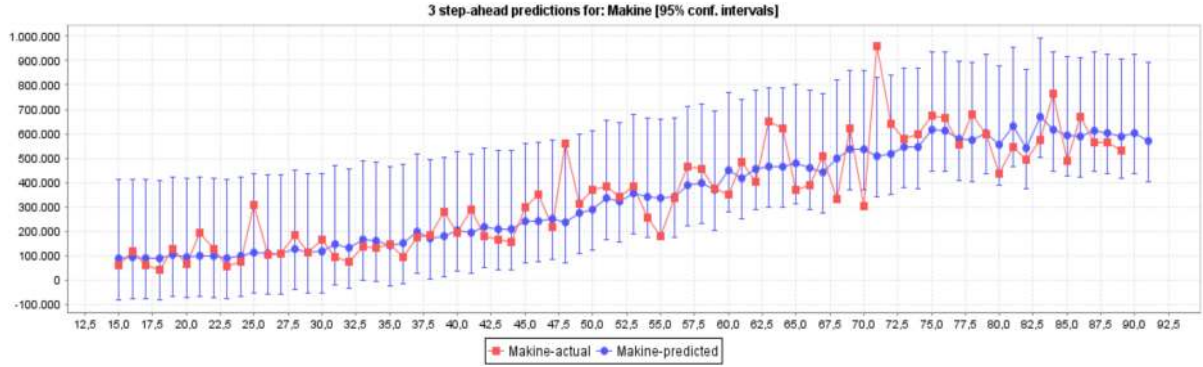
3.10.2.3 Gauss Süreç Regresyon Tahmini

GSR modeline ait eğitim aşaması sonuçları Tablo 3.28'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %24,9, %25 ve %25,39 olarak hesaplanmıştır.

Tablo 3.28 Makine Sektörü GSR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	69897,55	70568,81	71981,35
MAPE	24,90	25,00	25,39
RMSE	100123,92	100682,42	101838,27
MSE	10024799851,44	10136949139,78	10371032429,45

Eğitim Aşamasına ait Grafik Şekil 3.21'de yer almaktadır.



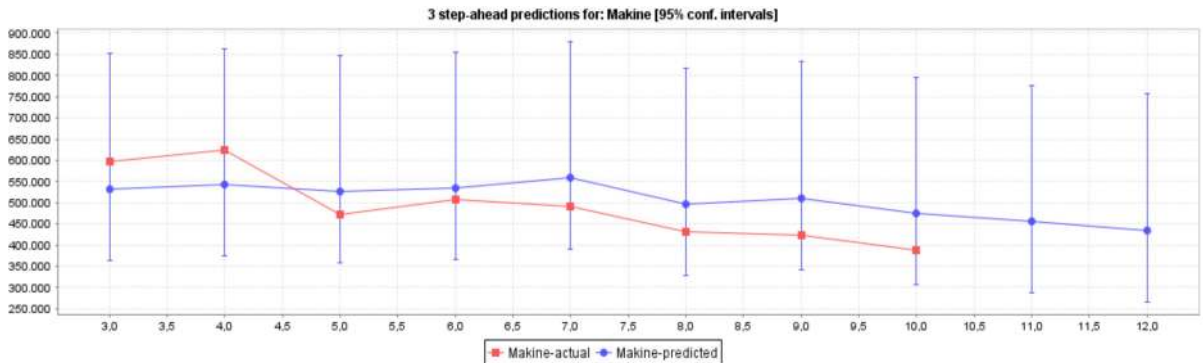
Şekil 3.21 Makine Sektörü GSR Eğitim Aşaması

GSR modeline ait test aşaması sonuçları Tablo 3.29'da yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %12,7, %13,53 ve %13,97 olarak hesaplanmıştır.

Tablo 3.29 Makine Sektörü GSR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	63784,25	66802,62	66450,29
MAPE	12,72	13,53	13,97
RMSE	65659,33	68441,76	68672,98
MSE	4311147815,43	4684274681,60	4715978619,53

MAPE değerinin %10'un üstünde kalması ve süreçte artmasından dolayı modelin başarılı olmadığı söylenebilir. Eğitim aşamasına ait zaman grafiği Şekil 3.22'de yer almaktadır.



Şekil 3.22 Makine Sektörü GSR Test Aşaması

3.10.3 Metal Sektörü Satış İçin Tahmin Modelleri

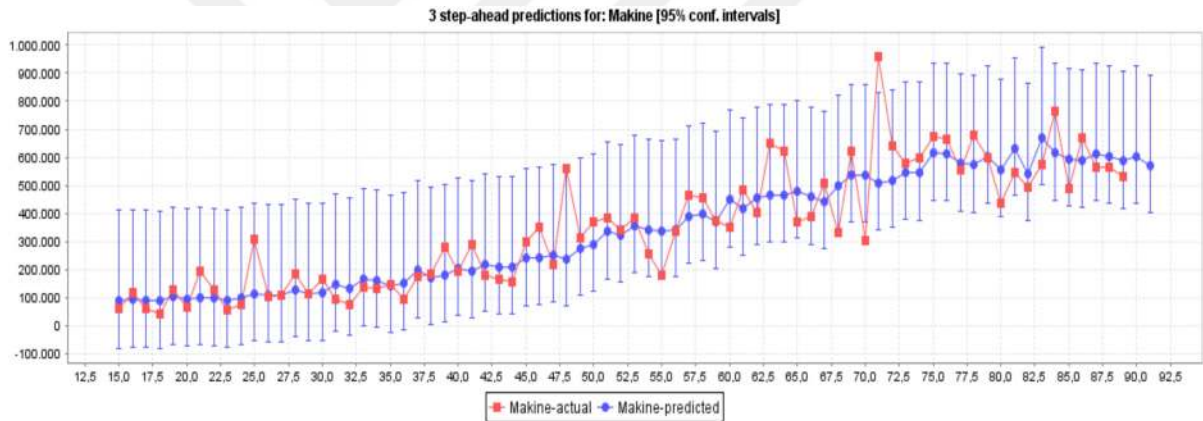
3.10.3.1 Destek Vektör Regresyon Tahmini

Destek Vektör Regresyon modeline ait eğitim aşaması sonuçları Tablo 3.30'da yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %16,84, %18,29 ve %18,59 olarak hesaplanmıştır. ($c=2$)

Tablo 3.30 Metal Sektörü Destek Vektör Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	61948,02	70029,95	73161,06
MAPE	16,84	18,29	18,59
RMSE	93692,58	98896,72	100443,59
MSE	8778298984,84	9780562018,10	1008891416,02

Eğitim aşamasına ait zaman grafiği Şekil 3.23'de yer almaktadır.



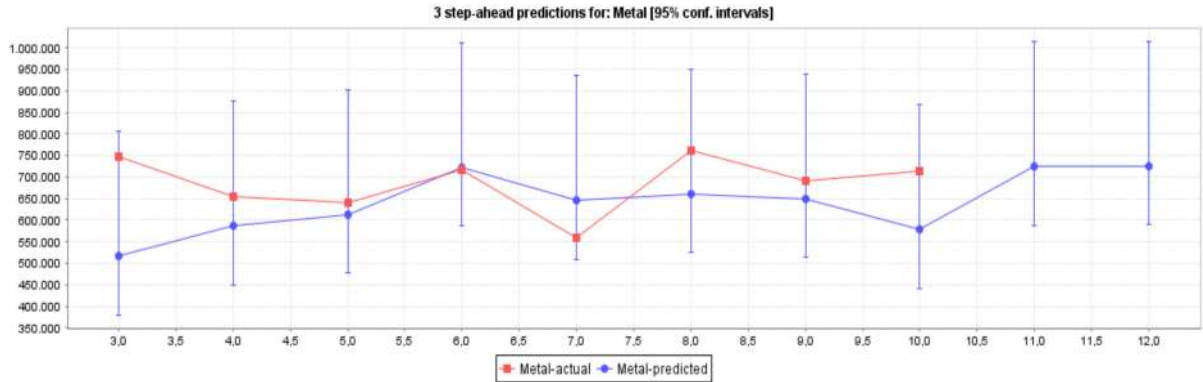
Şekil 3.23 Metal Sektörü Destek Vektör Eğitim Aşaması

Destek Vektör Regresyon modeline ait eğitim aşaması sonuçları Tablo 3.31'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %9,9, %11,3 ve %12,49 olarak hesaplanmıştır. ($c=2$)

Tablo 3.31 Metal Sektörü Destek Vektör Test Aşaması

	1.ay	2.ay	3.ay
N	10	9	8
MAE	68688,17	77241,03	86922,90
MAPE	9,90	11,30	12,49
RMSE	97935,42	98437,32	109962,32
MSE	9591346712,84	9689906599,87	1712330,111

MAPE kriterine göre model performansı 1. dönemden sonra başarılı olmasına rağmen 2. ve 3. dönemler için azalmaktadır. Modele ait grafik Şekil 3.24'de yer almaktadır.



Şekil 3.24 Metal Sektörü Destek Vektör Test Aşaması

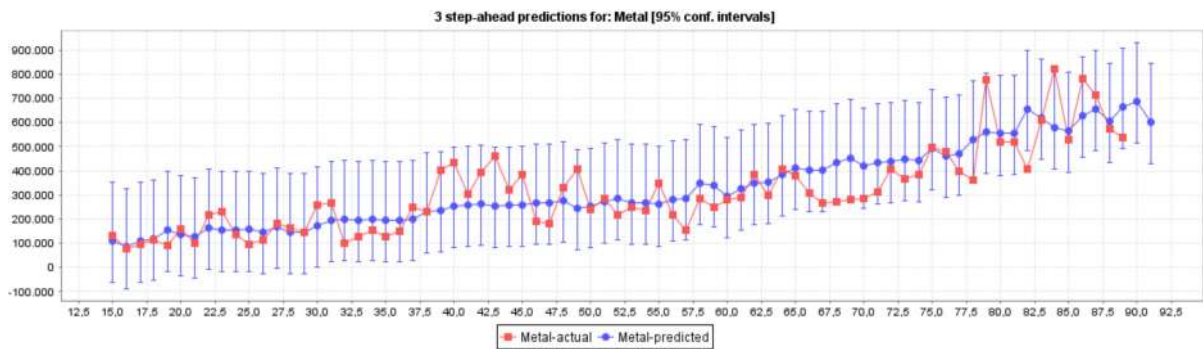
3.10.3.2 Çok Katmanlı Algılayıcı Tahmini

Çok Katmanlı Algılayıcı modeline ait eğitim aşaması sonuçları Tablo 3.32'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %24,01, %25,6 ve %26,4 olarak hesaplanmıştır. (LR=0.1,momentum=0.1)

Tablo 3.32 Metal Sektörü ÇKA Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	68290,82	70988,84	73276,85
MAPE	24,01	25,64	26,47
RMSE	89230,69	93760,88	94813,61
MSE	7962116631,36	8791102360,13	8989620497,77

Çok Katmanlı Algılayıcı modeline ait eğitim aşaması zaman grafiği Şekil 3.25'de yer almaktadır.



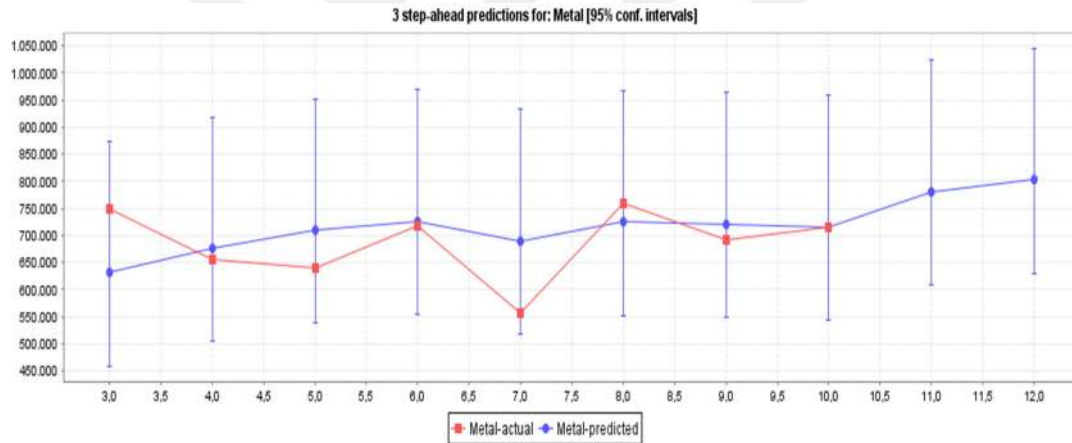
Şekil 3.25 Metal Sektörü ÇKA Eğitim Aşaması

Çok Katmanlı Algılayıcı modeline ait eğitim aşaması sonuçları Tablo 3.33'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %9,19, %7,65 ve %7,94 olarak hesaplanmıştır. (LR=0.1,momentum=0.1)

Tablo 3.33 Metal Sektörü ÇKA Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	59580,54	48406,88	51653,03
MAPE	9,19	7,65	7,94
RMSE	73436,61	66395,88	69387,48
MSE	5392935759,56	4408413408,42	4814622601,47

Çok Katmanlı Algılayıcı, MAPE kriterine göre tahmin sürecinde 2. ve 3. aylık dönemler 1. aylık dönemden başarılıdır. Eğitim aşamasına ait zaman grafiği Şekil 3.26'da yer almaktadır



Şekil 3.26 Metal Sektörü ÇKA Test Aşaması

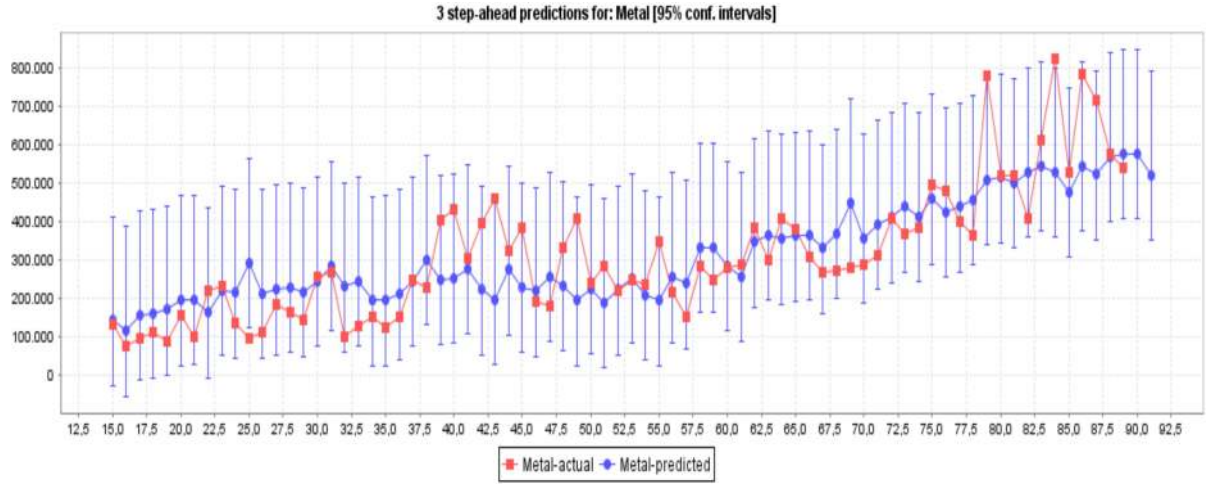
3.10.3.3 Gauss Süreç Regresyon Tahmini

GSR modeline ait eğitim aşaması sonuçları Tablo 3.34'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %31,49, %31,52 ve %31,73 olarak hesaplanmıştır.

Tablo 3.34 Metal Sektörü GSR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	75932,54	77158,99	78098,63
MAPE	31,49	31,52	31,73
RMSE	101205,03	102362,29	103238,87
MSE	10242457866,29	478039425,50	10658264210,33

Modele ait grafik Şekil 3.27'de yer almaktadır.



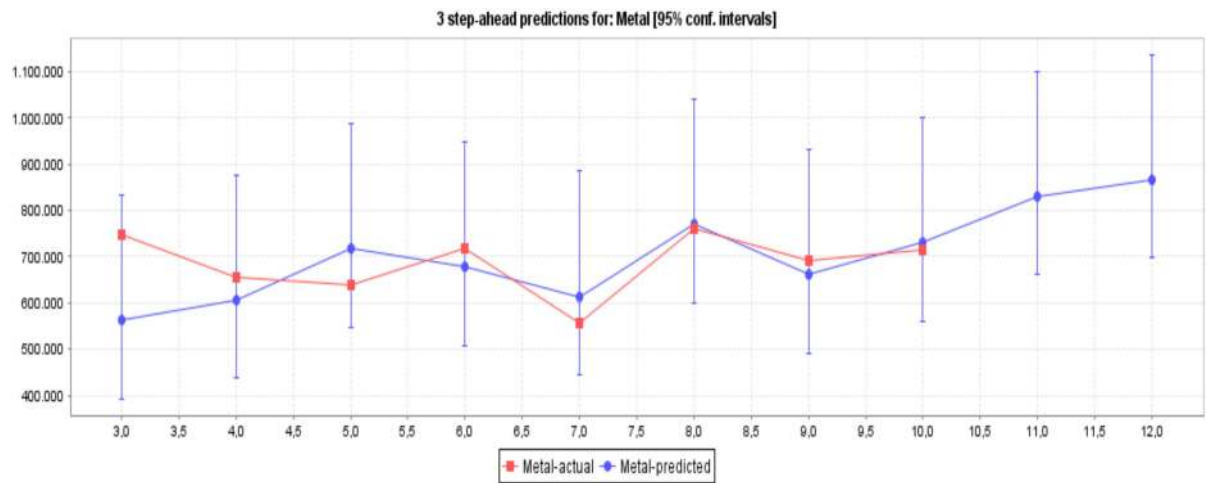
Şekil 3.27 Metal Sektörü GSR Eğitim Aşaması

GSR Eğitim aşamasına ait Test Aşaması değerleri Tablo 3.35'te yer almaktadır.

Tablo 3.35 Metal Sektörü GSR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	54466,16	52475,86	57775,08
MAPE	8,06	7,72	8,46
RMSE	73722,91	73587,48	78015,52
MSE	5435067122,35	5415116964,91	6086421933,89

Eğitim Aşamasında MAPE kriterleri %30 değerlerinde olmasına rağmen test aşamasında MAPE değerleri sırasıyla %8,06 , %7,72 ve %8,46 'dır. Test Aşamasına ait zaman grafiği Şekil 3.28'de yer almaktadır.



Şekil 3.28 Metal Sektörü GSR Test Aşaması

3.10.4 Endüstriyel Mutfak Sektörü Satış İçin Tahmin Modelleri

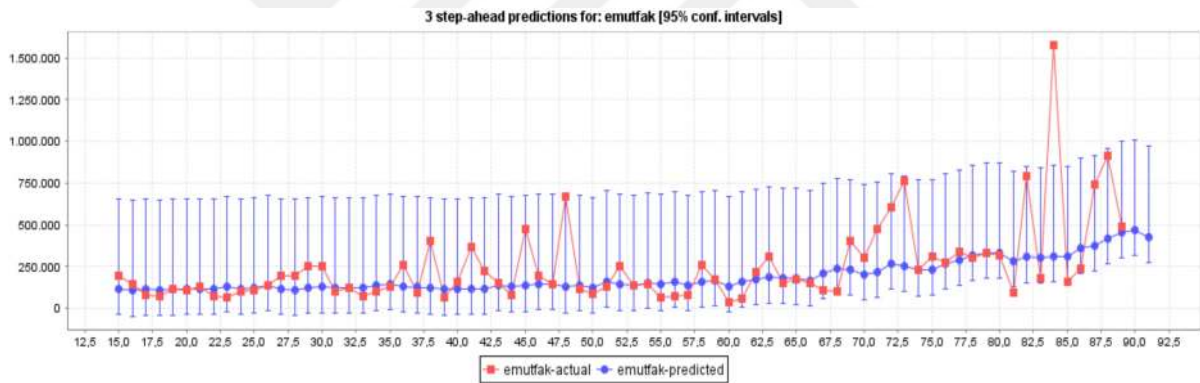
3.10.4.1 Destek Vektör Regresyon Tahmini

Destek Vektör Regresyon modeline ait eğitim aşaması sonuçları Tablo 3.36'da yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %47,41, %48,28 ve %47,28 olarak hesaplanmıştır. (c=1)

Tablo 3.36 Endüstriyel Mutfak Sektörü DVR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	114066,76	115859,84	116547,67
MAPE	47,41	48,28	47,28
RMSE	215876,53	216918,83	217693,72
MSE	46602677779,29	47053778909,87	47390553965,73

DVR yönteminde MAPE değeri oldukça yüksek çıkmıştır. Eğitim aşamasına ait zaman grafiği Şekil 3.29'da yer almaktadır.



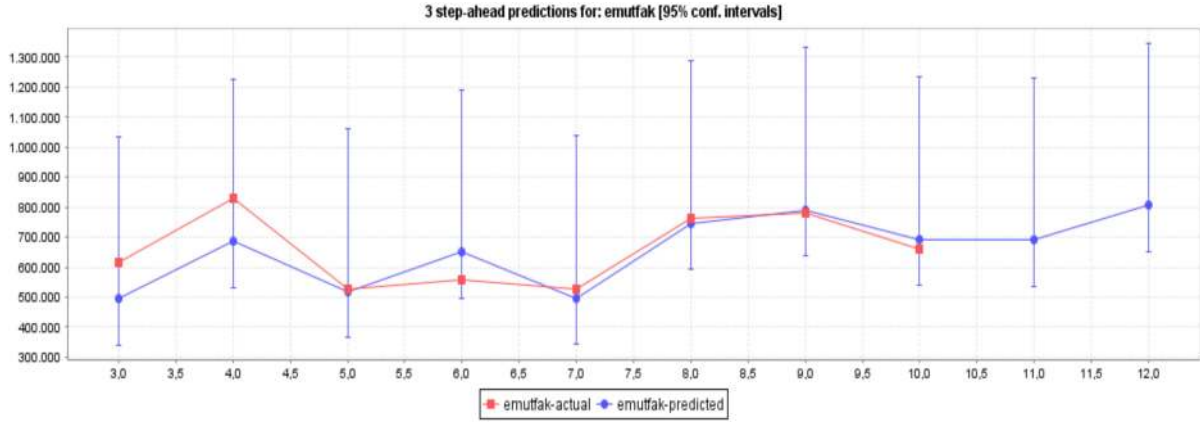
Şekil 3.29 Endüstriyel Mutfak DVR Eğitim Aşaması

Destek Vektör Regresyon modeline ait eğitim aşaması sonuçları Tablo 3.37'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %8,75, %8,74 ve %8,68 olarak hesaplanmıştır. (c=1)

Tablo 3.37 Endüstriyel Mutfak DVR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	54124,09	55357,94	56896,75
MAPE	8,75	8,74	8,68
RMSE	69948,08	73050,84	75716,85
MSE	4892733205,32	5336424607,58	5733041942,56

Eğitim aşamasına göre Test aşaması MAPE kriterine göre oldukça başarılıdır. Sürece ait zaman grafiği Şekil 3.30'da yer almaktadır.



Şekil 3.30 Endüstriyel Mutfak DVR Test Aşaması

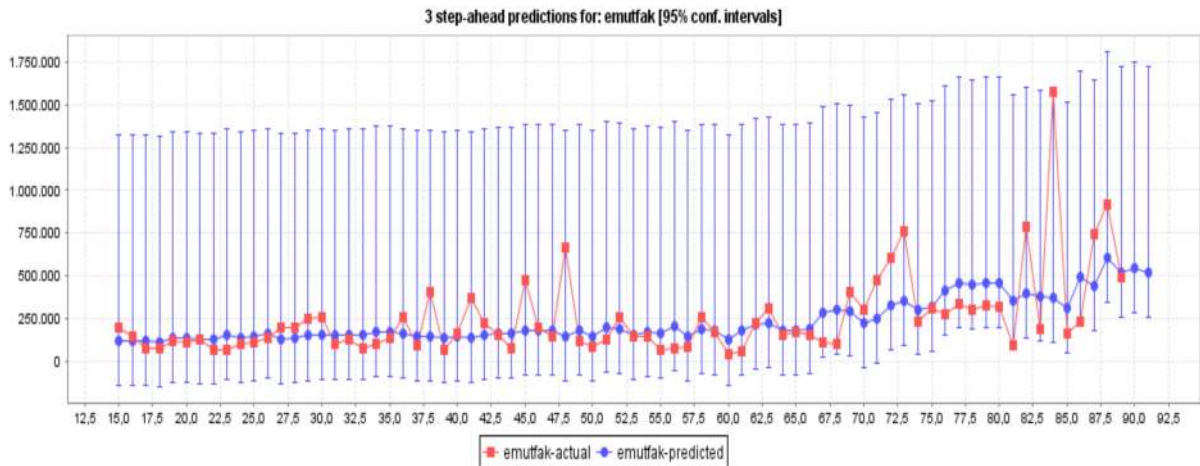
3.10.4.2 Çok Katmanlı Algılayıcı Tahmini

Çok Katmanlı Algılayıcı modeline ait eğitim aşaması sonuçları Tablo 3.38'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %59,56, %60,15 ve %58,93 olarak hesaplanmıştır. (LR=0.1,momentum=0.1)

Tablo 3.38 Endüstriyel Mutfak ÇKA Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	120169,17	122169,75	122958,74
MAPE	59,56	60,15	58,93
RMSE	201133,65	203741,34	204686,88
MSE	40454746147,75	41510534593,68	41896720480,79

Eğitim aşamasına ait zaman grafiği Şekil 3.31'de yer almaktadır.



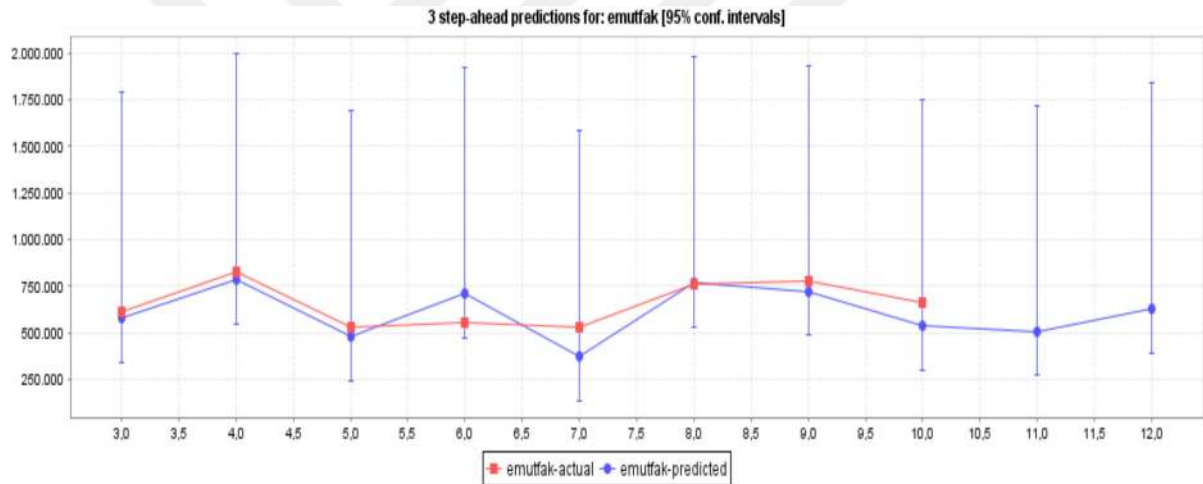
Şekil 3.31 Endüstriyel Mutfak ÇKA Eğitim Aşaması

Çok Katmanlı Algılayıcı modeline ait test aşaması sonuçları Tablo 3.39'da yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %7,85, %8,32 ve %8,09 olarak hesaplanmıştır. (LR=0.1,momentum=0.1)

Tablo 3.39 Endüstriyel Mutfak ÇKA Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	45043,60	48226,69	48217,83
MAPE	7,85	8,32	8,09
RMSE	65684,20	69791,16	71977,57
MSE	4314413644,63	4870806383,92	5180770166,17

Tablo 3.32'de görüldüğü gibi MAPE ve diğer kriter değerleri eğitim aşamasına göre oldukça başarılıdır. Sürece ait zaman grafiği Şekil 3.28'de yer almaktadır.



Şekil 3.32 Endüstriyel Mutfak ÇKA Test Aşaması

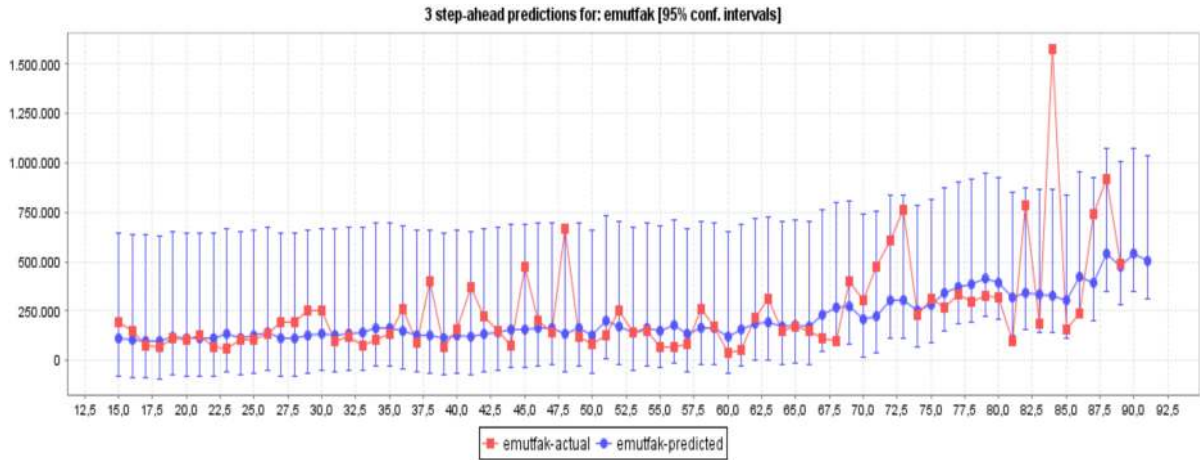
3.10.4.3 Gauss Süreç Regresyon Tahmini

GSR modeline ait eğitim aşaması sonuçları Tablo 3.40'da yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %50,87 %52,01 ve %51,05 olarak hesaplanmıştır.

Tablo 3.40 Endüstriyel Mutfak GSR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	115605,13	117426,15	118147,17
MAPE	50,87	52,01	51,05
RMSE	208423,95	208999,31	210286,83
MSE	43440543213,67	43680710733,41	44220550002,90

Eğitim aşamasına ait zaman grafiği Şekil 3.33'de yer almaktadır.



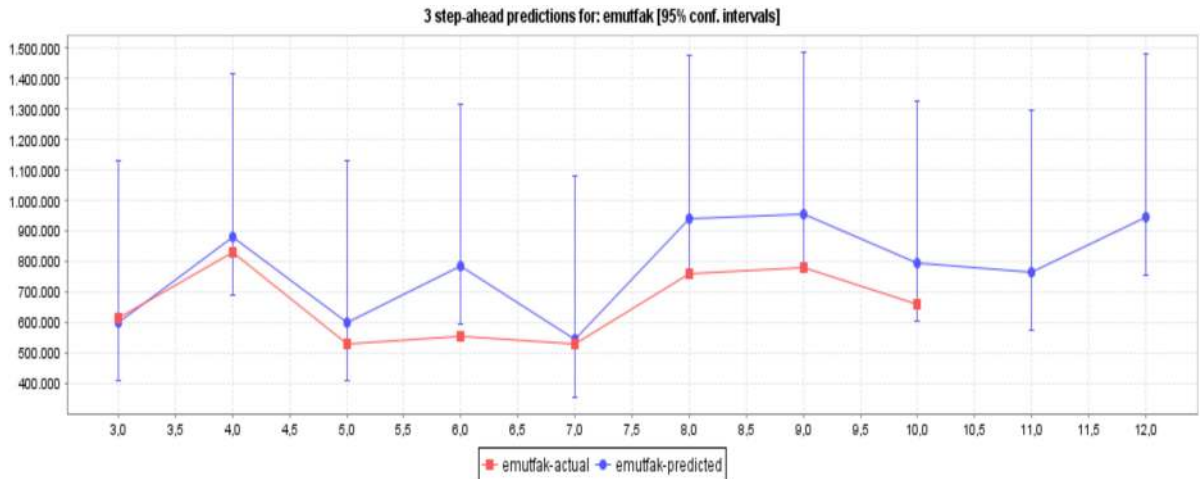
Şekil 3.33 Endüstriyel Mutfak GSR Eğitim Aşaması

GSR modeline ait eğitim aşaması sonuçları Tablo 3.41'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %13,91 %15,42 ve %16,58 olarak hesaplanmıştır.

Tablo 3.41 Endüstriyel Mutfak GSR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	89848,19	100163,67	109000,29
MAPE	13,91	15,42	16,58
RMSE	116071,98	125433,77	132545,17
MSE	13472704676,55	15733631524,55	17568223122,63

Test aşamasına ait zaman grafiği Şekil 3.34'de yer almaktadır.



Şekil 3.34 Endüstriyel Mutfak GSR Test Aşaması

3.10.5 Yapı-İnşaat Sektörü Satış İçin Tahmin Modelleri

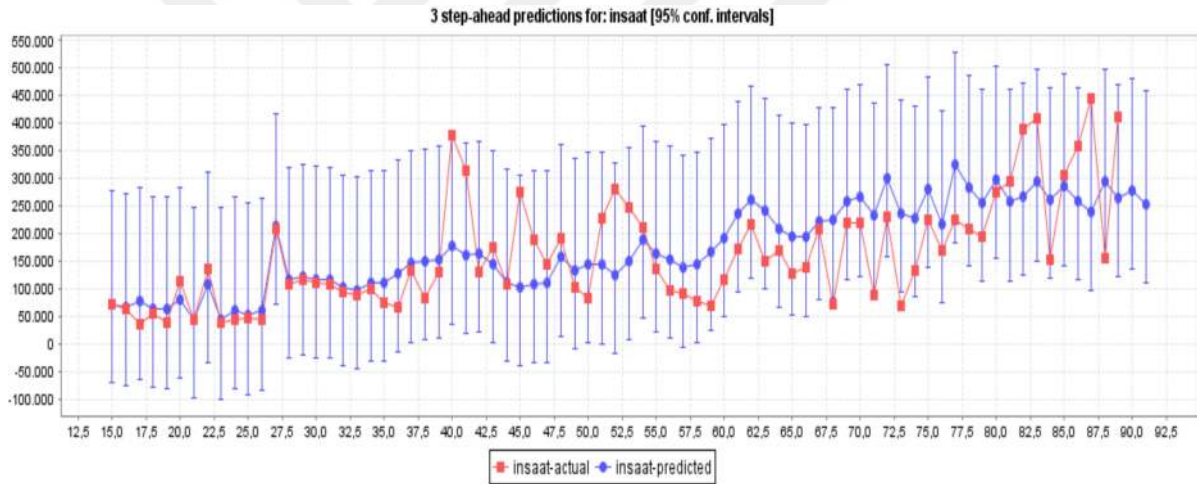
3.10.5.1 Destek Vektör Regresyon Tahmini

Destek Vektör Regresyon modeline ait eğitim aşaması sonuçları Tablo 3.42'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %33,01, %36,01 ve %41,48 olarak hesaplanmıştır. ($c=1$, $\epsilon=0,009$, $\text{expon}=3$, RegSMO)

Tablo 3.42 Yapı-İnşaat Sektörü DVR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	49358,65	53234,55	59243,63
MAPE	33,01	36,01	41,48
RMSE	72507,13	75695,66	78583,45
MSE	5257284029,85	5729832236,39	6175358488,48

Test aşamasına ait zaman grafiği Şekil 3.35'de yer almaktadır.



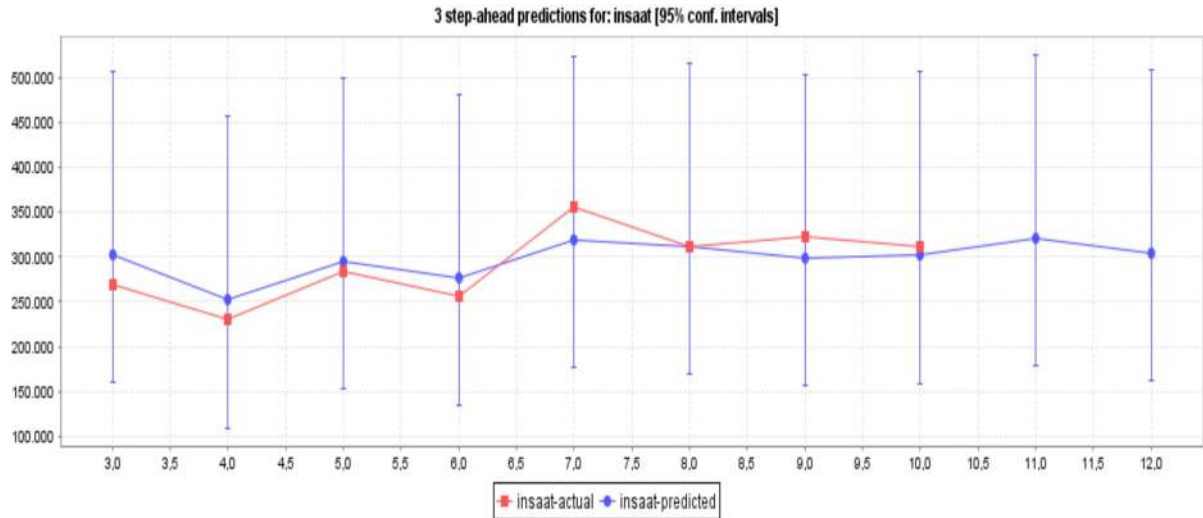
Şekil 3.35 Yapı-İnşaat Sektörü DVR Eğitim Aşaması

Destek Vektör Regresyon modeline ait test aşaması sonuçları Tablo 3.43'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %8,47, %6,96 ve %6,82 olarak hesaplanmıştır. ($c=1$, $\epsilon=0,009$, $\text{expon}=3$, RegSMO)

Tablo 3.43 Yapı-İnşaat Sektörü DVR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	21745,46	19292,79	19583,50
MAPE	8,47	6,96	6,82
RMSE	27230,61	21924,14	22514,04
MSE	741505848,36	480668108,60	506882158,47

Modele ait zaman grafiği Şekil 3.36'da yer almaktadır. Grafik incelendiğinde Modelin başarılı olduğu söylenebilir.



Şekil 3.36 Yapı-İnşaat Sektörü DVR Test Aşaması

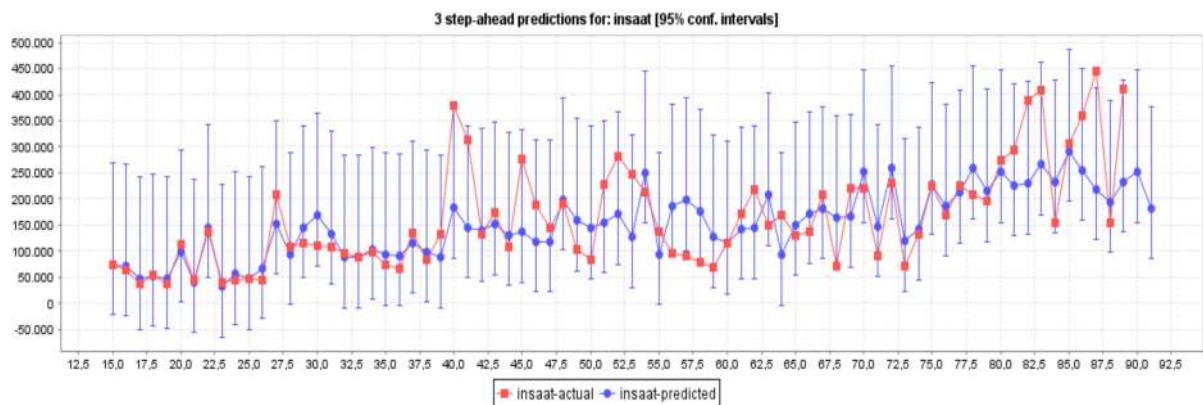
3.10.5.2 Çok Katmanlı Algılayıcı Tahmini

Çok Katmanlı Algılayıcı modeline ait eğitim aşaması sonuçları Tablo 3.44'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %60,8, %61,7 ve %60,58 olarak hesaplanmıştır.

Tablo 3.44 Yapı-İnşaat Sektörü ÇKA Eğitim Aşaması

N	77,00	76,00	75,00
MAE	39952,51	46191,51	49361,93
MAPE	24,41	28,38	30,73
RMSE	60748,31	69220,27	70445,34
MSE	3690356977,71	4791445671,77	4962545782,10

Modele ait zaman grafiği Şekil 3.37'de yer almaktadır. (LR=0.1,momentum=0.1)



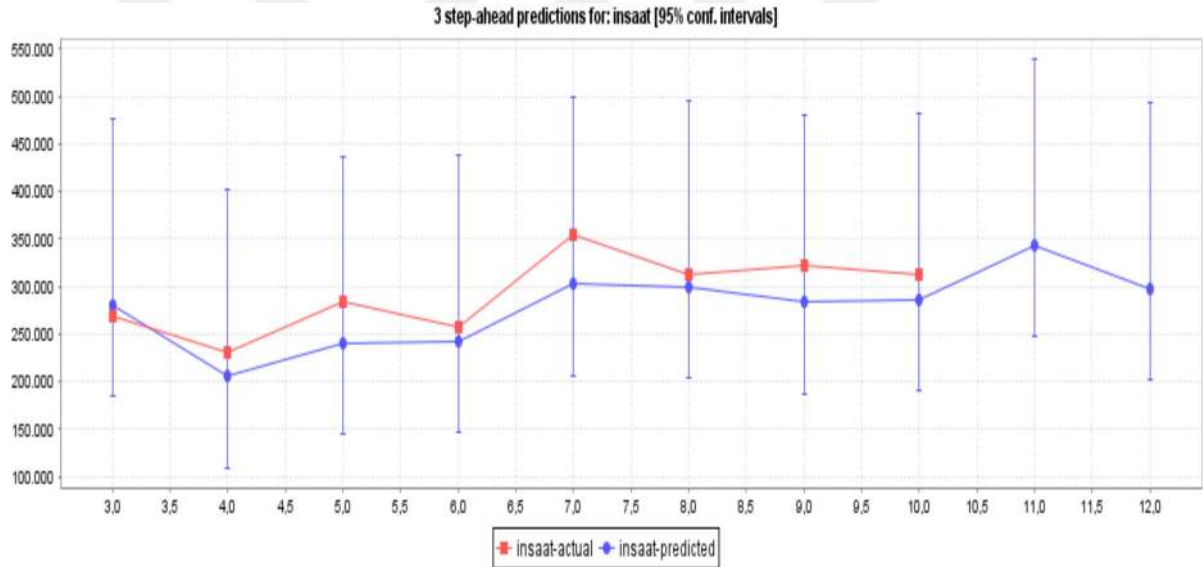
Şekil 3.37 Yapı-İnşaat Sektörü ÇKA Eğitim Aşaması

Çok Katmanlı Algılayıcı modeline ait test aşaması sonuçları Tablo 3.45'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %9,14, %9,34 ve %9,36 olarak hesaplanmıştır. (LR=0.1,momentum=0.1)

Tablo 3.45 Yapı-İnşaat Sektörü ÇKA Algılayıcı Test Aşaması

N	10,00	9,00	8,00
MAE	25046,11	26541,14	27938,39
MAPE	9,14	9,34	9,36
RMSE	31768,90	31247,53	31495,80
MSE	1009262924,74	976408336,83	991985147,70

Test aşamasına ait zaman grafiği Şekil 3.38'de yer almaktadır. MAPE değerlerinin %10 altında olması ve grafik değerleri incelendiğinde modelin başarılı olduğu söylenebilir.



Şekil 3.38 Yapı-İnşaat Sektörü ÇKA Test Aşaması

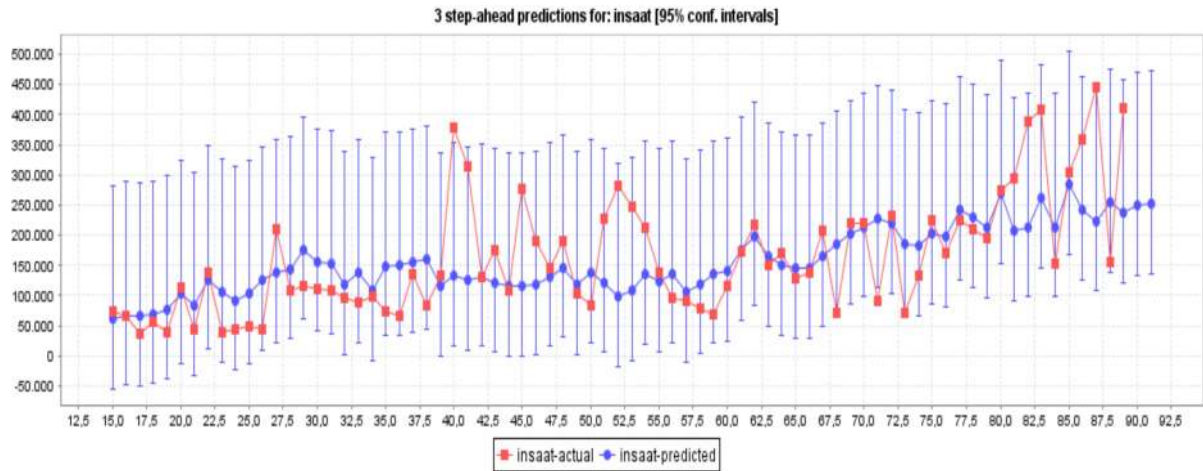
3.10.5.3 Gauss Süreç Regresyon Tahmini

GSR modeline ait eğitim aşaması sonuçları Tablo 3.46'da yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %42,45, %42,78 ve %44,42 olarak hesaplanmıştır.

Tablo 3.46 Yapı-İnşaat Sektörü GSR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	56888,24	56516,05	58085,02
MAPE	42,45	42,78	44,42
RMSE	78896,79	79576,44	80918,57
MSE	6224703173,42	332410223,65	6547814261,94

Eğitim aşamasına ait zaman grafiği Şekil 3.39'da yer almaktadır.



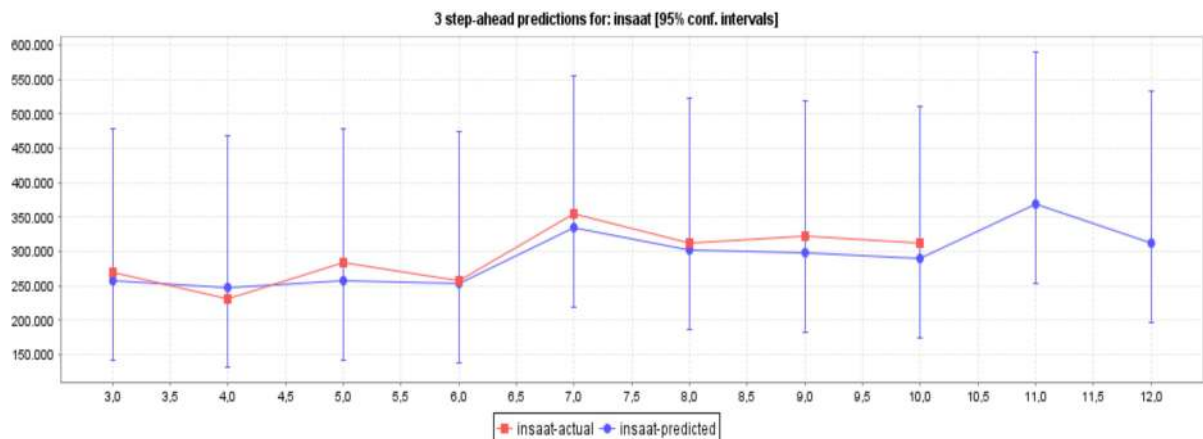
Şekil 3.39 Yapı-İnşaat Sektörü GSREğitim Aşaması

GSR modeline ait eğitim aşaması sonuçları Tablo 3.47'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %7,6, %6,59 ve %5,73 olarak hesaplanmıştır.

Tablo 3.47 Yapı-İnşaat Sektörü GSR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	19899,50	18325,77	16870,76
MAPE	7,60	6,59	5,73
RMSE	22221,93	20114,28	18494,95
MSE	493814178,90	404584454,05	342063103,13

MAPE değerlerinin %7,6 dan %5,73 seviyesine gerilemesi modelin başarılı sayılabileceğini göstermektedir. Modele ait zaman grafiği Şekil 3.40'da yer almaktadır.



Şekil 3.40 Yapı-İnşaat Sektörü GSRTTest Aşaması

3.10.6 Diğer Sektör Satış İçin Tahmin Modelleri

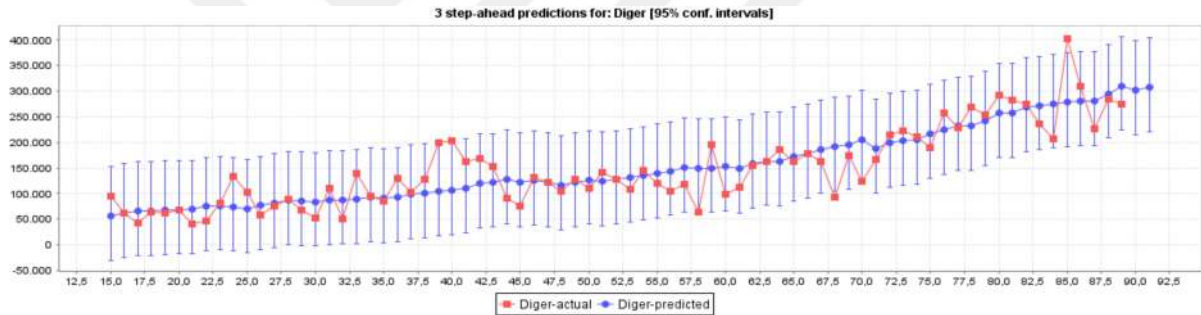
3.10.6.1 Destek Vektör RegresyonTahmini

Destek Vektör Regresyon modeline ait eğitim aşaması sonuçları Tablo 3.48'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %23,24, %23,56 ve %23,83 olarak hesaplanmıştır. ($c=1, \epsilon=0,009, \text{expon}=3, \text{RegSMO}$)

Tablo 3.48 Diğer Sektör DVR Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	27680,86	28457,39	28841,84
MAPE	23,24	23,56	23,83
RMSE	38036,28	38413,97	38958,87
MSE	1446758468,70	1475633240,06	1517793289,94

Modele ait zaman grafiği Şekil 3.41'de yer almaktadır.



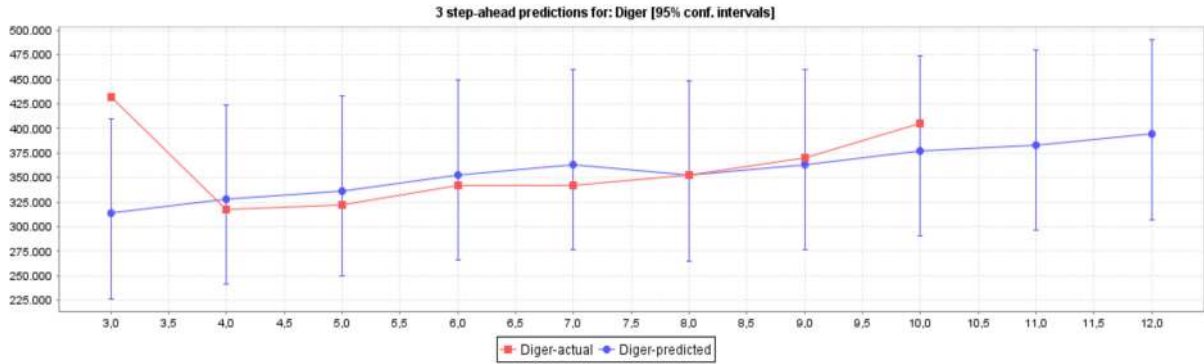
Şekil 3.41 Diğer Sektör DVR Eğitim Aşaması

Destek Vektör Regresyon modeline ait eğitim aşaması sonuçları Tablo 3.49'da yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %8,37, %6,31 ve %6,6 olarak hesaplanmıştır. ($c=1, \epsilon=0,009, \text{expon}=2, \text{RegSMO}$)

Tablo 3.49 Diğer Sektör DVR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	31474,84	24326,71	26277,27
MAPE	8,37	6,31	6,66
RMSE	46067,43	40247,81	44356,06
MSE	2122208103,48	1619886552,98	1967460193,90

Modele ait zaman grafiği Şekil 3.42'de yer almaktadır.



Şekil 3.42 Diğer Sektör DVR Test Aşaması

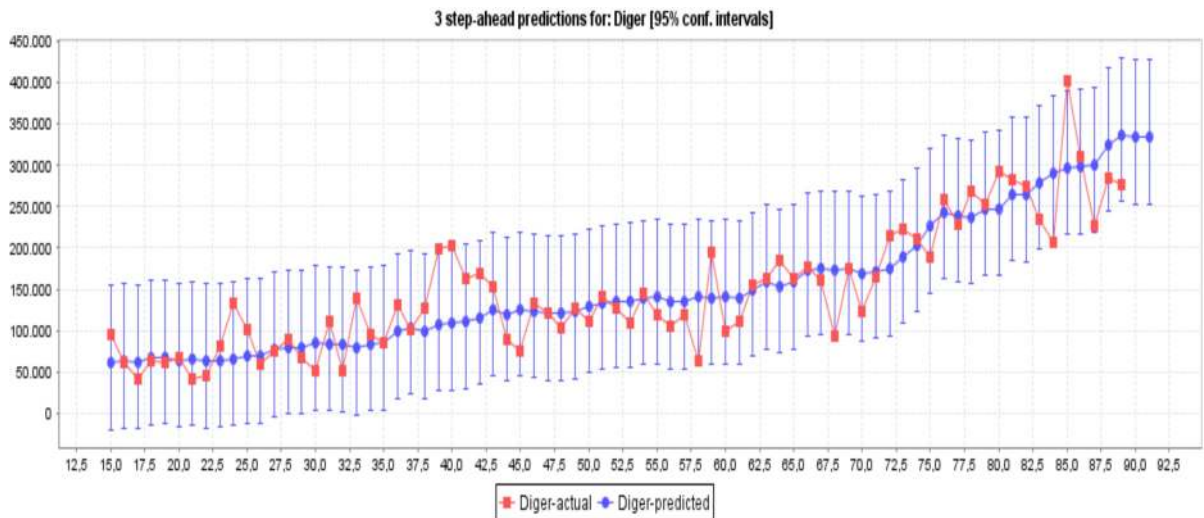
3.10.6.2 Çok Katmanlı Algılayıcı Tahmini

Çok Katmanlı Algılayıcı modeline ait eğitim aşaması sonuçları Tablo 3.50'de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %21,6, %22,15 ve %22,28 olarak hesaplanmıştır. (LR=0.1,momentum=0.1)

Tablo 3.50 Diğer Sektör ÇKA Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	26908,36	27720,82	27985,49
MAPE	21,66	22,15	22,28
RMSE	36367,86	37543,22	38106,28
MSE	1322621396,16	1409493137,93	1452088456,82

Eğitim aşamasına ait zaman grafiği Şekil 3.43'de yer almaktadır.



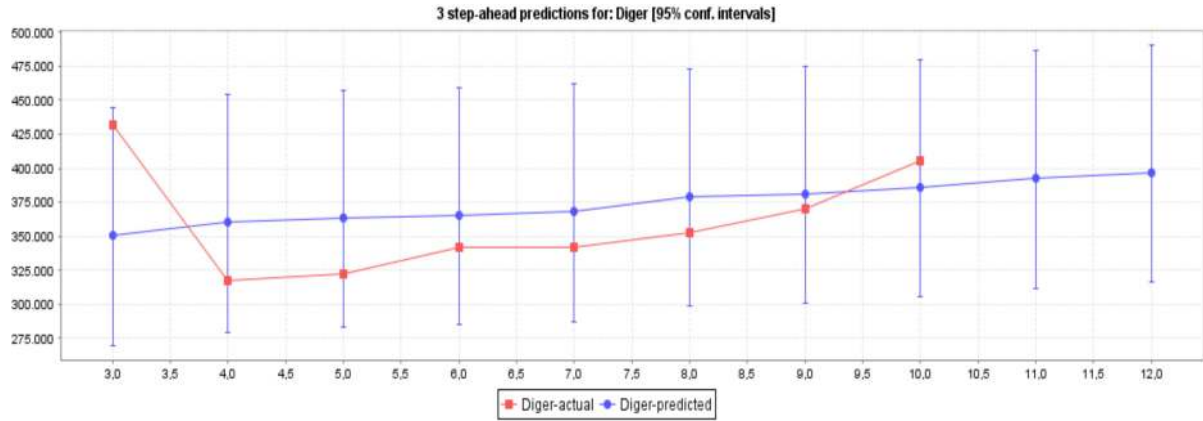
Şekil 3.43 Diğer Sektör ÇKA Eğitim Aşaması

ÇKA modeli test aşamasına ait sonuçlar Tablo 3.51’de yer almaktadır. Tablo 3.51 incelendiğinde MAPE kriterlerine göre Modelin başarılı söylenebilir.

Tablo 3.51 Diğer Sektör ÇKA Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	34822,55	34056,45	32790,40
MAPE	9,85	9,67	9,02
RMSE	38588,91	38715,92	38005,99
MSE	1489103743,03	1498922510,19	1444454916,12

ÇKA modeli test aşamasına ait zaman grafiği 3.44'de yer almaktadır.



Şekil 3.44 Diğer Sektör ÇKA Test Aşaması

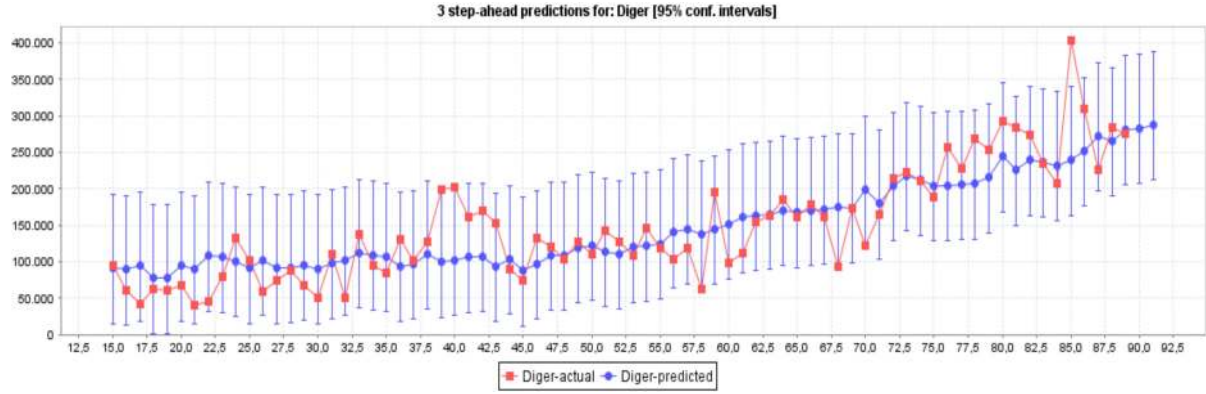
3.10.6.3 Gauss Süreç Regresyon Tahmini

GSR modeline ait eğitim aşaması sonuçları Tablo 3.52’ de yer almaktadır. MAPE değerleri 1,2 ve 3 aylık tahmin dönemlerinde sırasıyla %29,35, %29,77 ve %28,11 olarak hesaplanmıştır.

Tablo 3.52 Diğer Sektör GSR Tahmini Eğitim Aşaması

	1.ay	2.ay	3.ay
N	77,00	76,00	75,00
MAE	32024,51	32512,71	31666,65
MAPE	29,35	29,77	28,11
RMSE	42958,49	43301,64	42385,25
MSE	1845432068,11	1875031924,38	1796509271,64

Modele ait zaman grafiği Şekil 3.45'de yer almaktadır.



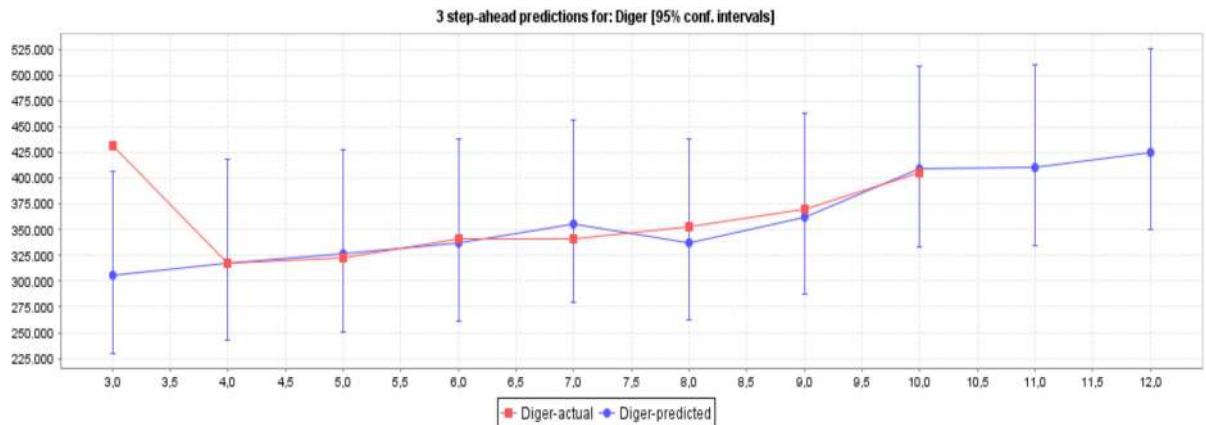
Şekil 3.45 Diğer Sektör GSR Eğitim Aşaması

GSR modeline ait test aşaması sonuçları Tablo 3.53'de yer almaktadır. MAPE değerleri sırasıyla % 6,97, %5,05 ve %5,41 değerlerini almıştır.

Tablo 3.53 Diğer Sektör GSR Test Aşaması

	1.ay	2.ay	3.ay
N	10,00	9,00	8,00
MAE	27301,87	20288,15	21927,52
MAPE	6,97	5,05	5,41
RMSE	50523,76	44064,91	45223,83
MSE	2552650453,26	1941715877,38	2045194702,56

Modele ait zaman grafiği Şekil 3.46'da yer almaktadır grafik, incelendiğinde modelin başarılı olduğu söylenebilir.



Şekil 3.46 Diğer Sektör GSR Test Aşaması

3.11 Modellerin Karşılaştırılması

3.11.1 Zaman Serisi Modellerinin Karşılaştırılması

Modellere ait değerlendirme kriterlerine göre sonuçları içeren bilgiler Tablo 3.54'de yer almaktadır.

Tablo 3.54 Zaman Serisi Modelleri

No	Yöntem	Sektör	MAPE	ME	RMSE	MAE
1	Optimal_ÜDY	Toplam	6,25	-17323,7	174251,5	150350,6
2	Optimal_Arima	Toplam	7,1	-85197,8	205003,3	168426,9
3	Tümevarım_ÜDY	Toplam	7,5	82473,27	235822,7	186834,6
4	Optimal_ÜDY	A	8,93	-792,49	189860,3	159763,5
5	Tümevarım_Arima	Toplam	10	-128874	296072,7	237089,7
6	Optimal_Arima	A	10,3	-56833,5	216337,8	182224,1
7	Optimal_Arima	Metal	11,77	-30758,2	89069,61	70865,87
8	Tümevarım_Arima	Metal	12,1	-40426,9	88161,94	71885,14
9	Tümevarım_ÜDY	Metal	12,11	-40487,1	88189,57	71895,18
10	Tümevarım_Arima	B	12,2	-32399,6	83993,79	67271,93
11	Tümevarım_Arima	Diger	12,27	44221,99	60498,83	46059,42
12	Tümevarım_ÜDY	A	12,28	85404,07	256718,5	224041,5
13	Tümevarım_ÜDY	B	12,38	-2930,8	82819,32	71576,29
14	Optimal_ÜDY	Metal	12,74	-61491,4	96760,88	74014,57
15	Optimal_Arima	B	12,77	-28364,3	90129,61	70501,6
16	Optimal_ÜDY	B	12,98	-16531,2	86223,85	73701,23
17	Optimal_Arima	Diger	13,1	45492,32	62938,54	48859,2
18	Tümevarım_Arima	A	14,32	-96474,3	311074,9	250304,4
19	Optimal_ÜDY	Diger	18,25	67261,96	80957,46	67261,96
20	Tümevarım_ÜDY	Makine	19,14	-64789	107094,2	88665,77
21	Tümevarım_ÜDY	Diger	19,55	71693,67	84835,29	71693,67
22	Optimal_ÜDY	Makine	20,75	-88780,8	117608,2	95307,07
23	Optimal_Arima	Makine	22,84	-98568,9	132476,7	103823,3
24	Optimal_Arima	Emutfak	22,92	72493,61	207392,2	156591,5
25	Tümevarım_Arima	Makine	25,68	-109467	147397,9	116696
26	Optimal_ÜDY	emutfak	26,42	149479,6	216093	181307,3
27	Tümevarım_Arima	emutfak	27,52	53419,85	245075,1	182917,6
28	Tümevarım_ÜDY	emutfak	34,44	190680,2	261025,3	228049,9
29	Tümevarım_ÜDY	insaat	37,32	-74624,5	99820,74	84378,88
30	Optimal_Arima	insaat	37,88	-73856,7	101699,7	84801,08
31	Tümevarım_Arima	insaat	37,93	-76621,5	101322,4	85785,29
32	Optimal_ÜDY	insaat	40,66	-83793,2	108084,6	92427,91

Toplam Satış 0. seviye, A ve B düğümleri 1.seviye , A düğümünü oluşturan Makine, Metal ve Endüstriyel Mutfak, B düğümünü oluşturan Yapı İnşaat ve Diğer sektör olmak üzere toplam 8 düğüm (zaman serisi) için ARIMA ve ETS yöntemleriyle 32 model oluşturulmuştur.

MAPE değerlerine göre Toplam Satış ve A düğümü için (ilk 6 model) başarılı olduğu söylenebilir. Optimal Birleşim yöntemiyle elde edilen sonuçlar Tümevarım birleşimine göre daha başarılı olduğu söylenebilir. Hiyerarşik zaman serisi tahmininde en başarılı modeller toplam satış ve 1. seviyede yer alan A düğümünde gerçekleşmiştir. B düğümü ve alt düzeylerde başarı oranı azalmaktadır. Alt düzey tahminde en başarılı model %11 Metal Sektör satışına ait tahmindir. B düğümü (Yapı-İnşaat ve Diğer sektör) satış tahmini Tümevarım-ARIMA modeliyle MAPE değeri %12,2 olarak gerçekleşmiştir.

3.11.2 Veri Madenciliği Modellerinin Karşılaştırılması

Modellerin karşılaştırılmasında Tablo 3.55'te yer alan ölçütlere göre model değerleri hesap tablosu formatında raporlanmıştır. Model sonuçları, başka bir modelle karşılaştırılması gerekmeyen “Mean Absolute Error” MAPE ölçütüne göre artan sıralamada aşağıdaki tabloda yer almıştır. Literatürde MAPE değeri %10'un altında kalan modeller çok başarılı,%10-%20 arasında kalan modeller başarılı olarak değerlendirilmektedir (Karaatlı, 2012).Tablo 3.55'de bir sonraki ayın tahmin edilmesine yönelik modeller, MAPE ölçütüne göre artan sıralamada yer almaktadır. Oluşturulan 18 modelin, 15'inde MAPE değeri %10'un değerinin altında kalmıştır. Buna göre Veri Madenciliği Yöntemlerinin genel olarak başarı olduğu söylenebilir.

Tablo 3.55 Veri Madenciliği Tahmin Modelleriyle 1. Ay Tahmin Sonuçları

Sıra	Sektör	Yöntem	MAPE	MAE	RMSE	MSE
1.	Toplam	ÇKA	5,53	136785,9	163459	26718845452
2.	Toplam	DVR	6,21	158845,08	237224,75	56275582630
3.	Toplam	GSR	6,75	172318,77	219741,36	48286264241,07
4.	Diğer	GSR	6,97	27301,87	50523,76	2552650453,26
5.	Yapı İnşaat	GSR	7,60	19899,50	22221,93	493814178,90
6.	Endüstriyel Mutfak	ÇKA	7,85	45043,60	65684,20	4314413644,63
7.	Metal	GSR	8,06	54466,16	73722,91	5435067122,35
8.	Makine	DVR	8,25	50015,39	72092,83	5197375881
9.	Diğer	DVR	8,37	31474,84	46067,43	2122208103
10.	Yapı İnşaat	DVR	8,47	21745,46	27230,61	741505848,4
11.	Endüstriyel Mutfak	DVR	8,75	54124,09	69948,08	4892733205
12.	Yapı İnşaat	ÇKA	9,14	25046,11	31768,9	1009262925
13.	Metal	ÇKA	9,19	59580,54	73436,61	5392935760
14.	Diğer	ÇKA	9,85	34822,55	38588,91	1489103743
15.	Metal	DVR	9,9	68688,17	97935,42	9591346713
16.	Makine	GSR	12,72	63784,25	65659,33	4311147815,43
17.	Makine	ÇKA	13,8	81088,26	105907,45	11216388261
18.	Endüstriyel Mutfak	GSR	13,91	89848,19	116071,98	13472704676,55

Tablo 3.55'de Toplam Satış MAPE değeri %5,53 değeri ile en iyi ÇKA tahmin etmiştir. DVR ve GSR yöntemlerine ait MAPE değerleri sırasıyla %6,21 ve %6,75 dir.

Diğer Sektör Satışlarında en başarılı model %6,75 MAPE değeri ile GSR modelidir. DVR ve ÇKA MAPE değerleri sırasıyla %8,37 ve %9,85 'dir.

Yapı-İnşaat Sektör Satışlarında en başarılı model %7,6 MAPE değeri ile GSR modelidir. DVR ve ÇKA, MAPE değerleri sırasıyla %8,47, %9,14 olarak hesaplanmıştır.

Endüstriyel Mutfak Sektör Satışlarında en başarılı model %7,85 MAPE değeri ile ÇKA modelidir. DVR ve GSR, MAPE değerleri sırasıyla %8,75,%13,91'dir.

Metal Sektör Satışlarında en başarılı model %8,06 MAPE değeri ile GSR modelidir. ÇKA ve DVR MAPE değerleri sırasıyla %9,19, %9,9 'dur.

Makine Sektör Satışlarında en başarılı model %8,25 MAPE değeri ile DVR modelidir. GSR ve ÇKA, MAPE değerleri sırasıyla %12,72 ve %13,8 olarak hesaplanmıştır.

Toplam ve sektörel satışların 1 Ay sonrasına yönelik oluşturulan 18 tahmin modelinin MAPE kriterine göre 15'inin başarılı olduğu söylenebilir.

İki aylık tahminde veri madenciliği yöntemlerine ait, kriterlere göre sonuçlar aşağıdaki Tablo 3.56'da yer almaktadır. Bu dönemde en başarılı tahmin Diğer sektör satışlarında %5,05 MAPE değeri ile GSR yöntemiyle elde edilmiştir.

Tablo 3.56 Veri Madenciliği Tahmin Modelleriyle 2. Ay Tahmin Sonuçları

Sıra	Sektör	Yöntem	MAPE	MAE	RMSE	MSE
1.	Diğer	GSR	5,05	20288,15	44064,91	1941715877
2.	Toplam	DVR	5,58	145159,7	232413,6	54016086356
3.	Toplam	ÇKA	5,99	148232,2	172275,7	29678914959
4.	Diğer	DVR	6,31	24326,71	40247,81	1619886553
5.	Yapı İnşaat	GSR	6,59	18325,77	20114,28	404584454,1
6.	Yapı İnşaat	DVR	6,96	19292,79	21924,14	480668108,6
7.	Metal	ÇKA	7,65	48406,88	66395,88	4408413408
8.	Metal	GSR	7,72	52475,86	73587,48	5415116965
9.	Makine	DVR	7,84	47490,07	75057,08	5633565291
10.	Endüstriyel Mutfak	DVR	8,74	55357,94	73050,84	5336424608
11.	Makine	ÇKA	9,33	55039,79	79428,97	6308961622
12.	Yapı İnşaat	ÇKA	9,34	26541,14	31247,53	976408336,8
13.	Diğer	ÇKA	9,67	34056,45	38715,92	1498922510
14.	Metal	DVR	11,3	77241,03	98437,32	9689906600
15.	Endüstriyel Mutfak	ÇKA	11,57	69364,42	89553,16	8019768962
16.	Toplam	GSR	11,74	278666,2	306265,3	93798446743
17.	Makine	GSR	13,53	66802,62	68441,76	4684274682
18.	Endüstriyel Mutfak	GSR	15,42	100163,7	125433,8	15733631525

Diğer Sektör Satışlarında en başarılı model %5,05 MAPE değeri ile GSR modelidir. DVR ve ÇKA, MAPE değerleri sırasıyla %6,31,%9,67'dir.

Toplam Satış tahmininde en başarılı model %5,58 MAPE değeri ile DVR modelidir.ÇKA ve GSR, MAPE değerleri sırasıyla %5,99,%11,74'dir.

Yapı-İnşaat sektörü satış tahmininde en başarılı model %6,59 MAPE değeri ile GSR modelidir. DVR ve ÇKA MAPE değerleri sırasıyla %6,96,%9,34'tür.

Metal sektörü satış tahmininde en başarılı model %7,65 MAPE değeri ile ÇKA modelidir.GSR ve DVR, MAPE değerleri sırasıyla %7,72 ve %11,3 olarak hesaplanmıştır.

Makine sektörü iki aylık satış tahmininde en başarılı model %7,84 MAPE değeri ile DVR modelidir. ÇKA ve GSR, MAPE değerleri sırasıyla %9,33 ve %13,53'tür.

Endüstriyel Mutfak sektörü satış tahmininde en başarılı model %8,74 MAPE değeri ile DVR modelidir. ÇKA ve GSR, MAPE değerleri sırasıyla %11,57 ve %15,42 olarak hesaplanmıştır.

Toplam ve sektörel satışların 2 Ay sonrasına yönelik oluşturulan 18 tahmin modelinin MAPE kriterine göre 13'ünün başarılı olduğu söylenebilir.

Üç aylık tahminde veri madenciliği yöntemlerine ait, kriterlere göre sonuçlar aşağıdaki Tablo 3.57'de yer almaktadır. Bu dönemde en başarılı tahmin Diğer sektör satışlarında %5,41 MAPE değeri ile GSR yöntemiyle elde edilmiştir. DVR ve ÇKA MAPE değerleri sırasıyla %6,66 ve %9,02 olarak hesaplanmıştır.

Tablo 3.57 Veri Madenciliği Tahmin Modelleriyle 3. Ay Tahmin Sonuçları

No	Sektör	Yöntem	MAPE	MAE	RMSE	MSE
1.	Diğer	GSR	5,41	21927,52	45223,83	2045194703
2.	Yapı İnşaat	GSR	5,73	16870,76	18494,95	342063103,1
3.	Toplam	ÇKA	5,88	148263,2	174823,3	30563200711
4.	Toplam	DVR	5,92	155570,2	245273,6	60159119750
5.	Diğer	DVR	6,66	26277,27	44356,06	1967460194
6.	Yapı İnşaat	DVR	6,82	19583,5	22514,04	506882158,5
7.	Makine	DVR	7,21	41823,87	69056,08	4768742591
8.	Metal	ÇKA	7,94	51653,03	69387,48	4814622601
9.	Metal	GSR	8,46	57775,08	78015,52	6086421934
10.	Endüstriyel Mutfak	DVR	8,68	56896,75	75716,85	5733041943
11.	Makine	ÇKA	8,97	50534,9	73331,59	5377521630
12.	Diğer	ÇKA	9,02	32790,4	38005,99	1444454916
13.	Yapı İnşaat	ÇKA	9,36	27938,39	31495,8	991985147,7
14.	Toplam	GSR	11,95	287633,5	324883,3	1,05549E+11
15.	Metal	DVR	12,49	86922,9	109962,3	1712330,1105
16.	Endüstriyel Mutfak	ÇKA	12,98	77821,79	94838,42	8994325500
17.	Makine	GSR	13,97	66450,29	68672,98	4715978620
18.	Endüstriyel Mutfak	GSR	16,58	109000,3	132545,2	17568223123

Yapı İnşaat Sektör Satışlarında en başarılı model %5,73 MAPE değeri ile GSR modelidir. DVR ve ÇKA, MAPE değerleri sırasıyla %6,82, %9,36'dır.

Toplam Satış tahmininde en başarılı model %5,88 MAPE değeri ile ÇKA modelidir. DVR ve GSR, MAPE değerleri sırasıyla %5,92, %11,95 olarak hesaplanmıştır.

Makine Sektör Satışlarında en başarılı model %7,21 MAPE değeri ile DVR modelidir. ÇKA ve GSR, MAPE değerleri sırasıyla %8,97 ve %13,97 olarak hesaplanmıştır.

Metal Sektör Satışlarında en başarılı model %7,94 MAPE değeri ile ÇKA modelidir. GSR ve DVR, MAPE değerleri sırasıyla %8,46 ve %12,49 olarak hesaplanmıştır.

Endüstriyel Mutfak sektörü satış tahmininde en başarılı model %8,68 MAPE değeri ile DVR modelidir. ÇKA ve GSR, MAPE değerleri sırasıyla %12,98 ve %16,58 olarak hesaplanmıştır.

Toplam ve sektörel satışların 3 Ay sonrasına yönelik oluşturulan 18 tahmin modelinin MAPE kriterine göre 13'ünün başarılı olduğu söylenebilir.

SONUÇ

Ülkemizde sanayi sektörü ve hayat standartları geliştiği oranda ikamesi olmayan paslanmaz çelik ihtiyacının artması beklenmektedir. ISSF verilerine göre ülkemiz, dünya genelinde paslanmaz çelik sektöründe 2012 rakamlarına göre 5. büyük ithalatçıdır. Paslanmaz çelik yurt içi talebin yanı sıra ülkemizin ihracatında önemli yeri olan Makine, Metal ve Yapı-İnşaat ve Endüstriyel Mutfak gibi alanlarda kullanılmaktadır. Bu nedenle ithalatçı firmaların gerek kendi sürdürülebilir büyümeleri gerekse ülke ekonomisine destek sağlamaları açısından talep tahminleri etkin bir şekilde yapmalıdırlar.

Firma satış verileri Makine,Metal, Endüstriyel Mutfak ve Diğer sektörlerle satış verilerinden oluşmaktadır.Model Ağaçlar yöntemiyle toplam ve sektörel satışlar için tahmin edici geçmiş ve gelecek dönemler için projeksiyon tutan doğrusal modeller elde edilmiştir. Toplam Satış için sırasıyla SDR değeri en yüksek olan Dolar Kuru, Reel Efektif Döviz Kuru, Üretici Fiyat Endeksi kritik değerlerine göre 5 adet regresyon denklemi elde edilmiştir. Bu değişkenler dışında toplam satışa etki eden diğer değişkenler ise Demir Madeni fiyatı, Sanayi Üretim Endeksi, İmalat Sanayi Kapasite Kullanım Oranıdır

Firmanın en çok satış yaptığı sektör toplam satışın %27 'ile Makine sektörüdür. 2012 yılında %36 oranıyla en yüksek değerini elde etmiştir. Makine sektörüne ait 5 adet tahmin edici modelsırasıyla SDR değeri en yüksek ÜFE,Sanayi Üretim Endeksi, Reel Efektif Döviz Kuru kritik değerlerine göre elde edilmiştir. Bu değişkenler dışında sektör satışlarına Sanayi Ciro endeksinin etkili olduğu gözlenmiştir.

Metal Sektörü 2009 ve 2015 yıllarında toplam satışın %27'sini oluşturarak en yüksek satış yapılan sektör olmasına rağmen toplam satışın %25'ini oluşturarak en fazla satış yapılan ikinci sektör olmuştur. Metal Sektörüne ait iki farklı tahmin edici doğrusal modeller ÜFE üzerinden elde edilmiştir. Sektör satışlarında etkili olan diğer değişkenler ise Dolar Kuru ve İmalat Sanayi Kapasite Kullanım Oranı olmuştur.

Endüstriyel Mutfak sektör satışların toplam satışın %22'ini oluşturarak en fazla satış yapılan üçüncü sektördür. Sanayi Ciro Endeksi değerine göre iki adet tahmin edici doğrusal modeller elde edilmiştir. Modele etki eden diğer değişkenler ise Dolar Kuru, Üfe,Sanayi Ciro Endeksi, İmalat Sanayi Kapasite Kullanım Oranıdır.

Yapı ve inşaat sektör satışları için toplam satışın %14'ünü oluşturarak en fazla satış yapılan dördüncü sektör olmuştur. Üfe ve Dolar Kurudeğerlerine göre iki adettahmin edici doğrusal model elde edilmiştir. Modelde yer alan diğer değişkenler ise Demir ,Krom ve Nikel

maden fiyatları, Sanayi Ciro ve Sanayi Üretim Endeksleri, Reel Efektif Döviz Kuru endeksi satışlar üzerinde etkilidir.

Satıştaki payı küçük çeşitli sektörlerin toplamından oluşan "Diğer" sektör 2009-2011 döneminde satış oranını en fazla artıran sektör olmuştur. Toplam Satıştaki payını %6' dan payı %12 düzeyine artıran Diğer Sektör Satışları için Dolar Kuru üzerinden iki farklı tahmin edici doğrusal model elde edilmiştir. Modelde yer alan diğer değişken ise Sanayi Üretim Endeksidir.

Hiyerarşik Kümeleme analizi sonucunda ilk kümede Makine, Metal, Endüstriyel Mutfak ikinci kümede ise Yapı-İnşaat ve Diğer Sektör yer almaktadır. Elde edilen yapıya ait 0.-1.-2. düzeyler için diğer bir ifade ile toplam 8 düğümü kapsayan (zaman serisi) hiyerarşik zaman serisi tahmini yapılmıştır. Tüme varım ve Optimal birleşim yöntemleriyle, iki farklı (ARIMA ve Üstel Düzgünleştirme) zaman serisi yöntemi kullanılarak 32 adet model oluşturulmuştur. Uygulama yapılan veri setimize göre modellerin sadece 6' sında MAPE değeri %10 değerinin altındadır. Başarılı modellerin 4'ü Toplam Satış 2' ise Makine, Metal ve Endüstriyel Mutfak sektörlerinin birleşimi olan A düğüm Tahmini modelidir. MAPE değerine göre en iyi değer %6,25 ile Toplam Satış tahmininde Optimal Birleşim Üstel Düzgünleştirme Modelidir.

Destek Vektör Regresyon yöntemiyle 1. aya ait tahminde Toplam Satış %6,21, Makine %8,25, Diğer %8,37 , Yapı İnşaat %8,47, Endüstriyel Mutfak %8,75, Metal %9,9 olarak hesaplanmıştır. DVR'nin tüm sektörler için başarılı olduğu gözlenmiştir. 2. Aya ait tahminde Toplam satış %5,58, Diğer sektör %6,31, Yapı İnşaat %6,96, Makine %7,84 , Endüstriyel Mutfak %8,74, Metal Sektörü %11,3 MAPE değerleriyle tahmin etmiştir. 3. Aya ait tahminde Toplam satış %5,92, Diğer sektör %6,66, Yapı İnşaat %6,82, Makine %7,21 , Endüstriyel Mutfak %8,68, Metal Sektörü %12,49 MAPE değerleriyle tahmin etmiştir. DVR yöntemi Metal sektörüne ait 2. ve 3. ay dışındaki tüm modellerde başka bir ifadeyle 18 modelin 16'sında başarılı olduğu gözlenmiştir.

Çok Katmanlı Algılayıcı yöntemiyle 1. aya ait tahminde Toplam Satış %5,53, Endüstriyel Mutfak %7,85, Yapı İnşaat %9,14, Metal %9,19, Diğer %9,85, Makine %13,8 olarak hesaplanmıştır. 2. Aya ait tahminde Toplam satış %5,99, Metal Sektörü %7,65, Makine %9,33, Yapı İnşaat %9,34, Diğer sektör %9,67, Endüstriyel Mutfak %11,57, MAPE değerleriyle tahmin etmiştir. 3. Aya ait tahminde Toplam satış %5,88, Metal Sektörü %7,94, Makine %8,97, Diğer sektör %9,02, Yapı İnşaat %9,36, Endüstriyel Mutfak %12,98, MAPE değerleriyle tahmin etmiştir. Çok Katmanlı Algılayıcı modeli kurulan 18 modelin 15'inde başarılı olduğu gözlenmiştir.

Gauss Süreç Regresyon yöntemiyle 1. aya ait tahminde Toplam Satış %6,75, Diğer %6,97, Yapı İnşaat %7,6, Metal %8,06, Makine %12,72, Endüstriyel Mutfak %13,91, olarak hesaplanmıştır 2. Aya ait tahminde Diğer sektör %5,05, Yapı İnşaat %6,96, Metal Sektörü %7,72, Toplam satışı %11,74, Makine %13,53 ,Endüstriyel Mutfak %15,42, MAPE değerleriyle tahmin etmiştir. 3. Aya ait tahminde Diğer sektör %5,41, Yapı İnşaat %5,73, Metal Sektörü %8,46, Toplam satışı %11,95, Makine %13,97, Endüstriyel Mutfak %16,58, MAPE değerleriyle tahmin etmiştir. Kurulan 18 modelin 10'unda başarılı olduğu gözlemlenmiştir.

Veri Madenciliği tahmin yöntemlerinin sektörlere göre uyumu dikkate alındığında başarı sıralaması DVR(16/18), ÇKA(15/18), GSR(10/18) yöntemidir. DVR yöntemi çalışmamıza en düşük MAPE değeri ve sektörlere uyum açısından en başarılı yöntemdir. GSR yöntemi DVR yöntemiyle eşit sayıda (7/18 modelde) en başarılı sonuç üretmesine rağmen farklı sektörler için başarı sıralamasında 3. sırada yer almaktadır. ÇKA ise en başarılı tahmin üretmede 3. yöntem olmasına rağmen, sektörlere uyum açısından (15/18) 2. sırada başarılı olmuştur. DVR,ÇKA,GSR yöntemlerinin ürettiği tahmin sonuçları zaman serileri yöntemlerine göre başarılı olduğu sonucuna ulaşılmıştır.

Çalışmada Model ağaç yöntemleriyle satış yapılan sektörlerin hangi makro ekonomik değişkenlerden ne derece etkilendiği tespit edilmiş, hiyerarşik zaman serisi ve veri madenciliği tahmin yöntemleriyle sektörlere göre satış hareketleri tahmin edilmiştir. Böylece makro ekonomik değişkenlerin sektörel satışlar üzerindeki etkisinin tespiti, özellikle veri madenciliği yöntemleriyle başarılı bir şekilde sektör satışlarının tahmin edilmesi, yöneticilerin stok, lojistik gibi çeşitli maliyetlerin azaltılmasına yönelik stratejik kararlar almasını kolaylaştırabileceği ifade edilebilir.

Geleceğe yönelik çalışmalarda veri madenciliği tahmin yöntemleri paslanmaz çelik satışı yapılan farklı sektörlerde uygulanabilir, satışlara etki eden bağımsız değişkenlere yönelik bulguların Türkiye’de paslanmaz çelik sektöründe faaliyet gösteren şirketlerin ve kurumların planlama çalışmalarına katkı sağlayacağı söylenebilir.

KAYNAKÇA

- Akdag, R. (2016). "Yapay Sinir Ağları, Destek Vektör Makineleri ve Box-Jenkins Yöntemleriyle Kentsel İçme Suyu Talebi Tahmini ve Karşılaştırmalı Analizi." *Business and Economics Research Journal* 7(1): 123.
- Akgül, I. (2003a). *Geleneksel Zaman Serisi Yöntemleri*. Der Yayınları İstanbul
- Akgül, I. (2003b). *Zaman Serilerinin Analizi ve Arima Modelleri*: Der Yayınları İstanbul
- Allen, P. G. ve Fildes, R. (2001). *Principles of Forecasting*. Springer US.
- Alpaydın, E. (2011). *Yapay Öğrenme*. Boğaziçi Üniversitesi Yayınevi, İstanbul
- Akpınar, H. (2014). *DATA Veri Madenciliği – Veri Analizi*, Papatya Yayınevi, İstanbul
- Armstrong, J. S. (ed.). (2001). *Principles of Forecasting: a Handbook for Researchers and Practitioners* (Vol. 30). Springer Science & Business Media.
- Athanasopoulos, G., Ahmed, R. A. ve Hyndman, R. J. (2009). "Hierarchical Forecasts for Australian Domestic Tourism". *International Journal of Forecasting*, 25(1): 146-166.
- Azadi, S., Amiri, H. ve Rakhshandehroo, G. R. (2016). Evaluating the Ability of Artificial Neural Network and PCA-M5P Models in Predicting Leachate COD Load in Landfills. *Waste Management*.
- Azofra, D., Martínez, E., Jiménez, E., Blanco, J., Azofra, F. ve Saenz-Díez, J. C. (2015). Comparison of the Influence of Photovoltaic and Wind Power on the Spanish Electricity Prices by Means of Artificial Intelligence Techniques. *Renewable and Sustainable Energy Reviews*, 42: 532-542.
- Bartolomei, S. M., Swee Barros, R. C., Basgalupp, M. P., Ruiz, D. D., de Carvalho, A. ve Freitas A. (2010). Evolutionary Model Tree Induction. In *Proceedings of the 2010 ACM Symposium on Applied Computing (SAC '10)*. ACM, New York, NY, USA: 1131-1137
- Bilen, H. (2009). Bankacılık Sektöründe Personel Seçimi ve Performans Değerlendirilmesine İlişkin Veri Madenciliği Uygulaması, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- Bolt, G. J. (1981). *Market and Sales Forecasting, A Total Approach*. (Second Edition). London: Kogan Page Limited.
- Box, G. (1970). EP, and Jenkins, G. M. "Time Series Analysis, Forecasting, and Control" San Francisco: Holden-Day.
- Bowerman B. L. ve O'Connell, R. T. (1993). *Forecasting and Time Series: An Applied Approach*, Third Edition, Duxbury Thomson Learning, Pacific Grove, CA, USA.

- Breiman, L., Friedman, J. H., Olshen, R. A. ve Stone C. J. (1984). *Classification and Regression Trees*, Wadsworth & Brooks, Monterey, CA.
- Brown, R. G. (1959) *Statistical Forecasting for Inventory Control*, McGraw-Hill, New York.
- Brown, R. G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series*, Englewood Cliffs, Prentice-Hall, New Jersey, 1963.
- Chatzis, S. P. ve Yiannis, D. (2012). "Nonparametric Mixtures of Gaussian Processes With Power-Law Behavior." *Neural Networks and Learning Systems*, IEEE Transactions: 1862-1871.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. ve Wirth, R. (2000). *CRISP-DM 1.0: Step by Step Data Mining Guide*, SPSS, Chicago, IL, USA.
- Chase Jr, Charles W. (2013). *Demand-Driven Forecasting: A Structured Approach To Forecasting*. John Wiley & Sons.
- Chien, C. F. ve Chen, L. F. (2008). Data Mining to Improve Personnel Selection and Enhance Human Capital: a Case Study in High-Technology Industry, *Expert Systems with Applications*, 34: 280-290.
- Ching, W. K. ve Pong, M. K. (2002). *Advances in Data Mining and Modeling*, 1st ed., World Scientific, Hong Kong, China.
- Chuang, C. C. ve Lee, Z. J. (2008). *Hybrid Robust Support Vector Machines for Regression With Outliers*, Elsevier
- Çağlar, T. (2007). *Talep Tahmininde Kullanılan Yöntemler ve Fens Teli Üretimi Yapan Bir İşletmede Uygulanması*, Yüksek Lisans Tezi, Kırıkkale Üniversitesi, Fen Bilimleri Enstitüsü, Kırıkkale, 2007.
- Çakmak, Z., Uzgören, N. ve Keçek, G. (2005) Kümeleme Analizi Teknikleri ile İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerinin İncelenmesi, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 12: 15-36.
- Çalış, A. (2013). *Veri Madenciliği Yaklaşımı İle Bireysel Müşterilerin Kredi Ödeme Performanslarının Değerlendirilmesi*, Yüksek Lisans Tezi, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Kocaeli.
- Çelik, E., Atalay, M. ve Bayer, H. (2014) *Yapay Sinir Ağları ve Destek Vektör Makineleri İle Deprem Tahmininde Sismik Darbelerin Kullanılması*, 2014 IEEE 22nd Signal Processing and Communications Applications Conference (SIU 2014): 730-733.
- Çiftci, N. (2014). *Türkiye’de Cari Açık, Reel Döviz Kuru ve Ekonomik Büyüme Arasındaki İlişkiler: Eş Bütünleşme Analizi*. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*.

- Damla, C. ve Yalcin, A. (2007). Flood Prediction Using Time Series Data Mining. *Journal of Hydrology*, 333: 305–316.
- Elmas, Ç. (2011). *Yapay Zeka Uygulamaları*, Ankara: Seçkin Yayıncılık, 2. Baskı
- Emel, G. G. ve Taşkın, Ç. (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(2).
- Etemad-Shahidi, A. ve Mahjoobi, J. (2009). Comparison Between M5' Model Tree and Neural Networks for Prediction of Significant Wave Height in Lake Superior. *Ocean Engineering*, 36(15): 1175-1181.
- Fayyad U., Shapiro G. ve Smyth P. (1996C) From Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence, 17: 37-54.
- Fildes, R. (2006) "The Forecasting Journals and Their Contribution to Forecasting Research: Citation Analysis and Expert Opinion", *International Journal of Forecasting*, 22(3): 415–432.
- Fildes, R. ve Beard, C. (1992). "Forecasting Systems For Production and Inventory Control", *International Journal of Operations and Production Management*, 12(5): 4–27.
- Frank E., Wang Y., Inglis S., Holmes G. ve Witten I. H. (1998). Using Model Trees for Classification. *Mach. Learn.* 32(1): 63-76
- Gardner, E. S. Jr. ve McKenzie, E. (1985). Forecasting Trends in Time Series, *Management Science*, 31: 1237-1246.
- Göndör, M. ve Bresfelean, V. P. (2012). REPTree and M5P for Measuring Fiscal Policy Influences on the Romanian Capital Market During 2003–2010. *International Journal of Mathematics and Computers in Simulation*, 6(4): 378-386.
- Han, J. ve Kamber, M. (2012). *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, USA.
- Hand, D., Mannila, H. ve Smyth, P. (2001). *Principles of Data Mining*, The MIT Press, Cambridge, Massachusetts, USA.
- Hanke, J. E. ve Reitsch, A. G. (1998) *Business Forecasting*, Prentice-Hall, New Jersey.
- Hanke, J. E., Wichern, D. W. ve Reitsch, A. G. (2001) *Business Forecasting*, Prentice Hall.
- Hicham, A. ve Mohamed, B. (2012, November). A Model for Sales Forecasting Based on Fuzzy Clustering and Back-Propagation Neural Networks With Adaptive Learning Rate. In *Complex Systems (ICCS), 2012 International Conference on*, IEEE, 1-5.
- Holt, C. C. (1957). "Forecasting Seasonals and Trends By Exponentially Weighted Averages", *O. N. R. Memorandum*, 52(1).

- Holt, C. C. (2004). "Forecasting Seasonals and Trends By Exponentially Weighted Moving Averages", *International Journal of Forecasting*, 20(5).
- Hong, S. ve Zheng, Z. (2012). "Remaining Useful Life Prognosis of Bearing Based on Gauss Process Regression." *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*. IEEE.
- Hoonakker, F., Lachiche, N., Varnek, A. ve Wagner, A. (2010, June). A Representation To Apply Usual Datamining Techniques To Chemical Reactions. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*: 318-326.
- Hsu, C. C., Huang, Y. P. ve Chang, K. W. (2008) "Extended Naïve Bayes Classifier for Mixed Data", *Expert Systems with Applications*, 35: 1080-1083.
- Hu, C. (2002). *Advanced Tourism Demand Forecasting: Artificial Neural Network and Box-Jenkins Modeling*.
- Huang, C. F. (2012). A Hybrid Stock Selection Model Using Genetic Algorithms and Support Vector Regression. *Applied Soft Computing*, 12(2): 807-818.
- Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G. ve Shang, H. L. (2011). Optimal Combination Forecasts For Hierarchical Time Series. *Computational Statistics and Data Analysis*, 55(9): 2579-2589.
- Hyndman, R. J. ve Athanasopoulos, G. (2014). *Forecasting: Principles and Practice*. OTexts.
- Hyndman, R. J., Ahmed, R. A., Shang, H. L. ve Wang, E. (2014). *hts: Hierarchical and Grouped Time Series*. R Package Version, 4.
- Irmak, S. (2009). *Veri Madenciliği Yöntemleri İle Sağlık Sektörü Veritabanlarında Bilgi Keşfi: Tanımlayıcı Ve Kestirimci Model Uygulamaları*. Doktora Tezi, Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü, Antalya.
- İkiz, S. (2015). *Bina Kabuğunda Paslanmaz Çelik Kullanımının Sürdürülebilirlik Açısından İncelenmesi* (Doctoral dissertation, Fen Bilimleri Enstitüsü).
- Jenhani, I., Ben Amor, N. ve Elouedi, Z. (2008). "Decision Trees as Possibilistic Classifiers", *International Journal of Approximate Reasoning*, 48: 784-807.
- Jiang, H. ve He, W. (2012). Grey Relational Grade in Local Support Vector Regression for Financial Time Series Prediction. *Expert Systems with Applications*, 39(3): 2256-2262.
- Gu, J., Zhu, M. ve Jiang, L. (2011). Housing Price Forecasting Based on Genetic Algorithm and Support Vector Machine. *Expert Systems with Applications*, 38(4): 3383-3386.
- Kaluç, E. ve Taban E. (2007) "Paslanmaz Çelikler, Geliştirilen Yeni Türleri ve Kaynak Edilebilirlikleri", *TMMOB*, İstanbul.

- Kamruzzaman, J. (Ed.). (2006). Artificial Neural Networks In Finance and Manufacturing. IGI Global.
- Karaatlı, M., Helvacıoğlu, Ö. C., Ömürbek, N. ve Tokgöz, G. (2012). Yapay Sinir Ağları Yöntemi İle Otomobil Satış Tahmini. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 8(17): 87-100.
- Kaytez, F., Taplamacioglu, C., Cam, E. ve Hardalac, F. (2015) Forecasting Electricity Consumption: A Comparison of Regression Analysis, Neural Networks and Least Squares Support Vector Machines. *Electrical Power and Energy Systems*, 67: 431–438.
- Khah, M. S., Moghaddam, M. P. ve Sheikh-El-Eslami, M. K. (2011) “Price Forecasting of Day-ahead Electricity Markets Using a Hybrid Forecast Method”, *ScienceDirect Journal of Energy Conversion and Management*, 52: 2165-2169.
- Khashei, M., ve Bijari, M. (2010). “An Artificial Neural Network (P,D,Q) Model for Time Series Forecasting“, *ScienceDirect Journal of Expert System with Applications*, 37: 479-489.
- Kheirkhah, A., Azadeh, A., Saberi, M., Azaron, A. ve Shakouri, H. (2013). Improved Estimation of Electricity Demand Function by Using of Artificial Neural Network, Principal Component Analysis and Data Envelopment Analysis. *Computers & Industrial Engineering*, 64(1): 425-441.
- Kim, K. J. (2003). Financial Time Series Forecasting Using Support Vector Machines. *Neurocomputing*, 55(1): 307-319.
- Kobu, B. (1994). Üretim Yönetimi, Avcıol Basım-Yayım, 8. Baskı, İstanbul.
- Kuhn, M. ve Kjell, J. (2013). Applied Predictive Modeling. New York: Springer.
- Kuo, R. J. ve Xue, K. C. (1998) “An Intelligent Sales Forecasting System Through Integration of Artificial Neural Network and Fuzzy Neural Network”, *Computers In Industry*, 37: 1-15.
- Kusters, U., McCullough, B. D. ve Bell, M. (2006). “Forecasting Software: Past, Present and Future”, *International Journal of Forecasting*, 22(3): 599–615.
- Lantz, B. (2013). Machine Learning With R. Packt Publishing Ltd.
- Larose, D. T. (2005). Discovering Knowledge in Data: an Introduction in Data Mining, 1st ed., Wiley, USA.
- Lin, M. C., Tserng, H. P., Ho, S. P. ve Young. D. L. (2012). A Novel Dynamic Progress Forecasting Approach for Construction Projects. *Expert Systems with Applications* 39: 2247–2255.

- Lin, L. Q. W. ve Adel W. S. (2016) "A Combined M5P Tree and Hazard-Based Duration Model for Predicting Urban Freeway Traffic Accident Durations." *Accident Analysis & Prevention* 91: 114-126.
- Linoff, G. S. ve Berry, M. J. A. (2011). *Data Mining Techniques for Marketing, Sales and Customer Relationship Management*, 3rd.ed., Wiley, Canada.
- Lu, C. J. (2014). Sales Forecasting of Computer Products Based on Variable Selection Scheme and Support Vector Regression. *Neurocomputing*, 128: 491-499.
- Lu, C. J., Lee, T. S. ve Lian, C. M. (2012). Sales Forecasting For Computer Wholesalers: A Comparison of Multivariate Adaptive Regression Splines and Artificial Neural Networks. *Decision Support Systems*, 54(1): 584-596.
- Luxhøj, J. T., Riis, J. O. ve Stensballe, B. (1996). A Hybrid Econometric—Neural Network Modeling Approach for Sales Forecasting. *International Journal of Production Economics*, 43(2): 175-192.
- Makridakis, S. (1993) "Accuracy Measures: Theoretical and Practical Concerns", *International Journal of Forecasting*, 9(527).
- Makridakis, S. ve Hibon, M. (1991). Exponential Smoothing: The Effect of Initial Values and Loss Functions on Post-Sample Forecasting Accuracy. *International Journal of Forecasting*, 7(3): 317-330.
- Makridakis, S., Wheelwright, S. C. ve Hyndman, R. J. (2008). *Forecasting Methods and Applications*. John Wiley & Sons.
- Malhotra, N. ve Birks, D. (2007). *Marketing Research-An Applied Approach*, Pearson Education (3.ed).
- Marvuglia, A. ve Messineo, A. (2012) "Using Recurrent Artificial Neural Networks to Forecast Household Electricity Consumption", *Journal of Energy Procedia*, 14: 45-55.
- McKenzie, E. (1986) "Error Analysis for Winters' Additive Seasonal Forecasting Systems", *International Journal of Forecasting*, 2(373).
- Megahed, A., Yin, P. ve Nezhad, H. R. M. (2016, June). An Optimization Approach to Services Sales Forecasting in a Multi-Staged Sales Pipeline. In *Services Computing (SCC), 2016 IEEE International Conference on*: 713-719, IEEE.
- Mustakim, M., Buono, A. ve Hermadi, I. (2016). Performance Comparison Between Support Vector Regression and Artificial Neural Network for Prediction of Oil Palm Production. *Jurnal Ilmu Komputer dan Informasi*, 9(1): 1-8.
- Oğuzlar, A. (2003) "Veri Ön İşleme" *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 21.

- Özdemir, A. F., Yıldıztepe, E. ve Binar, M. (2010). İstatistiksel Yazılım Geliştirme Ortamı: R.XII. Akademik Bilişim Konferansı, 10-12.
- Özdemir, A. ve Özdemir, A. (2006) “Talep Tahminlemesinde Kullanılan Yöntemlerin Karşılaştırılması: Seramik Ürün Grubu Firma Uygulaması”, Ege Academic Review: 105-114.
- Özer, A. ve Bahçeci, E. (2009). “Aisi 410 Martensitik Paslanmaz Çeliklerin Kesici Takım Ve Kaplamasına Bağlı İşlenebilirliği”. *Journal of the Faculty of Engineering & Architecture of Gazi University*, 24(4).
- Özkan, Y. (2008). Veri Madenciliği Yöntemleri, Papatya Yayıncılık Eğitim, İstanbul.
- Öztemel, E. (2006). Yapay Sinir Ağları, Papatya Yayıncılık, İstanbul.
- Pyle, D. (1999) Data Preparation for Data Mining, Academic Press, San Diego, CA.
- Petelin, D. et al. (2011a.) "Financial Modeling Using Gaussian Process Models." Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2011 IEEE 6th International Conference on. Vol. 2. IEEE.
- Quinlan, J. R. (1986). “Induction of Decision Trees”, *Machine Learning*, 1, 81-106.
- Quinlan, J. R. (1993). C4.5: Programs for Machine Learning, Morgan Kaufmann, San Francisco, CA.
- Pektaş, A. O. (2013). SPSS İle Veri Madenciliği, Dikeyksen Yayın Dağıtım, İstanbul.
- Racine, J. ve Hyndman, R. (2002). Using R to Teach Econometrics. *Journal of Applied Econometrics*, 17(2): 175-189.
- Romero, E. ve Alquézar, R. (2012). Comparing Error Minimized Extreme Learning Machines and Support Vector Sequential Feed-Forward Neural Networks. *Neural Networks*, 25: 122-129.
- Sevüktekin, M. ve Naregeleçekenler, M. (2010) "Ekonometrik Zaman Serileri Analizi, Geliştirilmiş 3. Baskı, Ankara: Nobel Kitabevi.
- Shmueli, G, Patel, N. R. ve Bruce, P. C. (2007). Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner, John Wiley & Sons, Hoboken, NJ, USA.
- Shuhaida, I., Ani, S. ve Rhaidah, S. (2011). A Hybrid Model of Self-Organizing Maps (SOM) and Least Square Support Vector Machine (LSSVM) For Time-Series Forecasting. *Expert Systems with Applications*, 38: 10574–10578.
- Shmueli, G, Patel, N. R. ve Bruce, P. C. (2007) Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner, John Wiley & Sons, Hoboken, NJ, USA.

- Silahtaroglu, G. (2013) "Veri Madenciliği" Papatya Yayınları, İstanbul.
- Snyder, R. D. (1985) "Recursive Estimation Of Dynamic Linear Statistical Models", Journal of the Royal Statistical Society, 47(272).
- Solazzi, M. ve Uncini, A. (2004) "Regularising Neural Networks Using Flexible Multivariate Activation Function", Neural Networks, 17: 247-260.
- Sun, Z. L., Choi, T. M., Au, K. F. ve Yu, Y. (2008). Sales Forecasting Using Extreme Learning Machine With Applications In Fashion Retailing. Decision Support Systems, 46(1): 411-419.
- Suner, A. ve Çelikoğlu, C. C. (2010). Toplum Tabanlı Bir Çalışmada Çoklu Uygunluk Analizi ve Kümeleme Analizi ile Sağlık Kurumu Seçimi, Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 25: 43-55.
- Sönmez, U., Çavuşoğlu, N. ve Ceyhun, V. (2016). Orbital Tig Kaynak Yöntemiyle Kaynak Edilmiş Dupleks Paslanmaz Çeliklerin Mekanik, Metalurjik Ve Korozyon Özellikleri. Engineer & the Machinery Magazine, (673).
- Tan, P. N. vd., (2006) Introduction to Data Mining, 1st ed., Pearson International Edition, USA.
- Tanyaş, M. ve Baskak, M. (2008). Üretim Planlama ve Kontrol, 3. Baskı, İrfan Yayıncılık, İstanbul.
- Tarı, R. (2011). Ekonometri. Umuttepe Yayınları.
- Taşkın, A. G. D. Ç. ve Emel, G. G. (2010). Veri Madenciliğinde Kümeleme Yaklaşımları ve Kohonen Ağları İle Perakendecilik Sektöründe Bir Uygulama. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 15(3).
- Tekin, M. (2009). Üretim Yönetimi Cilt 1, 6. Baskı, Günay Ofset, Konya.
- Thomassey, S. ve Fiordaliso, A. (2006). A Hybrid Sales Forecasting System Based on Clustering and Decision Trees. Decision Support Systems, 42(1): 408-421.
- Tokpınar, İ. (2014). Tedarik Zinciri Yönetiminde Talep Tahmin Doğruluğunu Arttırmak İçin Radar Diyagramının Kullanımı. Yüksek Lisans Tezi, Başakşehir Üniversitesi Fen Bilimleri Enstitüsü İstanbul.
- Tolun, S. (2008). Destek Vektör Makineleri: Banka Başarısızlığın Tahmini Üzerine Bir Uygulama, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 2008
- Tuffery, S., (2011) Data Mining and Statistics For Decision Making, 1st ed., Wiley, USA
- Uçkun, N. (2010). Paslanmaz Çelik Sektöründe Fiyat Riskinden Korunmak İçin Krom Nikel Ham Çelik ve EURO/USD Paritesinin Fiyatlara Etkilerinin İncelenmesi. Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 11(1).

- Xiao, T. ve Yang, D. (2008). Price and Service Competition of Supply Chains With Risk-Averse Retailers Under Demand Uncertainty. *International Journal of Production Economics*, 114(1): 187-200.
- Wang, F. K. vd. (2011). "Using Adaptive Networkbased Fuzzy Inference System to forecast Automobile Sales", *Journal of Expert System with Applications*, 38: 10587-10593.
- Wang, F. K. ve Du, T. (2014) Implementing Support Vector Regression With Differential Evolution to Forecast Motherboard Shipments. *Expert Systems with Applications*. 41: 3850–3855.
- Wilson, J. H. ve Keating, B. (2002). *Business Forecasting with Accompanying Excel-Based Forecast XTM Software*. Boston: McGraw-Hill.
- Winters, P. R. (1960). "Forecasting Sales by Exponentially Weighted Moving Averages", *Management Science*, (6): 324.
- Written, I. H. ve Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition, Morgan Kaufmann Publishers, San Francisco, CA, USA, 2005.
- Yang, L. ve Baolin, L. (2016). "The Combination Forecasting Model of Auto Sales Based on Seasonal Index and RBF Neural Network." *Decision Support Systems* 9.1.
- Yıldız, B., Doğanay, M. ve Aktaş, R. (2003). "Mali Başarısızlığın Öngörülmesi: İstatistiksel Yöntemler ve Yapay Sinir Ağı Karşılaştırılması", *Ankara Üniversitesi Siyasal Bilgiler Fakültesi Dergisi*, 58(4).
- Vhatkar, S. ve Dias, J. (2016). Oral-Care Goods Sales Forecasting Using Artificial Neural Network Model. *Procedia Computer Science*, 79: 238-243.
- Yu, X., Qi, Z. ve Zhao, Y. (2013) Support Vector Regression for Newspaper/Magazine Sales Forecasting. *Procedia Computer Science* 17: 1055 – 1062
- Zhan, C., Gan, A. ve Hadi, M. (2011). Prediction of Lane Clearance Time of Freeway Incidents Using the M5P Tree Algorithm. *IEEE Transactions on Intelligent Transportation Systems*, 12(4): 1549-1557.
- Zhang, G. P., (2003) "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model", *Neurocomputing*, 50: 159-175.
- Zhang, G. (2004). "Business Forecasting with Artificial Neural Networks: An Overview, Neural Networks in Business Forecasting", İdea Group Inc, USA: 4.
- Zhang, G., Patuwo, G. E. ve Hu, M. Y. (1998). "Forecasting with Artificial Neural Networks: The State of The Arts", *International Journal of Forecasting*, 4(1): 35.

Zhu, B. ve Wei, Y. (2013). Carbon Price Forecasting With a Novel Hybrid ARIMA and Least Squares Support Vector Machines Methodology. 41(2013): 517–524

CRISP-DM, SAS Enterprise Miner,

<http://www.sas.com/offices/europe/uk/technologies/analytics/atamining/miner/semma.htm>,

(erişim tarihi: 01.01.2015)

InfoMine USA, Inc., <http://www.infomine.com/>, (erişim tarihleri:01.01.2014-06.06.2016)

The International Stainless Steel Forum (ISSF) <http://www.worldstainless.org/>, (erişim tarihleri: 01.01.2014-06.06.2016)

Türkiye Cumhuriyet Merkez Bankası, Elektronik Veri Dağıtım Sistemi, <http://evds.tcmb.gov.tr/>, (erişim tarihleri: 01.01.2014-06.06.2016).

Türkiye İstatistik Kurumu, <http://www.tuik.gov.tr/>, (erişim tarihleri:01.01.2014-06.06.2016).

TMMOB Metalurji ve Malzeme Endüstrisi Portalı,

http://www.metalurji.org.tr/dergi/dergi162/d162_5761.pdf, (erişim tarihi: 01.01.2014).

ÖZGEÇMİŞ

Adı ve SOYADI : Orhan ECEMİŞ

Eğitim Durumu

Mezun Olduğu Lise : Fevzi Çakmak Lisesi, Kayseri, 1994

Lisans Diploması : Gazi Üniversitesi, Endüstriyel Sanatlar Eğitim Fakültesi,
Bilgisayar Eğitimi Bölümü, Ankara, 2000

Yüksek Lisans Diploması : Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Eğitimi
Ana Bilim Dalı, Ankara, 2004

Tez Konusu : Bir Çevrim İçi Sınav Sistemi Hakkında Öğretmen Görüşleri

Doktora Diploması : Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri
Ana Bilim Dalı, Antalya, 2016

Tez Konusu : Veri Madenciliği Yöntemleriyle Paslanmaz Çelik Sektöründe
Satış Tahmini

Yabancı Dil : İngilizce

E-Posta : oecemis@gantep.edu.tr