



AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



Nedret TOSUN

OECD ÜLKELERİNİN YAŞ MEYVE SEBZE İTHALATININ VERİ MADENCİLİĞİ
YÖNTEMLERİ İLE TAHMİNLENMESİ

Ekonometri Ana Bilim Dalı
Doktora Tezi

Antalya, 2020



AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



Nedret TOSUN

OECD ÜLKELERİNİN YAŞ MEYVE SEBZE İTHALATININ VERİ MADENCİLİĞİ
YÖNTEMLERİ İLE TAHMİNLENMESİ

Danışman

Doç. Dr. Sezgin IRMAK

Ekonometri Ana Bilim Dalı

Doktora Tezi

Antalya, 2020

Akdeniz Üniversitesi
Sosyal Bilimler Enstitüsü Müdürlüğüne,

Nedret TOSUN'un 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 : Doç. Dr. Hacer GÜNER GÖREN (İmza)

Üye : Doç. Dr. Kenan Oğuzhan ORUÇ (İmza)

Tez Başlığı: OECD ÜLKELERİNİN YAŞ MEYVE SEBZE İTHALATININ VERİ MADENCİLİĞİ YÖNTEMLERİ İLE TAHMİNLENMESİ
--

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

Tez Savunma Tarihi : 20/11/2020

Mezuniyet Tarihi : 17/12/2020

(İmza)
Prof. Dr. Suat KOLUKIRIK
Müdür

AKADEMİK BEYAN

Doktora Tezi olarak sunduğum “OECD Ülkelerinin Yaş Meyve Sebze İthalatının Veri Madenciliği Yöntemleri ile Tahminlenmesi” adlı bu çalışmanın, akademik kural ve etik değerlere uygun bir biçimde tarafımdan yazıldığını, yararlandığım bütün eserlerin kaynakçada gösterildiğini ve çalışma içerisinde bu eserlere atıf yapıldığını belirtir; bunu şerefimle doğrularım.

İmza

Nedret TOSUN





AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
TEZ ÇALIŞMASI ORJİNALLİK RAPORU
BEYAN BELGESİ



SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE

ÖĞRENCİ BİLGİLERİ	
Adı-Soyadı	Nedret TOSUN
Öğrenci Numarası	20118605202
Enstitü Ana Bilim Dalı	Ekonometri
Programı	Doktora
Programın Türü	() Tezli Yüksek Lisans (X) Doktora () Tezsiz Yüksek Lisans
Danışmanın Unvanı, Adı-Soyadı	Doç. Dr. Sezgin IRMAK
Tez Başlığı	OECD ÜLKELERİNİN YAŞ MEYVE SEBZE İTHALATININ VERİ MADENCİLİĞİ YÖNTEMLERİ İLE TAHMİNLENMESİ
Turnitin Ödev Numarası	1467284413

Yukarıda başlığı belirtilen tez çalışmasının a) Kapak sayfası, b) Giriş, c) Ana Bölümler ve d) Sonuç kısımlarından oluşan toplam 177 sayfalık kısmına ilişkin olarak, 07/12/2020 tarihinde tarafımdan Turnitin adlı intihal tespit programından Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nda belirlenen filtrelemeler uygulanarak alınmış olan ve ekte sunulan rapora göre, tezin benzerlik oranı;

alıntılar hariç % 5,

alıntılar dahil % 6'dır.

Danışman tarafından uygun olan seçenek işaretlenmelidir:

(X) Benzerlik oranları belirlenen limitleri aşmıyor ise;

Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporu'nun doğruluğunu onaylarım.

() Benzerlik oranları belirlenen limitleri aşıyor, ancak tez/dönem projesi danışmanı intihal yapılmadığı kanısında ise;

Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporu'nun doğruluğunu onaylar ve Uygulama Esasları'nda öngörülen yüzdelik sınırlarının aşılmasına karşın, aşağıda belirtilen gerekçe ile intihal yapılmadığı kanısında olduğumu beyan ederim.

Gerekçe:

Benzerlik taraması yukarıda verilen ölçütlerin ışığı altında tarafımca yapılmıştır. İlgili tezin orijinallik raporunun uygun olduğunu beyan ederim.

07/12/2020

(imzası)

Doç. Dr. Sezgin IRMAK

İÇİNDEKİLER

ŞEKİLLER LİSTESİ.....	v
TABLOLAR LİSTESİ.....	vi
KISALTMALAR LİSTESİ	vii
ÖZET.....	ix
SUMMARY	x
TEŞEKKÜR.....	xi
ÖNSÖZ.....	xii

BİRİNCİ BÖLÜM VERİ MADENCİLİĞİ

1.1. Veri Madenciliği Kavramları.....	1
1.2. Veri Madenciliği Sürecinin Veri Kaynakları.....	3
1.2.1. Veritabanları.....	3
1.2.2. Veri Ambarı.....	3
1.3. Veri Madenciliğinin Tarihsel Gelişimi.....	4
1.4. Veri Madenciliği Süreci	6
1.4.1. Problemin Tanımlanması ve Verinin Anlaşılması.....	7
1.4.2. Verilerin Hazırlanması (Ön İşleme)	9
1.4.2.1. Veri Kalitesi ve Veri Önışlemenin Önemi.....	9
1.4.2.2. Veri Ön İşlemenin Başlıca Görevleri	10
1.4.2.3. Veri Toplama	11
1.4.2.4. Veri Değerlendirme	11
1.4.2.5. Veri Temizleme ve Bütünleştirme	11
1.4.2.6. Veri Seçimi	12
1.4.2.7. Veri Dönüştürme	12
1.4.3. Modelleme.....	14
1.4.4. Modelin Değerlendirilmesi	14
1.4.4.1 Modelin Performansını Değerlendirme Yöntemi	15
1.4.5. Modelin Kullanılması	16

İKİNCİ BÖLÜM DIŞ TİCARET

2.1. İhracat.....	17
2.1.1. İhracatın Önemi ve İşletmelere Sağladığı Faydalar.....	18
2.2. İthalat.....	19

2.3. Talep Tahmini (Pazar Araştırması)	19
2.3.1. Talep Tahmini ve Pazar Araştırmasının Önemi	21

ÜÇÜNCÜ BÖLÜM

MODELLEME TEKNİKLERİ VE ALGORİTMALAR

3.1. Modelleme Teknikleri	23
3.1.1. Tanımlayıcı Modeller	24
3.1.2. Tahminleyici Modeller	24
3.2. Makine Öğrenmesi Algoritmaları	25
3.2.1. Karar Ağaçları	27
3.2.1.1. Karar Ağacı Algoritmaları	29
3.2.1.2. CART (Classification and Regression Tree-Sınıflandırma ve Regresyon Ağaçları) Algoritması	29
3.2.1.3. Karar Ağaçlarının Avantaj ve Dezavantajları	30
3.2.2. Torbalama (Bagging)	32
3.2.3. Rastgele Orman Algoritması	34
3.2.3.1. Rastgele Orman Algoritmasının Parametreleri	34
3.2.3.2. Rastgele Orman Algoritmasının İşleyişi	35
3.2.3.3. Rastgele Orman Algoritmasının Avantajları ve Dezavantajları	36
3.2.3.4. Rastgele Orman Algoritmasında Özniteliklerin Önemi	38
3.2.4. Artırma (Boosting)	38
3.2.5. XGBoost (Extreme Gradient Boosting) Algoritması	40
3.2.5.1. XGBoost Algoritmasının Parametreleri	41
3.2.5.2. XGBoost Algoritmasının İşleyişi	42
3.2.5.3. XGBoost Algoritmasının Avantaj ve Dezavantajları	45
3.2.5.4. XGBoost Algoritmasında Özniteliklerin Önemi	45
3.2.6. Yapay Sinir Ağları	46
3.2.6.1. Yapay Sinir Ağlarının Genel Özellikleri ve Avantajları	46
3.2.6.2. Yapay Sinir Ağlarının Önemli Dezavantajları	47
3.2.6.3. Yapay Sinir Ağlarının Yapısı ve Temel Elemanları	47
3.2.6.4. YSA’larda Katman Sayısı	51
3.2.6.5. Çok Katmanlı Algılayıcılar (Geri Yayılım Ağları)	51
3.3. Literatürde Yapılmış Uygulama Örnekleri	55
3.3.1. Veri Madenciliği Uygulamaları	55
3.3.2. Rastgele Orman Algoritması Uygulamaları	57
3.3.3. XGBoost Algoritması Uygulamaları	59

3.3.4. Yapay Sinir Ağları Uygulamaları.....	60
3.3.5. İhracat ve İthalat ile İlgili Diğer Çalışmalar ve Araştırmalarda Kullanılan Öznitelikler	64

DÖRDÜNCÜ BÖLÜM

OECD ÜLKELERİNİN YAŞ MEYVE SEBZE İTHALATININ VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ

4.1. Problemin Tanımlanması.....	69
4.1.1. Çalışmanın Gerekçesi	69
4.1.2. Kullanılacak Yöntemlerin Belirlenmesi	72
4.1.3. Kullanılacak Yöntemlerin Değerlendirme Ölçütleri	73
4.2. Verinin Anlaşılması	75
4.2.1. Çalışmada Kullanılacak Özniteliklerin ve Veri Kaynaklarının Belirlenmesi	75
4.2.2. Çalışmanın Kısıtları ve Risklerinin Belirlenmesi.....	79
4.2.3. Çalışmada Kullanılacak Yazılımın ve Donanımın Belirlenmesi	79
4.3. Verinin Hazırlanması	80
4.3.1. Verinin Toplanması, Değerlendirilmesi ve Bütünleştirilmesi.....	80
4.3.2. Veri Dönüştürme	82
4.3.3. Veri Setinin Eğitim ve Test Verisi olarak Ayrılması.....	85
4.4. Model Oluşturma	85
4.4.1. K-Kat Çapraz Doğrulama	86
4.4.2 Parametre Değerlerinin Belirlenmesi.....	87
4.5. Modellerin Değerlendirilmesi.....	88
4.5.1. XGBoost Algoritması ile Modelleme	88
4.5.2. Rastgele Orman Algoritması ile Modelleme.....	92
4.5.3. Yapay Sinir Ağları ile Modelleme.....	97
4.6. Yöntemlerin Kıyaslanması	100
SONUÇ	103
KAYNAKÇA	108
EKLER	122
Ek 1-OECD Ülkeleri ve Ülke Kodları Listesi	122
Ek 2-Yaş Meyve Sebze Ürünleri GTİP'leri ve Açıklamaları	123
Ek 3-Yapay Sinir Ağları Sonuçları	125
Ek 4-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında R2 Kriteri Açısından Ortalama Performansları	134

Ek 5-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında <i>RMSE</i> Kriteri Açısından Ortalama Performansları	136
Ek 6-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında <i>MAE</i> Kriteri Açısından Ortalama Performansları	139
Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri.....	142
ÖZGEÇMİŞ	157



ŞEKİLLER LİSTESİ

Şekil 1.1 Veritabanlarında Bilgi Keşfi Sürecini Oluşturan Adımlar	2
Şekil 1.2 Veritabanı Sistemi Teknolojisinin Gelişimi	5
Şekil 1.3 Veri Madenciliği Süreci (CRISPDM), ...	6
Şekil 3.1 Denetimli Makine Öğrenmesinin Aşamaları.....	26
Şekil 3.2 Karar Ağacı Örneği	28
Şekil 3.3 İşçi Görüşmeleri Verisine İlişkin Örnek Bir Karar Ağacı.....	28
Şekil 3.4 Karar Ağacı Algoritmalarının Topluluk Algoritmalarına Evrimi.....	32
Şekil 3.5 Torbalama Yaklaşımı	33
Şekil 3.6 Rastgele Orman Algoritması Akış Şeması	36
Şekil 3.7 Torbalama ve Artırma Algoritmalarının Çalışma Prensibi	38
Şekil 3.8 Schapire (1990) Tarafından Önerilen İlk Artırma Yaklaşımının Görsel İfadesi	39
Şekil 3.9 XGBoost İçin Akış Şeması.....	40
Şekil 3.10 Yapay Sinir Hücresi Gösterimi	48
Şekil 3.11 İki Gizli Katmandan Oluşan Bir Çok Katmanlı Algılayıcı	52
Şekil 3.12 İleri Beslemeli Geri Yayılımlı Yapay Sinir Ağı Örneği.....	53
Şekil 3.13 Eğitim ve Doğrulama Hatalarının Eğitim Süresine Göre Değişimi.....	55
Şekil 4.1 Analiz Süreci İş Akışı.....	68
Şekil 4.2 “Weight” Kriterine Göre XGBoost Algoritmasında Özniteliklerin Önem Düzeyleri	89
Şekil 4.3 Rastgele Orman Algoritmasında Özniteliklerin Önem Düzeyleri	94
Şekil 4.4 İki Katmanlı (25,25) YSA Topografisi	98

TABLOLAR LİSTESİ

Tablo 1.1 Veri Ambarı ve Veri Tabanı Özelliklerini Kıyaslama Tablosu	4
Tablo 3.1 Torbalama Algoritması.....	33
Tablo 3.2 Rastgele Orman Algoritması	35
Tablo 3.3 Artırma Algoritması	39
Tablo 3.4 XGBoost Algoritması.....	42
Tablo 3.5 Yapay Sinir Ağlarında Kullanılan Bazı Toplama Fonksiyonları.....	49
Tablo 3.6 Yapay Sinir Ağlarında Kullanılan Bazı Aktivasyon Fonksiyonları.....	50
Tablo 4.1 Türkiye'nin Üretiminde Önemli Ürünler ve 2018 Yılı Dünya Üretiminden Aldığı Pay.....	70
Tablo 4.2 Türkiye'nin İhracatında Önemli Ürünler ve 2019 Yılı Dünya İhracatından Aldığı Pay.....	71
Tablo 4.3 Veri Seti Öznitelikleri	76
Tablo 4.4 Veri Setindeki Değişkenlere Ait Tanımlayıcı İstatistikler	81
Tablo 4.5 Veri Setinde Ülke Kodlarında Gerçekleştirilen Dönüşüm Örneği	82
Tablo 4.6 Veri Setinde Ürün Kodlarında Gerçekleştirilen Dönüşüm Örneği	82
Tablo 4.7 Veri Seti Öznitelikleri	83
Tablo 4.8 Veri Setinde Ülke ve Ürün Kodlarında Gerçekleştirilen Dönüşüm Sonrası Veri Setinin Durumuna ilişkin Örnek.....	84
Tablo 4.9 “5-Kat” Çapraz Doğrulama Örnek Gösterim Tablosu	86
Tablo 4.10 Analiz Aşamasında Kullanılan Modellerin Parametrelerinin Optimizasyonu İçin Geliştirilen Deney Tasarımı	87
Tablo 4.11 XGBoost Algoritması Performans Değerleri.....	89
Tablo 4.12 XGBoost Algoritması ile OECD Ülkelerinin Mandarin İthalatı Tahminleri	91
Tablo 4.13 Rastgele Orman Algoritması Performans Değerleri	93
Tablo 4.14 Rastgele Orman Algoritması ile OECD Ülkelerinin Mandarin İthalatı Tahminleri.....	96
Tablo 4.15 Yapay Sinir Ağları Denemelerinin Performans Değerleri	98
Tablo 4.16 Seçilen Yapay Sinir Ağının Performans Değerleri	99
Tablo 4.17 YSA (25, 25) ile OECD Ülkelerinin Mandarin İthalatı Tahminleri	99
Tablo 4.18 Algoritmaların Performanslarının Kıyaslanması (1).....	101
Tablo 4.19 Algoritmaların Performanslarının Kıyaslanması (2).....	101
Tablo 4.20 Algoritmaların En İyi Performans Gösterdiği Tahminleme Değerleri.....	102

KISALTMALAR LİSTESİ

AHP	Analytic Hierarchy Process (Analitik Hiyerarşi Süreci)
ANN	Artificial Neural Network (Yapay Sinir Ağı)
ARIMA	Autoregressive Integrated Moving Average (Durağan Hale Getirilmiş Otoregresif Hareketli Ortalama)
CART	Classification and Regression Tree (Sınıflandırma ve Regresyon Ağaçları)
CRISPDM	The Cross-Industry Standard Process for Data Mining (Veri Madenciliği için Çapraz Endüstri Standardı Süreci)
DEİK	Dış Ekonomik İlişkiler Kurulu
FAO	Food and Agriculture Organization of the United Nations (Birleşmiş Milletler Gıda ve Tarım Organizasyonu)
GSYİH	Gayri Safi Yurt İçi Hasıla
GTİP	Gümrük Tarife İstatistik Pozisyonu
ICC	International Chamber of Commerce (Uluslararası Ticaret Odası)
IMF	International Monetary Fund (Uluslararası Para Fonu)
MAE	Mean Absolute Error (Ortalama Mutlak Hata)
MAPE	Mean absolute percentage error (Ortalama Mutlak Yüzde Hata)
MARS	Multivariate Adaptive Regression Splines (Çok Değişkenli Uyumlu Regresyon Şeritleri)
OECD	Organisation for Economic Co-operation and Development (İktisadi İşbirliği ve Gelişme Teşkilatı)
OLAP	Online Analytical Processing (Online Analitik İşleme)
OMI	Open Market Index

	(Açık Pazar Endeksi)
OMOI	Overall Market Opportunity Index (Genel Pazar Fırsat İndeksi)
PE	Processing elements (İşleme Elemanı)
ReLU	Rectified Linear Unit (Doğrultulmuş Doğrusal Birim)
RMSE	Root Mean Squared Error (Hata Karelerinin Ortalama Kökü)
SAW	Simple Additive Weighting (Basit Artırılabilir Ağırlıklandırma)
SOM	Self-Organizing Maps tipinde yapay sinir ağları
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution (İdeal Çözüme Benzer Bir Yapıda Tercihler Sıralaması Tekniği)
TÜFE	Tüketici Fiyatları Endeksi
ÜFE	Üretici Fiyatları Endeksi
XGBoost	Extreme Gradient Boosting (Aşırı Artırılmış Gradyan)
XML	Extensible Markup Language (Genişletilebilir İşaretleme Dili)
YSA	Yapay Sinir Ağları

ÖZET

Dünya yaş meyve sebze üretiminde ülkemiz ön sıralarda yer almakla birlikte ihracata bu potansiyelini yeterince yansıtamamaktadır. Günümüz dünyasında uluslararası alanda yaşanan rekabet her geçen gün artmakta olup, hedef pazarlar belirlenirken mevcut imkan ve kaynakların en etkin şekilde değerlendirilmesi ve ihracat işlemlerinde başarılı olmak için doğru bir hedef pazar çalışmasının gerçekleştirilmesi oldukça önemlidir.

Ayrıca, gelecek dönemde üretim planının hazırlanabilmesi ve bu doğrultuda stratejik kararlar alınabilmesi için ithalatçı ülkelerinin gelecek dönem alım taleplerinin belirlenmesi gerekmektedir.

Bu nedenle, tez çalışması ile Türkiye'nin yaş meyve sebze ihracatında önemli bir potansiyele sahip olan OECD ülkelerinin ürün ve ülke bazında gerçekleştireceği yaş meyve sebze ithalatının tahmin edilmesine yönelik bir analiz gerçekleştirilmiş ve ithalat değerlerini etkileyen öznitelikler olarak “Dönem”, “Alıcı ülke kodu”, “Ürün kodu”, “Enflasyon oranı”, “Nüfus”, “GSYİH”, “Beş yıl sonraki GSYİH tahmini değeri”, “İş yapma kolaylığı endeksi”, “Açık pazar endeksi”, “Alıcı ülkenin dolar döviz kuru”, “Üretimi” ve “Tarımda işgücü istatistiği” değerleri esas alınmıştır. XGBoost, Rastgele Orman ve YSA'nın ithalat değişkenini tahmin etmedeki performans değerleri kıyaslanmış ve XGBoost Algoritmasının diğer yöntemlere kıyasla daha iyi tahminler gerçekleştirdiği sonucu elde edilmiştir.

Anahtar Kelimeler: Makine Öğrenmesi, XGBoost, YSA, Rastgele Orman

SUMMARY

PREDICTING FRESH FRUIT AND VEGETABLES IMPORT OF OECD COUNTRIES' BY DATA MINING METHODS

Although, Turkey is one of the world's most important fresh fruit and vegetable producer countries, it can not adequately reflect this potential to export. In today's world, since the competition in the international trade is increasing day by day, it is very important to evaluate the available opportunities and resources in the most effective way while determining the target markets. In addition, in order to be successful in export process, target market study should be performed correctly.

On the other hand, to prepare a future production plan and to make strategic decisions in this direction, it is necessary to determine the purchase demands of the importer countries for the next period.

OECD countries have a significant potential in Turkey's fresh fruit and vegetable exports. Therefore, in this study, an analysis is carried out to predict OECD countries' fresh fruit and vegetable import by country and product base. For this purpose, the attributes that influence imports of OECD countries are taken as "Period", "Country code", "Product code", "Inflation rate", "Population", "GDP", "5 year later prediction of GDP", "Ease of doing business index", "Open market index", "Dolar exchange rate", "Production" and "Labor statistics in agriculture". The performance values of XGBoost, Random Forest and ANN in predicting the import variable were compared and the result of XGBoost Algorithm shows better performance compared to other methods.

Keywords: Machine Learning, XGBoost, ANN, Random Forest

TEŞEKKÜR

Bu çalışmaya değerlendirme ve önerileri ile katkıda bulunan, çalışma boyunca desteğini esirgemeyen değerli hocam Sayın Doç. Dr. Sezgin IRMAK, Doktora tezimin izleme komitelerinde yer alan ve kıymetli görüşleri ile tez süresince yönlendirmelerde bulunan değerli hocalarım Prof. Dr. Can Deniz KÖKSAL ve Doç. Dr. Emre İPEKÇİ ÇETİN'e en derin teşekkürlerimi sunarım. Ayrıca bana desteklerini esirgemeyen sevgili eşim Ömür TOSUN'a ve aileme her zaman yanımda oldukları için teşekkür ederim.

Nedret TOSUN

Antalya, 2020

ÖNSÖZ

Veri Madenciliği her geçen gün daha fazla önem kazanmakta ve daha geniş alanlarda kullanılmaktadır. Büyük veri yığınları içerisinde örüntülerin ve bilgilerin keşfi süreci olarak tanımlanabilen veri madenciliğinden, sağlık sektöründen, eğitim, bankacılık, sigortacılık ve hatta uzay bilimine kadar pek çok alanda faydalanılmaktadır.

Veri madenciliği disiplinler arası bir çalışma alanına sahiptir ve analiz sürecinde istatistik, makine öğrenmesi ve yapay zeka disiplinleri altında kullanılan yöntemlerden yararlanır.

Makine öğrenmesi, insanlara oldukça doğal gelen deneyimlerden öğrenme sürecini, probleme ait veriler üzerinden bilgisayarlara öğreten bir veri analitiği tekniğidir. Makine öğrenmesi algoritmaları, herhangi bir denklem ihtiyacı olmadan doğrudan veriden bilgiyi öğrenmeyi sağlayan hesaplama yöntemlerini kullanır. Ana araştırma alanı ise, karışık örüntüleri tanımak ve veriye dayalı akıllı kararları vermek için otomatik öğrenme işlemidir.

Tez çalışmasında veri madenciliği ve makine öğrenmesi algoritmaları üzerinde durulmuş ve seçilen makine öğrenmesi algoritmalarının performansına yönelik bir çalışma yürütülmüştür. Antalya ilinin Türkiye yaş meyve sebze üretim ve ihracatında en fazla paya sahip illerden olması nedeniyle bu çalışmada yaş meyve sebze ihracatının geliştirilmesine yönelik bir analiz çalışmasına odaklanılmıştır.

Türkiye yaş meyve sebze üretiminde ve ihracatında dünya genelinde önemli bir konumdadır. FAO'nun 2018 yılı verilerine göre Türkiye, dünya ayva, kayısı, kiraz ve incir üretiminde birinci; kavun ve pırasa üretiminde ikinci; biber, hıyar-kornişon, karpuz, mandalina ve havuç-şalgam üretiminde üçüncü; domates, elma, patlıcan, fasulye, çilek, ıspanak ve vişne üretiminde dördüncü; kuru soğan, şeftali-nektarin, lahana, kabak ve armut üretiminde beşinci; üzüm, erik ve limon üretiminde ise altıncı sırada yer almaktadır. Ancak, bazı ürünler haricinde dünya ihracat sıralamasında aynı performans sergilenememektedir.

Birleşmiş Milletler veri kaynağı “UN Comtrade Data Base”in 2019 yılı ihracat verilerine göre, Türkiye dünya genelinde ayva ve incir ile “Satsuma” cinsi mandalınanın ihracatında birinci; kabuklu kestane ihracatında üçüncü konumda iken, kayısı, kiraz, diğer mandalina türleri ve greyfurt ihracatında dördüncü; şeftali-nektarin, kabak, patlıcan ve limon ihracatında ise beşinci sırada yer almaktadır. Türkiye, üretim miktarı açısından üstünlüğü olan diğer ürünlerde ise dünya ihracatı sıralamasında daha gerilerde bulunmaktadır.

Ülkemiz üretim potansiyelinin ihracata da yansıtılabilmesi için ihracatın geliştirilmesine yönelik çalışmalar yapılmalıdır. Günümüz dünyasında uluslararası alanda

yaşanan rekabet her geçen gün daha da artırmaktadır. Firmalar yurt dışında ürünleri için hedef pazarları belirlerken ellerindeki imkan ve kaynakları en verimli şekilde değerlendirmek durumundadırlar. İhracat işlemlerinde başarılı olmak ve doğru pazara yönelebilmek için doğru bir hedef pazar çalışmasının gerçekleştirilmesi oldukça önemlidir.

Bu noktada, alınması gereken stratejik kararlar ile gelecek dönemde üretilmesi planlanan ürünlerin tür ve çeşitlerinin belirlenmesi, bir başka ifade ile üretim deseni ve planının hazırlanabilmesi için ithalatçı ülkelerinin gelecek dönem alım taleplerinin belirlenmesi yapılabilecek çalışmaların en başında gelmektedir.

OECD ülkeleri, Gayri Safi Yurt İçi Hasıla (GSYİH) düzeyi yüksek olan ve toplamda dünya yaş meyve sebze ithalatının %68'ini tedarik eden ülkelerden oluşmaktadır. Dolayısıyla, ülkemizin yaş meyve sebze ihracatını artırmak ve dünya ihracatından daha fazla pay alabilmek için OECD ülkelerinin yaş meyve sebze ithalatı incelemeye alınmıştır.

Dış ticaret verilerinin tahmin edilmesine yönelik literatür çalışmalarının incelenmesi neticesinde çok kriterli karar verme yöntemleri ile regresyon ve zaman serisi analizleri gibi ekonometrik yöntemlerin yoğun bir şekilde kullanıldığı tespit edilmiştir. Veri madenciliği yöntemleri açısından ise dış ticaret verilerine yönelik çalışmalarda Yapay Sinir Ağları ile kümeleme analizlerine sıklıkla başvurulduğu görülmektedir.

Diğer taraftan, son dönemlerde veri madenciliği alanında tahminleme çalışmalarında XGBoost Algoritmaları ile Rastgele Orman Algoritmalarının daha fazla tercih edilmeye başlandığı gözlenmiştir.

Bu nedenle, tez çalışması ile Türkiye'nin yaş meyve sebze ihracatında önemli bir potansiyele sahip olan OECD ülkelerinin ürün ve ülke bazında gerçekleştireceği yaş meyve sebze ithalatının tahmin edilmesine yönelik bir analiz gerçekleştirilmesi ve makine öğrenmesi algoritmalarından XGBoost, Rastgele Orman Algoritması ile Yapay Sinir Ağlarının performanslarının değerlendirilmesi amaçlanmıştır. En iyi performansı sergileyen algoritma kullanılarak ülkelerin ithalat değerlerinin öngörülmesi ve bu doğrultuda yürütülecek pazarlama çalışmalarına katkı sağlanması hedeflenmiştir.

Ayrıca, dış ticaret verilerinin tahminlenmesi çalışmalarında daha önce kullanılmayan XGBoost ve Rastgele Orman Algoritmaları ile analiz sürecini gerçekleştirip hem bu yöntemlerin tahminleme alanındaki başarısını dış ticaret alanına yansıtmak hem de akademik anlamda bir farklılık yaratmak amaçlanmıştır.

Tezde kullanılan algoritmaların performansının analiz edildiği çalışmanın sonucunda en iyi performansı sergileyen algoritma kullanılarak bundan sonraki dönemler için hangi ülkenin, hangi ürünü, ne kadar ithal edeceği sorularına yanıt bulunabilecektir.

Bunun için öncelikle, yaş meyve sebze sektöründe dış ticarete konu olan 66 ürün grubu açısından Türkiye dışında kalan 35 OECD ülkesinin 2012-2019 yılları arasındaki ithalat verileri, ikincil veri kaynağı olan Birleşmiş Milletler veri kaynağı “UN COMTRADE DATA BASE”den (<https://comtrade.un.org/data>, erişim tarihi: 20.05.2020) derlenmiştir. Değerlendirmeye alınan “Dönem”, “Alıcı ülke kodu”, “Ürün kodu”, “Alıcı ülkenin ilgili yıl enflasyon oranı”, “İlgili ülkenin nüfusu (1.000 kişi)”, “İlgili ülkenin GSYİH (milyar \$)”, “İlgili ülkenin 5 yıl sonraki tahmini GSYİH (milyar \$)”, “İlgili ülkede iş yapma kolaylığı endeksi”, “İlgili ülkenin açık pazar endeksi”, “Alıcı ülkenin ilgili yıldaki söz konusu üründen üretimi”, “Alıcı ülkenin ilgili yıl tarımda işgücü istatistiği”, “Alıcı ülkenin ilgili yıl para biriminin \$ cinsinden değeri” ile “İthalatçı ülkenin ilgili üründen ithalat değeri (\$)” öznitelikleri açısından 14.499 gözlem elde edilmiştir. “İthalatçı ülkenin ilgili üründen ithalat değeri (\$)” olan çıktının, girdi olarak alınan diğer öznitelikler ile tahmin edilmesi amaçlanmıştır. Ancak, üretim miktarı ile üretim sürecine dayalı bilgilerinin temin edildiği ikincil veri kaynağında (FAO) en güncel verilerin 2018 yılına ait olması nedeniyle “İthalatçı ülkenin ilgili üründen ithalat değeri (\$)” yerine “İthalatçı ülkenin ilgili üründen bir yıl sonraki ithalat değeri (\$)” esas alınmış, girdi değerleri ilgili yıl verileri iken, bir yıl sonraki ithalat değeri çıktı olarak değerlendirilmiştir. Böylece güncel veriler analistlerin elinde mevcut iken ülkelerin bir sonraki yıl hangi üründen ne kadarlık bir ithalat yapabileceğinin tahmin edilmesi amaçlanmıştır.

Bu doğrultuda tezin birinci bölümünde veri madenciliğinin tanımı, kavramları, veri kaynakları, veri madenciliğinin süreci, modelin kurulması ve değerlendirilmesi aşamaları açıklanmıştır.

İkinci bölümde ihracat ve ithalat terimleri açıklanmış, dış ticaretin ve talep tahmini çalışmalarının önemine değinilmiştir.

Üçüncü bölümde modelleme teknikleri ve algoritmalar üzerinde durulmuş, performanslarının kıyaslanacağı Yapay Sinir Ağları, XG Boost ve Rastgele Orman Algoritması ile bu algoritmaların temeli olan Karar Ağaçları ele alınmıştır. Ayrıca, söz konusu algoritmaların kullanıldığı geçmiş dönem uygulama örneklerine yer verilmiştir.

Dördüncü bölümde ise OECD ülkelerinin yaş meyve sebze ithalatının tahminleme çalışması için işletilen veri madenciliği süreci ile ilgili bilgi verilmiş, kullanılan üç farklı makine öğrenmesi yönteminin performanslarına ilişkin kıyaslama yapılmıştır.

Sonuç bölümünde, tezde gerçekleştirilen çalışmalar ile ilgili bilgi verilmiş, elde edilen sonuçlara ilişkin yorumlar, ilgili makamlara sunulan öneriler ve daha sonraki çalışmada izlenebilecek yöntemler hakkında görüşler sunulmuştur.

BİRİNCİ BÖLÜM

VERİ MADENCİLİĞİ

1.1. Veri Madenciliği Kavramları

Her geçen gün daha fazla önem kazanan ve daha geniş alanlarda kullanılan veri madenciliği, karar verme ve strateji geliştirme gibi konularda yöneticilere destek olmaktadır. Sağlık bilimlerinden, bankacılık, sigortacılık, eğitim, güvenlik, istihbarat ve müşteri ilişkileri yönetimi alanına kadar geniş bir kullanım ağına sahip olan veri madenciliği bilgi keşfi sürecinin bir parçasıdır (Akın, 2008: 34-36; Akgöbek ve Çakır, 2009:802; Özcan, 2014: 42-45).

Büyük, karmaşık ve çok fazla değişken içeren veri setlerindeki gizli ve kullanışlı bilgiyi çıkarabilme ve bu bilgi üzerinde değerlendirme yapabilme yeteneği dolayısıyla veri madenciliği her geçen gün daha önemli hale gelmektedir. Günümüzde veriye erişimin kolaylaşması, veri işleme maliyetlerinin düşmesi, veri depolama kapasitelerinin ve üretilen veri hacminin artması ile veri madenciliği yazılımlarındaki gelişim sayesinde veri madenciliği daha fazla kullanılmaktadır (Gupta, 2016: 4-5).

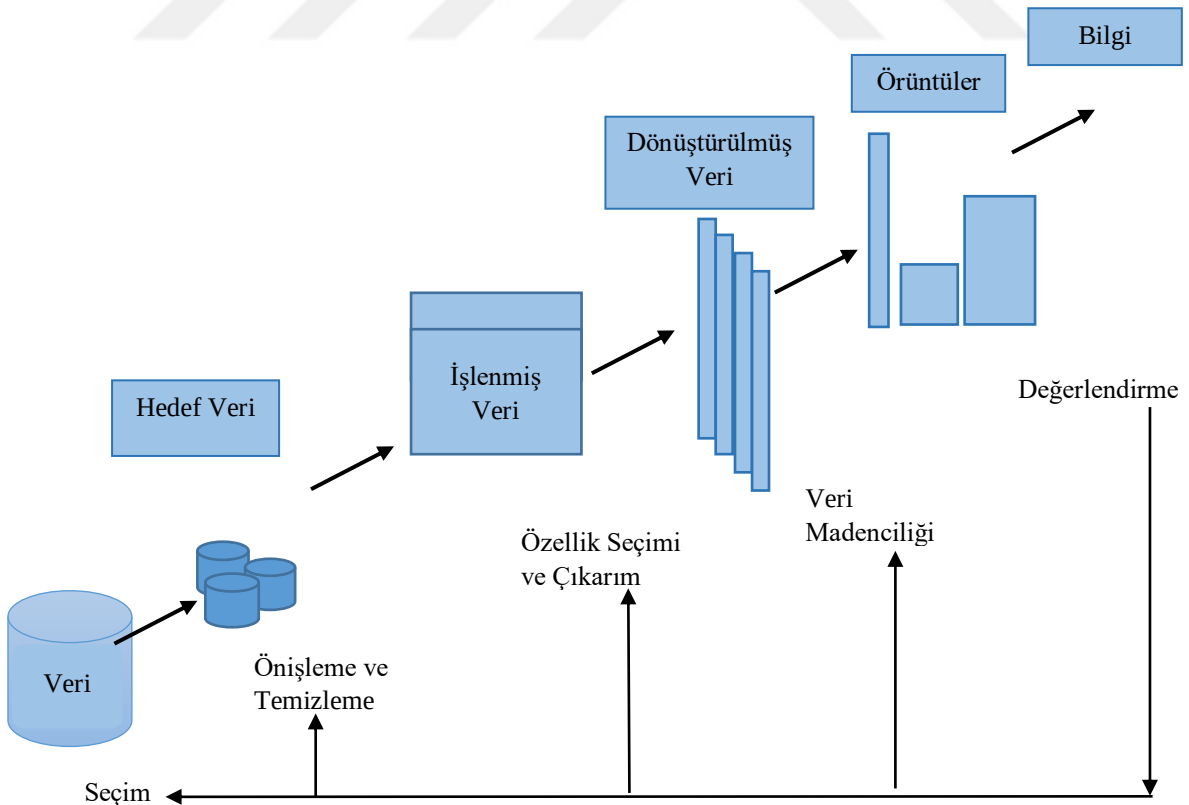
Veri madenciliği konusunda literatürde pek çok farklı tanımlama mevcuttur. Bu tanımlamalardan bazıları aşağıda sunulmaktadır:

- Veri madenciliği, genellikle büyük boyutlardaki gözlemsel veri setlerinin, beklenmeyen ilişkileri bulmak ve hem anlaşılabilir hem de kullanışlı olabilecek yeni yollarla özetlemek için analiz edilmesidir (Hand vd., 2001: 6).
- Veri madenciliği, istatistiksel ve matematiksel tekniklerin yanı sıra örüntü tanıma teknolojilerini kullanarak, depolarda saklanan büyük miktardaki veriden eleme yaparak anlamlı yeni korelasyon, örüntü ve eğilimleri keşfetme sürecidir (Larose ve Larose, 2014: xi).
- Nisbet vd. (2018: 22) veri madenciliğini, büyük, yoğun ve gürütülü veri setlerindeki veri elemanları arasındaki bilinmeyen ilişkileri bulmak için makine öğrenmesi algoritmalarının kullanılması şeklinde ifade etmektedir.
- Han ve Kamber (2012:8), veri madenciliğini, büyük miktardaki verilerden ilginç örüntüleri ve bilgileri keşfetme süreci olarak tanımlamaktadır.
- Veri madenciliği en basit şekilde büyük boyuttaki veriden değerli olan bilgiyi elde etme yöntemidir. Böylece, gerektiğinde geleceğe yönelik çıkarsamalarda bulunmak da mümkün olmaktadır (Özkan, 2016: 12).

- Veri madenciliği, büyük veritabanları içerisindeki bilinmeyen, geçerli, yeni, kullanışlı ve anlaşılabilir örüntülerin etkili keşif tekniklerinin bir koleksiyonudur (Gupta, 2016:3).
- Veri madenciliği, veri içerisindeki örüntülerin keşfidir ve bu süreç otomatik veya genellikle yarı otomatik olmalıdır (Witten vd., 2016:5).
- Veri madenciliği, büyük veri tabanlarından bilgiyi çıkarmak için makine öğrenmesi, örüntü tanıma, istatistik, veritabanları ve görselleştirme tekniklerini bir araya getirerek kullanmaktadır (Güldal ve Çakıcı, 2017: 1356).

Veri madenciliği, veriler arasındaki ilişkiler ile ileriye dönük öngörülerde bulunulabilmektedir. Bu nedenle, karar destek sistemlerinde veri madenciliği oldukça önemlidir (Özkan, 2016:12).

Veri madenciliği, bilgi keşfi sürecinin bir parçasıdır ve tekrarlayan bilgi keşfi sürecinin içinde bir adım olarak düşünülebilmektedir. Nisbet vd. (2018: 22) bilgi keşfini, veriye erişim, veri inceleme, veri önileme, modelleme, modelin uygulanması ve izlenmesi süreci olarak belirtmişler, bu geniş bilgi keşfi sürecini veri madenciliğini kapsayan bir konumda olduğunu vurgulamışlardır. Bilgi keşfi süreci Şekil 1.1’de özetlenmektedir.



Şekil 1.1 Veritabanlarında Bilgi Keşfi Sürecini Oluşturan Adımlar (Fayyad vd., 1996: 29).

1.2. Veri Madenciliği Sürecinin Veri Kaynakları

1.2.1. Veritabanları

Veritabanları, bilgilerin depolandığı alanlardır ve içerdiği bilgiler üzerinde güncelleme, sorgulama yapma ve kaydetmeye imkan veren yapıdadırlar. Veritabanlarının en temel yapıtaşı olan tablolarda bilgiler tutulmaktadır. Veritabanlarında genelde birden fazla bulunan tablolar ise birbirleri ile ilişkilendirilebilmektedir. Çok sayıda tablonun birbirleri ile bir mantıksal ilişki çerçevesinde depolandığı veri tabanlarına “İlişkisel Veri Tabanı” denilmektedir. Her tablo bir dizi öznitelikten (sütun veya alan) oluşur ve genellikle büyük bir dizi kayıt (veya satır) depolar. İlişkisel bir tablodaki her kayıt, benzersiz bir anahtar tarafından saptanan ve öznitelik değerleri kümesi tarafından tanımlanan bir nesneyi temsil eder (Han ve Kamber, 2012: 9; Özkan, 2016: 13; Özcan, 2014: 8).

Veritabanı yönetim sistemi olarak da adlandırılan veritabanı sistemi, birbiriyle ilişkili verilerin bir koleksiyonu olup düzenli bir şekilde veri depolamasının yapıldığı ve yönetildiği ortamdır. Bu yönetim sürecinde çeşitli yazılımlardan destek alınmaktadır. Bu süreçte yararlanılan yazılım programları, veritabanı yapılarını ve veri depolamasını tanımlamakta; eşzamanlı, paylaşılan veya dağıtılmış veri erişimini belirlemekte ve yönetmekte; ayrıca, sistem çökmelerine veya yetkisiz erişime yönelik girişimlere rağmen depolanan bilgilerin tutarlılığı ve güvenliği için mekanizmalar sağlamaktadır (Özcan, 2014: 8; Han ve Kamber, 2012: 9).

1.2.2. Veri Ambarı

Veri ambarları günümüzde giderek daha yaygın bir şekilde kullanılmaktadır. Veri tabanlarından elde edilen özet bilgi işleminden geçirilerek veri ambarında depolanmaktadır. Bir başka ifade ile veri ambarı, yararlı bilgilerin bulunduğu (madencilik işleminin uygulandığı), verilerin işlenip (gürültüden arındırılıp, veri temizleme, veri entegrasyonu, veri dönüşümü yapıp) statik hale getirildiği, veri yükleme ve periyodik veri yenileme işleminin gerçekleştirildiği alt veri kümesidir (Han ve Kamber, 2012: 9-10).

Veri ambarı, analitik işlemlerin yapılabilmesi için gerekli olan bilgiyi sağlamaktadır. Karar vericiye göre düzenlenmiş, bütünleşik durumda bulunan, ilgili konulara göre derlenmiş ve sadece okunabilir yapıdadır. Veri tabanı ile veri ambarının kıyaslanması Tablo 1.1’de sunulmaktadır (Özcan, 2014: 9-10).

Tablo 1.1 Veri Ambarı ve Veri Tabanı Özelliklerini Kıyaslama Tablosu

Veri Ambarı	Veri Tabanı
Off-Line bir yapıda çalışır.	On-Line bir yapıda çalışır.
Veri değişimi yapılmaz, veri sorgulaması yapılır.	Yoğun bir şekilde veri değişimi gerçekleştirilir.
Eski verilerin de depolanması nedeniyle içerdiği veri miktarı çoktur.	Sürekli olarak güncel veriler depolandığı için daha az veri miktarına sahiptir.
Üst kademe yöneticiler ile analiz yapan yetkililerin kullanımına açıktır. Bu nedenle daha az kullanıcı bulunmaktadır.	Veriye erişmek isteyen tüm kullanıcılara açıktır.

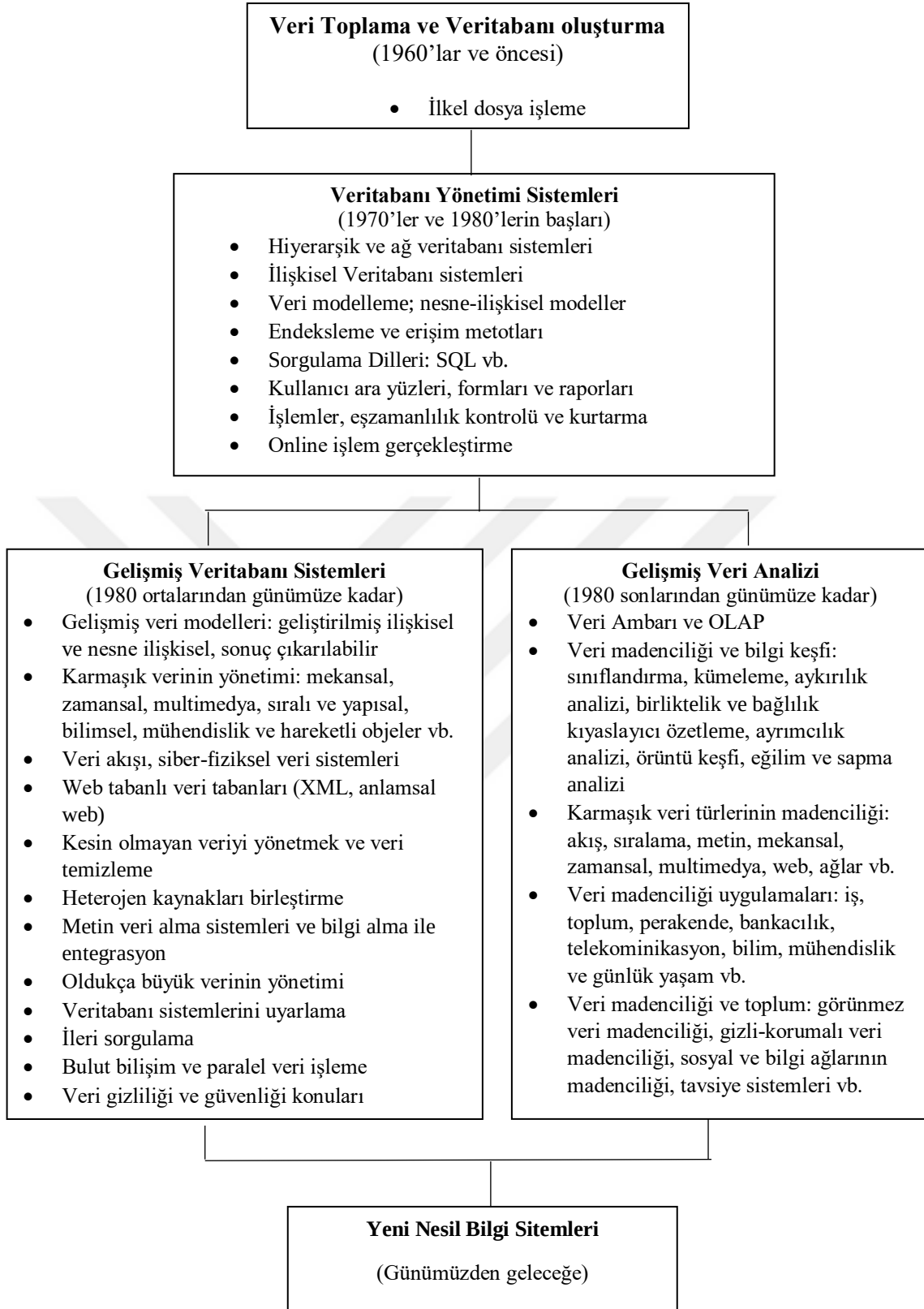
Kaynak: Özcan, 2014: 10.

1.3. Veri Madenciliğinin Tarihsel Gelişimi

Verinin kullanımı ile ilgili çalışmaların geçmişi 1960'lı yıllara dayanmaktadır. 1960'larda veriler elektronik kaynaklardan depolanmaya ve bilgisayarlar ile analiz edilmeye başlanmıştır. 1980'lerde ilişkisel veritabanları modelleri kullanılırken, 1990'larda, depolanan verinin hacmi çok büyük boyutlara ulaştığı için veri ambarlarından faydalanılmaya başlanmıştır. 2000'li yıllarda ise veri madenciliği tüm alanlarla kullanımda yerini almıştır (Akın, 2008: 30-38). Verinin işlenmesi süreci ile ilgili tarihsel gelişimi Şekil 1.2'deki gibi özetlenmektedir (Han ve Kamber, 2012: 3).

Veritabanı sisteminin gelişiminde ve veri madenciliği yöntemlerinde istatistik uygulamalarından da yararlanılmıştır. Ancak, veri madenciliği yöntemleri, klasik istatistiksel uygulamalardan aşağıdaki konularda ayrılmaktadır (Özkan, 2016:13-15).

1. Veri madenciliği yöntemleri nispeten daha büyük veriler üzerinde çalışmaktadır.
2. Klasik istatistiksel uygulamalardaki gibi varsayımlarda bulunmak çoğunlukla gerekmemektedir.
3. Hipotezler kurup, test etmek yerine eldeki verilerle yararlı sonuçlara ulaşmaya çalışılmaktadır.
4. Yeni verilere ulaştıkça, üzerinde çalışılan modelin eğitilmesine devam edilebilmektedir.

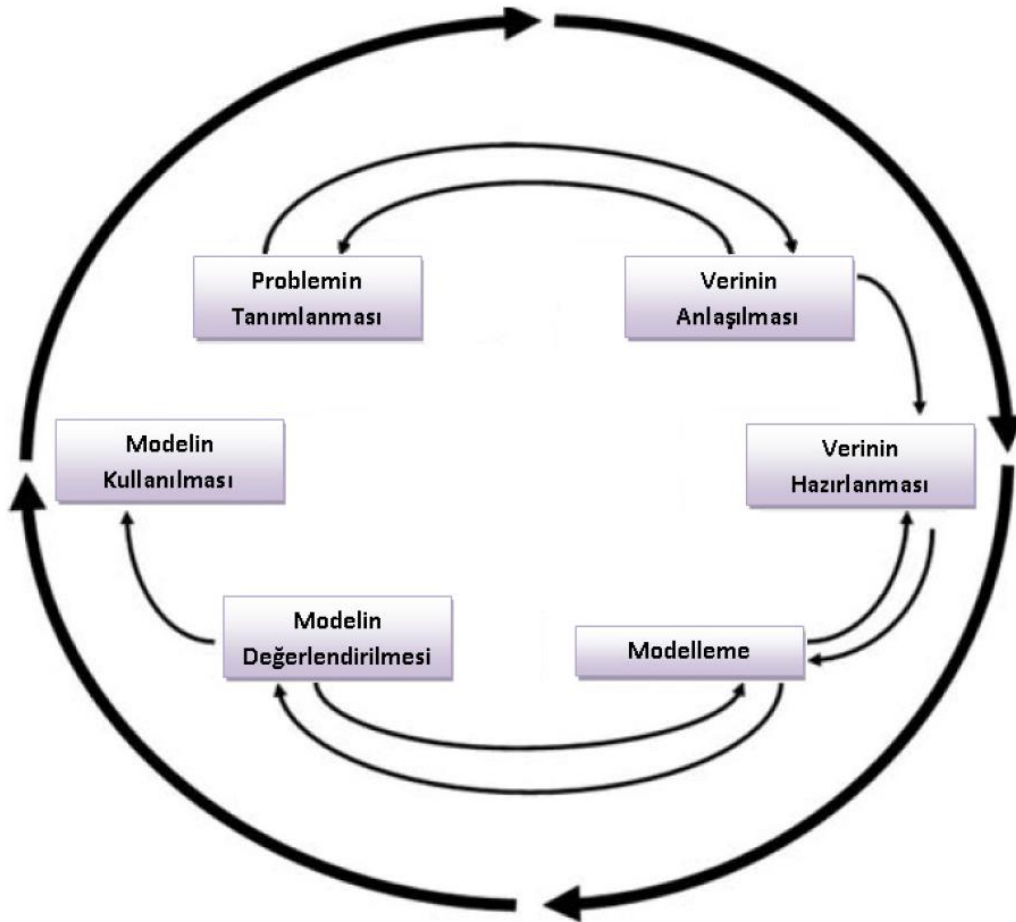


Şekil 1.2 Veritabanı Sistemi Teknolojisinin Gelişimi (Han ve Kamber, 2012: 3)

1.4. Veri Madenciliği Süreci

Veri madenciliği, veri ambarlarında depolanan büyük hacimli verilerden gizli bilgilerin ortaya çıkarılması, verilerin işlenmesi ve faydalı bilginin keşfedilmesi sürecidir. Dolayısıyla veri madenciliği sürecinde bir önceki aşamanın çıktısı bir sonraki aşamanın girdisi olmaktadır. Bu nedenle, her aşamanın dikkatle değerlendirilmesi gerekmektedir (Akgün, 2012: 46-47).

Kullanılan yazılım ve modelden bağımsız bir veri madenciliği süreci modeli olan Veri Madenciliği için Çapraz Endüstri Standardı Süreci (The Cross-Industry Standard Process for Data Mining (CRISPDM)) altı aşamadan oluşmaktadır. Söz konusu aşamalar Şekil 1.3'te sunulmaktadır.



Şekil 1.3 Veri Madenciliği Süreci (CRISPDM), (Larose ve Larose, 2014:4)

Veri madenciliği, kullanıcı tarafından pek çok kararın verildiği çok sayıda adımları içeren, interaktif ve iteratif bir süreçtir. Verinin anlaşılmadığı, özelliklerinin bilinmediği durumlarda, ne kadar etkin bir algoritma kullanılırsa kullanılsın başarıya ulaşılması mümkün değildir. Bu nedenle, Şekil 1.3'te de görüleceği üzere, veri madenciliği uygulamalarında öncelikle verinin anlaşılması ve problemin tanımlanması gerekmektedir.

Sonrasında sırasıyla verinin hazırlanması ve modelin kurulması aşamaları gelmektedir. Daha sonra kurulan model değerlendirilir, izlenir, doğrulanır ve kullanılır (Fayyad vd., 1996: 42).

Veri Madenciliği adımları konusunda farklı tanımlamalar da yer almaktadır. Örneğin,

- Haglin vd. (2005: 41-42) veri madenciliği adımlarını yedi aşamada tanımlamıştır: (1) Hedefin belirlenmesi; (2) Hedef verinin oluşturulması; (3) Veri önileme; (4) Veri dönüştürme; (5) Veri madenciliği; (6) Tahminleme ve değerlendirme; (7) Harekete geçme.
- Gorunescu (2011: 7-14) veri madenciliği adımlarını beş aşamalı olarak tanımlamıştır: (1) Problemin tanımı; (2) Verinin hazırlanması; (3) Modelin kurulması; (4) Modelin kullanılması; (5) Modelin izlenmesi
- Han ve Kamber (2012: 7-8), veri madenciliği için yedi aşamalı bir süreç tanımlamıştır: (1) Hedef veri setinin oluşturulması; (2) Veri temizleme ve önileme; (3) Veri azaltma ve dönüştürme; (4) Veri madenciliği fonksiyonunun seçimi; (5) Veri madenciliği algoritmasının seçimi; (6) Değerlendirme ve elde edilen bilginin sunumu; (7) Keşfedilen bilginin kullanılması.

Bu çalışmada ise Veri Madenciliği süreci beş aşamada değerlendirilmektedir:

1. Problemin tanımlanması ve verinin anlaşılması
2. Verinin hazırlanması (önileme süreci)
 - a. Veri toplama
 - b. Veri değerlendirme
 - c. Veri temizleme ve bütünleştirme
 - d. Seçim
 - e. Veri dönüştürme
3. Modelleme
4. Modelin değerlendirilmesi
5. Modelin kullanılması

1.4.1. Problemin Tanımlanması ve Verinin Anlaşılması

Veri madenciliğinde farklı amaçlar için farklı veri ve tekniklerin kullanması gerekebilmektedir. Veri madenciliği sürecinin beklenen başarıyı elde etmesi için amaçların dikkatli bir şekilde belirlenmesi önemlidir. Bu nedenle, amaçların belirlenmesi ve

problemin tanımlanması aşaması etkin bir veri madenciliği sürecinin ilk aşamasını oluşturmaktadır (Gupta, 2016: 12).

Problemin yanlış tanımlanması yanlış sonuçlara ulaşılmasına sebep olacaktır. Dolayısıyla, problemin tanımlanması aşamasında amaç açık bir şekilde belirtilmiş olmalı, hangi problemlerin çalışma ile çözülmesinin beklendiği öngörülmesi ve elde edilecek sonuçların başarısının nasıl bir ölçüt üzerinden değerlendirileceği belirlenmelidir. Analizi yapan kişinin süreci etkileyebilecek önemli faktörleri tespit etmesi gerekmektedir. Bu noktada, bütçe veya en azından ön hazırlık bütçesi ve çalışmadan elde edilecek faydalar da değerlendirilmelidir (Akgün, 2012: 48; Gupta, 2016:12-13).

Problemin tanımlanması aşaması, çalışmanın amacının ve gerekliliklerinin anlaşılması süreci ile bu bilginin daha sonra veri madenciliği problem tanımına dönüştürülmesi ve amaçların gerçekleştirilmesi için bir taslak planın hazırlanması adımlarına kadar olan süreci içermektedir.¹ Adım adım değerlendirilecek olursa problemin tanımlanması aşaması

- Çalışmanın amaçları ile beklenen çıktıların tanımlanması,
- Çalışmanın durumu ve ortamın değerlendirilmesi,
- Varsayımların listelenmesi,
- Kullanılacak terminolojinin açıklanması ve listelenmesi,
- Verinin temin edileceği veri kaynaklarının ve çalışmada faydalanılacak insan kaynağının belirlenmesi,
- Çalışmanın kısıtları ve risklerinin listelenmesi,
- Hangi yazılımın kullanılacağı belirlenmesi,
- Veri madenciliği çalışmasının amacı, başarı kriterleri, kullanılacak teknikler gibi detayların yer alacağı taslak planının hazırlanması,

adımlarından oluşmaktadır (Gupta, 2016: 13).

Verinin anlaşılması aşamasında öncelikle hangi özneliliklerin gerekli olduğu ve hangi verilerin toplanacağı, veri türleri (coğrafi veri, nicel veya nitel veri vb.), verileri çıkartmak veya eklemek için işletmedeki bilgisayar sistemlerinin sayısı ve miktar olarak yeterli olup olmadığı, işletmenin gerekli bilgileri içeren bir veri ambarının olup olmadığı gibi sorular değerlendirilmektedir (Gupta, 2016: 13). Daha sonra, verideki problemlerin (gürültülü ve kayıp veri vb.) belirlenmesi, verideki ilk izlenimlerin keşfedilmesi ve gizli bilgiden ilginç detaylar tespit etmek için veriye aşina olma aşaması ile devam etmektedir.¹

¹ https://docs.oracle.com/cd/B19306_01/datamine.102/b14339/5dmtasks.htm (erişim tarihi: 25.05.2020).

Verinin anlaşılması aşamasında ise aşağıdaki adımlar izlenmektedir (Gupta, 2016: 13-14):

- Veriler toplanır ve ilk veri toplama raporu hazırlanır.
- Toplanan veri tanımlanır ve veri tanımlama raporu hazırlanır.
- Toplanan veri incelenir ve inceleme raporu hazırlanır.
- Veri kalitesi değerlendirilir ve raporlanır.

1.4.2. Verilerin Hazırlanması (Ön İşleme)

Verinin anlaşılması aşamasının ardından verinin hazırlanması veya önışlenmesi aşaması gelmektedir. Bu aşamada, veri öznitelikleri yakından incelenmektedir. Gerçek dünya verileri gürültülü, büyük hacimli (birçok gigabayt veya daha fazla), karmakarışık ve heterojen kaynaklardan elde edilmiş olabilmektedir. Bu yapıları nedeniyle de gürültülü, kayıp ve tutarsız veri içerme olasılığı yüksektir. Bu durum verinin kalitesini düşürmekle birlikte düşük kalitede madencilik sonuçlarına sebep olmaktadır (Han ve Kamber, 2012: 39).

Veri hakkında edinilen bilgi, veri madenciliği sürecinin en önemli ve ilk aşaması olan veri önışleme aşamasında yararlıdır. Verinin hazırlanması aşaması, ilk ham veri setinden, modelleme aracına sunulacak nihai veri setini oluşturmak için gerekli olan tüm aktiviteleri içerir (Gupta, 2016: 14).

1.4.2.1. Veri Kalitesi ve Veri Önışlemenin Önemi

Veri, gerçekleştirilecek çalışmaya yönelik gereklilikleri sağlayabiliyorsa kalitelidir. Gerçek dünyada ise veri, mükemmellikten uzaktır.

Verinin kalitesini etkileyen faktörler doğruluk, tam olması, tutarlılık, zamanca uygunluk, güvenilirlik, yorumlanabilme (açıklanabilme) durumudur. Güvenilirlik, kullanıcılar tarafından veriye ne kadar güvenildiğini ifade ederken; yorumlanabilirlik, verinin ne kadar kolay anlaşılabilildiğini ifade etmektedir (Han ve Kamber, 2012:84-85).

Doğru olmayan veya kayıp verilerin pek çok olası nedeni vardır. Bunlar (Han ve Kamber, 2012:84-85; Tan vd., 2006: 37):

- Veri toplama araçlarının hatalı olması,
- Kullanıcıların zorunlu alanlara kişisel bilgilerini paylaşmak istememesi nedeniyle kasten yanlış veri değerleri girmiş olması (bu durum gizlenmiş kayıp veri olarak da bilinir) veya önemli bulunmayan verileri girmemiş olması,

- İlgili verinin, yanlış anlama veya ekipmanların arızası nedeniyle kaydedilememiş olması,
- Senkronize veri transferini ve tüketimini koordine etmek için limitli geçici bellek boyutu gibi teknolojik kısıtlamaların bulunması,
- İsimlendirme dönüşleri, veri kodları veya girdi alanları için tutarsız formatlar olması

gibi nedenlerden kaynaklanabilmektedir.

Pek çok veri madenciliği tekniği verideki kalite sorunlarını tolere edebilmektedir. Ancak, verinin anlaşılması ve kalitesinin iyileştirilmesi analiz sonuçlarının kalitesini de iyileştirmektedir (Tan vd., 2006: 19). Bu nedenle, modele girdi olacak gereksiz verileri ve modelin ürettiği gereksiz çıktıları azaltmak asıl amaçtır. Veri madenciliğinin amaçlarına uygun olması için veritabanının veri temizleme ve veri dönüştürme formunda ön işleme tabi tutulması gerekmektedir (Larose ve Larose, 2014:17).

1.4.2.2. Veri Ön İşlemenin Başlıca Görevleri

Veri temizleme rutinleri kayıp verileri tamamlayarak, gürültülü verileri düzgünleştirerek, aşırılık gösteren değerleri belirleyerek veya çıkararak ve tutarsızlıkları çözümleyerek veriyi temizleme işlemidir. Kullanıcılar verinin kalitesiz olduğuna inanırlarsa, uygulanan veri madenciliğinin sonuçlarına güvenemezler. Ayrıca kalitesiz veri, madencilik sürecinde karışıklığa ve güvenilmez çıktılara neden olur (Han ve Kamber, 2012:85).

Verilerin ön işleme tabi tutulduğu bu aşama, veri madenciliğin en uzun zaman gerektiren sürecidir. Ayrıca, söz konusu aşama için gerekli olan zaman, verinin boyutu ile doğru orantılıdır. Büyük hacimli veri setlerinde, veri ile ilgili problemlere genellikle daha fazla rastlanmaktadır (Akgün, 2012: 48-49).

Verilerin hazırlanması aşaması, veri toplama, değerlendirme, bütünleştirme ve temizleme, seçme ve dönüştürme adımlarından oluşmaktadır. Bir sonraki aşama olan modelin kurulması aşamasında rastlanacak sorunlar tekrar veri hazırlama aşamasına dönülmesine ve verinin yeniden işlem den geçirilmesine sebep olabilmektedir. Bu durum dolayısıyla bir analist, veri madenciliği sürecinde harcadığı toplam zaman ve çabanın %50-%85'ini verinin hazırlanması ve modelin kurulması aşamasına ayırmaktadır (Ayık vd., 2007: 442-443).

1.4.2.3. Veri Toplama

Veri toplama adımı, gerekli verinin toplanacağı kaynakların belirlenmesi ve verilerin derlenmesi işlemlerinden oluşmaktadır (Ayık vd., 2007: 443). Burada, gerekli olandan daha az veri ile analiz yapmak çalışmayı eksik bırakırken, gereğinden fazla veri kullanmak da sonuca ulaşmayı uzatabilmektedir (Akgün, 2012: 49).

1.4.2.4. Veri Değerlendirme

Veri madenciliği sürecinin işletileceği verilerin farklı kaynaklardan derlenmiş olması durumunda kodlama hataları ve farklı ölçü birimleri ile değerlendirme yapılması gibi nedenlerle veride uyumsuzluklar olabilmektedir. Bu nedenle, verinin kendi içinde ne kadar uyumlu olduğu bu adımda değerlendirilir (Ayık vd., 2007: 443).

1.4.2.5. Veri Temizleme ve Bütünleştirme

Bu adımda, bir önceki adımda tespit edilen uyumsuzluklar giderilmeye çalışılmakta, gürültülü ve tutarsız veriler uzaklaştırılmakta ve veriler tek bir veritabanında toplanmaktadır (Ayık vd., 2007: 443).

Veri temizleme aşaması büyük ve kapsamlı bir iştir. Veri temizlemede ilk süreç tutarsızlıkların tespitidir. Veri temizleme rutini, kayıp verileri tamamlama, aşırılık gösteren değerleri belirlerken gürültüleri düzgünleştirme ve tutarsızlıkları doğrulamaya çalışır. Tutarsızlık; birçok opsiyonel alanın olduğu zayıf bir veri giriş formu oluşturulması, kasıtlı hatalar, veri bozulmaları ve veri girişlerinde insan hatasının olması gibi pek çok faktörden kaynaklı olabilir (Han ve Kamber, 2012: 88-91).

Kayıp verileri tamamlama işlemlerinden bir kaç farklı yol önerilmektedir (Han ve Kamber, 2012: 88; Raschka ve Mirjalili, 2019: 109-112; Lee, 2019: 107-112; Gupta, 2016; 75).

- 1) Demetleri gözardı etme: Bu işlem, sınıf etiketi kayıp olduğunda uygulanır. Veri demetinin kayıp değer içeren çok sayıda özniteliği olmadığı sürece bu metod çok etkili bir yöntem değildir.
- 2) Kayıp verileri tek tek doldurmak: Bu yaklaşım zaman kaybettirmektedir. Çok fazla kayıp değer olduğu büyük veri setlerinde uygulanabilir değildir.
- 3) Kayıp değeri doldurmak için sabit bir değer atama: Sabit bir değer atama yaklaşımı basit olsa da kusursuz değildir.

- 4) Kayıp değere özniteliğin merkezi eğilim (ortalama veya medyan gibi) değeri atama: Normal (simetrik) dağılımlarda ortalama kullanılabilirken, simetrik olmayan dağılımlarda medyan kullanılması önerilir.
- 5) En fazla tekrarlanan değeri atama.

Gürültülü veri ise veri analistlerinin çalışmalarını olumsuz etkileyen ve analistlerin karşılaşmak istemedikleri bir durumdur. Gürültü, ölçülen bir değer içinde rassal hata veya varyansdır. Bu durum, veri setinin analizler için tutarsız olmasına neden olur. Veriden gürültüyü uzaklaştırma sürecinde eksik değerlerin bulunduğu veriler veri setinden silinebilir (Akgün, 2012: 18). Diğer taraftan, gürültülü verinin uzaklaştırılması için depolama, regresyon ve aşırılık analizi gibi yöntemler kullanılabilir (Han ve Kamber, 2012: 88-90):

- Depolama (binning) metodu: Bu metod (küçükten büyüğe veya büyükten küçüğe) sıralanmış verileri düzeltme amacıyla uygulamaktadır. Komşu değer, yani en yakın değere danışarak sıralanmış veriyi düzeltmektedir.
- Regresyon: Veri değerlerini bir fonksiyona uyduran yöntem olan regresyon veri düzeltirmede kullanılabilmektedir.
- Aşırılık (outlier) analizi: Aşırı / uç noktalar, benzer değerlerin bir grup veya küme içerisinde toplandığı, kümeleme yöntemi ile bulunabilmektedir.

Veri temizleme işlemi sonrasında farklı veri kaynaklarındaki veri setleri bütünleştirme işlemi ile tek bir veritabanında toplanır. Veri bütünleştirme sürecinde gereksiz değerlerin tespiti için nominal nitelik değerleri için χ^2 (chi-square) testi; sayısal nitelik değerleri için bir niteliğin değerlerinin diğer bir niteliğin değerlerinden ne kadar farklı olduğunu gösteren korelasyon katsayısı ve kovaryans kullanılır. Bilgi demetleri düzeyinde tekrarlar tespit edilmeli ve çözülmelidir (Han ve Kamber, 2012: 98-99).

1.4.2.6. Veri Seçimi

Bu adımda, bilgi keşfi için problem tanımlama adımında belirlenen amaç ve hedeflere ulaşmayı sağlayacak verilerin kullanımı önemlidir. Kullanılacak modele bağlı olarak seçilecek veriler de değişiklik gösterebilir. Bu adımda, çalışmada hangi verilerin kullanılacağı belirlenmektedir (Han ve Kamber, 2012: 8; Gemici, 2012: 6).

1.4.2.7. Veri Dönüştürme

Bu adımda veriler, özetleme veya toplama gibi işlemlerden geçirilerek madenciliğe uygun formlara dönüştürülüp birleştirilmektedir. Veri madenciliği sürecinin daha etkin

sonuçlanması ve elde edilen sonuçların daha iyi anlaşılması için verilerin dönüştürülmesi ve birleştirilmesi işlemlerinden oluşmaktadır (Han ve Kamber, 2012: 111-112).

Veri dönüştürme sürecinde kullanılabilecek stratejiler aşağıdaki gibidir (Han ve Kamber, 2012: 112; Oğuzlar, 2003:74; Coşlu, 2013: 616)

1. Değişken (öznitelik/özellik) ekleme: Yeni özniteliklerin oluşturulup madencilik sürecine yardımcı olması için mevcut özniteliklere yeni öznitelikler eklenmesi işlemidir.
2. Toplama/Birleştirme: Özetleme veya birleştirme operasyonlarının veriye uygulanması, gereksiz özelliklerin çıkarılması veya kümeleme ile veri boyutunun azaltılmasıdır. Örneğin, günlük satış verileri aylık veya yıllık değerlerin elde edilmesi için toplanabilir.

Veri azaltma aşamasında kullanılan stratejiler boyut azaltma, sayısal azaltma ve veri sıkıştırma adımlarını içerir. Boyut azaltma değerlendirilen niteliklerin veya rassal değerlerin sayısının azaltılmasıdır. Boyut azaltma metodları, orijinal veriyi daha küçük boyuta indirgeyen temel bileşen analizi tekniklerini içerir. Ayrıca, öznitelik alt kümesi seçimi de alakasız, zayıf alakalı veya gereksiz özniteliklerin veya boyutların tespit edildiği ve kaldırıldığı boyutsallık azaltma yöntemidir. Sayı azaltma, alternatif, daha küçük temsilleri kullanılarak parametrik (regresyon, log-linear modeller gibi) veya parametrik olmayan (histogramlar, kümeler, örnekleme veya veri toplama gibi) modeller ile verinin değiştirilmesidir.

Veri sıkıştırmada, orijinal verinin sıkıştırılmış veya azaltılmış bir temsilini elde etmek için dönüştürme işlemi uygulanır. Eğer, orijinal veri sıkıştırılmış veriden herhangi bir bilgi kaybı yaşanmadan yeniden yapılabiliyorsa bu veri azaltımına kayıpsız azaltım, orijinal verinin yaklaşık bir örneği yaratılabiliyorsa bu veri azaltımına da kayıplı azaltım denmektedir.

3. Normalizasyon: Değişkenler arasındaki ölçek büyüklükleri farkını gidermek için kullanılmaktadır. Farklı veri kaynaklarından elde edilen verilerin heterojen yapıda olmaları, ortalama ve varsyanlarının birbirlerinden önemli ölçüde farklı olmaları durumunda büyük değerlere sahip olan veriler model üzerinde diğerlerine kıyasla daha baskın olmaktadır. Bu durumun önüne geçmek için veri üzerinde normalizasyon işlemi uygulanmaktadır.
4. Veriyi Kesikleştirme: Bazı veri madenciliği algoritmalarının yalnızca kategorik değerleri ele alması veya analizin gerektirdiği durumlarda sürekli verilerin kesikli değerlere dönüştürülmesi işlemidir. Sayısal özniteliklerin (yaş gibi) değerlerinin

aralık etiketler (örneğin 0-10, 11-20 vb.) veya kavramsal etiketler (genç, yetişkin, yaşlı gibi) ile ifade edilmesi işlemidir. Bu işlem ile orijinal veri setinden daha düşük hacimli veriye ulaşılabilmesi de mümkün olmaktadır.

5. Kavram hiyerarşisi oluşturulması: Nominal verilerin üst düzey kavramlarla (örneğin, sokak gibi özniteliklerin şehir veya ülke gibi kavramlarla) genellenmesidir.

1.4.3. Modelleme

Modelin kurulması veya modelleme aşamasında, ön işlemden geçirilmiş, hazırlanmış verinin üzerinde hangi modelin kullanılacağı değerlendirilmektedir. Model kurma aşamasında, probleme ve öğrenme stratejisine uygun olan algoritmayı kullanarak girdilerden arzulanan sonuçlara ulaşılması süreci takip edilmektedir (Balaban ve Kartal, 2015: 44).

Modelleme aşaması aşağıdaki adımlardan oluşmaktadır (Gupta, 2016:14):

- Modelleme tekniği seçilir ve varsayımların listelendiği bir rapor hazırlanır.
- Bir test tasarım oluşturulur ve test tasarım raporu hazırlanır.
- Model kurulur ve parametreler seçilir. Parametreleri listeleyen ve modeli tanımlayan bir rapor hazırlanır.
- Model değerlendirilir ve değerlendirme raporu hazırlanır.

Bu aşamada, birçok modelleme tekniği seçilip uygulanır ve optimal değerlere göre parametreler kalibre edilir. Genellikle aynı veri madenciliği problemi türü için çok sayıda teknik bulunmaktadır. Bazı teknikler veride değişiklik yapılmasını ve veri hazırlama aşamasına sıklıkla dönmeyi gerektirebilmektedir. Bu sebeple, veri hazırlama, ön işleme ve modelleme aşamaları, en iyi modele ulaşıncaya kadar tekrarlanan bir süreçtir (Chapman vd., 2000: 11; Akpınar, 2000: 10).

Modelleme sürecinde çıktı değişkenine bağlı olarak farklı teknikler kullanılabilir. Çıktı değişkeni sürekli ise regresyon yöntemleri, çıktı değişkeni kesikli ise de sınıflandırma yöntemleri kullanılabilir.

1.4.4. Modelin Değerlendirilmesi

Modelin kurulması aşamasının ardından değerlendirme aşaması gelmektedir. Bu aşamada modelin, işin amaçlarını hangi ölçüde karşıladığı, yetersizliklerinin bulunup bulunmadığı değerlendirilir ve bu modelin yetersizliğinin nedenleri araştırılır. Bu aşamada, süreç gözden geçirilir, sonuçlar değerlendirilir, sonuçlar projenin başında tanımlanan

değerlendirme kriterleri ile kıyaslanır. Daha sonra, sonraki adımda yürütülecek çalışmalar belirlenir. Bu adımlar sonucunda ilk üç adımın tekrar gözden geçirilmesi de gerekebilir (Chapman vd., 2000:51; Gupta, 2016:15).

Kurulan modeller arasında en iyisinin seçilmesi için çeşitli performans değerlendirme yöntemleri bulunmaktadır.

1.4.4.1 Modelin Performansını Değerlendirme Yöntemi

Bir modelin performansı, kullanılan öğrenme algoritması, veri setinden ayrılan eğitim ve test kümelerinin boyutuna bağlı olabilmektedir. Modelin performansını değerlendirmek için kullanılan yöntemlerden bazıları aşağıda sunulmaktadır (Balaban ve Kartal, 2015: 45-47, Raschka ve Mirjalili, 2019: 195; Gupta, 2016: 186-188):

- **Holdout (Dışında Tutma) Yöntemi:** Bu yöntemde veri seti eğitim ve test verileri olarak ikiye ayrılır. Eğitim veri seti üzerinden modelin öğrenmesi sağlanır, test veri seti ile de öğrenmenin düzeyi ve modelin performansı test edilmektedir. Bu yöntemin dezavantajlarından biri gözlem sayısının az olması durumunda test veri setinde yer vermek üzere yeterli veriye sahip olunamayacak olması; bir diğeri ise verinin eğitim ve test veri seti olarak bir defa ayrılmış olması ve bunun sonucunda elde edilen model performansına tam bir güven sağlanamamasıdır.
- **Tekrarlı Holdout (Dışında Tutma) Yöntemi:** Bu yöntem, holdout yönteminin tekrarlanması işlemidir. Ancak, seçim rastgele yapılırsa da seçilen veri setlerinin tekrar yapılabilmesi ve üstüste binebilmesi ihtimali bir dezavantaj yaratmaktadır.
- **Tabakalı Örneklem Yöntemi:** Hedef veri setinin nominal değerlerden oluşması durumunda kullanılabilir, nümerik verilerde kullanıma uygun değildir. Veri seti eğitim ve test verileri olarak ayrılırken veri setindeki kategorilere ait oranların da korunmasına dikkat edilir.
- **Üçlü Ayırma Yöntemi:** Veri seti eğitim, doğrulama ve test verileri olmak üzere üçe ayrılır. Doğrulama verileri ile algoritmanın parametrelerinde düzeltmeler yapıp, test veri seti ile de modelin son performansı değerlendirilir.
- **Çapraz Doğrulama (Cross Validation) Yöntemi:** Bu yöntemde en fazla “k-kat çapraz doğrulama” ile “birini dışarda bırakarak çapraz doğrulama” kullanılmaktadır. K-kat çapraz doğrulamada, veri k eşit parçaya ayrılır. Her defasında ayrılan k veri setinden biri test, kalan $k - 1$ adet veri seti ise eğitim seti olur. Böylece k defa elde edilen performansın ortalaması modelin performansını verir. Birini dışarda bırakma yönteminde bir veri seti test için ayrılırken kalan setler

eğitim seti olarak kullanılmaktadır. Veri setinde örnek/gözlem sayısının az olması durumunda bu yöntem tercih edilmektedir.

- **Bootstrap Örnekleme Yöntemi:** Bu yöntemde, n adet örnekten oluşan bir veri setinde rastgele n defa eğitim seti için örnek seçilmekte, eğitim setinde yer almayan örnekler de test veri setinde yer almaktadır.

1.4.5. Modelin Kullanılması

Modelin değerlendirilmesi aşamasında modelin performansı yönetim tarafından da uygun bulunursa sonraki aşama olan modelin kullanılması aşamasında geçilir (Gupta, 2016: 15).

Modelin amacı, veri ile ilgili bilgiyi artırmak olsa da elde edilen bilginin müşterinin kullanabileceği bir şekilde düzenlenmesi ve sunulması gerekmektedir. Bu nedenle, modelin kullanılması aşaması, bir kurumun karar verme sürecinde canlı modeller uygulanması şeklinde olmalıdır. Örneğin, gerçek zamanlı kişiselleştirme Web sayfalarında pazarlama veritabanlarında tekrarlanan puanlamalar gibi. Gereksinimlere bağlı olarak uygulama aşaması bir rapor oluşturabilecek kadar basit veya kurumun genelinde tekrarlanabilir veri madenciliği süreci uygulayabilecek kadar karmaşık olabilmektedir (Chapman vd., 2000: 11).

Modelin kullanılması aşamasında nihai rapor hazırlanmaktadır. Bu rapor, bundan önceki aşamalarda hazırlanan raporlarının tamamını içerecek düzeyde kapsamlı bir rapor olmalıdır. Çalışma ve dokümanlar yeniden gözden geçirilmekte, çalışma süresince nelerin doğru, nelerin yanlış gittiği tespit edilmekte ve çalışmanın daha iyi sonuçlar vermesi için ne gibi iyileştirmeler yapılabileceği konuları raporlanmaktadır (Gupta, 2016: 15).

İKİNCİ BÖLÜM

DIŞ TİCARET

Dış ticaret, bir mal veya hizmetin (bedelsiz ihracatlar hariç olmak üzere) belirli bir ücret karşılığında ülke sınırları dışına satımı veya yurt dışından alımı işlemlerinin bütünüdür. İhracat ve ithalat olarak iki önemli olgusu bulunmaktadır. Ülke ekonomisinin gelişimine oldukça önemli katkısı olması dolayısıyla ülkeler ihracatın ithalata kıyasla daha fazla olmasına yönelik dış ticaret politikaları izlemektedir. Bunlar, gümrük vergileri, ithalat yasaklamaları, sübvansiyon ve teşvik uygulamaları ile serbest ticaret anlaşmaları gibi politikaları kapsamaktadır.

Bu bölümde, öncelikle ihracat ve ithalat kavramları ve önemleri üzerinde durulmakta, ayrıca, pazar araştırması ve talep tahmini çalışmalarının içeriği ve işletmelere katkısı konularına değinilmektedir.

2.1. İhracat

Uluslararası pazarlamada ihracat için pek çok tanımlama bulunmaktadır. Bunlardan bazıları aşağıda sunulmaktadır:

- İhracat, doğrudan veya dolaylı olarak mal ve hizmetlerin bir ülkeden başka bir ülkeye transfer edilmesi veya dış pazarlara satışının gerçekleştirilmesi olarak tanımlanmıştır (Yücel, 2006: 4).
- İhracat, ticari bir malın Gümrük ve Kambiyo mevzuatına uygun bir şekilde ülkenin gümrük hattını terk edip yurtdışındaki alıcılara satılması ve (bedelsiz ihracatlar hariç olmak üzere) bedelinin yine söz konusu mevzuatlara uygun olarak yurda getirilmesidir.²
- İhracat için bir başka tanımlama serbest dolaşımda bulunan (söz konusu ülkede üretilmiş, yetiştirilmiş veya başka bir ülkeden ithal edilmiş) mal ve hizmetlerin diğer ülkelere satışının gerçekleştirilmesi şeklinde yapılabilir.³
- İhracat Yönetmeliğinde ihracat, “Bir malın, yürürlükteki ihracat mevzuatı ile gümrük mevzuatına uygun şekilde Türkiye gümrük bölgesi dışına

² Apak, T., (2003). İhracat Bedellerinin Tahsili-Kapatılması ve Kur Farkları (Gümrük-Kambiyo-Vergi-Muhasebe), <http://www.alomaliye.com/2003/06/30/ihracat-bedellerinin-tahsili-kapatilmasi-ve-kur-farklari-gumruk-kambiyo-vergi-muhasebe/> (erişim tarihi: 20.05.2020)

³ Adım Adım İhracat, TİMAkademî, 2014. <https://tr.scribd.com/doc/207549411/Adim-Adim-Ihracat-2014-Ilanli-Kitap3> (erişim tarihi: 21.11.2016)

veya serbest bölgelere çıkarılmasını veyahut Müsteşarlıkça ihracat olarak kabul edilecek sair çıkış ve işlemleri” olarak tanımlanmaktadır.⁴

Ekonomi teorisi ihracatı, uluslararası ekonomik performansın ana dengeleyici unsuru olarak ele almaktadır. Ayrıca ihracat, bir ulusun rekabet gücünün kamu tarafından algılanışını şekillendirmekte ve bir ülkenin katlanabileceği ithalat düzeyini (en azından uzun dönemde) belirlemektedir (Czinkota, 2002: 315-316).

2.1.1. İhracatın Önemi ve İşletmelere Sağladığı Faydalar

Rekabetin her geçen gün arttığı günümüz dünyasında ihracat, işletmelerin başarılı olmaları için oldukça önemli bir konumdur. Hem ülke ekonomisine hem de firmaların gelişimine oldukça önemli katkılar sunan ihracat, yeni iş olanakları yaratmakta ve ekonomiye hem dış ticaret dengesi hem de istihdam açısından katkı sağlamaktadır (Onur, 2012: 5). Bu nedenle, hem gelişmekte olan ülkelerde hem de gelişmiş ülkelerde ihracata büyük önem verilmekte ve farklı destek programları ile gelişimi yönünde çalışmalar yürütülmektedir (Yücel, 2006: 5).

Zou vd (1998: 37) ve Czinkota (2002:315)’ya göre, makroekonomik düzeyde ihracatın, ekonomik büyümenin makinesi olarak kabul edilmesi, döviz kuru değerlerini, maliye ve para politikalarını etkileyebilmesi nedenleriyle hükümetler firmalarının ihrac pazarlarındaki performansının artırılması hususuna önem vermektedir.

İhracatın işletmelere sağladığı faydalar ise aşağıdaki şekilde sıralanabilmektedir (Yücel, 2006: 5-6, Tosun, 2007: 36-37; Özbek, 2009: 20-21; Sürek, 2010: 42-43; Onur, 2012: 5):

- İşletmeler, ihracat ile satışlarını ve karlılığını artırır.
- Dış pazarlara açılan işletmeler, rakiplerini inceleme ve elde ettikleri bilgileri kendi üretim ve pazarlama stratejilerine uyarlama fırsatı da elde ederler. Böylece rekabet güçlerini artırır.
- İhracat, işletmelerin insan kaynakları ve ekipman açısından mevcut altyapısını etkin kullanmasına, atıl kapasitesinden kurtulmasına ve büyüyüp gelişmesine katkı sağlar. Ayrıca, işletmeler ihracat ile ömrünü tamamlamış veya tamamlamak üzere olan ürünlerini alternatif pazarlara yönlendirebilirler.

⁴ İhracat Yönetmeliği, T.C. Resmi Gazete, 06.06.2006, <https://www.resmigazete.gov.tr/eskiler/2006/06/2006060606-10.htm> , (erişim tarihi: 20.05.2020).

- İç pazardaki satışlarda ilerleme kaydedilememesi, yani pazarın doyuma ulaşması nedeniyle veya iç piyasada ömrünü tamamlayan ürünlerin pazarlanması amacıyla işletmeler ihracata yönelebilirler.
- Diğer taraftan, sadece iç pazara yönelik faaliyet göstermek çeşitli riskleri de beraberinde getirmektedir. İç pazarda oluşabilecek bir ekonomik kriz ve istikrarsızlık, enflasyon ve döviz kurlarındaki dalgalanmalar dolayısıyla oluşacak risk, dış pazarlara açılarak ve mümkün olduğunca çok sayıda pazara ihracat gerçekleştirilerek paylaştırılabilir.
- İç piyasanın yanı sıra ihracata yönelik olarak da faaliyet gerçekleştiren işletmeler, uluslararası piyasalardaki rekabet ortamına uyum sağlamak ve nihai tüketicilerin taleplerini karşılayabilmek için üretim ve pazarlama yapılarını yeniden düzenlemekte, rakiplerine karşı üstünlük sağlamak için araştırma ve geliştirme faaliyetlerine önem vermekte, uluslararası teknolojik yenilikleri takip edebilmekte ve bu yeniliklere uyum sağlamaya gayret göstermektedir. Bu durum da, sermaye ve işgücünün verimli kullanılmasına imkan sağlamaktadır.

2.2. İthalat

İthalat, yani dışalım, başka bir ülkeden mal ve hizmet tedariğidir. İhracat verileri ile birlikte bir ülkenin dış ticaret dengesini oluşturur. Bir ülke için ithalat, bir başka ülkenin söz konusu ülkeye gerçekleştirdiği ihracatın karşılığıdır. Bu doğrultuda, ülkemizin potansiyel ihracat değerlendirmelerini yapabilmek için öncelikle alıcı ülkelerin veya hedef pazarların ithalat verilerini incelemek gerekmektedir.

2.3. Talep Tahmini (Pazar Araştırması)

Talep, dış ticaret açısından bir mal veya hizmete sahip olma konusundaki istek olarak tanımlanabilmektedir. Talep tahmini (pazar araştırması) ise mevcut ve potansiyel alıcılarının gelecek dönemlerde talep edecekleri ürün veya hizmet miktarını belirleyerek işletmeye kapasite kullanımının belirlenmesi, rakiplerin pozisyonu, fiyat seviyeleri, tüketicilerin tercih ve alışkanlıkları ile ilgili bilgi vermesinin yanı sıra söz konusu talepler için gerekli işgücü ve hammadde kullanımına ilişkin hazırlık imkanı sağlamaktadır (Erdoğan vd., 2019:114). Talep tahmininden elde edilen bilgi, satış tahmini ve üretim planına temel hazırlamaktadır. Satış tahmini ise bir sektör veya işletmenin bir pazar bölümüne satmayı beklediği mal ve hizmet hacmini ifade etmektedir (Tek, 1999; 296).

Pazar araştırması süreci herhangi bir mal veya hizmete yönelik belirli bir ülke veya bölgede oluşması beklenen arz ve talebin belirlenmesi çalışmasıdır. Birçok uluslararası pazar seçimi çalışması potansiyel yabancı pazarın belirlenmesinde farklı yaklaşımlar ve kriterler önermiştir. Bu çalışmaların en eskilerinden olan Çavuşgil (1985:30) yurtdışı pazar potansiyelinin üç aşamada araştırılmasını önermiştir. Bu aşamaları şu şekilde sıralayabiliriz: (1) fiziksel, siyasi, ekonomik ve kültürel çevreyi incelemeye dayalı ön ülke taraması; (2) pazara giriş, ürün potansiyeli, yerel dağıtım ve üretim üzerinden incelenen ülkenin talebini belirlemeye yönelik pazar potansiyeli değerlendirmesi; (3) satış ve karlılık tahminlerine dayalı firma satış potansiyeli analizi.

Pazar araştırmasında dikkat edilmesi gereken hususlar (Sosyal ve Ömürgönülşen, 2010: 130-131; Erdoğan vd., 2019:114):

- Tahminleme araştırmasının amacının ve döneminin belirlenmesi,
- Araştırma yapılacak hedef pazar veya pazarların belirlenmesi,
- Araştırma için ne kadarlık verinin gerekli olduğu ile nereden temin edileceğine karar verilmesi,
- Elde edilebilecek verinin miktarı ve durumuna göre tahminleme çalışmasında kullanılacak yöntemin belirlenmesi,
- Araştırma konusu ürün için pazarda oluşan veya oluşabilecek talep miktarının, fiyat düzeyinin, tüketicilerin özelliklerinin (gelir durumları, cinsiyet, yaş, eğitim durumu vb.), ikame ürünlerin belirlenmesi,
- Pazardaki rakiplerin toplam talepten alacağı pay ile rakiplerin güçlü ve zayıf yönlerinin belirlenmesi

olarak değerlendirilmektedir.

Talep tahmini süreci genellikle dört adımda gerçekleştirilir: (1) Verinin toplanması, (2) Talep tahminin gerçekleştirileceği dönemin belirlenmesi, (3) Tahminlemede kullanılacak yöntemin seçimi, (3) Talep tahminin sonuçlarının değerlendirilmesi (Erdoğan vd., 2019:115).

Tahminleme sürecinde kullanılan teknikler kalitatif (kişisel görüş ve değerlendirmeye dayalı) ve kantitatif (sayısal analize dayalı) teknikler olarak ikiye ayrılabilir. Kalitatif teknikler daha çok doğrudan bilgi edinilemeyen, veri toplanamayan, ölçülemeyen veya gözlemlenemeyen durumlarda kullanılmaktadır. Uygulamada en çok kullanılan kalitatif teknikler: (1) Pazarın araştırılmasına yönelik anket, mülakat ve telefon görüşmeleri, (2) Uzman görüşleri ve (3) Talep verilerinin az olması durumunda uzman görüşlerine yazılı olarak başvurup ortak bir görüşe ulaşıncaya kadar

devam eden bir çalışma yöntemi olan delphi tekniğidir (Erdoğan vd., 2019:116; Demir ve Gümüşoğlu, 2003; 497).

Tahminleme sürecinde kullanılan kantitatif teknikler ise regresyon analizi yöntemleri, zaman serisi analiz yöntemleri ve korelasyon analizi olarak üç grupta toplanabilmektedir (Tütek ve Gümüşoğlu, 2000; 197; Erdoğan vd., 2019:116):

- Nedensel teknikler (regresyon analizi) yöntemleri iki veya daha fazla değişken arasındaki sebep-sonuç ilişkisi gözönünde bulundurularak gerçekleştirilen tahminleme yöntemidir. Regresyon analizi, ekonometrik modeller, yapay zeka ve sezgisel algoritmalar bu yöntemler arasındadır.
- Zaman serisi, bir olayın kronolojik olarak değerlerinin sıralandığı dizidir. Zaman serisi analizleri ise geçmiş dönem verilerinin zaman içerisindeki eğilimleri değerlendirilerek gelecek döneme ilişkin tahminlerin yapıldığı yöntemlerdir. Bunlara hareketli ortalamalar yöntemi, üstel düzeltme yöntemi, Box-Jenkins gibi yöntemler örnek olarak verilebilmektedir.
- Korelasyon analizi, regresyon doğrusunun bağımlı değişkeni ne ölçüde açıkladığını ölçmek için kullanılan bir tekniktir. Ayrıca, iki değişken arasındaki ilişkinin ölçülmesinde de uygulanmaktadır.

2.3.1. Talep Tahmini ve Pazar Araştırmasının Önemi

Her geçen gün daha fazla üretici ve ihracatçıların dahil olduğu ve alıcıların sayısının ise sınırlı olduğu uluslararası piyasalarda başarılı olmak ve yoğun rekabet ortamında ihracat gerçekleştirebilmek için pazarlama odaklı çalışmak, pazardaki taleplere ilişkin araştırma yapmak gerekmektedir. Uluslararası pazarlara girmeden önce cevaplanması gereken öncelikli soruların başında hangi ülkelere odaklanılması gerektiği gelmektedir. Hedef pazarın seçilmesi süreci ihracat-pazarlama stratejisi oluşturmanın da ilk adımıdır (Onur, 2012: 3-13).

İşletmeler açısından talep tahmini ve pazar araştırmasının önemi çeşitli çalışmalarda şu şekilde özetlenebilmektedir:

- İşletmeler, ihracat stratejilerini geliştirmeleri için öncelikle pazar araştırması yapması; hedef pazar alıcılarının taleplerini, pazardaki tedarikçi, dağıtıcı ve aracı firmaları araştırmaları, fiyat düzeyleri, mevzuat ve düzenlemeler hakkında, kısacası pazarla ilgili detaylı bilgi edinmelidir (Onur, 2012: 3).
- İşletmeler uluslararası pazarlarda kalıcı olabilmek için yeni pazarlar araştırmak ve yeni pazarlara açılmak, hedef pazarların taleplerini tam olarak

anlamak, bu talepler doğrultusunda üretimlerini gerçekleştirmek ve yeni ürünler geliştirmek, fiyat politikalarını iyi belirlemek, pazara uygun dağıtım kanallarını seçmek ve pazardaki mevcudiyetlerini sağlamlaştıracak faaliyetlerde bulunmak durumundadırlar. İhracata sadece üretilen ürünlerin yurt dışına satışı olarak bakmak işletmelerin sadece belirli ürünlerle ve belirli pazarlarda faaliyetinin sınırlı kalmasına sebep olmaktadır (Yücel, 2006: 1).

- Firmalar, yeni bir yurt dışı pazara girmek ve hizmet, ürün, teknoloji, işgücü ve diğer kaynaklar açısından işlerini geliştirmek için pazara giriş stratejisi kullanırlar. Söz konusu strateji kapsamında, firmaların, hangi pazara gireceklerine (pazar seçim kararı), nasıl gireceklerine (pazara giriş yöntemi kararı) ve ne zaman gireceklerine (pazara giriş zamanı kararı) karar vermeleri gerekir. Dolayısıyla pazar seçim kararı, uluslararasılaşma sürecinin önemli boyutlardan biridir ve küresel pazarda genişleme sürecini zorlukları ve karışıklıklarını azaltmak için oldukça önemlidir (Mat Isaa vd., 2013: 4).

Gelecek dönemde ne kadarlık bir talep olacağının tahminlenmesi üretim faaliyetleri açısından oldukça önemlidir. Talebin karşılığı tedariktir ve talebi karşılayabilmek için üretim planının bu doğrultuda yapılması ve gerekli kaynakların temin edilmesi gerekmektedir. Bir işletmenin üretim planı, hangi üründen hangi dönemde ne kadar üretileceğini ifade etmektedir. Üretim planının yapılabilmesi için ihtiyaç duyulacak veri, satış tahmini ve mevcut siparişlerden temin edilmektedir. Dolayısıyla, hammadde, yarı mamul, işgücü ve makine kullanım ihtiyacının belirlenmesi konusunda ana veri talep tahmininden elde edilmektedir (Krajewski vd., 2013: 483; Kobu, 2013:455-456).

Talep tahminleri çalışmasına yeterince önem verilmemesi veya doğru bir çalışma yürütülmemesi firmaların talep fazlalığı veya azlığı ile karşılaşmasına, kapasitelerini doğru kullanamamasına, satış ve gelir kayıpları ile karşılaşmasına ve hatta iflas etmesine kadar varan sorunlara sebebiyet verebilmektedir. Diğer taraftan, pazar açısından değerlendirildiğinde ise arz bolluğu veya yetersizliğine ve piyasada ürün satış fiyatlarında dalgalanmalar yaşanmasına neden olabilmektedir. Bu açıdan, talep tahmini süreci titizlikle ele alınmalıdır.

ÜÇÜNCÜ BÖLÜM

MODELLEME TEKNİKLERİ VE ALGORİTMALAR

3.1. Modelleme Teknikleri

Veri madenciliği yöntemlerinde, veri genelde eğitim ve test verisi olarak ikiye ayrılmaktadır. Eğitim veri seti modele sunulmakta ve modelin veriyi öğrenmesi amacıyla kullanılmaktadır. Daha önce modele gösterilmeyen test veri seti ise modelin veriyi ne kadar iyi öğrendiğini test etmektedir. Eğer bir model, test verisi üzerinde doğru sonuçlar veriyorsa, modelin eğitim setinden test setini genelleyeabildiği söylenebilmektedir. Veri madenciliği çalışmalarında modelin mümkün oldukça doğru sonuçlara genelleme yapabilmesi amaçlanmaktadır.

Sahip olunan bilgi için fazla karmaşık bir model kurulması durumunda aşırı öğrenme gerçekleşmektedir. Aşırı öğrenme, eğitim setinin özelliklerine çok yakın ve eğitim setinde çok iyi sonuç veren ama yeni veride genelleme yapamayan modellerde gerçekleşmektedir. Diğer taraftan, kurulan model çok basitse veriyi tüm özellikleri ve değişkenleri açısından yakalamak mümkün olamayabilmektedir. Bu durumda eğitim setinde bile uyumsuz sonuçlar elde edilmektedir. İşte, bu şekilde çok basit bir model seçilmesine de yetersiz uyum veya eksik öğrenme denmektedir (Müller ve Guido, 2017: 28; Tan vd., 2006:173).

Aşırı öğrenme durumunda daha fazla örneklem toplamak, modelin öğrenme kabiliyetini geliştirmekte ve bu sayede ezberleme sorununu çözülebilmektedir. Genellikle, daha fazla veri noktası toplamak daha fazla esneklik yaratmaktadır. Ancak, yineleyen veri noktaları veya birbirine çok benzeyen veri noktaları bu konuda modelin öğrenme becerisini kolaylaştırmamaktadır (Wang vd., 2006: 736, Babyak, 2004:414). Eksik öğrenme durumunda modelin karmaşıklığı artırılmakta ve modele daha fazla öznelik dahil edilmektedir. Kullanılan veri madenciliği yönteminde parametrelerinin değerlerinin değiştirilmesi hem aşırı öğrenme hem de eksik öğrenme durumlarında tercih edilen bir uygulamadır.

Veri madenciliği yöntemleri, veri madenciliği görevinde bulunulacak örüntü çeşitlerinin belirlenmesi için kullanılmaktadır. Genel olarak bu tür görevler tanımlayıcı ve tahminleyici olarak ikiye ayrılmaktadır (Han ve Kamber, 2012: 15).

3.1.1. Tanımlayıcı Modeller

Tanımlayıcı model, bir hedef veri setinin özelliklerini karakterize etmekte, verideki örüntüleri (korelasyon, trend, küme ve anomali gibi) ve ilişkileri tanımlamakta ve çalışılan verideki özellikleri keşfetmektedir. Kümeleme, Özetleme ve Birliktelik Kuralları gibi yöntemler tanımlayıcı modellere örnek olarak verilebilmektedir (Han ve Kamber, 2012: 15; Dolgun, 2014:18; Tan vd., 2006: 7).

Kümeleme, birbirine yakın olan veya birbirleri ile ilgili gözlemlerin bulunması ve bir kümede gruplanması işlemidir. Kümeleme, sınıflandırmaya benzer bir yapıdadır, ancak kümelemede gruplar daha önceden tanımlanmamıştır (Tan vd., 2006: 9-10). Kümeleme, bir grup verinin sınıf etiketlerini oluşturmada kullanılabilmekte olup nesneler, sınıf içi benzerliği en fazla yapma ve sınıflararası benzerliği en az yapma prensipini temel alarak kümelenebilmektedir. Bu durum, kümelerin nesnelerin, küme içindekilerin benzerliğinin yüksek ancak, diğer kümedekilere kıyasla benzerliğinin düşük olacağı şekilde oluşturulmasıdır (Han ve Kamber, 2012: 19-20).

Özetleme, veriden özet bilginin sunulması tekniğidir (Oğuzlar, 2003:74). Birliktelik kuralları ise farklı özniteliklerin arasında birliktelikleri bulma ve verideki güçlü birliktelik özellikleri içeren örüntüleri keşfetme sürecidir (Tan vd., 2006:9).

3.1.2. Tahminleyici Modeller

Tahminleyici modellerde ise mevcut veri üzerinde tahminler elde etmek için çıkarsama yapılmaktadır. Tahminleme, özniteliklerin mevcut ve geçmiş durumunu analiz edip, bir veya bir kaç özniteliğin gelecekteki durumunu tahmin etme sürecidir. Burada tahmin edilen öznitelik hedef, çıktı veya bağımsız değişken olarak isimlendirilirken, tahminleme için kullanılan öznitelikler bağımlı değişken, girdi veya açıklayıcı olarak isimlendirilmektedir. Sınıflandırma, regresyon ve zaman serileri analizi gibi yöntemler tahminleyici modellere örnektir (Han ve Kamber, 2012: 15; Tan vd., 2006: 7; Bowles, 2015: 27).

Sınıflandırma, hedef verinin daha önceden tanımlanan grup veya sınıflara yerleştirilmesi tekniğidir. Hedef verilerin incelenmesinden önce sınıflar daha önceden tanımlanmış olduğundan, bu bir denetimli (danışmanlı) öğrenmedir. Model, eğitim veri setinin analizlerinden sınıf etiketleri bilinmeyen nesnelerin sınıf etiketlerini belirlemek için kullanılır. Elde edilen model, sınıflandırma kuralları, karar ağaçları, matematiksel formüller veya sinir ağları gibi pek çok farklı formda temsil edilebilmektedir (Han ve Kamber, 2012: 18).

Tahmin ifadesi hem sayısal tahmin hem de sınıf etiketi tahmini anlamına gelmektedir. Sınıflandırma kategorik (ayrık, sırasız) etiketleri tahmin ederken, regresyon ise sürekli değerli fonksiyonları modeller. Bir başka ifadeyle, regresyon sınıf etiketleri yerine sürekli değerlere sahip sayısal veri değerlerini tahmin etmek için kullanılır. Regresyon analizi, sayısal tahmin için en sık kullanılan yaklaşımdır. Regresyon, mevcut verilere dayalı olarak dağılım trendlerinin tanımlanmasını da kapsar (Han ve Kamber, 2012: 19, Bowles, 2015: 27).

Zaman serileri, veri setindeki her değer farklı bir zaman değeri ile bu değere karşılık özniteliğin verdiği değerden oluşmaktadır. Zaman serileri analizinde özniteliğin değerinin zamana göre değişimi incelenmektedir. Hava tahminleri veya stok tahminleri zaman serileri analizine örnek olarak verilebilmektedir (Witten vd., 2016: 330)

3.2 Makine Öğrenmesi Algoritmaları

Makine öğrenmesi insanlara oldukça doğal gelen deneyimlerden öğrenme sürecini bilgisayarlara öğreten bir veri analitiği tekniğidir (Han ve Kamber, 2012: 24).

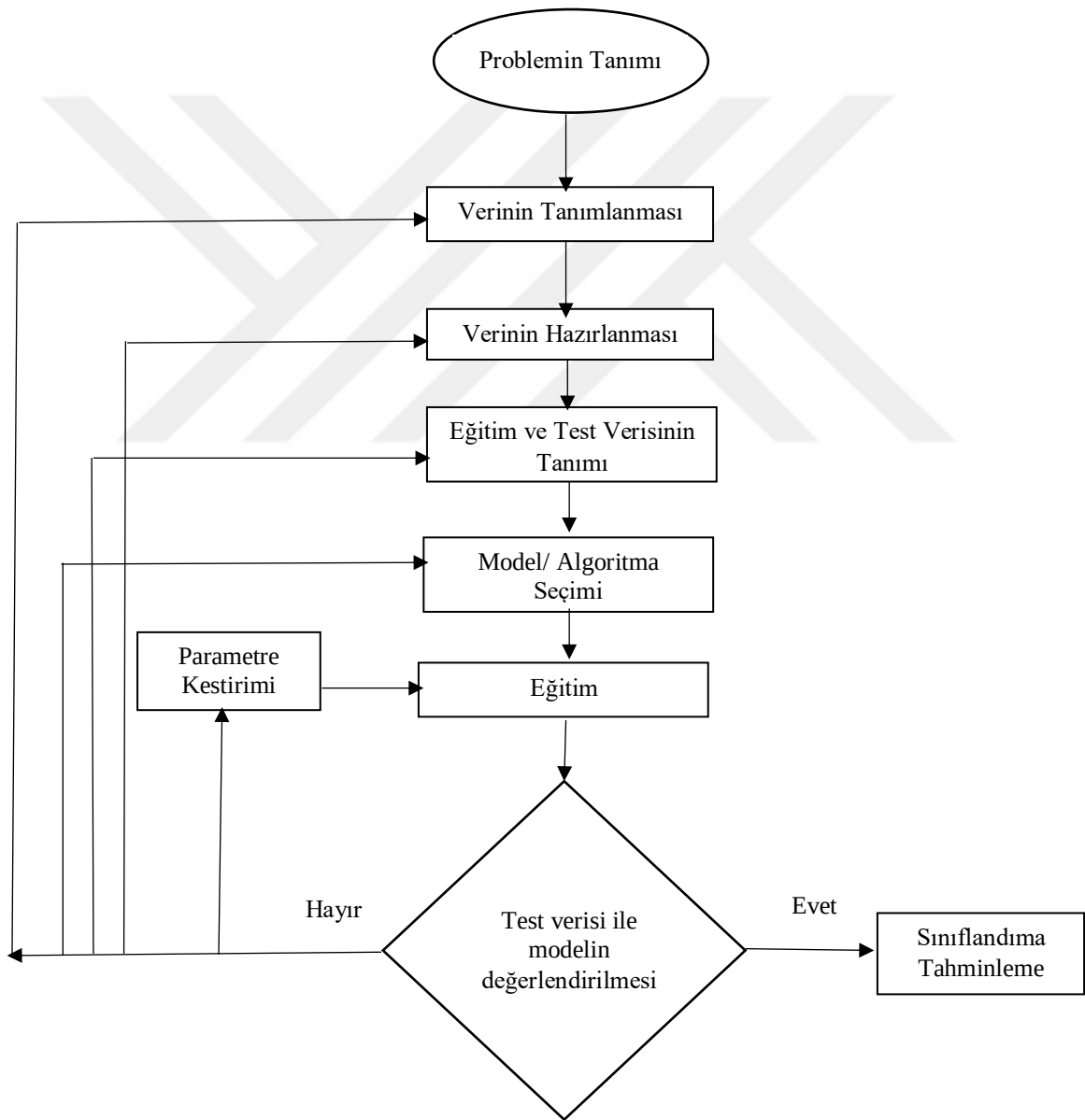
Das ve Behera (2017: 1301) ise makine öğrenmesini, geleceğe yönelik performansı iyileştirmek için geçmiş dönem deneyimlerinden öğrenme olarak tanımlamışlardır.

Makine öğrenmesi algoritmaları, veriden doğrudan bilgiyi öğrenmeyi sağlayan hesaplama yöntemlerini kullanmakta olup veriye bağlı olarak bilgisayarların nasıl öğrendiğini veya performanslarını nasıl artırdığını araştırmaktadır. Öğrenme, insan desteği olmaksızın geçmiş deneyimleri üzerine kurulu bir algoritmanın otomatik olarak iyileştirilmesi ve geliştirilmesi şeklinde gerçekleştirilmektedir. Ana araştırma alanı ise, karışık örüntüleri tanımak ve veriye dayalı akıllı kararları vermek için otomatik öğrenme işlemidir. Elektronik postaları filtreleme, el yazısı tanımlama, görüntü tanıma gibi çalışmalar makine öğrenmesi çalışmalarındandır. Örneğin, tipik bir makine öğrenme problemi, örnekler setinden öğrenmenin ardından mektuplar üzerindeki el yazılarını otomatik olarak tanıyabilecek şekilde bilgisayarı programlamaktır (Han ve Kamber, 2012: 24; Das ve Behera, 2017: 1301).

Makine öğrenmesi algoritmaları üç ana sınıfta incelenebilir. Bunlar denetimli (gözetimli) öğrenme, denetimsiz (gözetimsiz) öğrenme ve takviyeli öğrenmedir (Kotsiantis vd., 2007: 3-5; Das vd., 2015: 32-37; Das ve Behera, 2017: 130; Binici, 2019: 118; Avuçlu ve Ekmekci, 2020: 36).

- Denetimli öğrenme, etiketlenmiş örnekler yardımıyla yapılır. Sınıflandırma ve regresyon bu öğrenme biçimini kullanan tipik yaklaşımlardır. Denetimli

öğrenmede modele öğrenme aşamasında girdi bilgilerinin yanısıra çıktı bilgileri de sunulmaktadır. Model, girdi bilgisine göre üretmiş olduğu çıktı bilgisini beklenen çıktı değeri ile karşılaştırır, hata miktarını hesaplar, ağırlık değerlerini hata değerine göre yeniden düzenler. Bu süreç hata değeri daha önceden belirlenmiş olan orana ulaşana kadar devam eder. Denetimli makine öğrenmesi uygulamalarına, spam postaları tespit etme, yüz, ses ve el yazısı tanıma, metin filtreleme, verileri sınıflandırma gibi çalışmalar örnek olarak verilebilmektedir. Denetimli öğrenmenin aşamaları ile ilgili örnek görsel Şekil 3.1’de yer almaktadır.



Şekil 3.1 Denetimli Makine Öğrenmesinin Aşamaları (Kotsiantis, 2007: 250).

- Denetimsiz öğrenmede girdi örnekleri sınıf etiketli olmadığı için öğrenme süreci danışmansızdır. Tipik olarak, veri içerisindeki sınıfları keşfetmek için kullanılabilir. Denetimsiz öğrenmede, sadece eğitim veri setindeki girdiler sunulur. Çıktı değerlerini karşılaştırabileceği beklenen çıktı değerleri yoktur. Model, girdi verilerine göre eğitim veri setini gruplandırır ve kendi kurallarını oluşturur. Kümeleme analizi ve sepet analizi bu öğrenme türüne ilişkin örnek yöntemlerdir. Denetimsiz makine öğrenmesi uygulamalarına gen tanıma, pazar bölümlendirme ve kümeleme, sosyal ağ analizi gibi çalışmalar örnek olarak verilebilmektedir.
- Destekli veya takviyeli öğrenme ödül – ceza prensibine dayanmaktadır. Algoritma hareketleri doğrultusunda çevreden bir tepki beklemektedir. Bu tepki ödül veya ceza olabilir. Kazanılan ödül doğrultusunda eğitim süreci gerçekleşmekte, yaptığı eylemin ne kadar doğru veya yanlış olduğunu öğrenmektedir. Genellikle oyun yapay zekalarında kullanılmaktadır.

Bilgisayar programlarında olduğu gibi veri üzerinde geliştirilen bilgi keşfi çalışmalarında da algoritmalarından yararlanılmaktadır (Özdarıcı Ok vd., 2011:2). Bu bölümde uygulamada kullanılmakta olan makine öğrenmesi yöntemlerinden XGBoost, Rastgele Orman Algoritması ile Yapay Sinir Ağları ve söz konusu yöntemlerin temelini oluşturan Karar Ağaçlarına değinilmektedir.

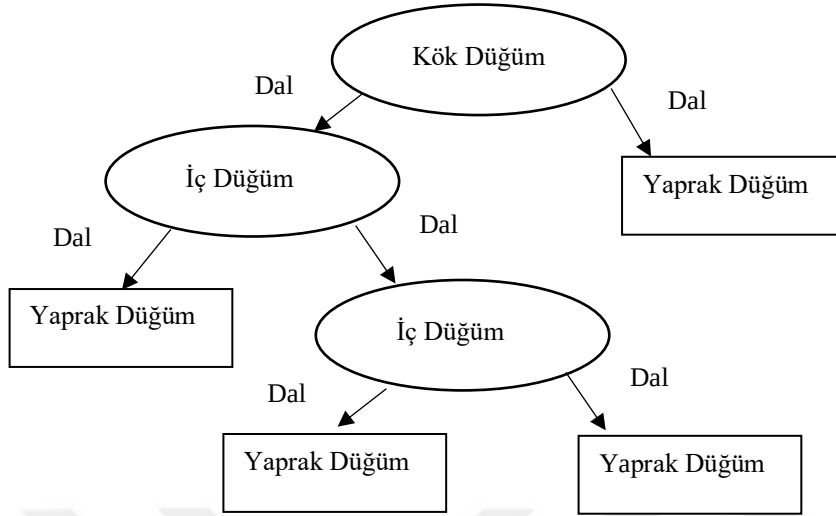
3.2.1. Karar Ağaçları

Karar ağaçları, hedef değeri hesaplamak için ikili kurallar seti kullanan kestirimci modellerdir. Hem sınıflandırmada hem de regresyon analizinde sıklıkla kullanılmaktadırlar. Evet/hayır, doğru/yanlış şeklindeki kurallar yardımıyla veriden bilgi keşfetmeyi amaçlamaktadırlar (Horning, 2010:2; Das ve Behera, 2017: 1305).

Ağaçtaki her düğüm, ya bir soruyu veya bir soruya ait yanıtı temsil etmektedir (Müller ve Guido, 2017: 70-71). Karar ağacı inşası algoritması, tüm yaprak düğümleri tek sınıf düğümü olana veya birden fazla sınıftan nesneler içeren bir düğümü bölmek için daha fazla öznitelik bulunmayana kadar devam eder (Gupta, 2016: 169).

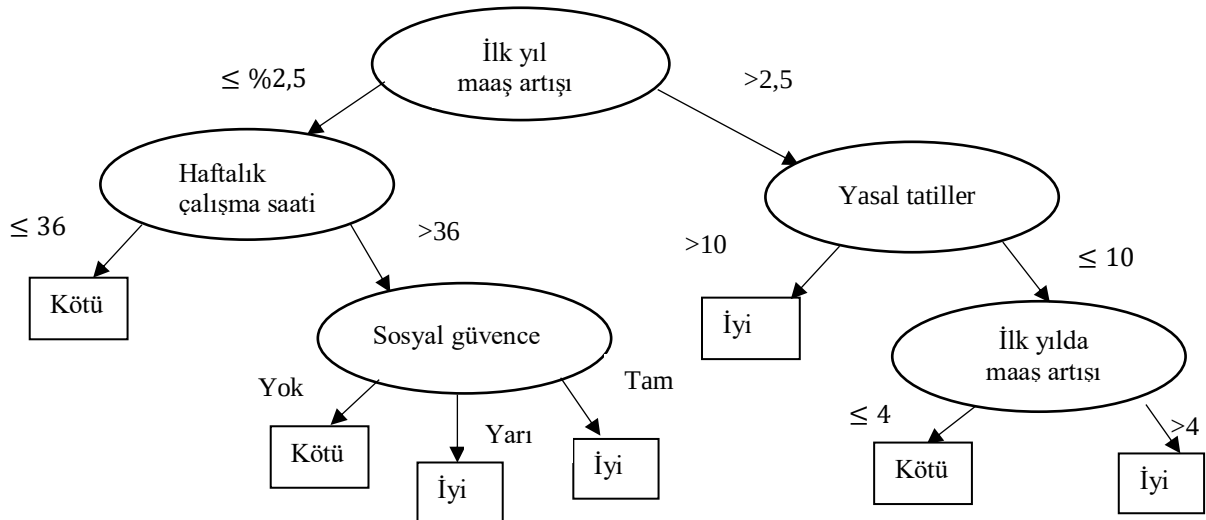
Kök, düğüm, iç düğüm, dal ve yapraktan oluşan bir karar ağacı örneği Şekil 3.2’de sunulmaktadır. Veri setinde yer alan her bir öznitelik bir “düğüm”de temsil edilmekte olup, en üstte “kök düğüm” yer almaktadır. Son düğüm olan “yaprak düğüm” ise bir sınıfı temsil etmektedir. Kök düğüm ile yaprak düğüm arasındaki ara düğümler “iç düğüm”, düğümlerin arasında kalan bağlantılar da “dal” olarak isimlendirilmektedir. Her düğüm,

ağaca sunulan veri setine verilen cevaba göre dal veya dallara bölünmektedir (Sabancı, 2019: 12).



Şekil 3.2 Karar Ağacı Örneği (Sabancı, 2019: 12)

Şekil 3.3’de işçi görüşmeleri verisine ilişkin örnek bir karar ağacı yer almaktadır. Bu örnekte birinci yılın sonunda yapılacak maaş artış oranı %2,5’den daha azsa ve haftalık çalışma saati otuzaltı saatin altındaysa bu işçi tarafından kötü bir durum olarak algılanmakta; ilk yılın sonunda verilecek maaş artışı %2,5’den fazla olması ve yasal tatilleri on günden fazla olması durumu ise işçi tarafından iyi bir durum olarak kabul edilmektedir.



Şekil 3.3 İşçi Görüşmeleri Verisine İlişkin Örnek Bir Karar Ağacı (Witten vd., 2016: 18)

Karar ağaçlarında aşırı öğrenmeyi engellemek için iki temel strateji kullanılmaktadır. Bunlar, ağacın oluşturulmasını erken durdurma (ön budama) ile ağacı oluşturup az bilgi içeren düğümleri çıkarma veya daraltma (geç budama) stratejileridir. Ön

budama için olası kriterler ağacın maksimum derinliğini (dal sayısını) sınırlamak, yaprak sayısını belirlenecek bir maksimum değerle sınırlamak, bölünmeye devam edebilmesi için bir düğümdeki nokta sayılarını minimumda tutmaktır (Müller ve Guido, 2017: 74). Geç budamada ise ağacın büyümesinin tamamlanmasının ardından, yeni sunulan verilerde karar ağacının doğruluğunu artırmak amacıyla ağaç daha sonra budanır. Geç budama, erken durdurma riski olan ön budamaya göre genellikle daha iyi sonuçlar vermektedir (Polat, 2008: 71; Gupta, 2016: 185).

3.2.1.1. Karar Ağacı Algoritmaları

Karar ağaçlarında kullanılan algoritmalarından bazıları aşağıda listelenmektedir.

- CART (Classification and Regression Tree-Sınıflandırma ve Regresyon Ağaçları)
- CHAID (Chi-squared automatic interaction detector- χ^2 Etkileşim Dedektörü)
- Ayrıntılı CHAID
- QUEST (Quick, Unbiased, Efficient Statistical Tree- Hızlı, Tarafsız, Verimli İstatistik Ağacı)
- C4.5 karar ağacı
- C5 karar ağacı
- ID3 ağacı
- MARS (Multivariate Adaptive Regression Splines-Çok Değişkenli Uyumlu Regresyon Şeritleri)

Bu tez çalışmasında CART temelli çalışan yöntemler kullanıldığı için CART algoritması açıklanmaktadır.

3.2.1.2. CART (Classification and Regression Tree-Sınıflandırma ve Regresyon Ağaçları) Algoritması

CART algoritması Breiman vd. (1984) tarafından geliştirilmiştir. Kök düğümünden başlayarak ağacı her defasında ikili bölünmeler ile dallara ayırarak çalışmaktadır. Hem regresyon hem de sınıflandırma problemlerinde kullanılmaktadır. Tahmin edilecek sonuç kesikli ise Sınıflandırma, sürekli ise Regresyon ağacı olarak kullanılmaktadır. Mümkün olduğunca homojen yapraklar elde etmek amacıyla çeşitli ölçümleri kullanarak bölünmeler gerçekleştirir. Bir düğümün saflık ölçütü homojenliğini ifade etmektedir. Saflık ölçütü “0” olduğunda düğüm tamamen homojen olmaktadır. Karar ağaçlarında en çok kullanılan bölünme ölçütleri “Gini” indeksi ile “Entropi” dir. En düşük “Gini” veya “Entropi”

değerini veren öznitelik dallara ayırma konusunda optimal öznitelik kabul edilmektedir (Li vd., 2019:194-195; Akkücüük, 2011:74-75).

Tezin analiz bölümünde kullanılan makine öğrenmesi algoritmalarında bölünme ölçütü olarak Gini indeksi kullanıldığı için Gini indeksinin formülasyonuna yer verilmektedir.

S düğümü için Gini indeksi, formülasyon (3.1)'deki gibi hesaplanmaktadır (Li vd. , 2019:194-195; Bezek Güre, 2019:71-72):

$$Gini(S) = 1 - \sum p_i^2 \quad (3.1)$$

Bu formülasyonda p_i , S düğümündeki i sınıfına ait nispi olasılığı ifade etmektedir. Gini indeksinin “0” olması için bir sınıfın olasılığının “1” olması gerekir. Bir başka ifadeyle, “saflık” durumunun söz konusu olması için Gini indeksi “0” olmalıdır.

3.2.1.3. Karar Ağaçlarının Avantaj ve Dezavantajları

Karar ağaçlarının avantajları aşağıda sunulmaktadır (Müller ve Guido, 2017: 82-83; Gupta, 2016:170, Albayrak ve Yılmaz, 2009:42):

- Parametrik olmayan yöntemlerden olması dolayısıyla istatistiksel varsayımlara ihtiyaç duymamaktadır.
- Sonuçlanan model, en azından küçük ağaçlar için kolay bir şekilde görselleştirilebilmekte ve uzman olmayan kişiler tarafından da anlaşılabilir olmaktadır.
- Her özelliğin ayrı bir şekilde işlenmesi ve verinin olası bölünmelerinin ölçeklendirmeye bağlı olmaması dolayısıyla normalizasyon veya standardizasyon gibi ön işlemlere ihtiyaç duymamaktadır. Veri tamamen farklı ölçeklerde özelliklere sahipse veya ikili ve sürekli özelliklere sahipse oldukça iyi sonuç verir.
- Hem sürekli hem de kategorik verileri ele alabilmekte ve anlaşılabilir sonuçlar üretmektedir.
- Karar ağacı yaklaşımı eğitim verisindeki kayıp verilerle baş edebilmekte ve verideki bazı hataları tolere edebilmektedir.

Bunun yanı sıra literatürde karar ağaçlarının dezavantajlarından aşağıdaki şekilde bahsedilmektedir (Kuzey, 2012: 77; Gupta, 2016:170; Müller ve Guido, 2017: 82-83):

- Hisse fiyatları veya faiz oranları gibi sürekli değişkenlerin değerini tahminlemeye yönelik çalışmalar için karar ağaçları çok uygun değildir.

- Sınıf başına eğitim veri setindeki gözlem sayısının düşük olması durumunda karar ağaçları büyük hatalar oluşmasına sebep olabilmektedir.
- Birçok özniteliğin ve büyük hacimli eğitim verisinin bulunduğu büyük problemlerde maliyetli olabilmektedir.
- Karar ağaçlarında karmaşıklığı kontrol eden en önemli parametre, henüz ağaç tam olarak geliştirilmeden ağacın oluşumunu durduran dal sayısı (max depth) parametresidir. Bu parametrenin çok büyük olması aşırı öğrenme, çok düşük olması ise eksik öğrenmeye yol açmaktadır.
- Birbiri ile etkileşimli çok sayıda öznitelik olması durumunda karar ağaçları iyi sonuç vermekle birlikte karmaşık etkileşime sahip veri setlerinde daha düşük performans göstermektedirler. Veri setindeki gürültüye karşı hassasiyet göstermektedir.
- Karar ağaçlarında ön budama kullanılsa bile aşırı öğrenmeye eğilimlidirler ve zayıf genelleme performansına sahiptirler.

Karar ağaçlarının aşırı öğrenme sorununa çözüm getirmek amacıyla topluluk öğrenmesi yöntemleri kullanılmaktadır. Eğitim setinden çok sayıda farklı eğitim setlerinin oluşturulduğu, modelin bu eğitim setleri üzerinden öğrenmesi sonucunda oluşan sonuçların değerlendirildiği ve daha güvenilir kararlar verebilmek için bir çok modelin çıktılarının birleştirildiği yöntemler topluluk öğrenmesi yöntemleri olarak adlandırılmaktadır (Witten, 2016: 351-352).

Topluluk öğrenmesi yaklaşımının genel prensibi, öğrenme modellerinin birleşiminin genel sonucun performansını artıracak yönündedir. Her bir ağaç göreceli olarak iyi tahminlerde bulunabilir, fakat öğrenme sırasında verinin bir bölümü için aşırı öğrenme gösterebilmektedir. Çoğu durumda, bir topluluk modelinden elde edilen sonuç, bir bireysel modelden elde edilen sonuçlardan daha iyi olmaktadır. İyi sonuç veren ve farklı açılardan aşırı öğrenme olan çok sayıda karar ağacı oluşturulursa sonuçların ortalamasını alarak aşırı öğrenme miktarı azaltılabilmektedir (Müller ve Guido, 2017: 83, Horning, 2010: 2; Bowles, 2015: 226).

Topluluk modellerinin en bilinenleri torbalama (bagging) ve artırma (boosting) yöntemleridir. Topluluk öğrenmesi algoritmalarının keşif aşamaları ile ilgili evrim süreci Şekil 3.4'te sunulmaktadır.



Şekil 3.4 Karar Ağacı Algoritmalarının Topluluk Algoritmalarına Evrimi

3.2.2. Torbalama (Bagging)

Torbalama yöntemi, Breiman (1996) tarafından geliştirilmiştir. Önyükleme (bootstrap) ve toplama (aggregating) ifadelerinin birleşiminden (bagging) isimlendirilmiştir. Sınıflandırma yöntemlerinin doğruluğunu artıran ve varyansı düşürerek aşırı öğrenme sorununu da ortadan kaldıran, makine öğrenmesi temelli bir topluluk modelidir (Breiman, 1996: 123-124).

Öncelikle önyüklemeli çok sayıda eğitim alt kümeleri oluşturulmakta ve oluşturulan bu eğitim setlerinin her biri için bir ağaç üretilmektedir. Üretilen ağaçlardan elde edilen sonuçlar bir arada değerlendirilmekte, sınıflandırma problemleri için en fazla oy alan sonuç, tahminleme problemleri için ağaç sonuçlarının ortalaması esas alınmaktadır (Aydemir vd., 2020: 187; Bühlmann ve Yu, 2002: 927-928).

Torbalama algoritmalarının genel işleyişine ilişkin algoritma Tablo 3.1’de sunulmaktadır.

Tablo 3.1 Torbalama Algoritması

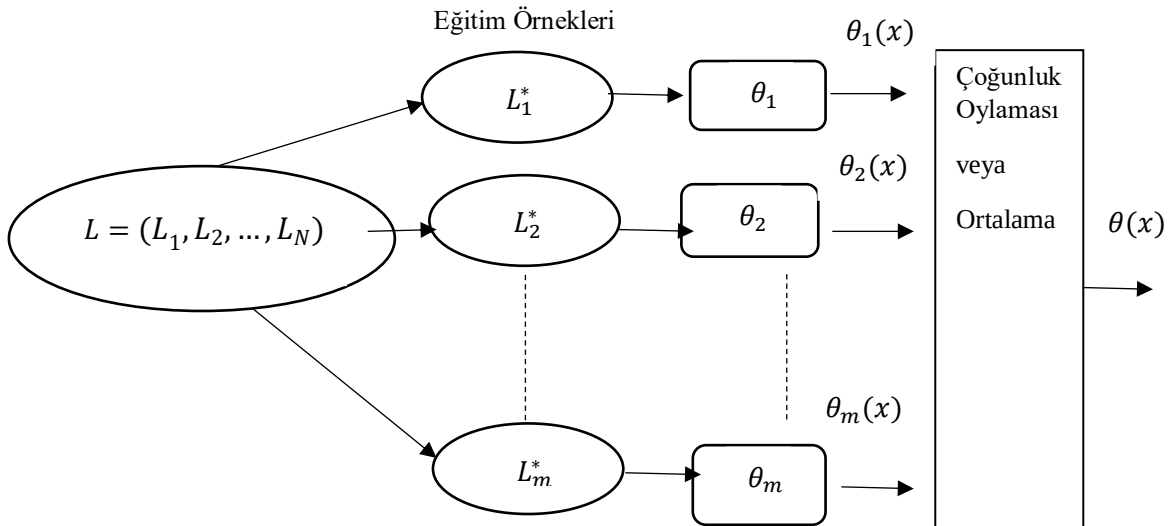
$i = 1, \dots, N$ için

- p boyutlu girdi değişkeni seti X_i ($X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})$)
- Y_i sınıf etiketi veya nümerik bir yanıt iken
- i gözlemine ait eğitim veri seti $L_i = (X_i, Y_i)$ olarak tanımlandığında
- x girdi değişkeni gözlemine karşılık tahminleme fonksiyonu $\hat{\Theta}_n(x) = h_n(L_1, \dots, L_n)(x)$ olarak tanımlandığında

1. N boyutlu eğitim veri setinden $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N)\}$ seçimler gerçekleştirerek ve yapılan seçimler yerine tekrar konularak (yani iadeli örneklemeyle) n çaplı ($n \leq N$) m adet önyükleme örnekleri $L_i^* = (X_i^*, Y_i^*)$, ($i = 1, \dots, m$) belirlenir. Her önyükleme örneğinin boyutu, en fazla orijinal eğitim veri seti kadardır. Ancak, iadeli seçimler yapıldığı için bazı değerler birden fazla defa önyükleme setlerinde yer alabilmektedir. n değerinin çok büyük olması (örneğin $n = N$) durumunda, önyükleme örnekleminde eğitim setinin %63,2'sinin olması beklenmektedir.
2. Önyükleme örneğinden $\hat{\Theta}_n^*(x)$ tahminleme fonksiyonu ile sonuçlar elde edilir.
3. m adet ön yüklem sonuçları değerlendirilir; sınıflandırma problemi için oylama yada tahminleme problemi için ortalama ile nihai sonuç elde edilir.

Kaynak: Breiman, 1996: 123; Bühlmann ve Yu, 2002: 927-928; Korkem, 2003:3; Namlı ve Özcan, 2017: 138

Torbalama baz sınıflandırıcıların varyansını azaltarak genelleme hatasını iyileştirmektedir. Her örneğin seçilme olasılığı eşit olduğundan torbalama eğitim verilerinin belirli bir örneğine odaklanmamaktadır. Bu sayede, torbalama gürültülü verilere uygulandığında aşırı öğrenme konusunda daha az duyarlı olmaktadır (Tan vd., 2006: 285).



Şekil 3.5 Torbalama Yaklaşımı (Ferreira ve Figueiredo, 2012:39).

Şekil 3.5'te torbalama süreci görsel olarak ifade edilmektedir. Torbalama ile elde edilen çok sayıda örneklem ile model eğitilmekte ve sonrasında çoğunluk oylaması veya ortalama ile sonuç elde edilmektedir.

3.2.3. Rastgele Orman Algoritması

Breiman (2001) tarafından geliştirilen Rastgele Orman Algoritmaları torbalama algoritmalarının gelişmiş bir modelidir. Rastgele Orman Algoritmasında çok sayıda karar ağacı oluşturulmakta olup, bu karar ağaçları orman olarak nitelenmektedir. Ormandaki her ağaç sınıflandırmada bir kategorik değer, regresyonda ise bir nümerik değer tahmin eder ve nihai tahmin bu ağaçların bireysel tahminlerinin birleştirilmesinden elde edilir (sınıflandırma problemlerinde çoğunluk oylaması, regresyonda ise ortalamasının alınmaktadır). Rastgele Orman Algoritmalarının gücü, iki farklı mekanizmayı; torbalama ve rastgele özellik seçimi mekanizmalarını bir arada kullanarak bir ağacı oluşturmak için veri noktalarının seçimi ve her bölünme noktası için özniteliklerin seçiminde rassallık sağlanmaktadır (Gislason vd., 2006: 295; Namlı ve Özcan, 2017: 137; Akman vd., 2011: 37; Müller ve Guido, 2017: 83).

Rastgele Orman Algoritmasının torbalama yönteminden farkı, ağacın oluşturulması sürecinde torbalama yönteminin yaptığı gibi veri kümesindeki tüm öznitelikleri kullanmak yerine her düğümde rastgele seçilen özniteliklerin arasından en iyi olanını kullanarak dallara bölünmesidir. Ağaçlar, rastgele seçilen özniteliklere göre inşa edilmekle birlikte ormanın gelişiminde her ağaç Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Tree-CART) standardında oluşmakta, Gini indeksini en düşük yapacak şekilde bölünmektedir (Boulesteix vd., 2012: 496; Akar, 2013: 28-29; Bezek Güre, 2019: 71).

3.2.3.1. Rastgele Orman Algoritmasının Parametreleri

Rastgele Orman Algoritmasında analiz sürecinde kullanılan pek çok parametre olmakla birlikte bu bölümde, sıklıkla kullanılan ve tezin analiz sürecinde de yer verilen *n_estimator*, *max_features*, *max_depth*, *min_sample_leaf* ve *min_samples_split* parametrelerine değinilmektedir (Liaw ve Wiener, 2002: 18; Gislason vd., 2006:295;⁵)

n_estimator: Ormandaki ağaç sayısını ifade etmektedir. Bu parametre, ne kadar büyükse o kadar iyi kabul edilmektedir. Ancak, ağaç sayısı ile paralel bir şekilde algoritmanın çalışma süresi de artmaktadır. Ayrıca, kritik sayının üzerindeki bir ağaç sayısı ile algoritmanın çok daha iyi sonuçlar elde etmeyi bırakacağı da unutulmamalıdır.

max_features: Bir düğümü bölerken dikkate alınan rastgele öznitelik alt kümesinin boyutudur. Bu değer, ne kadar düşükse varyansdaki düşüşün o kadar büyük olacağı, ne kadar yüksekse de yanlılığın o kadar artacağı düşünülmektedir. Varsayım değerleri olarak, regresyon problemlerinde *max_features* öznitelik sayısına eşit,

⁵ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> (erişim tarihi: 20.07.2020)

sınıflandırma problemlerinde ise öznitelik sayısının karkeköküne eşit değer alınmaktadır. Düğümlerin bölünmesi aşamasındaki öznitelik sayısının kısıtlanması hem algoritmanın hesaplama karmaşıklığını hem de ağaçlar arası korelasyonu düşürmektedir.

min_samples_split: Bir ara düğümün bölünmesi için gerekli olan minimum örnek sayısıdır. Hiçbir değer verilmediğinde varsayılan değer olarak “2” sayısı atanmaktadır.

max_depth: Ağacın maksimum derinliğini ifade etmektedir. Herhangi bir değer verilmediğinde tüm yapraklar saf olana kadar veya tüm yapraklar *min_samples_split* değerinin altına inene kadar düğümler bölünmeye devam etmektedir.

min_sample_leaf: Bir ağacın son noktası olan yaprak düğümünde olması gereken minimum örnek sayısıdır. Herhangi bir ağaç derinliğindeki noktada bölünme sağ ve sol dalların her birinde bıraktığı eğitim örneği en az *min_sample_leaf* kadarsa gerçekleştirilebilmektedir.

3.2.3.2. Rastgele Orman Algoritmasının İşleyişi

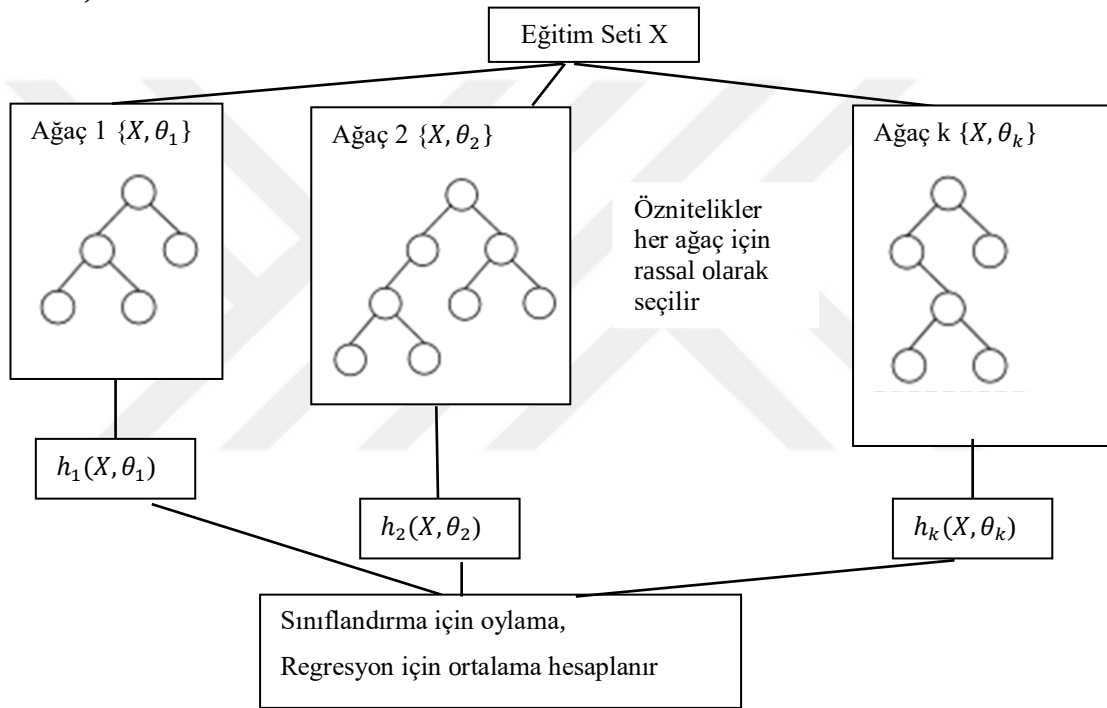
Rastgele Orman Algoritmasının işleyişi Tablo 3.2’deki adımlardan oluşmaktadır.

Tablo 3.2 Rastgele Orman Algoritması

$i = 1, \dots, N$ için p boyutlu girdi seti X_i ($X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})$) ve Y_i çıktı iken i gözlemine ait eğitim veri seti $L_i = (X_i, Y_i)$ olarak tanımlandığında x girdi değişkeni gözlemine karşılık tahminleme fonksiyonu $\hat{\theta}_n(x) = h_n(L_1, \dots, L_n)(x)$ olarak tanımlandığında
1. $b=1, \dots, n_{estimator}$ için
a. n boyutlu bir önyükleme örneği $L_b^* = (X_b^*, Y_b^*)$ alınır
b. T_b Rastgele Orman ağacı oluşturulur. Ağacın her düğümü için <i>max_depth</i> derinliğine ulaşılan kadar veya tüm yapraklar <i>min_samples_split</i> değerinin altına inene kadar bölünmeler devam eder. Bölünme işlemleri için aşağıdaki adımlar takip edilir.
i. p öznitelikten rassal olarak m adet öznitelik seçilir.
ii. m öznitelik arasından en iyi bölünmeyi sağlayan öznitelik seçilir.
iii. Ağaç dallara bölünür.
2. Oluşturulan Rastgele Ormanın sonuçları değerlendirilir.
Yeni bir x noktasında Rastgele Orman algoritmasının j ’ninci ağaç için tahmini $\hat{h}_j(x)$ için
a. $\hat{f}(x) = \frac{1}{n_{estimator}} \sum_{j=1}^{n_{estimator}} \hat{h}_j(x)$, regresyon (tahminleme) problemleri için.
b. $\hat{f}(x) = \operatorname{argmax}_y \sum_{j=1}^{n_{estimator}} (\hat{h}_j(x) = y)$, sınıflandırma problemleri için. (Burada $\hat{h}_j(x) = y$ ise 1, aksi halde “0” değerini almaktadır)

Kaynak: Cutler vd., 2012: 164; Müller ve Guido, 2017: 83; Akar, 2013: 29-30; Humphris, 2020; 6.

Rastgele Orman Algoritmasının oluşturulması için öncelikle ağaç sayısına ($n_{estimator}$) karar verilmesi gerekir. Bu ağaçlar birbirinden farklıdır ve algoritma her ağacın farklı olmasını sağlamak için rastgele seçimler yapar. Rastgele Orman Algoritmasında tahmin yaparken, algoritma ilk olarak ormandaki her ağaç için bir tahmin gerçekleştirir. Regresyon için nihai tahmini elde etmek üzere bu sonuçların ortalaması alınırken; sınıflandırma için her bir ağaç tarafından elde edilen tahminler üzerinden oylama yapılmaktadır. Rastgele Orman Algoritmasının işleyişi ile ilgili süreç Şekil 3.6'da gösterilmiştir. Rastgele orman diğer bireysel ağaçlara kıyasla daha az aşırı öğrenme gösterir ve daha sezgisel bir karar sınırı sağlar (Müller ve Guido, 2017: 84-85, Horning, 2010: 3-4).



Şekil 3.6 Rastgele Orman Algoritması Akış Şeması (Zhang vd., 2018: 21024).

3.2.3.3. Rastgele Orman Algoritmasının Avantajları ve Dezavantajları

Regresyon ve sınıflandırma işlemleri için rastgele orman oldukça yaygın bir şekilde kullanılan makine öğrenme metodudur. Temelde Rastgele Orman Algoritması, karar ağaçlarının yetersizliklerini giderirken tüm olumlu yönlerinden pay almıştır. Karar ağaçlarını kullanmasının nedeni karar verme sürecinin tam olarak yansıtılabilmesidir. Ancak, onlarca veya yüzlerce ağacı detaylı bir şekilde canlandırmak temelde mümkün değildir. Ayrıca, rastgele ormandaki ağaçlar, özellik altkümelerinin kullanılması dolayısıyla karar ağaçlarına kıyasla daha derin olma eğilimindedir (Müller ve Guido, 2017: 86-87, Özdemir, 2018, 54).

Rastgele Orman Algoritmasının avantajları literatürde aşağıdaki şekilde belirtilmektedir (Horning, 2010: 1; Müller ve Guido, 2017: 86-88; Özdemir, 2018, 52-54; Rahmati vd. 2019:12; Bezek Güre, 2019: 74):

- Güçlüdür, parametrelerin detaylı ayarlaması olmadan genelde iyi bir şekilde sonuç verir ve verinin ölçeklendirilmesine gerek duymaz
- Gürültülü ve aykırı verilere karşı dayanıklıdır.
- Modelleme sürecinde değişkenlerin önemlerini de tahmin eder.
- Değişken silme olmadan çok sayıda girdi değişkenini işleyebilir.
- Çok büyük veri setlerinde iyi çalışmakta ve eğitim, güçlü bir bilgisayar ile paralel olacak bir şekilde işlemci çekirdekleri arasında dağıtılabilmektedir.
- Hızlıdır, kısa sürede sonuç verir.
- Hem fen bilimleri hem de sosyal bilimler alanlarındaki çalışmalarda kullanılmaktadır. Kayıp verilere karşı daha duyarlı bir yapıda olması, büyük ölçekli veri setlerinin yanı sıra küçük ölçekli veri setinde de doğru sonuçlar vermesi sosyal bilimler alanında tercih edilmesini sağlamıştır. Diğer taraftan, kullanılan değişkenlerin önemlilik derecelerinin belirlenebilmesi, sürekli verilere ilave olarak kategorik verilerde de uygulanabilmesi ve karar ağaçlarında budama gereksinimi olmaması fen bilimleri alanında da tercih edilmesine katkı sağlamıştır.
- Kullanımı kolay ve esnek bir denetimli öğrenme algoritmasıdır. Sürekli ve kategorik veri setlerine uygundur.
- Rastgele çok sayıda karar ağacını içeren bir orman yaratması ve sonucu tüm karar ağaçları çıktılarını baz alarak hesaplaması dolayısıyla aşırı öğrenme probleminin üstesinden gelebilmektedir.
- Veriyi eğitirken uç değerlerle baş etme konusunda iyidir ve sınıflandırma hatası emi gibi yardımcı bilgiler üretebilmektedir.

Rastgele Orman Algoritmasının dezavantajları ise aşağıda özetlenmektedir (Müller ve Guido, 2017: 88; Özdemir, 2018, 54).

- Çok fazla boyutlu ve aralıklı veride, örneğin metin verisinde çok iyi performans gösterememektedir.
- Doğrusal modellere kıyasla daha fazla belleğe ihtiyaç duyar, eğitim ve tahmin işleminde de daha yavaştır. Bir uygulamada zaman ve bellek önemli ise doğrusal modellerin tercih edilmesi daha mantıklı olacaktır.

- Daha önce de belirtildiği üzere dal sayısı kontrol parametresi (max_features) her ağacın ne kadar rassal olduğunu belirlemektedir ve küçük dal sayısı değeri aşırı öğrenmeye yol açabilmektedir.
- Rastgele orman ile elde edilen sonuçta bir güven aralığı verilememektedir.

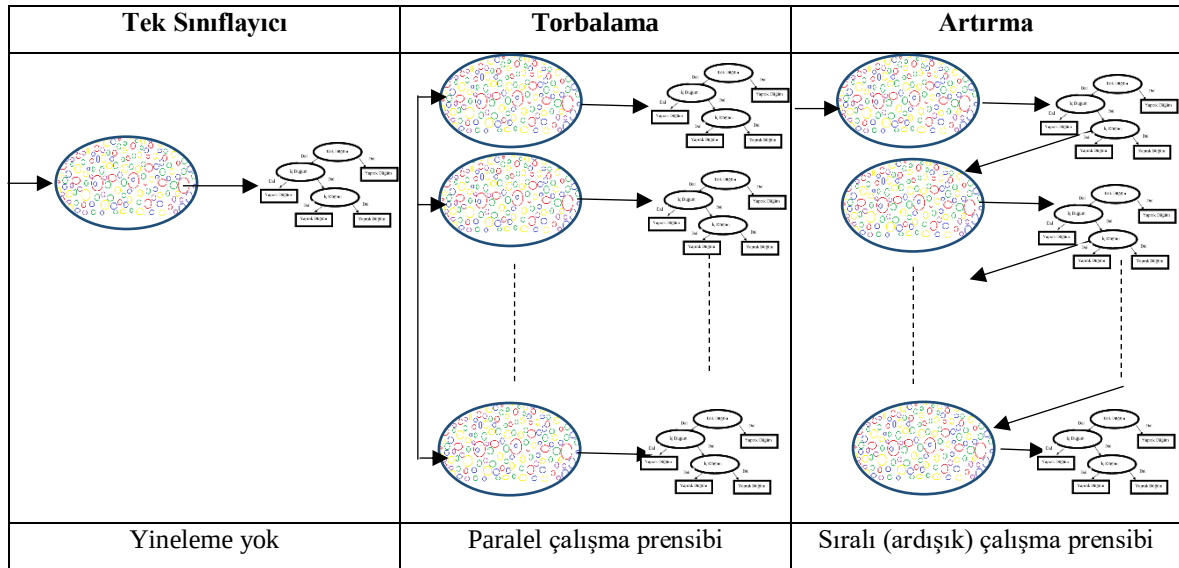
3.2.3.4. Rastgele Orman Algoritmasında Özniteliklerin Önemi

Rastgele Orman Algoritması öznitelikleri önem düzeylerine göre sıralamaktadır. Burada özneliliğin önemi ile tüm diğer öznitelikler sabitken değerlendirilen öznelilik için test verisinin değiştirilmesi durumunda tahminleme hatasındaki değişikliğin büyüklüğü kast edilmektedir (Liaw ve Wiener, 2002: 18).

3.2.4. Artırma (Boosting)

Artırma, öğrenme algoritmalarının doğruluğunu iyileştiren bir yöntemdir. İlk olarak Schapire (1990) tarafından kanıtlanabilir polinom zamanlı artırma algoritması ile ortaya atılmıştır. Daha sonra Freund (1995), belli bir miktarda optimallik verse de bazı dezavantajları olan etkili bir artırma algoritması önermiştir.

Torbalama yönteminde rastgele seçilen önyüklemeye örnekleri tekrarlanarak model eğitilirken, artırma yönteminde eğitim örneklerinin ağırlıkları ayarlanarak model eğitilmektedir (Bezek Güre, 2019:33).



Şekil 3.7 Torbalama ve Artırma Algoritmalarının Çalışma Prensibi (Sabancı, 2019: 18)

Torbalama ve Artırma algoritmalarının genel çalışma prensipleri Şekil 3.7’de yer almaktadır. Torbalama algoritmaları paralel bir yapıda karar ağaçları oluştururken, artırma algoritmaları sıralı bir şekilde ağaç oluşumuna gitmektedir.

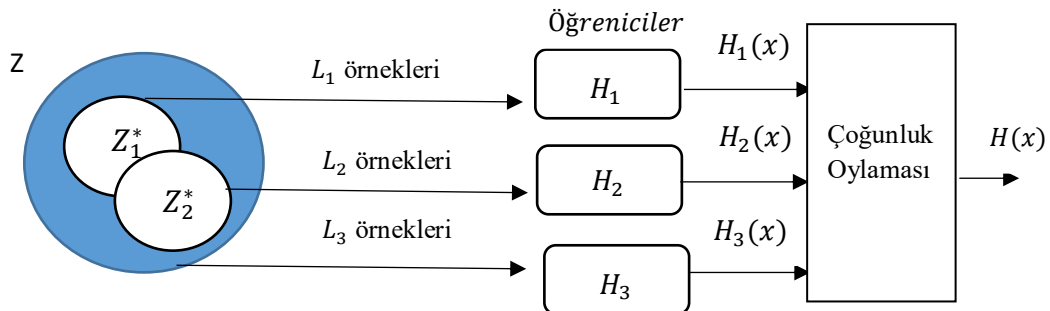
Artırma (Boosting) algoritması, yanlış sınıflandırılmış örneklere daha fazla ağırlık vererek modelde iyileştirmeler elde etmeye çalışan eğitim setinin tekrarlı bir şekilde sunulduğu iteratif bir yöntemdir. Çoğu durumda artırma, torbalamadan daha doğru sonuçlar vermektedir. Artırma sınıflandırmanın hem varyansını hem de yanlılığını (bias) düşürmekte olup, daha doğru sonuçlar veren bir sınıflandırma yöntemidir (Gislason vd., 2004: 1049).

Schapire (1990: 201-202) tarafından önerilen ilk Artırma yaklaşımını ifade eden artırma algoritması Tablo 3.3’de sunulmaktadır.

Tablo 3.3 Artırma Algoritması

Girdi: N adet gözlemden oluşan veri seti $Z = \{z_1, z_2, \dots, z_N\}$ olduğu varsayılırsa, $x_i \in \mathcal{X}$ ve $y_i \in \{-1, +1\}$ iken i gözlemine ait veri değişkenleri $z_i = (x_i, y_i)$ olarak;	
Çıktı: bir sınıf olan $H: \mathcal{X} \rightarrow \{-1, +1\}$ olarak ifade edildiğinde Artırma yönteminin sınıflandırma problemi işleyişini gösteren algoritma aşağıdaki adımlardan oluşmaktadır:	
1.	Z veri setinden Z_1^* alt veri kümesini oluşturmak için rastgele ve yerine koymadan (iadesiz bir şekilde) $L_1 < N$ örnekleme seçilir.
2.	Z_1^* alt veri kümesini kullanarak zayıf öğrenciler çalıştırılır ve H_1 sınıfı üretilir.
3.	Yarısı H_1 sınıfında yanlış sınıflandırılan örneklerin, diğer yarısı ise Z veri seti örnekleri arasından olacak şekilde $L_2 < N$ koşuluyla Z_2^* alt veri kümesi seçilir.
4.	Z_2^* alt veri kümesini kullanarak zayıf öğrenciler çalıştırılır ve H_2 sınıfı üretilir.
5.	H_1 ve H_2 sınıfları tarafından kabul edilmemiş olan Z veri setindeki tüm örneklerden Z_3^* alt veri kümesi oluşturulur.
6.	Z_3^* alt veri kümesini kullanarak zayıf öğrenciler çalıştırılır ve H_3 sınıfı üretilir.
7.	Çoğunluk oylaması ile nihai sınıflandırma yapılır: $H(x) = \text{sign}(\sum_{b=1}^3 H_b(x))$.

Kaynak: Ferreira ve Figueiredo, 2012:41.



Şekil 3.8 Schapire (1990) Tarafından Önerilen İlk Artırma Yaklaşımının Görsel İfadesi (Ferreira ve Figueiredo, 2012:41).

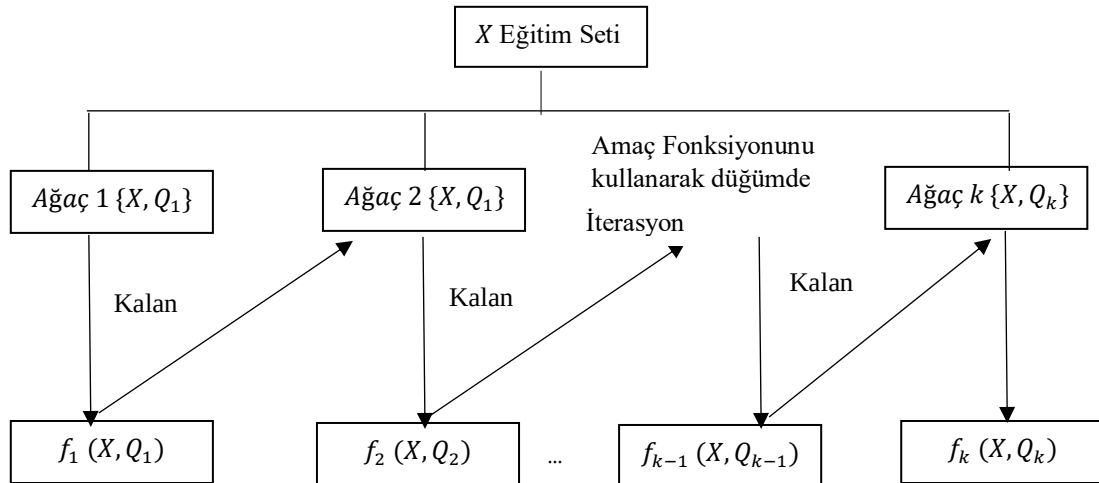
Şekil 3.8’de örnek görseli yer alan Schapire (1990) tarafından önerilen ilk artırma yönteminde eğitim seti, Z_1^* , Z_2^* ve Z_3^* olarak üç parçaya ayrılmaktadır. Eğer ilk iki

sınıflandırıcı olan H_1 ve H_2 sınıf etiketlerinden memnun ise bu sınıflandırıcılar üzerinden değerlendirilme yapılmakta ve Z_3^* alt veri kümesi oluşturulmamaktadır. Aksi halde, Z_3^* alt veri kümesi seçilip H_3 sınıfı da üretilmektedir (Ferreira ve Figueiredo, 2012:41).

Artırma algoritmalarında öncelikle eğitim verisi üzerinden örneklem oluşturulur ve her bir gözleme eşit ağırlık verilir. Daha sonra artırma algoritmasının iterasyonlarında temel öğrenici örneklem üzerinden eğitilir ve hata hesaplanır. Doğru sınıflandırılan gözlemlerin ağırlıkları düşürülürken, yanlış sınıflandırılan gözlemlerin ağırlığı artırılır. Artırma yöntemi ile ilk gözleme göre kurulan modelin performansı artırılmaktadır. Genel olarak artırma algoritmalarında Schapire (1990) tarafından önerilen ilk artırma algoritmasında 3 adımda gerçekleştirilen süreç, en güçlü performansı sergileyen tahminleme modeline ulaşıncaya veya yanlış sınıflandırılan gözlemler kalmayınca kadar devam etmektedir (Sabancı, 2019:16; Wang vd., 2011; 225; Bezek Güre, 2019:34).

3.2.5. XGBoost (Extreme Gradient Boosting) Algoritması

XGBoost, Gradyan Artırma Karar Ağaçlarının geliştirilmiş bir versiyonu olan makine öğrenme algoritmasıdır (Chen ve Guestrin 2016: 785). Hem regresyon ve hem de sınıflandırma problemleri için kullanılabilen XGBoost, Rastgele Orman Algoritması gibi çoklu karar ağaçlarını daha güçlü bir model yaratmak için birleştiren ve CART ağaçları temelli bir diğer topluluk öğrenmesi yaklaşımıdır. Ancak, Rastgele orman yaklaşımının aksine sıralı bir şekilde, her bir ağacın bir önceki ağacın hatalarını düzeltmeye çalıştığı ağaçlar inşa etmektedir (Müller ve Guido, 2017: 88).



Şekil 3.9 XGBoost için Akış Şeması (Zhang vd., 2018: 21025)

XGBoost, ağacın karmaşıklığını kontrol etmede kullandığı farklı düzenleme (regularizasyon) teknikleri ile diğer boosting algoritmalarına kıyasla daha iyi performans sergilemektedir. XGBoost, veri bilimcileri tarafından yaygın bir şekilde kullanılan ve pek çok problemde en gelişmiş sonuçları veren ölçeklendirilebilir artırılmış ağaç sistemidir.

(Chen ve Guestrin 2016: 785). Algoritmanın temel çalışma süreci ise Şekil 3.9'da yer almaktadır.

XGBoost'un etkisi çok sayıda makine öğrenmesi ve veri madenciliği çalışmalarında yoğun kabul görmüştür. Kaggle'ın blog'unda 2015 yılında düzenlenen makine öğrenmesi yarışmasında 29 kazanan çözümün 17 tanesinde XGBoost kullanılmıştır. Bu çözümlerin sekiz tanesi modeli eğitmek için XGBoost kullanırken, diğerleri XGBoost ile sinir ağları topluluk modellerini kullanmıştır (Chen ve Guestrin, 2016: 785).

3.2.5.1. XGBoost Algoritmasının Parametreleri

XGboost Algoritmasının pratikte kullanılan pek çok parametresi olmakla birlikte bu bölümde çoğunlukla kullanılan ve tezin analiz sürecinde de yer verilen *max_depth*, *learning_rate*, *n_estimator* ve *min_child_weight* parametreleri ile *Min_split_loss* parametresine değinilmiştir (Müller ve Guido, 2017: 89-92; ⁶; ⁷).

max_depth: Ağacın maksimum derinliğini ifade etmektedir. Bu değerin artırılması modeli daha karmaşık yapmaktadır. Bu değerin çok yüksek seçilmesi modeli aşırı öğrenmeye daha yatkın hale getirebilir.

learning_rate: Aşırı öğrenmeyi engellemek için kullanılan adım boyutudur. Her ağacın bir önceki ağacın hatalarını düzeltme derecesidir. Daha yüksek öğrenme oranı her ağacın daha güçlü düzeltici olduğu, daha kompleks (karmaşık) modellere izin verdiği anlamına gelir. Topluluğa daha fazla ağaç eklemek artan ağaç sayısı parametresi ile sonuçlanabilmekte ve modelin karmaşıklığını artırmaktadır. Böylece modelin eğitim setindeki hatalarını düzeltmeye daha fazla şansı olmaktadır.

n_estimator: Gradyan artırma ağaçlarının sayısını ifade etmektedir. Ağaç sayısı değerinin artması daha karmaşık modele yönlendirir. Zaman ve bellek maliyetine bağlı olan ağaç sayısı değerini belirlemek için pratikte yaygın olan uygulama farklı öğrenme oranlarının araştırılması şeklindedir.

min_child_weight: Bir karar düğümünde (çocuk) kaç gözlem kullanılacağına ilişkin bir kontrol parametresidir. Eğer ağaç bölünme adımındaki örnek ağırlığı toplamı *min_child_weight*'den daha düşük olan bir yaprak ile sonuçlanırsa ağacın oluşturulma süreci sonlandırılır, daha fazla bölünme olmaz. Doğrusal regresyonda *min_child_weight* her düğümde olması gereken minimum örnek sayısına karşılık gelmektedir.

⁶ <https://xgboost.readthedocs.io/en/latest/parameter.html>, (erişim tarihi: 20.07.2020)

⁷ <https://xgboost.readthedocs.io/en/latest/parameter.html#parameters-for-tree-boost> (erişim tarihi: 30.11.2019).

min_child_weight ne kadar büyükse algoritma o kadar korunmalı olur. Bu parametre aşırı öğrenmeyi kontrol altında tutmak için kullanılır.

Min_split_loss: XGboost algoritmasının gamma parametresidir. Ağacın yaprak düğümünde ek bir bölme yapmak için gereken minimum kayıp azaltımıdır. Daha büyük gama değerleri, daha korunmalı ve daha az karmaşık modellere yol açmaktadır.

3.2.5.2. XGBoost Algoritmasının İşleyişi

XGBoost Algoritmasında, Artırma algoritmalarının temel mantığı işlemekte olup, öncelikle bir regresyon fonksiyonu oluşturulmakta ve bu fonksiyonun gerçek değerlerden sapmaları için tahmin hataları hesaplanmaktadır. Bu hatalar göz önünde bulundurularak yeni bir regresyon fonksiyonu oluşturulmaktadır. Bu aşamada oluşan hatalar yeni oluşturulan regresyon fonksiyonu ile bir önceki regresyon fonksiyonunun bileşimi ile değerlendirilmektedir. Bu şekilde devam edilirken regresyon fonksiyonu ilk durumuna kıyasla daha fazla karmaşıklaşır ancak, RMSE (Hata Karelerinin Ortalama Kökü) değeri giderek düşer.

Tablo 3.4 XGBoost Algoritması

Girdi: Veri seti D ; Kayıp fonksiyonu $l(y, f_k(x))$; İterasyon sayısı t ; Öğrenme oranı v ,

ve Terminal düğümleri sayısı T için,

1. Model sabit bir değer ile kurulur.

$$f^0(x) = f_0(x) = \theta_0 = \operatorname{argmin} \sum_{i=1}^n l(y_i, \theta);$$

2. $t = 1, \dots, n_estimator$ için:

$$3. \quad g_i = \delta_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$$

$$4. \quad h_i = \delta_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$$

5. Amaç fonksiyonunu minimum yapacak bölünmeyi seçecek yapı belirlenir.

$$\mathcal{L}_{bölünme} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma;$$

6. Öğrenme yapısı için j yaprağının ağırlığı $\{w_j\}_{j=1}^T$ belirlenir.

$$w_j = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda};$$

$$7. \quad f_t(x) = v \sum_{j=1}^T w_{jt} I;$$

$$8. \quad f^{(t)}(x) = f^{(t-1)}(x) + f_t(x);$$

Çıktı: $f(x) = \sum_{t=0}^{n_estimator} f_t(x)$

Kaynak: Chen ve Guestrin, 2016: 786-789; Pan, 2017:2-3; Zhang vd., 2018: 21024-21025; Roine ve Holter, 2018:26-30.

XGBoost Algoritmasının işleyisi Tablo 3.4’de sunulmaktadır. XGBoost’un adımları temel olarak şu şekilde işlemektedir: öncelikle tahminleyici öğretilir, hata hesaplanır ve hatayı tahminlemek için model yeniden öğretilir.

XGBoost algoritmasının amacı, amaç fonksiyonunu minimum yapacak $f(x)$ yaklaşımı bulmaktır. XGBoost, optimum ağaç yapısı ile yaprak ağırlığını eş zamanlı bir şekilde sorgulamakta ve bulmaktadır. Algoritmanın adımları aşağıda açıklanmaktadır:

- n adet gözlemden oluşan m adet öznelikli veri seti $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ olduğu ve i gözlemine ait girdi x_i ve çıktı y_i olduğu varsayılırsa,
- f_k ’nin regresyon ağacını, $f_k(x_i)$ ise k ’ninci ağacın i ’ninci gözlem için ürettiği sonuç olduğu model ve modelin tahmin ettiği çıktı formülasyon (3.2)’deki gibi ifade edilmektedir.

$$\hat{y}_i = \theta(x_i) = \sum_{k=1}^K f_k(x_i) \quad (3.2)$$

Burada K , toplam ağaç sayısını, bir başka ifade ile $n_estimator$ parametresini ifade etmektedir. Düzenlenmiş (Regularize edilmiş) amaç fonksiyonu ise formülasyon (3.3)’deki gibi ifade edilmektedir.

$$\mathcal{L}(\theta) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (3.3)$$

- Burada, $l(y_i, \hat{y}_i)$, i ’ninci gözlem için gerçek değer y_i ile tahmin edilen değer $\hat{y}_i$ arasındaki farkı ifade eden kayıp fonksiyonunu; $\Omega(f_k)$ ise regularizasyon fonksiyonudur. $\Omega(f_k)$ fonksiyonunun amacı aşırı öğrenmeyi önlemek ve algoritma tarafından üretilen modelleri basitleştirmektir. Regularizasyon fonksiyonunun açılımı ise formülasyon (3.4)’de sunulmaktadır.

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (3.4)$$

Formülasyon (3.4)’de yer alan γ ve λ , sırasıyla her yaprağın karmaşıklığını kontrol eden ve yaprak ağırlığı w ’nin büyüklüğünü kontrol eden ceza parametreleri, terminaldeki düğüm sayısı T ise ağacın maksimum derinliğini max_depth parametresidir.

- Algoritmada iterasyon sayısı t ile ifade edilmektedir. Algoritma işledikçe tahmin edilen değer de revize edilmektedir. t ’ninci iterasyondaki tahmin değeri $\hat{y}_i^{(t)}$ formülasyon (3.5)’deki gibi ifade edilmektedir.

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3.5)$$

Bu durumda, t 'ninci iterasyondaki amaç fonksiyonu formülasyon (3.6)'daki gibi olmaktadır.

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (3.6)$$

Burada n , ağaçtaki yaprak sayısıdır.

- v , bir başka ifade ile *learning_rate* aşırı öğrenmeyi engellemek için kullanılan adım boyutu, her ağacın bir önceki ağacın hatalarını düzeltme derecesidir.
- Algoritmanın birinci adımında, model sabit bir $f^0(x)$ değeri ile kurulur.
- Ağaç sayısı (iterasyon sayısı) birden *max_depth* değerine ulaşana kadar üç-sekiz arasındaki adımlar tekrarlanır.
- Algoritmanın üçüncü adımında gradyan inişi uygulanmakta ve $f(x_i)$ 'ye göre kayıp fonksiyonunun türevi alınmaktadır. Bu adımın her yinelenmesinde, adımın yönü hakkında bilgi sağlayan negatif gradyanı tahmin etmek için bir regresyon ağacı oluşturulmaktadır.
- Dördüncü adımında, $f(x_i)$ 'ye göre kayıp fonksiyonunun ikinci dereceden türevi alınmaktadır.

Bu durumda amaç fonksiyonunun ikinci dereceden yaklaşımı aşağıda sunulmaktadır.

$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i))] + \Omega(f_t) \quad (3.7)$$

Sabit terimler çıkarıldığında t 'ninci iterasyondaki amaç fonksiyonu formülasyon (3.8)'deki gibi ifade edilebilmektedir.

$$\tilde{\mathcal{L}}^{(t)} \simeq \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (3.8)$$

- Beşinci adımda amaç fonksiyonunu minimum yapacak optimal bölünme bulunmaya çalışılır. Formülasyon (3.9)'da yer alan $\mathcal{L}_{bölünme}$ için minimum değeri verecek bölünme bulunmaya çalışılır.

$$\mathcal{L}_{bölünme} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (3.9)$$

Burada, I_R ve I_L , sırasıyla sağ ve sol bölümde yer alacak gözlem sayısıdır. I ise toplam gözlem sayısıdır ($I = I_R \cup I_L$).

- Altıncı adımda yaprakların ağırlıkları belirlenmekte, yedinci ve sekizinci adımda ise optimal ağırlıklılara ve öğrenme oranına göre model regüle edilerek güncellenmektedir.

3.2.5.3. XGBoost Algoritmasının Avantaj ve Dezavantajları

XGBoost Algoritmasının avantajları (Chen ve Guestrin, 2016: 785-786; Yangın, 2019: 33-37):

1. XGBoost'un başarısının ardındaki en önemli faktör her senaryoda ölçeklenebilir olmasıdır. Sistem, tek bir makine çözümlerinden on kat daha hızlı çalışmakta ve dağılmış veya sınırlı bellek yapılarında milyarlarca örneği ölçeklendirebilmektedir. Paralel ve dağıtılmış hesaplama öğrenmeyi hızlandırır. Böylece, modelin daha hızlı keşfini sağlar.
2. XGBoost çekirdek hesaplamadan yararlanır ve veri bilimcilerine yüz milyonlarca örneği işleme imkanı verir. Son olarak, oldukça büyük veriyi en az miktarda küme kaynağını kullanarak ölçeklendirebilen uçtan uca bir sistem oluşturabilmektedir.
3. Yüksek tahmin etme gücüne sahiptir.
4. Regularizasyon yapısı ile aşırı öğrenmeyi azaltır.
5. Kayıp gözlem değerlerini otomatik olarak atayabilmektedir. Gerçek verilerde yaşanan kayıp değer durumuna bu özelliği ile çözüm getirmektedir.

XGBoost Algoritmasının dezavantajları (Bai vd., 2017:4):

1. Kategorik verilerde iyi performans sergileyememektedir. Kategorik verilerde dönüşüm gerektirmektedir.
2. Parametrelerin iyi ayarlanamaması durumunda aşırı öğrenme gerçekleşebilmektedir. Rastgele Orman Algoritmasına kıyasla daha fazla ayar gerektirmektedir.

3.2.5.4. XGBoost Algoritmasında Özniteliklerin Önemi

XGBoost Algoritması, modelin performansını artırmak ve orijinal verinin boyutunu azaltmak amacıyla öznitelikleri önemlerine göre filtreleyebilmektedir. Bir öznitelikğin önemi rastgele bir gürültü verisi ile değiştirilmesi durumunda tahminleme performansında önemli bir değişiklik meydana getirip getirmeyeceğine bağlıdır. Bir öznitelik ağacın kurulumunda önemli kararların verilmesinde ne kadar fazla kullanılırsa, puanı o kadar

yükselir. “weight” kriterine göre önem düzeyi bir özneteliğin ağaçta kaç defa yer almasını; “gain” kriteri ise bir öznetelik kullanılırken kazanılan ortalama eğitim kaybını ifade etmektedir (Ji ve Lin, 2019: 148; Zheng vd., 2017: 4-5).

3.2.6. Yapay Sinir Ağları

Teknolojideki gelişmeler sonucu insan beyninin çalışma şeklini prensip edinen ve öğrenme işlemini bilgisayar sistemlerinde gerçekleştiren Yapay Sinir Ağları finans, tıp, bankacılık, pazarlama sektörlerinden mühendislik uygulamalarına kadar oldukça geniş bir alanda kullanılmaktadır. Yapay nöron modeli 1943 yılında Warren McCulloch ve Walter Pitts tarafından ilk olarak önerildiğinden bu yana Yapay Sinir Ağları üzerindeki araştırmalar son yıllarda daha fazla artmıştır. Ayrıca, Yapay Sinir Ağlarının doğrusal olmayan problemlerde güvenilir sonuçlar vermesi daha fazla tercih edilmesini sağlamıştır. YSA’lar tahminleme, kümeleme ve sınıflandırma gibi pek çok problemin çözümünde kullanılabilmektedir (Rojas, 1996:3; Öztemel, 2012: 11; Karaatlı vd., 2012: 90; Akpınar, 2014: 236; Kurt vd. 2017:100).

Literatürde Yapay Sinir Ağları ile ilgili pek çok farklı tanımlama mevcuttur:

- En iyi tanımlamalardan birini yapan Haykin (1999:24) Yapay Sinir Ağlarını bilgiyi çevreden öğrenme süreci ile elde eden ve bağlantılarında bilgiyi depolayan basit işleme birimlerinin çok büyük ölçekte paralel birleşimi olarak tanımlamıştır.
- Rojas (1996:3) Yapay Sinir Ağlarını (YSA) sinir sisteminin bilgi işleme becerisinin modellenmesi girişimi olarak nitelemiştir.
- Öztemel (2012:29) Yapay Sinir Ağlarını, insan beynine benzer bir şekilde örnekler üzerinden öğrenerek yeni bilgiler keşfeden, türetebilen ve bunu yardım almadan gerçekleştiren bilgisayar sistemleri olarak tanımlamaktadır.
- YSA’lar ile ilgili bir diğer tanımlama ise şöyledir: YSA’lar, birçok farklı işleme elemanından oluşan ve genelde doğrusal olmayan öğrenme makineleridir (Güresen ve Kayakutlu, 2011: 427).
- YSA’lar parametrik olmayan esnek modelleme yaklaşımlarıdır (Tang ve Chi, 2005: 246).

3.2.6.1. Yapay Sinir Ağlarının Genel Özellikleri ve Avantajları

Yapay Sinir Ağlarının genel özellikleri ve avantajları aşağıda özetlenmektedir (Akkurt, 2005: 8-9; Karaatlı vd., 2012: 90; Tang ve Chi, 2005: 246):

- Gürültülü ve kayıp verileri tolere edebilmekte, değişen ağ bağlantıları ile kendi kendine öğrenebilmekte, genelleme yapabilmekte ve girdi değişkenleri arasında karmaşık ilişkiler kurabilmektedir. YSA'ların en güçlü yanı ise tahminlemeye çalıştığı ilişki yapısı ile ilgili bir ön varsayıma ihtiyaç duymamasıdır.
- YSA'ların en önemli özelliklerinden biri geçmiş bilgiler üzerinden öğrenmesini gerçekleştirebilme ve veride herhangi bir değişiklik oluşması halinde öğrenme işlemini yenileyebilmesidir.
- YSA'larda işleme elemanları birbirlerinden bağımsız bir yapıda eşzamanlı olarak, paralel çalışmaktadır. Bileşenlerin hesaplamaları birbirinden bağımsız, eş zamanlı olarak çalışan çok sayıda nörondan oluşur. Böylece, zaman içerisinde herhangi bir nöron işlevini kaybederse bu durum ağın başarısını önemli düzeyde etkilemez. Paralellik özelliklerinin bir sonucu olarak girdilerin yapılarından kaynaklanan hatalara karşı dayanıklıdırlar. Eksik ve/veya hatalı veri içeren veri setlerinde de anlamlı sonuçlar verebilmektedirler.
- Örüntü/desen ilişkilendirme, tahminleme, sınıflandırma yapabilirler.
- Sadece nümerik veri üzerinde çalışabilirler.

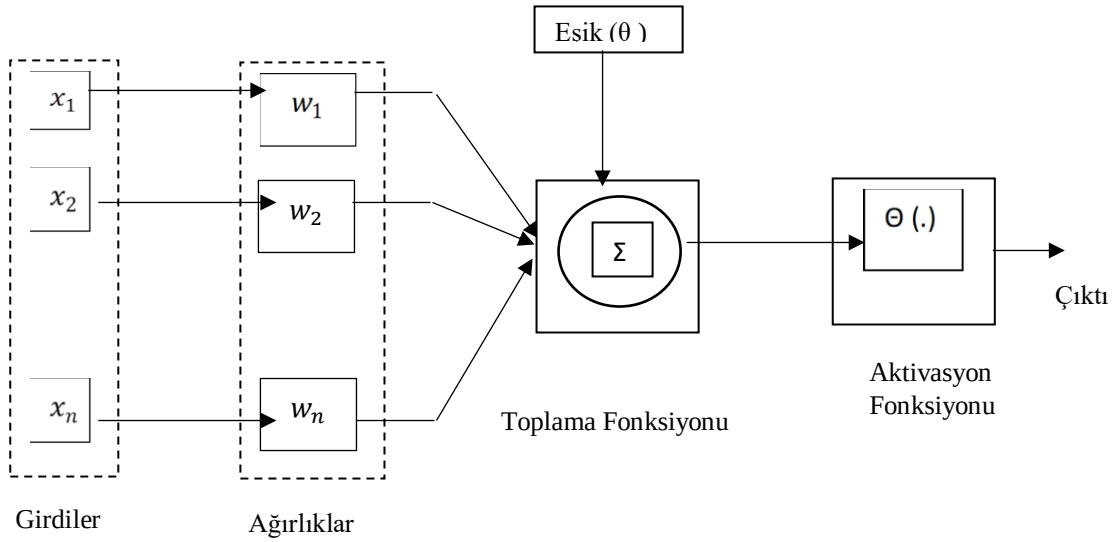
3.2.6.2. Yapay Sinir Ağlarının Önemli Dezavantajları

- YSA'larda uygun ağ yapısı ve parametre değerlerinin tespiti deneme yanılma yoluyla bulunmaktadır. Dolayısıyla YSA'ların ürettiği çözümler optimum garantisi vermemektedir.
- Ağın eğitime ne zaman son verileceği yönünde geliştirilmiş bir yöntem bulunmamakla birlikte eğitim verisindeki hata değerinin belli bir oranın altına inmesi halinde eğitim sonlandırılır.
- Ağın davranışları ve problemi nasıl çözdüğü konusunda bir bilgi yoktur. Bu durum ağa olan güveni azaltmaktadır.

YSA'ların tercih edilmesi için problemin başka teknikler ile etkin, pratik ve verimli bir çözüm elde edilememiş olması gibi durumların ortaya çıkması beklenir. Bir başka ifade ile her problemde YSA uygulamak mantıklı olmayabilir (Öztemel, 2012: 29-30).

3.2.6.3. Yapay Sinir Ağlarının Yapısı ve Temel Elemanları

Yapay Sinir Ağlarının temel elemanları girdi katmanı, gizli katman ve çıktı katmanından oluşur. Yapay Sinir Ağlarının yapısı ve temel elemanları Şekil 3.10'da yer almaktadır.



Şekil 3.10 Yapay Sinir Hücresi Gösterimi (Çuhadar, 2006: 117, Sarı, 2016: 38)

YSA’larda her işleme elemanı kendinden veya diğer işleme elemanından bağlantılar alır. YSA’ların genel olarak işleyişi aşağıda sunulmaktadır (Çuhadar, 2006: 117; Sarı, 2016: 38; Güresen ve Kayakutlu, 2011: 427; Öztemel, 2012: 45-52):

- Girdiler, bir yapay sinir hücresine sunulan verilerdir. Bu veri dışardan gelebileceği gibi, bir başka hücreden de gelebilir.
- Girdi katmanındaki proses elemanları hücreye dış dünyadan gelen verileri olarak gizli katmanlara yönlendirir.
- Gizli katmanda ise girdi katmanından iletilen bilgiler işlenip çıktı katmanına iletilmektedir.
- Bağlantılar üzerinde takip eden sinyaller ağırlık olarak adlandırılan ayarlanabilir parametrelerle ölçeklendirilir. Ağırlıklar, hücreye sunulan bilginin önemini ve bu hücre üzerinde ne kadar etkili olacağını gösterir.
- Çıktı katmanında yer alan işleme elemanları gizli katmandan iletilen bilgileri işleyerek toplar ve aktivasyon fonksiyonu sonucunda bir çıktı üretir.

Toplama fonksiyonu, hücreye iletilen net girdiyi belirler. Çok farklı toplama fonksiyonları kullanılmakla birlikte en yaygın kullanılanı ağırlıklı toplama fonksiyonudur. Her gelen girdi kendi ağırlığı ile çarpılmakta ve tüm girdiler için bu işlemin gerçekleştirilmesi sonrasında elde edilen toplam değer, net girdi (NET) değerini vermektedir.

$$NET = \sum_i^n G_i W_i$$

G : Girdileri

W : Ağırlıkları

n : Hücreye gelen toplam girdi sayısını gösterir.

Literatürde en çok kullanılan toplama fonksiyonlarından bazıları; toplam, çarpım, maksimum, minimum, çoğunluk ve kümülatif toplama fonksiyonları ile formülasyonları Tablo 3.5’de sunulmaktadır.

Tablo 3.5 Yapay Sinir Ağlarında Kullanılan Bazı Toplama Fonksiyonları

Net Giriş	Açıklama
Toplam $\text{Net Girdi} = \sum_i^n G_i W_i$	Ağırlık değerleri ile net girdi değerlerin çarpımlarının toplamıdır.
Çarpım $\text{Net Girdi} = \prod_i G_i W_i$	Ağırlık değerleri ile net girdi değerlerin çarpılması yöntemi uygulanır.
Maksimum $\text{Net Girdi} = \text{Max}(G_i W_i), \quad i = 1, \dots, N$	Ağırlık değerleri ile girdi değerlerinin çarpımında en büyük değeri veren değer net girdiyi oluşturur.
Minimum $\text{Net Girdi} = \text{Min}(G_i W_i), \quad i = 1, \dots, N$	Ağırlık değerleri ile girdi değerlerinin çarpımında en küçük değeri veren değer net girdiyi oluşturur.
Çoğunluk $\text{Net Girdi} = \sum_i \text{sgn}(G_i W_i)$	Girdiler ile ağırlıkları çarpıldıktan sonra pozitif ve negatif sonuçların sayısı bulunur. Büyük olan sayı hücrenin net girdi değerini verir.
Kümülatif Toplam $\text{Net Girdi} = \text{Net}(\text{eski}) + \sum_i^n G_i W_i$	Ağırlık değerleri ile girdi değerlerini çarpımları toplanır ve daha önce hesaplanmış olan girdi değerleri ile toplanır. Elde edilen değer Net girdiyi verir.

Kaynak: Öztemel, 2012: 45-52.

Aktivasyon Fonksiyonu, hücreye iletilen net girdiyi işler ve çıktıyı hesaplar. Literatürde en sık karşılaşılan aktivasyon fonksiyonları Tablo 3.6’daki gibidir. Burada, NET olarak ifade edilen değer toplama fonksiyonundan elde edilen net girdi değeri olup aktivasyon fonksiyonlarının genel kullanımları aşağıdaki gibidir (Sarı, 2016: 42):

1. Lineer (doğrusal) aktivasyon fonksiyonu: Hücreye gelen girdiler aynen hücrenin çıktısını oluşturur.
2. Hiperbolik tanjant aktivasyon fonksiyonu: NET girdi değerinin tanjant fonksiyonundan geçirilmesi ile hücrenin çıktısı elde edilir. Sonuç, $[-1,1]$ aralığındadır. Pratikte Hiperbolik tanjant (tanh) fonksiyonu sigmoid aktivasyon fonksiyonuna kıyasla daha fazla tercih edilmektedir.

Tablo 3.6 Yapay Sinir Ağlarında Kullanılan Bazı Aktivasyon Fonksiyonları.

Fonksiyonun Adı	Fonksiyonun Formülasyonu	Fonksiyonun Grafiği
Lineer (doğrusal) aktivasyon fonksiyonu	$f(NE T) = NE T$	
Hiperbolik tanjant aktivasyon fonksiyonu	$f(NE T) = \frac{e^{NE T} - e^{-NE T}}{e^{NE T} + e^{-NE T}}$	
Sigmoid aktivasyon fonksiyonu	$f(NE T) = \frac{1}{1 + e^{-NE T}}$	
Eşik (basamak) aktivasyon fonksiyonu	$f(NE T) = \begin{cases} 1, & NE T > t \\ 0, & NE T \leq t \end{cases}$	
ReLU (Rectified Linear Unit) Doğrultulmuş Doğrusal Birim aktivasyon fonksiyonu	$f(NE T) = \begin{cases} 0, & NE T < 0 \\ NE T, & NE T \geq 0 \end{cases}$	

Kaynak: Sarı, 2016: 41

3. Sigmoid aktivasyon fonksiyonu: Yapay Sinir Ağları çalışmalarında en çok kullanılan aktivasyon fonksiyonu türüdür. Türevinin alınabilmesi dolayısıyla geri yayılım algoritmalarında da sıklıkla kullanılmaktadır. Bu fonksiyon sonucunda elde edilen çıktı değeri $[0,1]$ aralığında yer alır.

Geri yayılım sırasında çıktı değeri 0 veya 1 olduğunda gradyanı 0 olmakta ve ağırlıklar değiştirilememektedir. Bu durumdaki nöronlar doyuma ulaşmış nöron olarak adlandırılmaktadır. Sadece bu nöron değil, bu nörona bağlı olan nöronların

ağırlıkları da yavaş güncellenmektedir. Eğer sistemde çok sayıda doyuma ulaşmış nöron varsa sigmoid aktivasyon fonksiyonu ile ağıın geri yayılımı mümkün olmamaktadır.

4. Eşik (basamak) aktivasyon fonksiyonu: Genelde tek katmanlı ağlarda kullanılmakta olup hücrenin çıktı değeri $[0,1]$ aralığında yer alır.
5. ReLU (Rectified Linear Unit) Doğrultulmuş Doğrusal Birim aktivasyon fonksiyonu: ReLU günümüzde başta evrişimli sinir ağları (convolutional neural networks) ve derin öğrenme olmak üzeri YSA mimarilerinde en fazla kullanılan aktivasyon fonksiyonudur. ReLU geri yayılımdaki kolaylığı ve hesaplama maliyetlerinin yüksek olmaması dolayısıyla tercih edilmektedir.

3.2.6.4. YSA’larda Katman Sayısı

YSA’lar ağdaki sinirlerin bağlantı ve bilgi akış yönüne göre ileri beslemeli ve geri beslemeli Yapay Sinir Ağları olarak ikiye ayrılmaktadır. İleri beslemeli Yapay Sinir Ağlarında nöronlar giriş katmanından çıkış katmanına doğru ilerlemekte, bir katman sadece kendinden sonra gelen katmana bağlanmaktadır. YSA’ya gelen bilgi öncelikle giriş katmanına, ardından gizli katmanlara ve son olarak çıkış katmanına geçerek çıktıyı oluşturur. Geri beslemeli Yapay Sinir Ağlarında ise hücrenin çıktısı sadece kendinden sonraki katmana değil, kendinden önceki ve kendi katmanındaki herhangi bir hücreye de girdi sunabilmektedir. Yapay Sinir Ağlarında öğrenme, danışmanlı, danışmansız ve takviyeli öğrenme olarak üç farklı şekilde gerçekleştirilebilir (Kurt vd., 2017: 101).

Nöronların dizilişi, bağlantı yapısı, katman sayıları (tek katmanlı, çok katmanlı) ve katmanlar arasındaki veri iletimi (ileri beslemeli, geri beslemeli) ağıın topolojisini, bir başka ifade ile ağıın mimarisini oluşturur. Ağıın mimarisi, performansı üzerinde oldukça etkilidir (Bayır, 2006: 43).

YSA’lar tahmin-öngörü, fonksiyon yaklaştırma, örüntü tanıma, veri ilişkilendirme, kümeleme, veri filtreleme, optimizasyon gibi pek çok işlev için kullanılmaktadır. Tahminleme çalışmalarında kullanılan Yapay Sinir Ağlarından bazıları şunlardır: Geri Yayılım (Back-Propagation), Yönlendirilmiş Rastsal Tarama (Directed Random Search), Yüksek Dereceli Sinir Ağları (Higher Order Neural Networks), Geriyayılım içinde SOM (Self Organizing Map into BackPropagation), Radyal Tabanlı Fonksiyon (Radial Basis Function, RBF) (Bayır, 2006: 37).

Bu tez çalışmasının analiz aşamasında danışmanlı öğrenmeye imkan veren veri setiyle, ileri beslemeli ağlardan yaygın bir kullanıma sahip olan Çok Katmanlı

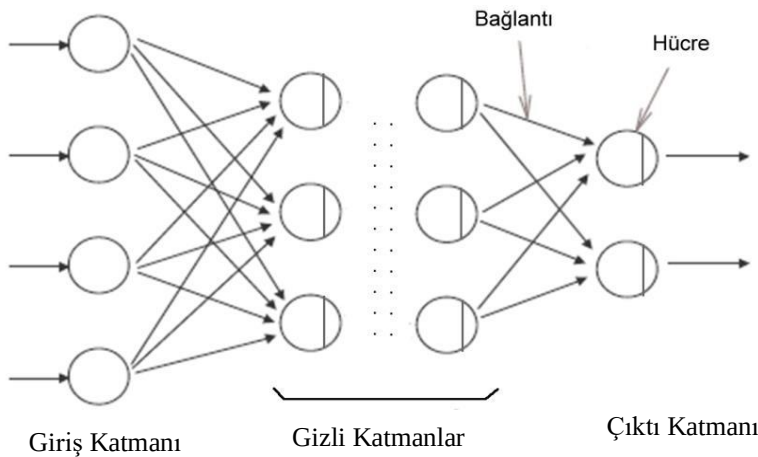
Algılayıcılar çalışılmıştır. Bu nedenle, Bölüm 3.2.6.5.'de Çok Katmanlı Algılayıcılar konusu detaylandırılmıştır.

3.2.6.5. Çok Katmanlı Algılayıcılar (Geri Yayılm Ağları)

Yapay Sinir Ağları, özellikle çok katmanlı algılayıcılar, geleneksel istatistiksel tekniklere kıyasla daha etkin alternatiflerdir. Çok katmanlı algılayıcıların, istatistik tekniklerindeki gibi verinin dağılımı ile ilgili varsayımda bulunmasına gerek yoktur. Doğrusal olmayan fonksiyonları modelleyebilir, görmediği, yeni veriyi genelleyecek şekilde eğitilebilir. Çok katmanlı algılayıcıların bu özellikleri onları nümerik modeller geliştirme konusunda cezbedici bir alternatif yapmaktadır. Çok katmanlı algılayıcılar gelecek trendlerini tahminleme, değişkenler arasındaki ilişkilerin modellendiği fonksiyon yaklaşımı ve verinin belirli sınıflara ayrıldığı örüntü sınıflandırması gibi pek çok alanda kullanılmaktadır (Gardner ve Sorling, 1998: 2627-2630).

Çok katmanlı algılayıcılar, Şekil 3.11'de görüldüğü üzere bir girdi vektörü ile bir çıktı vektörü arasında doğrusal olmayan eşleştirmeyi temsil eden bir model olan ve basit bağlantılı nöronlardan oluşan bir sistemdir. Nöronlar ağırlıklarla birbirine bağlanır ve nörona ulaşan girdilerin toplamının bir fonksiyonu olan çıktı sinyali bir transfer veya aktivasyon fonksiyonu ile değiştirilir. Bir nöronun çıktısı bağlantılı ağırlık ile ölçeklendirilir ve ağırlık bir sonraki katmanındaki nöronun girdisi olmak üzere ileri beslenir.

Çok katmanlı algılayıcının yapısı değişken olmakla birlikte genelde çok katmanlı nöronlardan oluşur. Girdi katmanı hesaplayıcı bir role sahip değildir. Sadece girdi vektörünü ağırlık sunar. Girdi ve çıktı vektörleri, çok katmanlı algılayıcının girdi ve çıktı katmanlarını ifade etmektedir. Çok katmanlı algılayıcının bir veya çok sayıda gizli katmanı olabilir ve sonunda bir çıktı katmanı vardır.

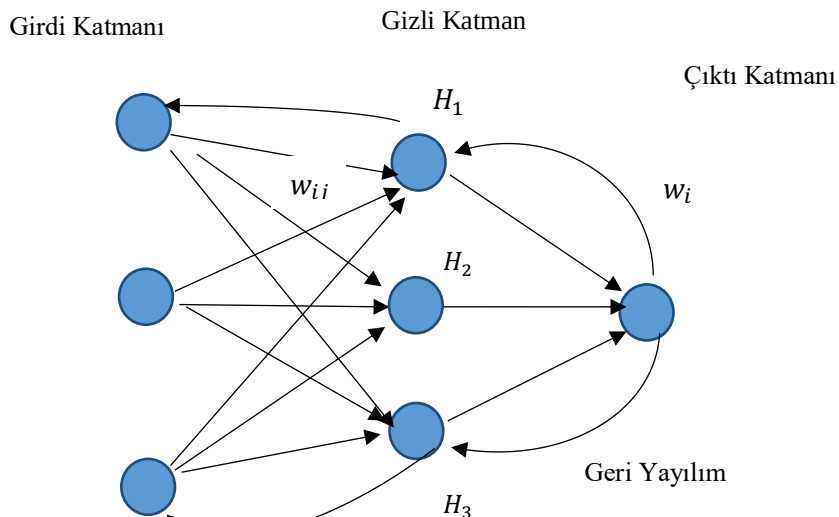


Şekil 3.11 İki Gizli Katmandan Oluşan Bir Çok Katmanlı Algılayıcı (Gardner ve Sorling, 1998: 2628).

Çok katmanlı algılayıcılar danışmanlı bir yapıdadır. Çok katmanlı algılayıcılarda eğitim aşamasında, bir seri girdi ve bu girdilere atanmış çıktı vektörlerinden oluşan bir eğitim setine ihtiyaç vardır. Eğitim süresince çok katmanlı algılayıcıya eğitim seti tekrarlayan bir şekilde sunulur ve ağıdaki ağırlıklar istenen girdi-çıkı eşleşmesi elde edilene kadar değiştirilmeye devam edilir (Gardner ve Sorling, 1998: 2629).

Eğitim süresince elde edilen çıktılar istenilen çıktılara eşit olmayabilir. İstenilen ve gerçekleşen çıktılar arasındaki fark hata sinyali olarak tanımlanır. Eğitim, çok katmanlı algılayıcının tüm hata düzeyini düşürmek ve ağıdaki ağırlıkların derecesinin ne kadar değiştirileceğini belirlemek için bu hata sinyallerinin büyüklüğünü kullanır. Eğitim sürecinin amacı en düşük hatayı veren ağırlıklar kombinasyonunun belirlenmesidir. Bu ağ hatalarının düzlem üzerinde işaretlenmesi ile hata yüzeyi elde edilir. En düşük hata düzeyini bulmak için kullanılan tekniklerden biri gradyan azaltmadır (gradient descent). Çok katmanlı algılayıcılarda, çıktı katmanındaki hatalar gizli katmanlara yayılarak ağırlıklar belirlendiğinden Geriye Yayılım Öğrenme Algoritması (Back-Propagation Learning Algorithm) olarak da tanımlanmaktadır. Geri yayımlı eğitim algoritmaları hata yüzeyinde global minimumu bulmak için dereceli azalma tekniğini kullanır. Geri beslemeli algoritma pek çok uygulamada iyi bir performans sergilemiştir (Gardner ve Sorling, 1998: 2629; Tan vd., 2006: 253).

Ağıdaki ağırlıklara ilk olarak küçük rassal değerler verilir. Geri yayılım algoritması daha sonra hata yüzeyinin yerel gradyanını hesaplar ve en hızlı yerel gradyan doğrultusunda ağırlıkları değiştirir. Oldukça düz bir hata düzeyinde ağırlıkların hata yüzeyinin global minimumuna yakınsaması beklenir. Geri yayılım algoritmasının işleyişinin görseli Şekil 3.12’de yer almaktadır (Gardner ve Sorling, 1998: 2629; Nisbet vd., 2018: 129-130).



Şekil 3.12 İleri Beslemeli Geri Yayımlı Yapay Sinir Ağı Örneği (Nisbet vd., 2018: 130)

İleri beslemeli geri yayılım algoritmasının aşamaları aşağıdaki gibidir (Gardner ve Sorling, 1998: 2629; Nisbet vd., 2018: 129-130):

1. İlk olarak ağırlıklarına rassal değerler verilir
2. Eğitim verisinden ilk girdi vektörü ağı sunulur.
3. Bir çıktı elde etmek için girdi vektörü ağı boyunca yayılır.
4. Gerçekleşen çıktı ile istenen değer karşılaştırılarak bir hata sinyali hesaplanır.
5. Ağda geriye doğru hata sinyali yayılır.
6. Genel hatayı minimum yapmak için ağırlıklar değiştirilir
7. Genel hata düzeyi tatmin edici düzeyde düşene kadar 2-7 aşamaları bir sonraki girdi vektörü ile devam ettirilir.

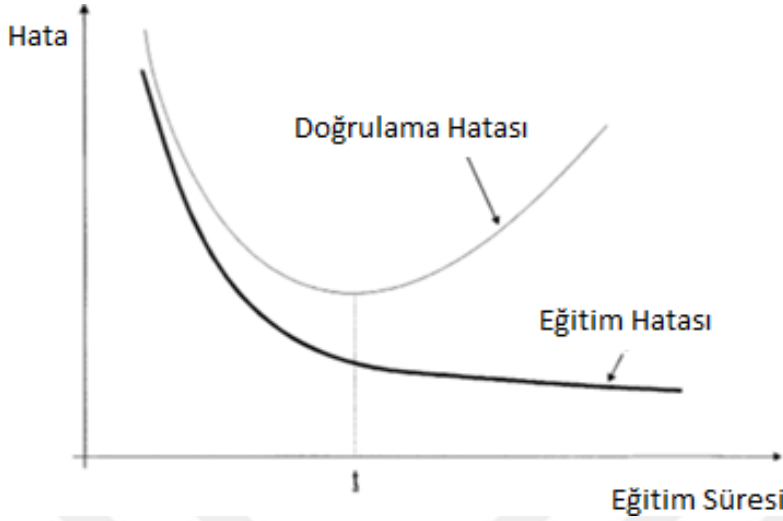
Eğitim algoritmasının yerel minimumda takılıp kalması arzulanan bir durumdur. Bunun için geri besleme algoritması iki ayarlanabilir parametre içerir: öğrenme oranı ve momentum. Bu parameterelerin değerleri probleme özgü olarak değişebilmektedir. Öğrenme oranı (λ = ağırlıkların değişim miktarı), yinelemeli gradyan azaltma öğrenme sürecinde alınan adım boyutunu tanımlar. Eğer çok büyük olursa, büyük ağırlık değişikliklerinden dolayı ağı hatası düzensizce değişir ve global minimumu atlayabilir. Çok küçük alınırsa eğitim çok uzun sürebilir. Momentum (α) ise yerel minimuma takılıp kalması durumunda gradyan azaltma sürecine yardım eder (Gardner ve Sorling, 1998: 2630; Tan vd., 2006: 253-254).

Le Cun vd. (1993: 160) en uygun öğrenme oranı değeri için şöyle bir süreç tanımlamıştır: ağı ürettiği çıktılar, gerçek çıktılara yakınsamıyorsa momentum artırılır, öğrenme oranı düşürülür. Eğer ağı bir sonuç etrafında gidip geliyorsa ve bir karara varamıyorsa öğrenme oranının düşürülmesi bir sonuca yakınsamasına yardımcı olur.

Öğrenme parametreleri için uygun değerlerin seçimi karar verme sürecinden geçilmesini gerektirir ve bu durum geri yayılım algoritmasını en cezbedici olmayan yönüdür (Gardner ve Sorling, 1998: 2634).

Çok katmanlı algılayıcıyı uygularken karşılaşılan ilk problem katman sayısı ve bu katmanlardaki nöron sayısından oluşan ağı mimarisine karar verebilmektedir. Bu sürece yardımcı olabilecek herhangi bir kural yoktur. Teknik olarak girdi ve çıktı arasındaki herhangi bir fonksiyonu yakınsamak için bir gizli katman gereklidir. Gizli katmandaki nöron sayısı ise girdi ve çıktı eşleştirmesinin karmaşıklığına, verideki gürültü miktarına ve eğitim verisinin miktarına bağlıdır. Gizli katmandaki nöron sayısı çok az olursa geri yayılım algoritması eğitim boyunca minimuma yakınsayamaz. Tam tersine gizli

katmandaki nöron sayısı çok fazla olursa ağ eğitim verisine aşırı öğrenme göstermekle birlikte düşük genelleme performansı sergiler (Gardner ve Sorling, 1998: 2630).



Şekil 3.13 Eğitim ve Doğrulama Hatalarının Eğitim Süresine Göre Değişimi

Şekil 3.13’de de görüleceği üzere, eğitim veri seti üzerinde en küçük hataya ulaşmaya kadar eğitilen bir ağın aşırı öğrenmiş (veriyi ezberlemiş) olabilmesi mümkündür.

3.3. Literatürde Yapılmış Uygulama Örnekleri

3.3.1. Veri Madenciliği Uygulamaları

Büyük veri yığınları içerisinde örüntülerin ve bilgilerin keşfi süreci olarak tanımlanabilen veri madenciliğinden, sağlık sektörü, eğitim, bankacılık, sigortacılık ve hatta uzay bilimine kadar pek çok alanda faydalanılmaktadır.

Veri madenciliği sağlık sektöründe güncel ve geçmiş dönem bilgilerine dayanarak karar verme sürecini kolaylaştıracak şekilde bir karar destek mekanizması oluşturmada kullanılmaktadır.

Doğan ve Türkoğlu (2008), kan biyokimya parametreleri ile demir eksikliği anemisi teşhisinde hekimlere yardımcı olacak bir karar destek sistemi oluşturmuş, oluşturulan karar destek sisteminin doktorun kararları ile birebir aynı sonuçlar verdiğini ifade etmişlerdir. Irmak (2009), 96 aylık dönemdeki hasta başvurularından oluşan verileri 3 farklı yöntem ile zaman serisi analizlerini yapmış ve gelecek dönem için hasta başvurularına ilişkin tahmin çalışması gerçekleştirmiştir. Danacı vd. (2010), örüntü tanıma programı ile doku hücreleri örnekleri hakkında veri temin edip meme kanseri hücrelerinin teşhisini yapmıştır. Yurtay vd. (2013), hasta verileri üzerinde hiyerarşik kümeleme

algoritmaları kullanılarak 36-45 yaş arası hastalarda HCT, HGB kan değerlerindeki azalmaların hastalık seviyelerini artırdığını tespit etmişlerdir.

Veri madenciliği, bankacılıkta, müşterilerin kredi risklerini tahmin etme ve sahte kredi kartı kullanımını tespit etme gibi çalışmalarda, sigortacılıkta ise riskli müşterilerin davranış kalıplarını ve bu müşteri tiplerine göre alternatif senaryolar üretilmesi için kullanılmaktadır.

Ata vd. (2008), kredi kartı sahiplerine ait veri seti üzerinde yaşam olasılıkları, hazard olasılıkları ve regresyon modelleri ile analiz yapmışlar ve yaş, gelir ve medeni durumun müşterilerin kredi kartı kullanımını bırakma konusunda önemli birer risk faktörleri olduğunu tespit etmişlerdir. Seyrek ve Ata (2010), mevduat bankalarının veri zarflama analizi ile etkinlik ölçümlerini yapmışlar ve bu verileri kullanarak bankanın etkinliğinde hangi finansal performansların önemli olduğunu karar ağaçları ile tespit etmişlerdir. Baynal ve Çalış (2016), Türkiye’de bir banka şubesinin müşterilerinin oniki farklı değişkene göre k-ortalamlar yöntemi ile kümelemesini yapmış, bu sayede söz konusu kümelerdeki müşteri profillerine göre satış stratejileri geliştirilmesine katkı sağlanmasını amaçlamışlardır.

Eğitim alanında öğrencilerin başarılarını etkileyen unsurlar ile öğrencilerin performanslarını tahminlemeye yönelik çalışmalarda da veri madenciliği kullanılmıştır. Çeşmeli vd. (2015) cinsiyet, yaş, dersin kategorisi, öğrenim türü, hazırlık sınıfı okuma, üniversite öncesi ikametgâh yeri, ceza durumu gibi özniteliklere göre öğrencilerin ders başarı durumlarını Yapay Sinir Ağları ve Uyarlamalı Sinir-Bulanık Mantık Sınıflayıcısı yöntemlerini kullanarak incelemiş ve öğrencinin hazırlık sınıfı okuması durumunun ve dersin kategorisinin öğrenci başarısında önemli etkiye sahip olduğunu keşfetmişlerdir. Özbay ve Ersoy (2017), lisans mezunu öğrencilerin Öğrenme Yönetim Sistemi hareketlilikleri ve akademik başarıları arasındaki ilişkileri incelemiş ve öğrencilerin başarılarını karar ağaçları ile tahmin etmeye çalışmışlardır.

Veri madenciliği ayrıca, ihracat verilerine dayanarak ülkelerin kümelenmesine yönelik çalışmalarda da kullanılmıştır. Ersöz vd. (2015), beş milyon ton ve üzeri Demir – Çelik ihracatı yapan ülkelerin 2003-2012 yılları arasındaki ihracat verilerine “Kareli Öklit Uzaklığı” ve “Ward Yöntemi”ne göre Hiyerarşik Kümeleme Analizi uygulamış, Türkiye’nin yüksek katma değerli çelik ihracatçısı İtalya, Fransa ve Belçika ile aynı kümede bulunduğunu gözlemiştir.

3.3.2. Rastgele Orman Algoritması Uygulamaları

Rastgele Orman Algoritması genel olarak diskriminant analizi, destek vektör makineleri ve yapay sinir ağlarına kıyasla daha iyi bir performans göstermektedir. Literatürde Rastgele Orman Algoritmasını diğer sınıflandırma yöntemleri ile kıyaslayan pek çok çalışma mevcuttur. Bu çalışmalar alanlarına göre aşağıdaki gibi ayrılmıştır:

Arazi örüntü sınıflandırılması:

Pal (2005), arazi örüntü sınıflandırılması alanında yürüttüğü çalışmada Rastgele Orman Algoritmasının Destek Vektör Makinelerine kıyasla sınıflandırma doğruluğu, eğitim süresi açısından eşit iyi bir performans sergilediği, kullanıcı tanımlı parametre sayısı açısından daha az parametre gerektirdiği; genel sınıflandırmada ise 100 ağaçlı sınıflandırmada %88,37, 1200 ağaçlı sınıflandırmada ise %88,02 oranında doğruluk ile daha başarılı olduğunu tespit etmiştir.

Akar vd. (2010), Trabzon ilini içeren 4 bantlı IKONOS uydu görüntülerinin sınıflandırılmasında Rastgele Orman Algoritması ile ECHO Spectral-Spatial, En Çok Benzerlik, Fisher Lineer Diskriminant, En Küçük Mesafe sınıflandırma yöntemleri kullanmış ve Rastgele Orman Algoritmasının diğer yöntemlere kıyasla daha iyi performans sergilediği, %96,8 genel sınıflandırma doğruluğu gösterdiği sonucuna varmışlardır.

Akar ve Güngör (2012) çalışmasında Karadenizde kırsal ve kentsel alandaki görüntülerin sınıflandırılmasında, Rastgele Orman Algoritmasının performansını Gentle AdaBoost, En Çok Benzerlik ve Destek Vektör Makineleri algoritmaları ile kıyaslamış, Rastgele Orman Algoritmasının %85,63 ile en yüksek doğruluk oranına sahip olduğunu, ikinci en iyi performansa sahip Destek Vektör Makinelerine kıyasla kentsel alan için ve kırsal alan için %15 daha iyi bir performans sergilediğini göstermişlerdir.

İklim senaryoları tahminlemesi:

Prasad vd. (2006), Regresyon Karar Ağaçları, Torbalama Ağaçlar, Rastgele Orman ve Çok değişkenli uyumlu regresyon şeritleri (Multivariate Adaptive Regression Splines-MARS) yöntemlerini kullanarak Kanada İklim Merkezi küresel dolaşım modeline göre mevcut ve gelecekteki iklim senaryoları altında 3 ağaç türü için bitki örtüsü haritasını tahmin etmeye çalışmışlardır. Kappa ve bulanık Kappa istatistikleri, değişken önemi ile uygun yaşam alanı gibi değerlendirme kriterleri açısından kıyaslanmaları neticesinde Rastgele Orman ile Torbalama Ağaç algoritmalarının daha iyi bir performans sergilediği, Rastgele Orman'ın ise Torbalama Ağaca kıyasla daha iyi sonuç verdiğini belirtmişlerdir.

Hisse senetleri tahminlemesi:

Quintana vd. (2017) hisse senetlerinin halka arz durumlarını tahmin etmek üzere Rastgele Orman Algoritmasını kullanmış ve örnek tabanlı öğrenme algoritmaları; kareler regresyonunun en az medyanı; yerel ağırlıklı öğrenme; M5 model ağaçları; M5 model kuralları; çok katmanlı algılayıcı; radyal temel fonksiyon ağları; minimum optimizasyon ve eğitilmiş destek vektör makineleri ile performansını kıyaslamışlardır. Tahminlerin ortalama ve medyan verileri açısından diğer yöntemlerden daha iyi sonuç verdiğini ifade etmişlerdir.

Bankacılık alanı:

Kalaycı (2018), çalışmasında bankacılık alanında Türkiye'deki KOBİ'lerin ödeme davranışlarına ilişkin verileri kullanılarak 6 ay sonraki davranışlarını Lojistik Regresyon, Karar Ağaçları, Destek Vektör Makineleri, Yapay Sinir Ağları, Rastgele Orman Algoritması ve Meyilli Hızlandırma yöntemleri ile tahmin etmeye çalışmış ve en iyi başarıyı Meyilli Hızlandırma yöntemi ile ikinci en iyi performansın ise Rastgele Orman Algoritması ile elde etmiştir.

Gıda ve sağlık alanı:

Ulutaş (2018), proteinlerin üç boyutlu yapılarının tahmin edilmesinde önemli bir adım olan protein parçacık seçimine yönelik bir çalışma yürütmüş; üç ve dokuz amino asitlik parçacıkların yapısal benzerliklerini tahmin etmek için lojistik regresyon, AdaBoost, karar ağacı, en yakın komşu, Naive Bayes, Rastgele Orman, destek vektör makinası ve çok-katmanlı algılayıcı yöntemlerini kullanmış ve regresyon modelleri arasında en yüksek tahminleme performansını Rastgele Orman yönteminin verdiği sonucuna ulaşmıştır.

Sabancı (2019), ebeveynlerin sigara içme alışkanlıklarının çocukların solunum hastalıkları üzerindeki etkisini incelemek üzere Çok Değişkenli Uyumlu Regresyon Şeritleri (MARS) modelini, Rastgele Orman Algoritması ile topluluk öğrenimi yöntemine dönüştürmüş ve 200 adet eğitim ve test verisi üzerinde performansını ölçmüştür. Genellikle büyük veri setlerinde iyi çalışan MARS modelinin önermiş olduğu karma yöntem sayesinde küçük boyutlu verilerde de iyi bir sonuç verdiğini ortaya koymuştur.

Eğitim alanı:

Chung ve Lee (2019), öğrencilerin okuldan ayrılma risklerini, 165.715 lise öğrencisinin ayrılma durumları ve izinsiz okula gelmemeleri gibi değişkenler üzerinden Rastgele Orman Algoritması ile tahmin etmişler ve %95 oranında doğruluk elde etmişlerdir.

Bezek Güre (2019), Türkiye'deki öğrencilerin matematik başarılarını etkileyen unsurları belirlemek amacıyla Rastgele Orman Algoritması, Çok Katmanlı Algılayıcı Yapay Sinir Ağları ve Radyal Tabanlı Fonksiyon Yapay Sinir Ağları yöntemlerini tahminleme çalışmasında kullanmıştır. Rastgele Orman Algoritmasının diğer iki yönteme kıyasla daha başarılı sonuçlar verdiğini tespit etmiştir.

Üretim alanı:

Benito Garzón vd. (2008), iklim değişikliğinin İber Yarımadasındaki bitki türü çeşitliliğine etkisini Hükümetlerarası İklim Değişikliği Panelinde öne sürülen senaryolar altında Rastgele Orman Algoritması kullanarak incelemiş dağ kozalaklı türlerde ciddi azalmalar gözlemlendiği, ılıman türlerde bazı bölgelerde azalma olacağı, Akdeniz altı türlerin ise kayda değer bir düşüş göstereceğini tahmin etmişlerdir.

Everingham vd. (2016), Avustralya'daki Tully bölgesinde şeker kamışı verimine ilişkin tahminde bulunmak için mevsimsel iklim tahmin endeksleri ve yağış oranları, maksimum ve minimum sıcaklık girdi olarak ele alıp Rastgele Orman Algoritması ile tahminleme çalışması yürütmüşlerdir. 1 Eylül, 1 Ocak ve 1 Mart'ta oluşturulan tahmin sonuçlarında Eylül ayında üretimin medyanın üzerinde olup olamayacağını % 86.36 oranında belirlemenin mümkün olduğunu, Ocak ayına kadarki dönem için de bu doğruluk oranının % 95,45'e yükseldiğini belirtmişlerdir.

3.3.3. XGBoost Algoritması Uygulamaları

Gradyan Artırmanın geliştirilmiş bir versiyonu bir makine öğrenme algoritması olan XGBoost, Chen ve Guestrin (2016) tarafından ortaya çıkarılmasından bu yana sağlık, biyoloji, çevre, finans ve yazılım gibi alanlarda tahminleme ve sınıflandırma çalışmasının odağında yer almıştır. Bu çalışmalardan bazıları bu bölümde alanlarına göre aktarılmaktadır.

Sağlık alanı:

Lei vd. (2017), solunum toksisitesi verilerine dayalı tahminleme çalışması için uygunluk vektör makinesi, destek vektör makinesi, düzenlenmiş rastgele orman, XGBoost, Naive Bayes ve doğrusal diskriminant analizi yöntemlerini kullanmış ve sayısal tahmin çalışmasında en iyi performansı XGBoost ile elde etmişlerdir.

Xu vd. (2018), insan sağlığı üzerinde olumsuz etkileri olan PM2.5 partiküllerinin havadaki konsantrasyonunu tahmin etmek amacıyla sekiz farklı algoritma kullanmışlar ve Cubist, Rastgele Orman ve XGBoost algoritmalarının en iyi sonuçları veren algoritmalar olduğunu tespit etmişlerdir.

Web sayfaları tıklanmaları ve Ağa izinsiz girişler:

Dhaliwal vd. (2018) ağa izinsiz girişlerin tespit edilmesine yönelik XGBoost ile bir model geliştirmişler ve farklı sınıflandırma modellerine kıyasla XGBoost'un %98,7 doğruluk ile daha iyi sonuç verdiği sonucuna ulaşmışlardır.

Çakmak vd. (2019) internet üzerinden reklam veren seyahat acentalarının web sitelerinin tıklanmasına yönelik bir tahminleme çalışmasında XGBoost, Rastgele Orman algoritmaları ile Destek Vektör Makinalarını kullanarak performanslarını kıyaslamışlar ve XGBoost algoritmasının diğer yöntemlerden daha iyi bir tahminleme sonucunu verdiğini ifade etmişlerdir.

Deligiannis vd. (2020) kampanyalar ile ilgili bilgilendirme mesajları ile ne kadar kişinin ilgilendiğine ilişkin tahminleme çalışmasında XGBoost algoritmasını kullanmış ve çalışma sonucunda bu algoritmanın gerçek değerlerine kıyasla %2,4 bir yakınsama ile tahminleme yapabildiğini ifade etmişlerdir.

Üretim ve rezerv tahminleri:

Merembayev vd. (2019), Kazakistan'daki uranyum yataklarının günlük jeofizik verilerine dayalı sınıflandırma problemi için Gradyan Artırma, Rastgele Orman, Lojistik Regresyon, K-en yakın komşu ve XGBoost Algoritmalarını kullanmışlar ve %97'lik doğruluk oranı ile en iyi performansın XGBoost tarafından sergilendiğini ifade etmişlerdir.

Bankacılık alanı:

Tozlu (2019), ticari kredi başvuru sürecinde şirket bilançolarında aktarmalar arındırma yapılacak kalemlerin tespiti için kullandığı makine öğrenmesi yöntemlerinden yığılmış genelleme, üst seviye öğrenici olarak Rastgele Orman ve temel öğreniciler olarak Light Gradyan Artırma, XGBoost ve CatBoost algoritmalarının kullanıldığı versiyonuyla en iyi performans gösteren model olduğunu belirtmiştir.

Zhang vd. (2019), rulman yatağına ilişkin arıza teşhisi çalışmasında XGBoost algoritması ile Rastgele Orman, Adaboost ve Gradyan Artırma Algoritmalarının performansları karşılaştırmışlar, XGBoost algoritmasının zaman ve doğruluk oranı açısından diğer yöntemlerden daha üstün olduğunu vurgulamışlardır.

3.3.4. Yapay Sinir Ağları Uygulamaları

Teknolojideki gelişmeler sonucu insan beyninin çalışma şeklini prensip edinen ve öğrenme işlemini bilgisayar sistemlerinde gerçekleştiren Yapay Sinir Ağları finans, tıp, bankacılık, pazarlama sektörlerinden mühendislik uygulamalarına kadar oldukça geniş bir

alandaki kullanılmıřtır. Yapay Sinir Ağları üzerindeki arařtırmalar son yıllarda daha fazla artmıřtır. Bu çalıřmalardan bazıları alanlarına göre ařağıda sunulmaktadır.

Üretim alanı:

Birkaç arařtırmacı, Yapay Sinir Ağlarının istatistiksel regresyon modelleriyle karřılařtırarak performansını test etmiřlerdir. Chryssolouris ve Guillot (1990), üretimde yüzey pürüzlülüğü durumu ile ilerleme hızı, kesme hızı ve takım eğim açısı gibi iřlem parametreleriyle iliřkilendiren bir model oluřturup, çoklu regresyon analizi, grup veri iřleme yöntemi ve Yapay Sinir Ağlarını kıyaslamıřlar ve YSA'ların regresyon modeline kıyasla daha üstün olduğunu gözlemiřlerdir.

Feng ve Wang (2003), tornalama iřleminin yüzey pürüzsüzlüğü düzeyini tahmin etme konusunda Çoklu Regresyon analizi ile Yapay Sinir Ağlarının eřit bir řekilde etkin olduğunu belirtmiřtir.

Bař (2006), Türkiye'nin sanayi üretim endeksinin tahmin edilmesine yönelik bir çalıřma yürütmüř, İleri Beslemeli Geri Yayılımlı YSA ile doğrusal regresyon yöntemini kıyaslamıř, YSA'nın daha iyi sonuç verdiğini belirtmiřtir.

Kurt vd. (2017), Türkiye kağıt-karton üretimi, atık kağıt, endüstriyel odun, tomruk üretimi, nüfus, GSYİH, TÜFE, ÜFE, döviz kurları ve ekonomik büyüme rakamlarını kullanarak Türkiye'nin Kağıt-karton ihracatını YSA ile tahmin etmeye çalıřmıřtır.

Temiz (2018), doğaltařların renk ve desen yapısına göre sınıflandırılmasında, Yapay Sinir Ağları, Destek Vektör Makineleri, k En Yakın Komřu, Karar Ağaçları, Naive Bayes yöntemlerini kullanmıř ve Yapay Sinir Ağları ile Destek Vektör Makinelerinin en yüksek başarıya sahip sınıfları oluřturduğunu ifade etmiřtir.

Sağılık alanı:

Pekel (2018) çalıřmasında, diyabet teřhisi üzerine Karar Ağacı, Destek Vektörü Makineleri, Naive Bayes, Yapay Sinir Ağlarının performanslarını bir sınıflandırma problemi üzerinde test etmiř ve analiz edilen veri setinde Naive Bayes Algoritmasının diğeri yöntemlere kıyasla daha iyi performans gösterdiğini tespit etmiřtir.

Nas (2020), acil servis için hasta geliř sıklığını tahmin etmek amacıyla bir çalıřma yürütmüř, karar ağacı, rastgele orman, destek vektör mekaniğı algoritması, en yakın k komřuluğı, gradyan artırımı, stokastik gradyan artırımı, adaboost, çoklu katmanlı algılayıcı, lojistik regresyon ve tekrarlayan yapay sinir ağları makine öğrenmesi yöntemlerini kullanmıřtır. Üzerinde çalıřtığı veri setinde Karar Ağacı, çok katmanlı algılayıcılar ve tekrarlayan yapay sinir ağlarının en iyi sonucu verdiğini vurgulamıřtır.

İhracat ve ithalat alanı:

Bayır (2006), imalat sanayiinin ihracat verileri aylık değerlerini tahmin etmek amacıyla girdi olarak ortalama döviz alış kuru (ABD Doları), sanayi sektörü üretim endeksi, ulusal sınai endeksi kapanış fiyatları ve aylık imalat sanayi kısmi verimlilik endeksi değişkenlerini kullanmıştır. İleri Beslemeli Geri Yayılımlı YSA ile çoklu regresyon modellerini ile yapılan çalışmanın kıyaslamasında YSA'ların sonuçlarının daha iyi bir performansa sahip olduğunu ifade etmiştir.

Bircan vd. (2006), Türkiye'nin ihracat gerçekleştirdiği ülkeleri kümelemek için SOM tipi YSA'lar uygulamış, ihracatın bazı ülkelerde yoğunlaştığı ve ihracatın ülkeler bazında homojen bir yapıda olmadığı sonucuna varmışlardır. Elde edilen 14 kümede, referans vektör değeri yüksek olan kümelerin Türkiye ile ticareti yoğun olan ülkelere, vektör değeri düşük olan kümelerin ise Türkiye ile ticareti az olan ülkelere olduğunu tespit etmişlerdir.

Zou vd. (2007), Çin'in buğday ihracatındaki fiyat seviyelerinin tahmin edilmesine yönelik olarak YSA'lar ile ARIMA modellerini kıyaslamış, YSA'ların daha hatasız sonuçlar verdiğini tespit etmiş ve Çin'in gelecekteki gıda-hububat fiyatlarını modellemede iyi bir alternatif olarak kullanılabileceğini ifade etmişlerdir.

Co ve Boosarawongse (2007), Tayland'ın pirinç ihracatını tahmin etmek için ARIMA ve YSA kullanmışlar, Holt-Winters ve Box-Jenkins modellerinin tatmin edici sonuçlar vermesine rağmen doğrulama aşamasında öngörülemeyen veriler açısından yeterince iyi olmadığını, YSA'nın sonuçlarının ARIMA modellerine kıyasla dinamik doğrusal olmayan trend ile mevsimselliği ve ikisi arasındaki etkileşimi izlemek açısından daha iyi sonuç verdiğini tespit etmişlerdir.

Özbek (2009), 1995-2008 yılları arasında denim pantolunun ihracatını değerlendirmiş ve girdi olarak maliyet, ihracatçı ülkelerdeki ihracat teşvikleri, ihracatçı ülkelerin sahip olduğu markalar, ihracatçı ülkelerin pazara yakınlığı, ihracatçı ülkelerin döviz kurları, ihracatçı ülkelerin sektörel uzmanlaşma durumu, ithalatçı ülkeler ve ithalat payları, ithalat yapan ülkelerin uygulamakta oldukları ticaret sınırlamaları, ithalatçı ülkelerde kişi başına düşen gelir, ithalatçı ülkelerdeki nüfus ve ithalatçı ülkelerin yapmış oldukları ticari anlaşmaları kullanmıştır. Çalışmada YSA'ların ARIMA modeline göre öngörülerde daha yüksek başarı elde ettiği sonucuna varılmıştır.

Karahan (2015), gerçekleştirilen ihracatın tarihleri, ABD doları döviz kuru, satış fiyatı, ihracat yapılan pazar sayısı ve mevsim etkilerinin neden olduğu zarar değişkenlerine göre Malatya'nın kayısı ihracatı miktarını tahmin etme çalışması gerçekleştirmiş,

çalışmada ARIMA ve Yapay Sinir Ağlarını kullanmış, YSA modelinin açıklayıcılığı %86 iken, ARIMA modelinin açıklayıcılığı %75; hata testleri açısından da YSA'nın ortalama mutlak yüzde hata (MAPE) değeri 0.151 iken, ARIMA modelinin 0.233 sonucunu bulmuştur. Sonuç olarak YSA'nın daha iyi bir sonuç verdiği karar vermiştir.

Ersen (2016), Türkiye'nin ağaç ve orman ürünleri sektörüne ait ihracat ve ithalat değerlerini tahminlemede Box-Jenkins ve Yapay Sinir Ağları yöntemlerinin performanslarını karşılaştırmıştır. Yöntemlerin tahminleme başarılarını Hata Karelerinin Ortalama Kökü, Ortalama Mutlak Hata ve Ortalama Mutlak Yüzde Hata istatistikleri ile değerlendirmiş; Box-Jenkins ve Yapay Sinir Ağı modellerinin başarılı olmasına rağmen YSA modellerinin performansının daha iyi olduğunu tespit etmiştir.

Tsai ve Huang (2017), GSYİH, faiz oranları, ihracat ve ithalat değeri, ihracat ve ithalat konteyner adedi ve iskele vinçlerin sayısını esas alarak konteyner akışlarını Yapay Sinir Ağları ile tahmin etmeye çalışmıştır. YSA ile konteyner akışlarının tahmin edilen değerleri ile gerçek değerlerinin birbirine çok yakın olduğu ve Asya limanlarında bu modelin kullanılabileceği yönünde önerilerini sunmuşlardır.

Alam (2019), 1968 ve 2017 yılları arasındaki verileri kullanarak, Suudi Arabistan'ın ithalat ve ihracatını YSA ve ARIMA modelleri ile tahmin etmeye çalışmış, YSA ile birlikte ARIMA (1, 1, 2) ve ARIMA (0, 1, 1) modellerinin Suudi Arabistan'ın ithalat ve ihracatını tahmin etmede uygun modeller olduğu tespit etmiştir.

Talep Tahmini, Pazar Araştırması ve Satış Tahmini alanı:

Hazır vd. (2015), GSYİH, reel kesim güven endeksi, tüketici güven endeksi, dolar kuru, nüfus, konut satış değerleri, yatırım harcamaları gibi faktörlere göre Türkiye mobilya talebini tahminleme çalışmasını Yapay Sinir Ağları ve Regresyon Analizi ile gerçekleştirmiştir. Yapay Sinir Ağlarının MAPE ve RMSE değerlerine göre daha iyi sonuç verdiğini tespit etmiştir.

Zontul ve Yangın (2017), öğrenci sayısı, hane başı eğitim harcaması, dolar kuru, ÜFE, TÜFE ve müşteri sayısı özniteliklerini kullanarak Yapay Sinir Ağları ile eğitim sektöründeki bir firmanın satış tahminine yönelik çalışması gerçekleştirmiştir.

Yücesan (2018), döviz kuru, tatil günleri, tüketici güven endeksi, üretici fiyat endeksi, bölgedeki konut satışları değişkenlerini kullanarak ARIMA, Yapay Sinir Ağları ve ARIMAX yöntemleriyle beyaz eşya sektöründe bir satış tahmini çalışması yürütmüş, Yapay Sinir Ağlarının diğer yöntemlere kıyasla daha başarılı olduğunu ifade etmiştir.

Demir ve Akkaş (2018), yem endüstrisindeki bir işletme için tahminleme çalışmasında zaman serisi, Yapay Sinir Ağları ve Destek Vektör Makinelerini kullanmış ve Destek Vektör Makineleri yöntemini diğer yöntemlere kıyasla daha başarılı bulmuştur.

Yılmaz vd. (2018), konutların fiziksel özellikleri, konumu, ulaşım araçları duraklarına mesafesi gibi kriterleri kullanarak konut fiyatlarının tahminine yönelik Yapay Sinir Ağı ile analiz gerçekleştirmiştir.

3.3.5. İhracat ve İthalat ile İlgili Diğer Çalışmalar ve Araştırmalarda Kullanılan Öznitelikler

İhracat ile iç piyasaya satış arasındaki en önemli farkı ihracatta, ithalatçı ve ithalatçı ülke ile ihracatçı ülkenin kültürel, bürokratik ve ekonomik farklılıklarının olmasıdır. İhracatçı ve ithalatçı pazar birbirinden söz konusu kriterler açısından farklılaştıkça ürün ve hizmetlerde farklılaşma sağlanması gerekebilmektedir (Onur, 2012: 14).

Bu açıdan, Onur (2012: 16), pazar araştırmasında dikkate alınması gerekli kriterleri demografik yapı, kültürel yapı ve yaşam tarzı, coğrafi konum, sosyal ve ekonomik yapı, kültürel yapı, yaşam şekli, teknolojik ve endüstriyel yapı, politik ve hukuki durum, tüketiciler, dağıtım kanalları ve rakipler ile ilgili veriler olarak sıralamıştır.

Bir diğer çalışmada Çavuşgil (1997) tarafından 25 alternatif ülke seti kullanılarak Genel Pazar Fırsat İndeksi (OMOI) geliştirilmiştir. Bu yaklaşımda, orta sınıf büyüklüğü, politik risk, ekonomik özgürlük, telekomünikasyon ve fiziksel altyapı gibi kriterleri içeren temel ekonomik, politik ve sosyal ölçüleri esas alarak pazar potansiyeli belirlenip sıralanmıştır. Bir sonraki adımda, her değerlendirme değişkeni için indeks oluşturulmuş ve son olarak OMOI indeksini elde etmek için bunlara ağırlıklar verilmiştir.

Çavuşgil vd. (2004), 29 farklı kritere göre ülkeleri değerlendirmiş, öncelikle kriterlere faktör analizi yöntemini uygulayarak 7 faktöre ayırmış ve her faktörün ağırlığını belirlemişlerdir. Daha sonra hiyerarşik kümeleme tekniğini kullanarak ülkeleri kümelere ayırmış ve 10 farklı küme elde etmiştir. Ülkeleri faktörler nezdinde aldığı genel değerlerine göre sıralamıştır. Böylece, kümeleme tekniğini kullanarak yapısal olarak benzer gruplar; sıralama ile de küme içerisinde piyasa potansiyellerine göre ülkelerin öncelikli sırası elde edilmiştir.

Sheng ve Mullen (2006) tarafından Çavuşgil (1997)'nin OMOI yöntemi daha büyük ülke setine uygulanarak pazar fırsatları sıralanmış ve ABD'nin gerçekleşen ticareti ile kıyaslanarak OMOI'nın geçerliliği değerlendirilmiştir. Bu çalışmada, OMOI'nın

istikrarlı bir araç olduğu ancak, sonuçlar ağırlıklara bağlı olduğu için verilecek ağırlıkların oldukça dikkatli bir şekilde değerlendirilmesi gerektiğine karar verilmiştir.

Bazı çalışmalar, ortaya çıkan hedef pazarın dinamizmi ve gelecek potansiyelini hesaba katma konusunda makroekonomik pazar seçim modellerinin yetersiz olduğunu belirtmişlerdir. Bu çalışmalarda, pazarlar değerlendirilirken uzun dönem pazar potansiyeli, kültürel uzaklık, yerel sektördeki rekabetin gücü, müşterilerin yabancı ürünlere ve işletmelere yönelik algısı gibi kriterlerin değerlendirilmesinin de gerektiği ortaya çıkmıştır (Çavuşgil, 1997).

Priya ve Venkatesh (2012), Hindistan'da bulunan bir demir-çelik firması için cezbedici pazar lokasyonlarını bulmak amacıyla çoklu regresyon yöntemi ile işleme kazancı, hedef pazarın iletişim altyapısı, hedef pazarın ekonomik büyüme oranı, hedef pazarda rekabet avantajı, hedef pazarın talebi, hukuk sisteminin hedef pazarda işleyişi, hedef pazarın diğer pazarlara yakınlığı, hedef pazar ve muafiyetli ticaret anlaşmaları seçim kriterlerini, AHP yöntemiyle bu kriterlerin ağırlıklarını belirledikleri entegre bir pazar seçim karar modeli kullanmışlardır.

Zhao vd.(2011), Avrupa Birliği, Amerika, Asya, Japonya, Bağımsız Devletler Topluluğu (Rusya Federasyonu, Kırgızistan, Kazakistan), Hong Kong alternatiflerini satın alma gücü, pazar büyüklüğü, rekabet koşulları, politik stabilite, kalite algısı kriterlerine göre bulanık AHP yöntemini kullanarak Çin tekstil sektörü için ihracat pazarlarını belirlemişlerdir.

Atalay (2012), tez çalışmasında beş ana kriter altında belirlenen 32 kriter ile AHP yöntemini kullanarak Türkiye orman ürünleri için hedef pazarları belirlemiştir. Coğrafi konum ve ulaşım ana kriterine göre ABD; demografik yapı ana kriterine göre Birleşik Arap Emirlikleri; dünya geneli dış ticareti ana kriterine göre Çin; Türkiye ile dış ticareti ana kriterine göre Almanya ve sosyal ekonomik yapı ana kriterine göre Singapur en yüksek öneme sahip ülkeler olarak ortaya çıkmıştır. Tüm kriterler birlikte değerlendirildiğinde ise en yüksek önem derecesine sahip ülke ABD olurken, sıralamayı Singapur, Norveç, Almanya, Çin, Birleşik Arap Emirlikleri, Fransa, Güney Afrika ve Romanya takip etmiştir.

Miecinskiene vd. (2014), mobilya sektöründe faaliyet gösteren bir firma için potansiyel olarak görülen 8 ülkeden ikisinin hedef ülke olarak belirlenmesi çalışmasını nüfus artış oranı, işsizlik düzeyi, enflasyon, gayri safi milli hasılasındaki büyüme oranı, kişi başı gelir düzeyi, kişi başı ihracat ve kişi başı ithalat değerleri kriterlerine göre değerlendirmiştir. Altı uzmanın bu kriterlere verdiği puanlamalar sonrasında oluşturulan

nihai kriter ağırlıkları ülkelerin söz konusu kriterler açısından normalize edilmiş değerleri ile çarpılarak hedef ülkelerin aldığı değerler sıralanmıştır.

Mat Isaa vd. (2013) çalışmalarında Malezya inşaat sektöründe faaliyet gösteren 129 firmaya uluslararası pazarlara girmek için tanımlanan firma, ülke, sektör ve proje özelinde yer alan toplam 27 kriter ile ilgili anketler göndermiş ve yanıtlanan anket verilerine faktör analizi uygulamıştır. Projeye özgü faktörlerin firmaların uluslararası pazar seçimde oldukça etkili olduğu ortaya çıkmıştır. Bu faktörler önem derecelerine göre, rekabetin yoğunluğu, yüksek kalite gereksinimleri, rakiplere yakınlık, hedef ülkeye mesafe ve firmanın uluslararası rekabet edebilirliği olarak ortaya çıkmıştır.

Mobin vd. (2014) tarafından, İran fıstığı için 18 ülke pazarı, yönetmelik-düzenlemeler, kültürel etkenler, nakliye mesafeleri, ekonomik durum, pazar potansiyeli, politik etkenler ve tespit edilen karşılaştırmalı avantajlar kriterleri üzerinden değerlendirilmiştir. Kriterleri entropi ağırlık metodu kullanarak ağırlıklandırmış ve bu ağırlıkları kullanarak SAW (Simple Additive Weighting), TOPSIS ve VIKOR yöntemleriyle sıralanmıştır.

Bir diğer farklı çalışma Öztürk vd. (2015) tarafından geliştirilmiştir. Bu çalışmada, üç aşamada ülkeler dört farklı kümeye ayrılmıştır. Öncelikle lineer regresyon yöntemiyle tüketici harcamaları bağımlı, gelir ve yıllar bağımsız değişken olarak ele alınıp ülkelerin gelir ve yıllar için regresyon katsayılarını belirlemiş, regresyon katsayılarının medyan değerlerine göre bazı ülkeler ürünle ilgili, bazıları ilgisiz olarak gruplanmıştır. İlgili bulunan ülkeler Robinson ülke risk skoru ile tüketim harcamasındaki büyüme oranlarına göre dört farklı kümeye ayrılmıştır: 1) küresel değerli (tüketici harcamalarındaki büyüme düşük, risk yüksek); 2) küresel endüstride kazananlar (tüketici harcama büyümesi yüksek, risk yüksek); 3) endüstriyel açıdan değerli (tüketici harcamaları yüksek, risk düşük); 4) durgun (tüketici harcamaları düşük, risk düşük).

Christian vd. (2016), çanta imalatı yapan bir firmanın Mısır, Fas, Nijerya, Gana ve Güney Afrika pazarları arasında seçim için çok kriterli karar verme metodu TOPSIS'i kullanmıştır. Söz konusu beş ülke, ekonomik özgürlük, eğitim düzeyi, coğrafi uzaklık, cep telefonu kullanım oranı, Gayri Safi Yurt içi Hasıla, ithalatı, ihracatı, ülke içi kuruluşlar arası güç farkı, bireysellik, erkek egemenliği, belirsizlik riski ve nüfus kriterlerine göre değerlendirilmiştir.

DEİK (2016) tarafından hazırlanan çalışmada Türkiye'nin Çin'e hangi kalemlerde ne kadar ihracat potansiyelinin olduğu detaylı bir analizle ürün bazında belirlenmiştir. Gümrük Tarife İstatistik Pozisyonlarında (GTİP) dört haneye inilerek 1.200 ürün

pozisyonu için Türkiye'nin ihracatı, Türkiye'nin ihracatında Çin'in payı, Çin'in söz konusu üründe dünyadan toplam ithalatı, Türkiye'nin Çin'e ilgili ürünü ihracatında son beş yıllık ortalama artış, Çin'in söz konusu ürünü ithalatında son beş yıllık ortalama artış ile Helmers ve Pasteels (2006)'nın çalışmasında önerdikleri ve Çin'in ithalatı açısından söz konusu ürünün ne kadar daha ihracat yapılabileceğini ifade eden gösterge ticaret potansiyeli ve nispi gösterge ticaret potansiyel oranı kriterleri değerlendirilmiştir. Türkiye'nin Çin'e İhracatında Potansiyel Taşıyan Hedef Ürünler beş gruba ayrılmıştır.

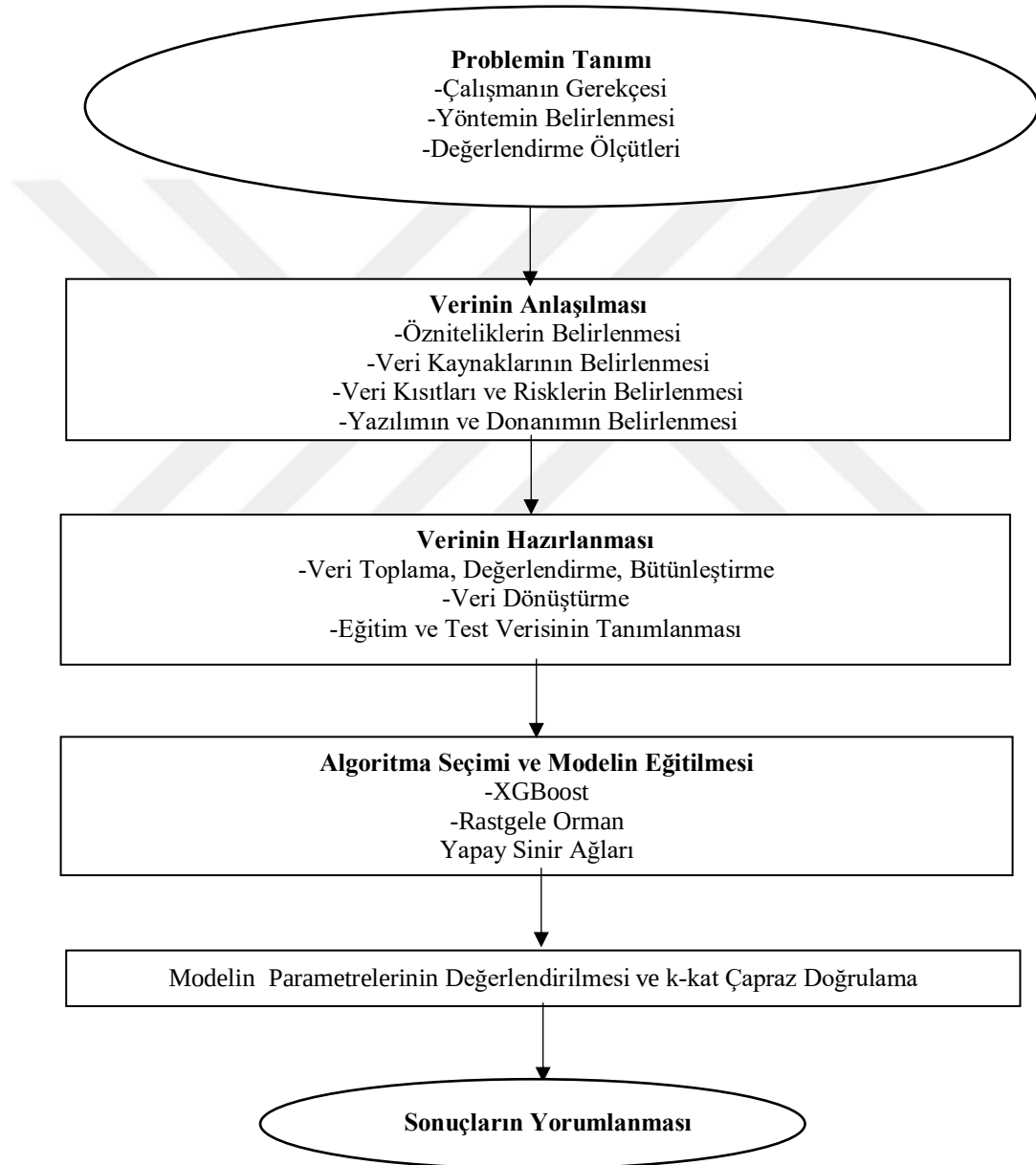
Gao vd. (2018), tüketici güven endeksi, tüketici fiyat endeksi, çelik üretimi ve benzin fiyatı kriterlerinden oluşan veri seti ile Çin'de otomobil üreticisi bir firmanın satış tahmini için bir ekonometrik model (Vektör Hata Düzeltme Modeli-VECM) geliştirmişler ve klasik zaman serilerine kıyasla daha fazla başarıya ulaştıklarını ifade etmişlerdir.



DÖRDÜNCÜ BÖLÜM

OECD ÜLKELERİNİN YAŞ MEYVE SEBZE İTHALATININ VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ

Tezin analiz bölümünü oluşturan bu bölümde OECD ülkelerinin yaş meyve sebze ithalatını tahmin etmek amacıyla gerçekleştirilen işlemler Şekil 4.1’de yer almaktadır.



Şekil 4.1 Analiz Süreci İş Akışı

Öncelikle problemin tanımlanması ve verinin anlaşılması aşamasında çalışmanın gerekçesine değinilmiş, kullanılacak yöntemler ve bu yöntemlerin tercih edilme nedenleri, yöntemlerin hangi ölçüm kriterlerine göre değerlendirileceği konuları açıklanmıştır.

Verinin anlaşılması aşamasında, kullanılacak öznitelikler ve bu öznitelik değerlerinin temin edileceği veri kaynakları belirlenmiş, çalışmanın kısıtları ile kullanılacak yazılım ve donanım ile ilgili bilgi verilmiştir.

Verinin hazırlanması aşamasında, verinin toplanması, değerlendirilmesi ve verilerin bütünleştirme işlemi ile tek bir veri seti haline getirilmesi sağlanmıştır. Veri setinde kategorik verilerin 0-1 şeklinde ikili değerlere dönüştürülmesi ile öznitelik eklemeleri ve standardizasyon işlemi ile veri dönüşümü gerçekleştirilmiştir. Daha sonra veri setinin ilk 5 yılı eğitim, son iki yılı ise test verisi olarak ayrılmıştır.

Model oluşturma aşamasında 10-kat çapraz doğrulama ile parametrelerin optimum değerleri elde edilmiştir.

Ulaşılan en uygun parametre değerleri ile algoritmalar için tahminleme süreci işletilmiş ve algoritmaların sonuçları kıyaslanmıştır.

4.1. Problemin Tanımlanması

4.1.1. Çalışmanın Gerekçesi

Antalya ili, Türkiye yaş meyve sebze üretim ve ihracatında en fazla paya sahip illerdendir. Bu nedenle, tez çalışmasında yaş meyve sebze ihracatının geliştirilmesine yönelik bir analiz çalışmasına odaklanılmıştır.

Türkiye yaş meyve sebze üretiminde ve ihracatında dünya genelinde önemli bir konumdadır. Tablo 4.1’de görüleceği üzere, FAO verilerine göre Türkiye, dünya ayva, kayısı, kiraz ve incir üretiminde birinci; kavun ve pırasa üretiminde ikinci; biber, hıyar-kornişon, karpuz, mandalina ve havuç-şalgam üretiminde üçüncü; domates, elma, patlıcan, fasulye, çilek, ıspanak ve vişne üretiminde dördüncü; kuru soğan, şeftali-nektarin, lahana, kabak ve armut üretiminde beşinci; üzüm, erik ve limon üretiminde ise altıncı sırada yer almaktadır. Ancak, bazı ürünler haricinde dünya ihracat sıralamasında aynı performans sergilenememektedir.

Tablo 4.1 Türkiye'nin Üretiminde Önemli Ürünler ve 2018 Yılı Dünya Üretiminden Aldığı Pay

Ürün	Türkiye Üretimi (Ton)	Dünya Üretimi (Ton)	Dünya Üretiminden Türkiye'nin Aldığı Pay	Dünya Üretim Sıralamasında Türkiye'nin Konumu
Domates	12.150.000	182.256.458	6,67%	4
Karpuz	4.031.174	103.931.337	3,88%	3
Üzüm	3.933.000	79.125.982	4,97%	6
Elma	3625960	86142197	4,21%	4
Biber	2.554.974	36.771.482	6,95%	3
Kuru soğan	1.930.695	96.773.819	2,00%	5
Portakal	1.900.000	75.413.374	2,52%	7
Hıyar-kornişon	1.848.273	75.219.440	2,46%	3
Kavun	1.753.942	27.349.214	6,41%	2
Mandalina	1.650.000	34.393.430	4,80%	3
Limon	1.100.000	19.368.838	5,68%	6
Patlıcan	836.284	54.077.210	1,55%	4
Şeftali ve nektarin	789.457	24.453.425	3,23%	5
Lahana ve diğer brassicalar	765.276	69.381.555	1,10%	5
Kayısı	750.000	3.838.523	19,54%	1
Fasulye	668.973	26.626.351	2,51%	4
Havuç ve şalgam	644.367	39.996.287	1,61%	3
Kiraz	639.564	2.547.944	25,10%	1
Kabak	616.777	27.643.932	2,23%	5
Armut	519.451	23.733.772	2,19%	5
Muz	498.888	115.737.861	0,43%	23
Marul ve hindiba	487.543	27.256.487	1,79%	7
Çilekler	440.968	8.337.099	5,29%	4
İncir	306.499	1.135.316	27,00%	1
Erik	296.878	12.608.678	2,35%	6
Karnabahar ve brokoli	294.743	26.504.006	1,11%	7
Pırasa	252.958	2.179.050	11,61%	2
Greyfurt	250.000	9.374.739	2,67%	7
İspanak	225.174	26.255.931	0,86%	4
Vişne	184.167	1.529.000	12,04%	4
Ayva	176.479	688.660	25,63%	1

Kaynak: FAO, 2018 Yılı Verileri (<http://www.fao.org/faostat/en/>, erişim tarihi: 16.07.2020)

Birleşmiş Milletler veri kaynağı “UN Comtrade Data Base”ın Tablo 4.2’de yer alan 2019 yılı ihracat verilerine göre, Türkiye dünya genelinde ayva ve incir ile “Satsuma” cinsi mandalınanın ihracatında birinci; kabuklu kestane ihracatında üçüncü konumda iken, kayısı, kiraz, diğer mandalina türleri ve greyfurt ihracatında dördüncü; şeftali-nektarin, kabak, patlıcan ve limon ihracatında ise beşinci sırada yer almaktadır. Türkiye, üretim miktar açısından üstünlüğünün bulunduğu diğer ürünlerde ise dünya ihracatı sıralamasında daha gerilerde bulunmaktadır.

Tablo 4.2 Türkiye'nin İhracatında Önemli Ürünler ve 2019 Yılı Dünya İhracatından Aldığı Pay

GTİP	Ürün	Türkiye'nin İhracat Değeri (USD)	Dünya İhracat Değeri (USD)	Dünya İhracatından Türkiye'nin Aldığı Pay	Sıra
070200	Domates	303.046.886	8.486.537.748	3,57%	8
080521	Mandarin, tangerin (klemantinler hariç) : satsuma, yerli mandarin, okitsu, mandarin ve vilking, tanjerin, mineola mandarin	297.911.750	1.681.836.528	17,71%	1
080420	İncir	286.517.656	450.932.582	63,54%	1
080550	Limon	243.788.003	3.100.001.828	7,86%	5
080929	Kiraz	183.838.813	3.537.544.933	5,20%	4
080610	Taze üzüm	150.107.207	7.396.204.778	2,03%	12
080390	Muz (plantainler hariç)	139.862.233	7.487.245.693	1,87%	13
070960	Biber	124.926.491	5.467.557.858	2,28%	7
081090	Demirhindi, mahun elması, ekmek ağacı meyvesi, sapodillo meyvesi/ çarkıfelek meyvesi. karambola ve pitahaya /nar/ muşmula/ kuşburnu ve diğer meyveler	98.166.382	3.795.740.364	2,59%	6
080510	Portakal	92.037.252	4.526.862.938	2,03%	10
080810	Elma	90.403.552	5.596.924.881	1,62%	13
080930	Şeftali-nektarin	89.784.014	1.914.573.978	4,69%	5
080540	Greyfurt	71.490.419	688.884.653	10,38%	4
070310	Soğan-şalotlar	54.034.547	3.347.598.607	1,61%	13
070993	Kabak	48.181.340	1.252.995.254	3,85%	5
080529	Mandalina diğerleri	45.117.889	697.254.013	6,47%	4
070190	Patates - taze / soğutulmuş	41.560.222	3.393.663.829	1,22%	15
080910	Kayısı (zerdali dahil)	38.096.089	447.806.810	8,51%	4
070700	Hıyarlar ve kornişonlar	36.774.200	2.390.898.500	1,54%	7
080241	Kabuklu-kestane	36.066.109	202.715.586	17,79%	3
081010	Çilek	24.866.821	2.701.002.871	0,92%	13
080830	Armut	23.938.688	1.890.225.581	1,27%	12
080940	Erik	17.018.262	812.728.623	2,09%	10
080410	Hurma	15.522.532	616.680.985	2,52%	8
070610	Havuç-şalgam	13.665.078	1.007.046.140	1,36%	17
070930	Patlıcanlar - taze / soğutulmuş	13.191.872	426.596.888	3,09%	5
080711	Karpuzlar	12.656.613	1.483.640.359	0,85%	17
080840	Ayva	12.002.802	30.567.188	39,27%	1
080719	Kavunlar	5.481.433	1.317.491.182	0,42%	17
070959	Cüce kız mantarları -kuzu mantarlar -diğer mantarlar	5.291.999	551.229.936	0,96%	16
070490	Beyaz lahana-kırmızı lahana-alabaşlar	4.345.750	1.142.265.279	0,38%	25

Kaynak: UN COMTRADE, 2019 Yılı Verileri (<https://comtrade.un.org/>, erişim tarihi: 16.07.2020)

Tablo 4.2 Türkiye'nin İhracatında Önemli Ürünler ve 2019 Dünya İhracatından Aldığı Pay (...devamı)

GTİP	Ürün	Türkiye'nin İhracat Değeri (USD)	Dünya İhracat Değeri (USD)	Dünya İhracatından Türkiye'nin Aldığı Pay	Sıra
070999	Salatalık sebzeler (marul (lactuca sativa) ve hindiba (cichorium spp.)/pazı ve yaban enginarı/ kebere rezene/ tatlı mısır/ bamya/ semizotu/ maydanoz/ asma yaprağı ve diğerleri	3.440.325	2.123.434.999	0,16%	34
081050	Kivi	3.357.168	2.802.753.123	0,12%	15
070390	Pırasa	2.929.753	267.825.256	1,09%	13
070320	Sarımsak	1.926.627	787.118.031	0,24%	22
070519	Diğer marullar	1.923.510	1.455.707.471	0,13%	22
080522	Klemantinler, monreale, king, freumont, nova, marisol mandarin	1.558.909	1.452.486.529	0,11%	23
080440	Avokado	1.204.352	6.018.299.961	0,02%	36
080430	Ananas	1.060.277	1.035.915.741	0,10%	31
070410	Karnabahar- brokoli	1.058.648	1.699.253.581	0,06%	26

Kaynak: UN COMTRADE (<https://comtrade.un.org/>, erişim tarihi: 16.07.2020)

Ülkemiz üretim potansiyelinin ihracata da yansıtılabilmesi için ihracatın geliştirilmesine yönelik çalışmalar önemlidir. Günümüz dünyasında uluslararası alanda yaşanan rekabet her geçen gün daha da artırmaktadır. Firmalar yurt dışında ürünleri için hedef pazarları belirlerken ellerindeki imkan ve kaynakları en verimli şekilde değerlendirmek durumundadırlar. İhracat işlemlerinde başarılı olmak ve doğru pazara yönelebilmek için doğru bir hedef pazar çalışmasının gerçekleştirilmesi oldukça önemlidir.

Bu noktada, alınması gereken stratejik kararlar, gelecek dönemde üretilmesi planlanan ürünlerin tür ve çeşitlerinin belirlenmesi, bir başka ifade ile üretim deseni ve planının hazırlanabilmesi için ithalatçı ülkelerinin gelecek dönem alım taleplerinin belirlenmesi yapılacak çalışmaların en başında gelmektedir.

OECD ülkeleri ise Gayri Safi Yurt İçi (GSYİH) düzeyi yüksek olan ve toplamda dünya yaş meyve sebze ithalatının %68'ini tedarik eden ülkelere olmaktadır. Dolayısıyla, ülkemizin yaş meyve sebze ihracatını artırmak ve dünya ihracatından daha fazla pay alabilmek için OECD ülkelerinin yaş meyve sebze ithalatı incelemeye alınmıştır.

4.1.2. Kullanılacak Yöntemlerin Belirlenmesi

Pazar araştırması ve talep tahmini çalışmalarında zaman serileri, regresyon analizleri gibi ekonometrik yöntemlerin yanı sıra Yapay Sinir Ağlarının da yoğun bir

şekilde tercih edildiği gözlenmiştir. Bununla birlikte yakın zamanda gerçekleştirilen tahminleme çalışmalarında Rastgele Orman ile XGBoost Algoritmalarının giderek daha fazla kullanıldığı gözlenirse de bu yöntemlerin dış ticarete ilişkin tahminleme çalışmalarına literatürde rastlanmamıştır.

Akademik anlamda farklı bir çalışma olması açısından tez çalışmasında OECD ülkelerinin yaş meyve sebze ithalatlarının tahminlenmesi çalışmasında, yoğun bir şekilde kullanılan Yapay Sinir Ağlarının yanı sıra Rastgele Orman ve XGBoost Algoritmalarının kullanılması ve oluşturulacak veri seti üzerinde performanlarının kıyaslanması hedeflenmiştir. Her üç teknik için de öncelikle parametre optimizasyonu yapılacak daha sonra tespit edilen en iyi parametre seti için tahmin değerleri bulunmaya çalışılacaktır.

4.1.3. Kullanılacak Yöntemlerin Değerlendirme Ölçütleri

Modellerin değerlendirilmesi için R^2 (determinasyon katsayısı), $RMSE$ (Root Mean Squared Error- Hata Karelerinin Ortalama Kökü), MAE (Mean Absolute Error- ortalama mutlak hata) ve $MAPE$ 'nin (Mean Absolute Percentage Error-Ortalama mutlak yüzde hata) performans kriteri olarak kullanılması planlanmıştır. Söz konusu değerlendirme ölçütlerinin hesaplamaları formülasyon (4.1) ile (4.4) arasında açıklanmaktadır (Gnananandarao vd., 2019: 55).

Determinasyon katsayısı R^2 , tahmin edilen değerlerin gerçek değerlere ne kadar uygun olduğunu gösteren bir performans kriteridir. R^2 değeri 1'e ne kadar yakınsa tahmin edilen değerler gerçek değerlere o kadar yakındır. Bir başka ifadeyle, model, veri setini o kadar doğru tahminlemektedir. R^2 değeri aşağıdaki formülasyon ile hesaplanmaktadır.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4.1)$$

Burada,

i: gözlem sayısı ($i=1, \dots, n$),

y_i : gerçek değer,

$\hat{y}_i$: tahminlenen değeri,

$\bar{y}_i$: tahminlenen değerlerin ortlamasını ifade etmektedir.

$RMSE$, gerçek değere ne kadar yakın bir tahminleme yapıldığını ifade eden bir performans kriteridir. Söz konusu kriterin hesaplama formülasyonu aşağıda sunulmaktadır.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.2)$$

MAE, performans kriterinin hesaplanması için bir gözlemin gerçek değer ile tahmini değer arasındaki farkının mutlak değerlerinin toplamı gözlem sayısına bölünür.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (4.3)$$

MAPE, ortalama mutlak yüzde hata olup, formülasyon (4.4)'deki gibi hesaplanmaktadır.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100 \quad (4.4)$$

Makine öğrenmesi sürecinde öncelikle eldeki veriler “eğitim verisi” ve “test verisi” olmak üzere iki parçaya ayrılmaktadır. Veriyi bölmek ile ilgili herhangi bir kural olmamakla birlikte literatürde yer alan pek çok çalışmada eldeki veri setinin %70'i rastgele olarak eğitim veri kümesine atanmaktadır. Bu sayede hem öğrenme süreci için yeterli gözleme sahip olunmakta, hem de öğrenilen verilerden genelleme yapabilmek için de test verisi kullanılabilir.

Öğrenme sürecinde, eğitim verisi yardımıyla oluşturulan modellere ait matematiksel modellerin optimizasyonu sağlanmaktadır. Bunun için her bir makine öğrenmesi algoritmasına ait parametrelerin en iyi değerleri bulunmaktadır. Optimal parametre değerleri seçilirken gerçek gözlem değerleri ile modelin hesapladığı sonuçlar (tahmin çıktısı) arasındaki fark olarak bilinen toplam hata değerinin minimizasyonu temel alınır. Toplam hata olarak adlandırılan ceza fonksiyonunu en küçükleyecek parametre değerleri seçilir.

Öğrenme aşamasını test aşaması izlemektedir. Bu aşamada örneklerden ayrılan test verisi yardımıyla makine öğrenmesi modelinin gerçek hayattaki gücü olarak ifade edilebilen genelleme gücü ölçülmektedir. Geliştirilen modelin test verisinde tek başına yüksek başarımlı göstermesi yeterli değildir, makine öğrenmesi algoritmasının eğitim ve test verilerinde gösterdiği performans değerlerinin de birbirine yakın olması beklenmektedir.

Başarılı bir makine öğrenmesi algoritmasından beklenen, eğitim verileri yardımıyla öğrendiği bilgiyi o alandaki tüm verilere yansıtabilmesidir. Bu, modelin daha önce hiç karşılaşmadığı veya ilk kez gördüğü veriler üzerinde başarılı tahminler yapabilmesi anlamına gelmektedir. Eğer eğitim verisinde yüksek performans göstermiş bir modelin test verisinde gösterdiği başarı oranı düşük ise bu durumda aşırı öğrenme (ezberleme) olduğu söylenebilir. Aşırı öğrenme bir makine öğrenmesi algoritmasının eğitim verisinin çok iyi öğrenmesi fakat modelin ilk kez karşılaştığı bir veriyi (test verisi) kullanarak elde ettiği tahminlerin başarısız veya düşük performans oranına sahip olma durumudur.

4.2. Verinin Anlaşılması

4.2.1. Çalışmada Kullanılacak Özniteliklerin ve Veri Kaynaklarının Belirlenmesi

OECD ülkelerinin yaş meyve sebze ithalat verilerini tahmin etmek için geçmiş dönem verileri ikincil veri kaynağı olan Birleşmiş Milletler veri kaynağı “UN COMTRADE DATA BASE”den (<https://comtrade.un.org/data>, erişim tarihi: 20.05.2020) derlenmiştir.

Söz konusu kaynakta veriler, ürünlerin dünya genelinde kullanılmakta olan Gümrük Tarife İstatistik Pozisyonlarına (GTİP) göre derlenmiştir. GTİP; Dünya Gümrük Örgütü’nün standart hale getirdiği armonize sistemden alınan ve tüm dünyada ürünlerin tanımlanması ve sınıflandırılması için kullanılan 12 haneli bir koddur. Bu kodun ilk altı hanesi armonize sistem tarafından belirlenir ve ülkeler tarafından herhangi bir değişiklik yapılmasına izin verilmez.⁸ Söz konusu ikincil veri kaynağında da istatistikler ilk altı hane üzerinden tutulmuştur. Yaş meyve sebze ihracatı dahilinde değerlendirilen altılı GTİP bazında 66 ürün grubu bulunmaktadır.

Türkiye’nin yaş meyve sebze ihracatına ilişkin öngörülerde bulunabilmek adına hedef pazar konumundaki OECD üyesi ülkelerin yaş meyve sebze ithalatı incelenmiştir. Bu doğrultuda, Türkiye dışında kalan OECD üyesi 35 ülkenin ithalatı mercek altına alınmıştır. Söz konusu ülkeler, veri analizinde değerlendirilebilmesi için UN COMTRADE ikincil veri kaynağında belirlenen ülke kodları ile ifade edilmiştir.

OECD üyesi ülkeler ve kodları listesi ile yaş meyve sebze ithalatında dikkate alınan ürün grupları listesi Ek-1 ve Ek-2’de sunulmaktadır.

Literatür taraması ve uzman görüşleri doğrultusunda belirlenen öznitelikler ve verilerin temin edildiği ikincil veri kaynakları Tablo 4.3’de yer almaktadır.

⁸ <https://www.gumruklememe.com.tr/gumruklememe-terimleri-sozlugu/gtip-nedir/> (erişim tarihi: 14.11.2019).

Tablo 4.3 Veri Seti Öznitelikleri

Öznitelik İsmi	Öznitelik Türü	Veri Kaynağı	Özniteliğin Durumu
Dönem	Sürekli	-	Girdi
Ülke Kodu	Kategorik	UN COMTRADE	Girdi
Ürün Kodu (GTİP)	Kategorik	-	Girdi
Alıcı ülkenin ilgili yıl enflasyon oranı	Sürekli	Dünya Bankası	Girdi
İlgili ülkenin nüfusu (1.000 kişi)	Sürekli	Birleşmiş Milletler	Girdi
İlgili Ülkenin Gayri Safi Yurt İçi Hasılası (milyar \$)	Sürekli	IMF	Girdi
İlgili Ülkenin 5 yıl sonraki tahmini Gayri Safi Yurt İçi Hasılası (milyar \$)	Sürekli	IMF	Girdi
İlgili Ülkede İş yapma Kolaylığı Endeksi	Sürekli	Dünya Bankası	Girdi
İlgili Ülkenin Açık Pazar Endeksi	Sürekli	Uluslararası Ticaret Odası (ICC)	Girdi
Alıcı ülkenin ilgili yıldaki söz konusu üründen üretimi	Sürekli	FAO	Girdi
Alıcı ülkenin ilgili yıl tarımda işgücü istihdamı	Sürekli	FAO	Girdi
Alıcı ülkenin ilgili yıl para biriminin \$ cinsinden değeri	Sürekli	FAO	Girdi
İthalatçı ülkenin ilgili üründen bir yıl sonraki ithalatı Değeri (\$)	Sürekli	UN COMTRADE	Çıktı

Özniteliklerin açıklamaları aşağıda sunulmaktadır:

1. **Dönem:** İthalatın gerçekleştirildiği yılı ifade etmektedir. 2012 yılı itibariyle yaş meyve sebze ürünlerine ait GTİP’lerde revizyon gerçekleştirilmiş olması nedeniyle veri seti oluşturulurken 2012-2018 yılları arasında OECD ülkelerinin yaş meyve sebze ürünleri açısından, ürün bazına verileri derlenerek oluşturulmuştur.
2. **Ülke kodu:** Her ülkenin dış ticaret işlemlerinde kullanılan bir ülke kodu bulunmaktadır. Çalışmada kullanılan ülke kodları UN COMTRADE web sayfasından temin edilmiştir.
3. **İlgili yaş meyve sebze ürününü ifade eden GTİP kodu:** Yaş meyve sebze ihracatı dahilinde değerlendirilen altılı GTİP bazında 66 ürün grubu bulunmaktadır.
4. **Alıcı ülkenin enflasyon oranı:** Tüketici fiyat endeksi olarak ölçülen enflasyon, tüketicilerin satın aldıkları ürün ve hizmetlerden oluşan bir ürün-hizmet sepetinin ortalama fiyatlarının yıllık değişim oranıdır. Alıcı ülkenin enflasyon oranı alım gücünü doğrudan etkileyen bir faktör olduğu için oluşturulacak veri setinden enflasyon oranı değişkenine de yer verilmiştir. 2012-2018 yıllarının her biri için ilgili alıcı ülkenin (OECD ülkesinin) enflasyon oranı verileri ikincil veri kaynağı olan Dünya Bankasının web sayfasından temin edilmiştir⁹.
5. **İlgili ülke nüfusu:** İlgili ülkenin tüketim kapasitesini doğrudan etkileyen önemli bir değişkendir. 2012-2018 yıllarının her biri için ilgili alıcı ülkenin (OECD ülkesinin) nüfusu 1000 kişi olarak ikincil veri kaynağı olarak Birleşmiş Milletlerin web sayfasından temin edilmiştir¹⁰.
6. **İlgili ülkenin Gayri Safi Yurt İçi Hasılası değeri:** Ülke geliri de nüfus gibi ithalat hacmini etkileyen önemli değişkenler arasındadır. 2012-2018 yıllarının her biri için ilgili alıcı ülkenin (OECD ülkesinin) Gayri Safi Yurt İçi Hasılası cari fiyatlar üzerinden milyar dolar olarak ikincil veri kaynağı olarak IMF (International Monetary Fund- Uluslararası Para Fonu) web sayfasından temin edilmiştir¹¹.
7. **İlgili ülkenin 5 Yıl sonraki tahmini Gayri Safi Yurt İçi Hasılası değeri:** 2017-2023 yıllarının her biri için ilgili alıcı ülkenin (OECD ülkesinin) Gayri Safi Yurt İçi Hasılası tahminleri cari fiyatlar üzerinden Milyar dolar olarak ikincil veri kaynağı olarak IMF

⁹ <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG> (erişim tarihi: 17.10.2019).

¹⁰ <https://population.un.org/wpp/Download/Standard/Population/> (erişim tarihi: 17.10.2019).

¹¹ <https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD/CHN> (erişim tarihi: 19.10.2019).

(International Monetary Fund- Uluslararası Para Fonu) web sayfasından temin edilmiştir¹².

8. **İlgili ülkede İş Yapma Kolaylığı Endeksi:** Söz konusu endeks iş yapma ortamı için yasal düzenlemelerin, farklı iş adamları tarafından algılanışını ölçmektedir. Veriler, Dünya Bankası tarafından yıllık olarak yayınlanmaktadır. 2012-2018 yıllarının her biri için ilgili alıcı ülkede (OECD ülkesinin) İş Yapma Kolaylığı Endeksi değeri ikincil veri kaynağı olarak Dünya Bankasının web sayfasından temin edilmiştir¹³.
9. **İlgili ülkenin Açık Pazar Endeksi (OMI):** Uluslararası Ticaret Odası (ICC) tarafından oluşturulan söz konusu endeks, hükümetlerin açık ekonomiler yaratmak amacıyla verdikleri taahhütlerin ne kadarını yerine getirdiklerini ifade eder. 75 ülke değerlendirilmekte olup, ilk yayımı 2011, ikincisi 2013, üçüncüsü 2015 ve dördüncüsü 2017 yılında yayınlanmıştır. 2011 yılı verileri 2012 yılı için, sonraki veriler ise iki yıl için geçerli olduğu öngörülerek kullanılmıştır¹⁴.
10. **Alıcı ülkenin ilgili yıldaki söz konusu üründen üretimi:** İlgili ithalatçı ülkenin incelemeye alınan yıldaki ilgili üründen gerçekleştirdiği üretim miktarıdır. Ton biriminden ifade edilmiş olan veriler ikincil veri kaynağı olan FAO'nun (Food and Agriculture Organization of the United Nations: Birleşmiş Milletler Gıda ve Tarım Organizasyonu) web sayfasından temin edilmiştir¹⁵.
11. **Alıcı ülkenin ilgili yıl tarımda işgücü istihdamı:** İlgili ithalatçı ülkenin değerlendirme yapılan yılda tarım sektöründe çalıştırdığı işgücü sayısıdır. 2012-2018 yıllarının her biri için ilgili alıcı ülkenin (OECD ülkesinin) nüfusu 1000 kişi olarak ikincil veri kaynağı olan FAO'nun web sayfasından temin edilmiştir¹⁶.
12. **Alıcı ülkenin ilgili yıl para biriminin \$ (USD) cinsinden değeri:** Alıcı ülkenin USD karşısındaki değeri söz konusu ülkenin alım gücünü etkileyen önemli bir değişken olduğu için söz konusu değişken alıcı (OECD) ülkelerin 2012-2018 yıllarının her biri için para birimlerinin USD karşılığı değeri ikincil veri kaynağı FAO'nun web sayfasından temin edilmiştir¹⁷.

Hedef Özniteliği (Çıktı değişkeni)

İthalatçı ülkenin ilgili üründen bir yıl sonraki ithalat değeri (\$): Bu değişken hedef ülkenin ithalat kapasitesinin belirlenmesi açısından önemlidir. İthalatçı ülkenin 2013-

¹² <https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD/CHN> (erişim tarihi: 19.10.2019).

¹³ <https://www.doingbusiness.org/en/custom-query> (erişim tarihi: 19.10.2019).

¹⁴ <https://iccwbo.org/publication/icc-open-markets-index-2017/> (erişim tarihi: 19.10.2019).

¹⁵ <http://www.fao.org/statistics/en/> (erişim tarihi: 19.10.2019).

¹⁶ <http://www.fao.org/statistics/en/> (erişim tarihi: 22.10.2019).

¹⁷ <http://www.fao.org/statistics/en/>, erişim tarihi: 22.10.2019.

2019 yıllarının her biri için ilgili üründen gerçekleştirdiği ithalatının dolar olarak değerini ifade etmektedir. Söz konusu değişken tahmin edilecek çıktı niteliğindeki değişkendir. Söz konusu değişken ikincil veri kaynağı UN COMTRADE web sayfasından temin edilmiştir.

4.2.2. Çalışmanın Kısıtları ve Risklerinin Belirlenmesi

Veri setinin oluşturulması sürecinde karşılaşılan kısıtlar aşağıda sunulmaktadır:

1. Yaş meyve sebze ürünleri GTİP'lerinde 2012 yılında revizyon gerçekleştirilmiş ve bazı ürünlerin sadece GTİP kodu değişmiş iken bazı ürünlere ise ayrı GTİP'ler oluşturulmuştur. Bu nedenle, veri seti oluşturulurken 2012-2019 yılları arasında OECD ülkelerinin yaş meyve sebze ürünleri açısından, ürün bazına verileri derlenerek oluşturulmuş, 2012 yılı öncesi değerlendirilmeye alınamamıştır. 35 OECD üyesi ülke ve 66 GTİP bazında veri seti 14.499 satırdan oluşmuştur.
2. FAO'dan alınan verilerde en güncel değerler 2018 yılına aittir. Bu nedenle, her period yılındaki girdi değişkenlerine karşılık bir sonraki yıla ait ithalat değerleri çıktı değeri olarak esas alınmıştır. "İthalatçı ülkenin ilgili üründen bir yıl sonraki ithalatı Değer (\$)" kriteri tahmini yapılacak çıktı konumundadır. Böylece ilgili yıl verileri kullanılarak bir sonraki yılın ithalat verileri tahmin edilmeye çalışılmıştır.
3. Çıktı olan, ithalat değerinde "0" olan değerler olduğu için Ortalama Mutlak Yüzde Hata (MAPE) değerleri hesaplanamamaktadır. Bu nedenle, MAE değerleri veri setindeki çıktının ortalama ve standart sapmaları ile birlikte yorulanmaktadır.

4.2.3. Çalışmada Kullanılacak Yazılımın ve Donanımın Belirlenmesi

Çalışmadaki tüm uygulamalar Python dili ile yazılmış olup bütün algoritmalar sekiz çekirdekli işlemciye sahip onaltı GB RAM'i olan bir bilgisayarda çalışılmıştır.

Python programlama dili, Hollandalı bir programcı olan Guido Van Rossum tarafından 1990'lı yılların başında oluşturulmuştur. Python, yorumlayıcı, interaktif, nesnel (object-oriented) bir programlama dilidir. Modülleri, istisnaları, dinamik yazımı, çok yüksek düzeyde veri türlerini ve sınıfları içerir. Python çok net söz dizimi ile dikkat çekici gücünü birleştirir. Çeşitli windows sisteminin yanı sıra pek çok sistem çağırısı ve kütüphaneyle ara yüzü vardır ve C veya C++'da da genişletilebilir. Unix varyasyonları ile Mac ve Windows 2000'de de çalışabilmektedir.¹⁸

¹⁸What is Python?, <https://docs.python.org/3/faq/general.html#what-is-python> (erişim tarihi: 04.08.2019).

Python, daha hızlı ve entegre olarak çalışabilen, öğrenmesi kolay ve güçlü bir programlama dilidir. Python programının net söz dizimi, dinamik yazımı ve yorumlanabilme özelliği, bu programı komut oluşturma ve hızlı uygulama geliştirme işlemleri için ideal bir programlama dili olarak değerlendirilmesini sağlamaktadır.¹⁹

Python; C ve C++ gibi programlama dillerine kıyasla aşağıdaki üstünlüklere sahiptir:

- Daha kolay öğrenilebilmektedir.
- Daha basit ve daha okunaklı bir söz dizimine sahiptir.
- Net ve basit sözdizimi sayesinde program geliştirme süresi kısalmış ve hızlanır.

Python'un bu özellikleri dolayısıyla dünya çapındaki büyük kuruluşlar (Google, Yahoo ve Dropbox vb) Python dilini bilen programcıları da bünyelerinde bulundurmaya tercih etmektedirler.²⁰

4.3. Verinin Hazırlanması

4.3.1. Verinin Toplanması, Değerlendirilmesi ve Bütünleştirilmesi

Veri ön analizi aşamasında, her değişkene ait veri farklı kaynaklardan veya aynı veri kaynağının farklı bölümlerinden temin edildiği için öncelikle veri toplama ve veri birleştirme adımlarından geçirilmiştir. Söz konusu değişkenlerin işlenmesi neticesinde 14.499 satır ve 13 öznitelikli bir veri seti oluşmuştur. Bir başka ifadeyle, veri seti 14.499 satır ve 13 sütundan (12 girdi, bir çıktı (ithalat değeri)) oluşmaktadır. Veri setindeki özniteliklere ait tanımlayıcı istatistikler ise Tablo 4.4'de yer almaktadır.

OECD ülkelerine ait verilerde hem girdi değişkenleri hem de çıktı değişkenleri açısından herhangi bir kayıp veri bulunmamaktadır.

¹⁹The Python Tutorial, <https://docs.python.org/3/tutorial/index.html#the-python-tutorial> (erişim tarihi: 04.08.2019).

²⁰ What is Python?, <https://docs.python.org/3/faq/general.html#what-is-python> (erişim tarihi: 04.08.2019).

Tablo 4.4 Veri Setindeki Değişkenlere Ait Tanımlayıcı İstatistikler

	Dönem	Ülke kodu	Ürün kodu	Alıcı ülkenin ilgili yıl enflasyon oranı	İlgili ülkenin nüfusu (1.000 kişi)	İlgili ülkenin GSYİH (milyar \$)	İlgili ülkenin 5 yıl sonraki tahmini GSYİH (milyar \$)	İlgili ülkede iş yapma kolaylığı endeksi	İlgili ülkenin açık pazar endeksi	Alıcı ülkenin ilgili yıldaki söz konusu üründen üretimi	Alıcı ülkenin ilgili yıl tarımda işgücü istihdamı	Alıcı ülkenin ilgili yıl para biriminin \$ cinsinden değeri	İthalatçı ülkenin ilgili üründen bir yıl sonraki ithalat değeri (\$)
Gözlem sayısı	14.499	14.499	14.499	14.499	14.499	14.499	14.499	14.499	14.499	14.499	14.499	14.499	14.499
Ortalama	-	-	-	1,351	34.915,90	1.417,08	1.614,502	76,54	4,09	230.380,40	613,40	58,98	42.411.690
Standart sapma	-	-	-	1,291	60.224,17	3.204,92	3.878,319	5,42	0,42	1.039.973	1.192,80	199,98	140.450.800
Minimum değer	2012	-	-	-1,736	325.642	14,72	23,918	60,50	2,80	0,00	2,00	0,60	0,00
%25	-	-	-	0,346	5.296,32	186,98	214,012	72,24	3,80	0,00	73,00	0,77	916.620
%50	-	-	-	1,301	10.325,54	417,72	456,166	76,62	4,10	1.762	166,76	0,90	5.717.191
%75	-	-	-	2,126	46.671,91	1.379,10	1.508,48	80,54	4,40	67.943	698,00	6,60	28.026.070
Maksimum değer	2018	-	-	6,041	327.096,26	20.580,25	24.881,41	89,29	5,00	21.090.680	6.826,48	1.160,43	2.864.125.000

4.3.2. Veri Dönüştürme

Veri setindeki özniteliklerde “Ürün kodu (GTİP)” ile “Ülke kodu” öznitelikleri kategorik yapıdadır. Makine öğrenmesi algoritmalarında herhangi bir veri dönüşümü yapılmaz ise kullanılan gözlemleri sürekli bir değer gibi düşünecek ve dolayısıyla büyük değer daha önemlidir yaklaşımı ile veriyi modelleyecektir. Bu yaklaşım kategorik gözlemlerde hatalı yorumlara yol açabilmektedir. Veri setinde yer alan Ürün Kodu ve Ülke Kodu değişkenlerinin birbirlerine göre herhangi bir üstünlükleri yoktur. İlgili özniteliklerde sadece “Var” veya “Yok” bilgisi ile ilgili ürün veya ülke bilgisi modele doğru bir şekilde aktarılmalıdır. Bu dönüşüm ikili değişkenler yardımıyla yapılmıştır.

İkili dönüşümde, her bir kategorik değişkende yer alan farklı gözlem sayısı kadar yapay öznitelik oluşturulur ve bu bilgi 1 veya 0 şeklinde ifade edilir. Ülke ve ürün kodları için dönüşüm süreci örneği ve bazı örnek değerler Tablo 4.5 ve Tablo 4.6’da yer almaktadır.

Tablo 4.5 Veri Setinde Ülke Kodlarında Gerçekleştirilen Dönüşüm Örneği

Gözlem No	Eski Veri	Yeni Veri						
	Ülke kodu	Ülke kodu_36	Ülke kodu_40	Ülke kodu_56	Ülke kodu_124	Ülke kodu_152	...	Ülke kodu_842
1	36	1	0	0	0	0		0
2	40	0	1	0	0	0		0
3	56	0	0	1	0	0		0
4	124	0	0	0	1	0		0
5	152	0	0	0	0	1		0
...	...	...	...	...	...	...	...	...
14499	842	0	0	0	0	0	...	1

Tablo 4.6 Veri Setinde Ürün Kodlarında Gerçekleştirilen Dönüşüm Örneği

Gözlem No	Eski Veri	Yeni Veri						
	GTİP	Ürün kodu _070110	Ürün kodu _070190	...	Ürün kodu _070511	...	...	Ürün kodu _080520
1	070110	1	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...
33	070190	0	1	0	0	0	0	0
...	...	...	...	...	...	...	...	...
310	070511	0	0	0	1	0	0	0
...	...	...	...	...	...	...	...	...
14499	080520	0	0	0	0	0	0	1

Dönüşümden sonra 66 adet Ürün Kodu ve 35 adet ülke değişken sayısı dahil toplam 101 yeni öznitelik elde edilmiştir. Diğer öznitelikler ile birleştirildiğinde 111 girdi ve 1 çıktı olmak üzere toplam 112 öznitelik ve 14.499 gözlemden oluşan bir veri seti elde edilmiştir. Oluşan yeni veri setine ilişkin girdi-çıkıtı öznitelikleri Tablo 4.7’de ve örnek görsel Tablo 4.8 yer almaktadır.

Tablo 4.7 Veri Seti Öznitelikleri

Girdiler	1. Dönem 2. Ülke kodu_36 3. Ülke kodu_40 4. Ülke kodu_56 5. 36. Ülke kodu_842 37. Ürün kodu_070110 38. Ürün kodu_070190 39. 102.Ürün Kodu_081090 103.Alıcı ülkenin ilgili yıl enflasyon oranı 104.İlgili ülkenin nüfusu (1.000 kişi) 105.İlgili ülkenin Gayri Safi Yurt İçi Hasılası(milyar \$) 106.İlgili ülkenin 5 yıl sonraki tahmini Gayri Safi Yurt İçi Hasılası (milyar \$) 107.İlgili ülkede iş yapma kolaylığı endeksi 108.İlgili ülkenin açık pazar endeksi 109.Alıcı ülkenin ilgili yıldaki söz konusu üründen üretimi 110.Alıcı ülkenin ilgili yıl tarımda işgücü istihdamı 111.Alıcı ülkenin ilgili yıl para biriminin \$ cinsinden değer
Çıktı	İthalatçı ülkenin ilgili üründen bir yıl sonraki ithalatı değeri (\$)

Tablo 4.8 Veri Setinde Ülke ve Ürün Kodlarında Gerçekleştirilen Dönüşüm Sonrası Veri Setinin Durumuna ilişkin Örnek

Dönem	ithalatçı ülkenin ilgili üründen bir yıl sonraki ithalatı Değer (\$)	Ülke Kodu _36	Ülke Kodu _40	Ülke Kodu _56	Ülke Kodu _124	...	Ürün kodu_ 070110	...	Alıcı ülkenin ilgili yıl enflasyon oranı	İlgili ülkenin nüfusu (1.000 kişi)	İlgili Ülkeni n GSYİ H (milya r \$)	İlgili Ülkenin 5 yıl sonraki tahmini GSYİH (milyar \$)	İlgili Ülkede İş yapma Kolaylığı Endeksi	İlgili Ülkenin Açık Pazar Endeksi	Alıcı ülkenin ilgili yıldaki söz konusu üründen üretimi (Ton)	Alıcı ülkenin ilgili yıl tarımda işgücü ist.	Alıcı ülkeni n ilgili yıl para birimi nin \$ cinsind en değer
2012	2.470	1	0	0	0		1		1,763	22904	1.569	1.387	81	3,8	1.288.186	317	0,97
2012	5.078.410	0	1	0	0		1		2,486	8502	410	418	76	4,2	665.416	191	0,78
2012	68.116.226	0	0	1	0		1		2,840	11085	498	496	72	4,7	2.929.800	53	0,78
2012	3.509.833	0	0	0	1		1		1,516	34922	1.828	1.650	81	3,8	4.579.136	306	1,00
....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

XGBoost ile Rastgele Orman Algoritmaları normalizasyona ihtiyaç duymamakla birlikte, YSA’larda normalizasyon işlemi önemlidir. Değişkenlerin birimlerinin birbirinden oldukça farklı (heterojen yapıda) olması nedeniyle sadece YSA’da kullanılmak üzere girdi özniteliklerine normalizasyon uygulanmıştır. Normalizasyon işlemleri açısından literatürde en sık başvurulan yöntem ise standardizasyondur. Standardizasyon formülü ise,

$$z = \frac{x - \mu}{\sigma} \quad (4.5)$$

şeklindedir.

Formülasyon (4.5)’te;

z: normalize edilmiş değeri,

x: girdi değerini,

μ : girdi değişkeninin ortalamasını,

σ : girdi değişkeni setinin standart sapmasını

ifade etmektedir.

4.3.3. Veri Setinin Eğitim ve Test Verisi olarak Ayrılması

Veri setinde 2012-2016 dönemine ilişkin veriler eğitim; 2017 ve 2018 yılına ilişkin veriler ise test veri seti olarak ayrılmıştır. Bu durumda eğitim veri seti 10.428 adet gözlemden; test veri seti ise 4.071 adet gözlemden oluşmuştur. Dolayısıyla, veri setinin %72’si eğitim, %28’i test verisi olarak kullanılmıştır.

4.4. Model Oluşturma

Makine öğrenmesi modelinin kararlılığının doğrulanmasına her zaman ihtiyaç bulunmaktadır. Modelin verideki örüntüleri doğru bir şekilde kapsamasını ve verideki değişimlere karşı duyarlı olmasını garanti edecek bir sistem kurulmalıdır.

Genel olarak eğitim aşamasının ardından modelin tahminleme hatası belirlenir. Bu süreçte, tahmin edilen ile gerçek çıktı arasındaki sayısal farklar, bir diğer ifade ile eğitim hatası tahmin edilir. Ancak bu değerlendirme bize sadece modelin eğitim verisi üzerinde ne kadar iyi çalıştığını gösterir. Bu değerlendirme tekniği ile ilgili sorun, henüz sunulmamış veri seti üzerinde modelin ne kadar iyi bir genelleme yapabileceğine ilişkin bir yanıt elde edemememizdir. Modelin bu konudaki yeterliliği hakkında çapraz doğrulama ile bilgi edinilebilmektedir.

4.4.1. K-Kat Çapraz Doğrulama

Modelinizi eğitmek için her zaman yeterli veri olmadığından, verinin bir kısmını doğrulama için ayırmak eksik öğrenme sorunlarına neden olmaktadır. Eğitim verisini azaltmak ise veri setindeki önemli trendler ve örüntülerin kaybedilmesi/gözden kaçırılması riski alınmış olur ki bu durum da sapmalar nedeniyle hataların oluşumunu artırır. Bu nedenle, eğitim için geniş bir veri sağlayan ve doğrulama için geniş miktarda veri bırakan bir metoda ihtiyaç bulunmaktadır. K-kat çapraz doğrulama işte bu sorunun çözümüne imkan sağlamaktadır (Raschka ve Mirjalili, 2019: 195, Müller ve Guido, 2017: 252.)

Çapraz doğrulama makine öğrenmesi modelinin performansını test etmek için kullanılan istatistiksel bir tekniktir. İyi bir çapraz sorgulama metodu ile tüm veri setinin üzerinde modelin performansı kapsamlı bir şekilde ölçülebilmektedir.

K-kat çapraz doğrulama ile veri k adet alt kümeye bölünmektedir. Sonrasında ise yukarıda belirtilen çapraz doğrulama adımları k defa tekrarlanmaktadır. Her defasında k adet alt veri setinden biri test veri seti diğerleri ise eğitim veri seti olarak kullanılmaktadır. Modelin genel etkinliğini elde etmek için k adet denemeden elde edilen tahminleme hatalarının ortalaması alınır. Söz konusu metotta her bir veri noktası bir doğrulama veri setinde bir defa ve bir eğitim kümesinde $k - 1$ defa bulunur. Bu durum, verinin çoğunluğunu uyum için kullandığımız için sapmaları azaltır ve verinin çoğunluğu doğrulama veri setinde de kullanılmış olduğu için varyansı önemli miktarda azaltır (Raschka ve Mirjalili, 2019: 195, Müller ve Guido, 2017: 252).

Genel kural olarak $k = 5$ veya $k = 10$ olarak uygulanması tercih edilir. Ancak, k 'nın değeri sabit değildir ve herhangi bir değeri de alabilir.

Analiz sürecinde 10-kat çapraz doğrulama uygulanmış olmakla birlikte, $k = 5$ için örnek alındığında veri seti bölümlendirmesi Tablo 4.9'da sunulmaktadır. Burada, veri seti beş eşit parçaya bölünür ve aşağıdaki işlemler test veri seti her defasında değiştirilerek beş defa uygulanır.

Tablo 4.9 “5-Kat” Çapraz Doğrulama Örnek Gösterim Tablosu

İterasyon 1	Test	Eğitim	Eğitim	Eğitim	Eğitim
İterasyon 2	Eğitim	Test	Eğitim	Eğitim	Eğitim
İterasyon 3	Eğitim	Eğitim	Test	Eğitim	Eğitim
İterasyon 4	Eğitim	Eğitim	Eğitim	Test	Eğitim
İterasyon 5	Eğitim	Eğitim	Eğitim	Eğitim	Test

1. Grup bir test veri ve dört ($k-1$) adet eğitim veri seti olarak ayrılır.
2. Eğitim seti üzerinde model uygulanır ve test veri seti üzerinde değerlendirilir.

3. Tahmin hatası hesaplanır ve test veri seti değiştirilip 2-3 adımları tekrarlanır.
4. Ortalama tahmin hatası modelin uyum başarısını verir.

4.4.2 Parametre Değerlerinin Belirlenmesi

Her makine öğrenmesi modeli farklı sayıda kontrol parametrelerine sahiptir ve bunların eş zamanlı olarak en iyilerinin belirlenmesi gerekmektedir. Herhangi bir makine öğrenmesi modelinin performansının doğrudan doğruya seçilen kontrol parametrelerine (hiper parametre) bağlı olması nedeniyle öncelikle doğru parametrenin seçilmesine yönelik çalışma gerçekleştirilmiştir. Modellerde öncelikle parametre optimizasyonu yapılmıştır. Parametre optimizasyonu yapmak amacıyla eğitim veri seti üzerinde 10-kat çapraz doğrulama yapılmıştır. Daha sonra bulunan en iyi parametre değerleri için modellerin genelleme gücü test veri seti ile ölçülmüştür.

Tablo 4.10 Analiz Aşamasında Kullanılan Modellerin Parametrelerinin Optimizasyonu İçin Geliştirilen Deney Tasarımı

Model	Parametre	Değerler
XGBoost	Maksimum Ağaç Derinliği (Max_depth)	5, 7, 9, 11, 13, 15
	Öğrenme Katsayısı (Learning_rate)	0,05; 0,08; 0,1; 0,15; 0,25; 0,50; 0,75
	Tahminleyici (Ağaç) Sayısı (n_estimator)	150, 200, 250, 300, 350
	Karar düğümünde kullanılacak gözlem sayısı (min_child_weight)	1, 3, 5, 7
Rastgele Orman	Tahminleyici (Ağaç) Sayısı (n_estimator)	150, 200, 250, 300, 350
	Maksimum Ağaç Derinliği (Max_depth)	5, 7, 9, 11, 13, 15
	Bir yaprak düğümünde olması gereken minimum örnek sayısı (min_sample_leaf)	3, 5, 7, 9, 11
	Bir iç yaprak düğümünü bölmek için gereken minimum örnek sayısı (min_samples_split)	5, 7, 10, 12, 15
Yapay Sinir Ağları	Gizli Katman Sayısı	2
	Gizli Katmandaki Nöron Sayısı	5, 10, 15, 20, 25
	Aktivasyon Fonksiyonu	ReLU

Analiz çalışmasında kullanılacak XGBoost, Rastgele Orman ve YSA için üzerinde çalışılacak parametre değerlerinden oluşan deney tasarımı Tablo 4.10'da sunulmaktadır. Özellikle YSA için yapılan denemelerde tek katmanlı ağ mimarilerinde sürekli eksik öğrenme sorunu ile karşılaşmıştır. Bu sebepten dolayı çalışmada iki gizli katman içeren ağ mimarisinin kullanılmasına karar verilmiştir. Ayrıca aktivasyon fonksiyonu için yapılan

denemelerde “logistic” ve “Tanh” fonksiyonları kullanıldığında işlem süresi oldukça uzamıştır. “ReLU” aktivasyon fonksiyonu ile yapılan denemelerde tespit edilen en uzun işlem süresi 1570 saniye olmasına rağmen gerek “logistic” gerekse de “tanh” ile yapılan denemeler sekiz saati aşmasına rağmen öğrenme süreci tamamlanamamıştır. Bu sebepten dolayı çalışmada sadece “ReLU” aktivasyon fonksiyonu tercih edilmiştir. Yine yapılan ilk denemeler sonucunda modellerin başarı düzeyine katkı sunmayan “dönem” değişkeni uygulamada kullanılmamıştır.

Bu durum neticesinde 110 girdi ve 1 çıktı olmak üzere toplam 111 öznitelik ve 14.499 gözlem üzerinden analizler gerçekleştirilmiştir.

4.5. Modellerin Değerlendirilmesi

4.5.1. XGBoost Algoritması ile Modelleme

Belirlenmeye çalışılan parametreler için,

- Her ağacın bir önceki ağacın hatalarını düzeltme derecesini kontrol eden öğrenme oranı parametresi için: 0,05; 0,08; 0,1; 0,15; 0,25 0,50; 0,75
- Maksimum ağaç derinliği parametresi için: 5, 7, 9, 11, 13, 15
- Değerlendirmeye alınacak ağaç sayısı parametresi için: 150, 200, 250, 300, 350
- Her bir karar düğümünde kullanılacak minimum gözlem sayısı değişkeni için: 1, 3, 5, 7

değerleri analiz edilmiştir.

Söz konusu değerler göz önünde bulundurulduğunda oluşabilecek 840 farklı kombinasyonun 10-kat çapraz doğrulama ile test edilebilmesi için toplam 8400 deney yapılmıştır. Elde edilen en iyi parametre değerleri ise;

- Her ağacın bir önceki ağacın hatalarını düzeltme derecesini kontrol eden öğrenme oranı parametresi için: 0,5
 - Maksimum ağaç derinliği parametresi için: 7
 - Değerlendirmeye alınacak ağaç sayısı parametresi için: 250
 - Her bir karar düğümünde kullanılacak minimum gözlem sayısı değişkeni için: 1
- şeklindedir.

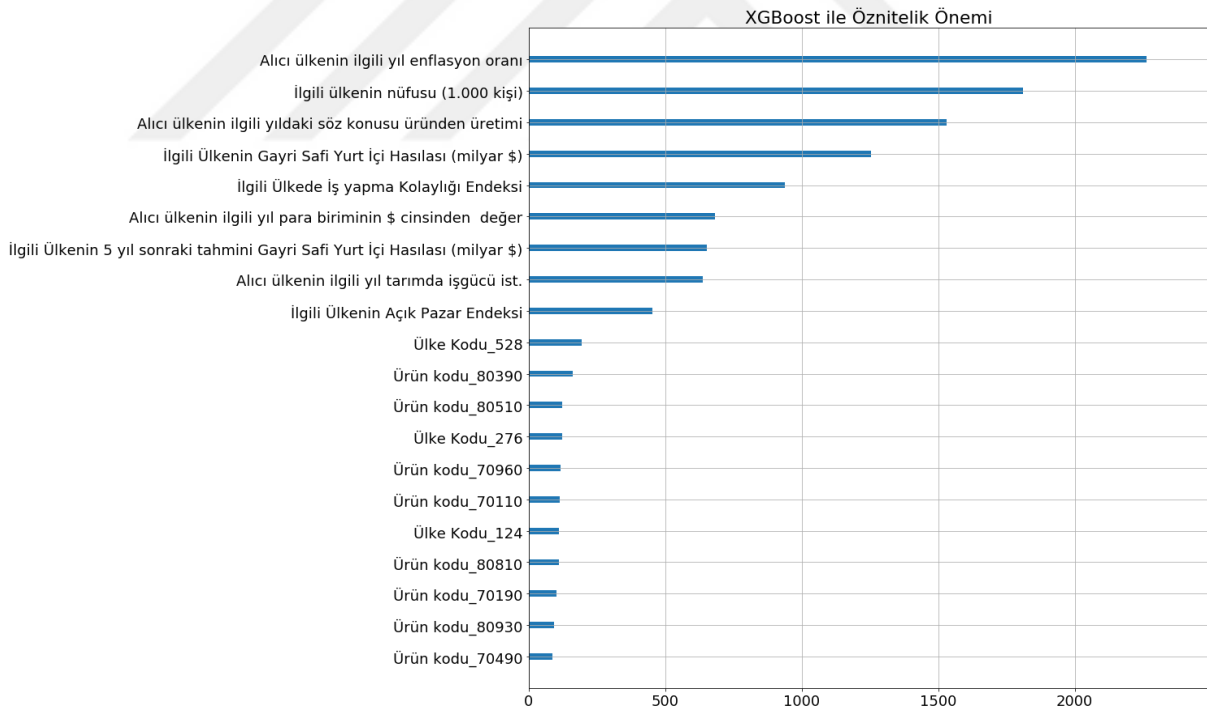
Elde edilen en iyi parametreler ile XGBoost algoritmasının performansına yönelik olarak elde edilen sonuçlar Tablo 4.11’de sunulmaktadır.

Tablo 4.11 XGBoost Algoritması Performans Değerleri

	R^2	$RMSE$	MAE
Eğitim Veri Seti	0,99	920.046,74	1.371.282,62
Test Veri Seti	0,89	52.375.909,39	15.700.976,44
Süre (sn.)	7,09		

Ağaç artırmada ölçeklendirilebilir bir makine öğrenme algoritması olan XGBoost (Extreme Gradient Boosting) özniteliklerin önem düzeyleri değerlendirebilmektedir. XGBoost Algoritması, modelin performansını artırmak ve orijinal verinin boyutunu azaltmak amacıyla öznitelikleri önemlerine göre filtreleyebilmektedir. Bir özneliğin önemi bir özelliğin rastgele bir gürültü verisi ile değiştirilmesi durumunda tahminleme performansında önemli bir değişiklik meydana getirip getirmeyeceğine bağlıdır. “weight” kriterine göre önem düzeyi bir özneliğin ağaçta kaç defa yer aldığını ifade etmektedir (Ji ve Lin, 2019: 148; Zheng vd., 2017: 5).

XGBoost Algoritması kullanılarak bir özneliğin ağaçta yer alma sayısı olan “weight” kriterine göre önem düzeyleri Şekil 4.2’de gösterilmiştir.

**Şekil 4.2 “Weight” Kriterine Göre XGBoost Algoritmasında Özniteliklerin Önem Düzeyleri**

Şekil 4.2’nin incelenmesinden, ithalat değerini etkileyen en önemli değişkenin “Alıcı ülkenin ilgili yıl enflasyon oranı” olduğu anlaşılmaktadır. Daha sonraki önemli değişkenler ise sırasıyla “İlgili ülkenin nüfusu”, “Alıcı ülkenin ilgili yılda söz konusu üründen üretimi”,

“İlgili ülkenin GSYİH”si, “İlgili ülkede iş yapma kolaylığı endeksi” ve “İlgili ülkenin 5 yıl sonraki tahmini GSYİH” olmuştur.

Çalışmanın özniteliklerin belirlenmesi aşamasında ithalatı değerini etkilediği öngörülen özniteliklerin XGBoost tarafından da önemli bulunması, söz konusu özniteliklerinin seçiminin ne denli isabetli olduğunun da bir göstergesidir. Bu özniteliklerin XGBoost tarafından da önemli bulunmasının nedenleri şu şekilde açıklanabilmektedir:

- Ülkedeki enflasyon oranı doğrudan o ülkede oluşan fiyat artışlarının yansıması olduğu için döviz kuru ve GSYİH değerleri ile birlikte alım gücünü etkilemekte ve dolayısıyla tüketimi etkilemektedir.
- Ülkedeki iş yapma kolaylığı ilgili ülke ile ticaret yapabilirliğin bir göstergesi olması nedeniyle söz konusu endeks değeri yüksek olan ülkelerin daha fazla ithalatı ve dış ticaret hacmi olması beklenen bir durumdur.
- Ülkedeki nüfusun artması tüketim ihtiyacını ve ithalatı artırırken, üretim miktarındaki artış ise tüketimin iç piyasadan sağlanabilmesi dolayısıyla ithalat miktarını düşürmektedir.
- Ülke ekonomisinin taahhütlerini ne kadarını yerine getirebildiğini ifade eden açık pazar endeksi ülkenin güvenilirliği ile ilgili olup, yüksek olmasının ticaret hacmine olumlu etkisi bulunmaktadır.
- Ülkeler açısından önemli öznitelikler değerlendirilecek olursa, “528” (Hollanda), “276”(Almanya) ve “124”(Kanada) ülkeleri modelin veri setini açıklama gücüne daha fazla katkı sağladığı şeklinde yorulanabilmektedir. OECD ülkeleri içinde en fazla yaş meyve sebze ithalatı ABD, Almanya, Hollanda, Birleşik Krallık, Kanada ve Fransa tarafından gerçekleştirilmektedir. 2013-2019 yılları ithalat verilerinden oluşturulan veri setinde ise Hollanda ile Kanada en fazla gözlem sayısına sahip ülkelerdendir. İthalat değerini etkileyen önemli öznitelikler arasında söz konusu ülkelerin önemli bulunmasında bu durumların da etkisi olduğu düşünülmektedir.
- Ürünler açısından ise “080390-Muz”, “080510-Portakal”, “070960-Biber”, “070110-Patates tohumluk”, “080810-Elma”, “070190-Patates”, “080930-Şeftali-nektarin” ve “070490-Lahana (kırmızı, beyaz, alabaşlar)”ın diğer ürünlere kıyasla modele daha fazla katkı sağladığı anlaşılmaktadır. OECD ülkelerinin 2019 yılında en fazla ithal ettiği ilk beş yaş meyve sebze ürünü sırasıyla muz (plantainler hariç), domates, avokado, taze üzüm ve biberdir. Yedi yıllık döneme ait veri seti genelinde ise en fazla ithalat değerine sahip

olan ürünler muz, domates, taze üzüm, incir ve biberdir. XGBoost'ta çıktı değerini etkileyen önemli ürün grupları olarak belirtilen özniteliklerin aynı zamanda veri setinde en fazla ticarete konu ürünler arasında olması önemli bulunmalarının nedenlerindendir.

Modelin performansını artırmak ve orijinal verinin boyutunu azaltmak amacıyla sadece XGBoost tarafından önemli bulunan öznitelikler ile analiz yapılması mümkündür. Ancak, tezin analiz çalışmasında sonuca hiç bir etkisi olmayan “dönem” özniteliği dışında hiçbir öznitelik için eleme işlemi yapılmamış, mevcut öznitelikler üzerinden analiz gerçekleştirilmesi tercih edilmiştir.

RMSE, *MAE* ve R^2 değerleri açısından algoritmaların test veri seti üzerinde gösterdiği performanlar Ek-4, Ek-5 ve Ek-6'da yer almaktadır. *RMSE* ve *MAE* kriterleri açısından minimum değerler en iyi tahminleme olarak yorumlanmaktadır. Ek-5 ve Ek-6'da yer alan tablolarda minimum *RMSE* ve *MAE* değerlerinin, ürün bazında ortalama ithalat değerine oranlanması sonucunda en düşük değeri veren ürün, algoritmaların performanslarını değer olarak sergileyebilmek açısından seçilmiştir. Bu itibarla, *RMSE* ve *MAE* sonucunun ortalama ithalata oranında en düşük değeri “080520- Mandarin, tangerin” verdiği ve söz konusu ürüne ilişkin kriterleri açısından en düşük R^2 değerleri de tatmin edici düzeyde olduğu için bu ürünle ilgili XGBoost Algoritmasının tahminleme sonuçları Tablo 4.12'de sunulmaktadır.

Mandarin ürünü için gerçek değere en yakın tahminleme 2019 yılında İsviçre'nin gerçekleştirdiği ithalat için yapılmıştır. 2019 yılında İsviçre'nin mandarin ithalatı 64.209.000 USD iken %0,12'lik bir sapma ile XGBoost Algoritması 64.132.964 USD tahmin etmiştir.

Tablo 4.12 XGBoost Algoritması ile OECD Ülkelerinin Mandarin İthalatı Tahminleri

Gözlem No	Dönem	Ülke Kodu	Ülke	Ürün kodu	Gerçek değer (USD)	Tahmin değeri (USD)	% fark
14435	2018	36	Avustralya	080520	8.927.412	909.364	-89,81%
14467	2019	36	Avustralya	080520	7.706.000	0	-100,00%
14436	2018	40	Avusturya	080520	49.588.066	56.743.880	14,43%
14468	2019	40	Avusturya	080520	51.039.000	56.328.344	10,36%
14437	2018	56	Belçika	080520	96.941.885	109.797.704	13,26%
14469	2019	56	Belçika	080520	89.734.000	106.190.580	18,34%
14438	2018	124	Kanada	080520	221.983.819	202.611.650	-8,73%
14470	2019	124	Kanada	080520	223.383.000	202.938.560	-9,15%
14439	2018	152	Şili	080520	1.810.113	1.083.749	-40,13%
14471	2019	152	Şili	080520	1.713.000	1.333.648	-22,15%
14440	2018	203	Çekya	080520	46.985.481	46.402.428	-1,24%
14472	2019	203	Çekya	080520	45.129.000	45.609.750	1,07%
14441	2018	208	Danimarka	080520	27.126.349	30.116.274	11,02%
14473	2019	208	Danimarka	080520	22.588.000	27.227.530	20,54%
14442	2018	233	Estonya	080520	9.573.491	13.579.758	41,85%
14474	2019	233	Estonya	080520	8.698.000	13.098.096	50,59%

Tablo 4.12 XGBoost Algoritması ile OECD Ülkelerinin Mandarin İthalatı Tahminleri (...devamı)

Gözlem No	Dönem	Ülke Kodu	Ülke	Ürün kodu	Gerçek değer	Tahmin değeri	% fark
14443	2018	246	Finlandiya	080520	45.980.469	43.872.812	-4,58%
14475	2019	246	Finlandiya	080520	43.036.000	39.829.090	-7,45%
14444	2018	251	Fransa	080520	427.329.000	370.943.200	-13,19%
14476	2019	251	Fransa	080520	417.071.000	398.550.430	-4,44%
14445	2018	276	Almanya	080520	439.411.424	438.207.100	-0,27%
14477	2019	276	Almanya	080520	392.490.000	426.427.780	8,65%
14446	2018	348	Macaristan	080520	13.482.455	17.953.094	33,16%
14478	2019	348	Macaristan	080520	14.970.000	22.563.814	50,73%
14447	2018	352	İzlanda	080520	2.062.297	2.559.887	24,13%
14479	2019	352	İzlanda	080520	1.802.000	2.945.772	63,47%
14448	2018	372	İrlanda	080520	31.643.189	34.430.604	8,81%
14480	2019	372	İrlanda	080520	29.949.000	42.301.290	41,24%
14481	2019	376	İsrail	080520	8.000	18.731.660	234045,75%
14449	2018	381	İtalya	080520	95.371.752	95.506.136	0,14%
14482	2019	381	İtalya	080520	89.480.000	98.873.430	10,50%
14450	2018	392	Japonya	080520	33.395.241	114.861.480	243,95%
14483	2019	392	Japonya	080520	31.576.000	29.632.920	-6,15%
14451	2018	428	Letonya	080520	12.831.300	10.886.722	-15,15%
14484	2019	428	Letonya	080520	13.742.000	10.776.316	-21,58%
14452	2018	440	Litvanya	080520	24.121.574	36.721.984	52,24%
14485	2019	440	Litvanya	080520	25.345.000	37.078.450	46,29%
14453	2018	442	Lüksemburg	080520	4.170.578	6.198.177	48,62%
14486	2019	442	Lüksemburg	080520	5.110.000	7.240.631	41,70%
13620	2019	484	Meksika	080520	3.211.000	1.898.332	-40,88%
14454	2018	484	Meksika	080520	3.801.898	473.756	-87,54%
14487	2019	484	Meksika	080520	3.211.000	1.898.332	-40,88%
14455	2018	528	Hollanda	080520	217.230.085	244.870.370	12,72%
14488	2019	528	Hollanda	080520	209.359.000	231.710.660	10,68%
14456	2018	554	Yeni Zelanda	080520	11.751.453	16.753.324	42,56%
14489	2019	554	Yeni Zelanda	080520	12.097.000	15.005.424	24,04%
14457	2018	579	Norveç	080520	49.857.128	51.913.970	4,13%
14490	2019	579	Norveç	080520	51.583.000	51.410.240	-0,33%
14458	2018	616	Polonya	080520	113.628.615	102.357.740	-9,92%
14491	2019	616	Polonya	080520	0	101.665.900	
14459	2018	620	Portekiz	080520	27.413.441	17.322.086	-36,81%
14492	2019	620	Portekiz	080520	27.312.000	21.949.160	-19,64%
14460	2018	410	Güney Kore	080520	20.000	56.182.076	280810,38%
14493	2019	410	Güney Kore	080520	452.000	55.641.996	12210,18%
14461	2018	703	Slovakya	080520	24.902.420	27.373.836	9,92%
14494	2019	703	Slovakya	080520	26.208.000	26.784.132	2,20%
14462	2018	724	İspanya	080520	30.430.142	23.517.852	-22,72%
14495	2019	724	İspanya	080520	26.548.000	21.212.154	-20,10%
14463	2018	752	İsveç	080520	53.602.317	50.080.788	-6,57%
14496	2019	752	İsveç	080520	54.273.000	51.205.756	-5,65%
14464	2018	757	İsviçre	080520	64.525.578	69.083.570	7,06%
14497	2019	757	İsviçre	080520	64.209.000	64.132.964	-0,12%
14465	2018	826	Birleşik Krallık	080520	379.305.137	385.280.770	1,58%
14498	2019	826	Birleşik Krallık	080520	372.879.000	398.014.700	6,74%
14466	2018	842	ABD	080520	555.970.191	392.743.500	-29,36%
14499	2019	842	ABD	080520	488.391.000	418.458.940	-14,32%

4.5.2. Rastgele Orman Algoritması ile Modelleme

Çalışmada incelenen ikinci makine öğrenmesi algoritması olan Rastgele Orman Algoritmasında, optimum değeri belirlenmeye çalışılan parametreler ve üzerinde çalışılan değerleri aşağıdaki gibidir:

- Maksimum ağaç derinliği parametresi için: 5, 7, 9, 11, 13, 15
 - Bir yaprak düğümünde olması gereken minimum örnek sayısı parametresi için: 3, 5, 7, 9, 11
 - Değerlendirmeye alınacak ağaç sayısı parametresi için: 150, 200, 250, 300, 350
 - Bir iç yaprak düğümünü bölmek için gereken minimum örnek sayısı parametresi için: 5, 7, 10, 12, 15
- değerleri dikkate alınmıştır.

Söz konusu değerler göz önünde bulundurulduğunda oluşabilecek 750 farklı kombinasyonun 10-kat çapraz doğrulama ile test edilebilmesi için toplam 7500 deney yapılmıştır. Elde edilen en iyi parametreler aşağıda gösterilmiştir.

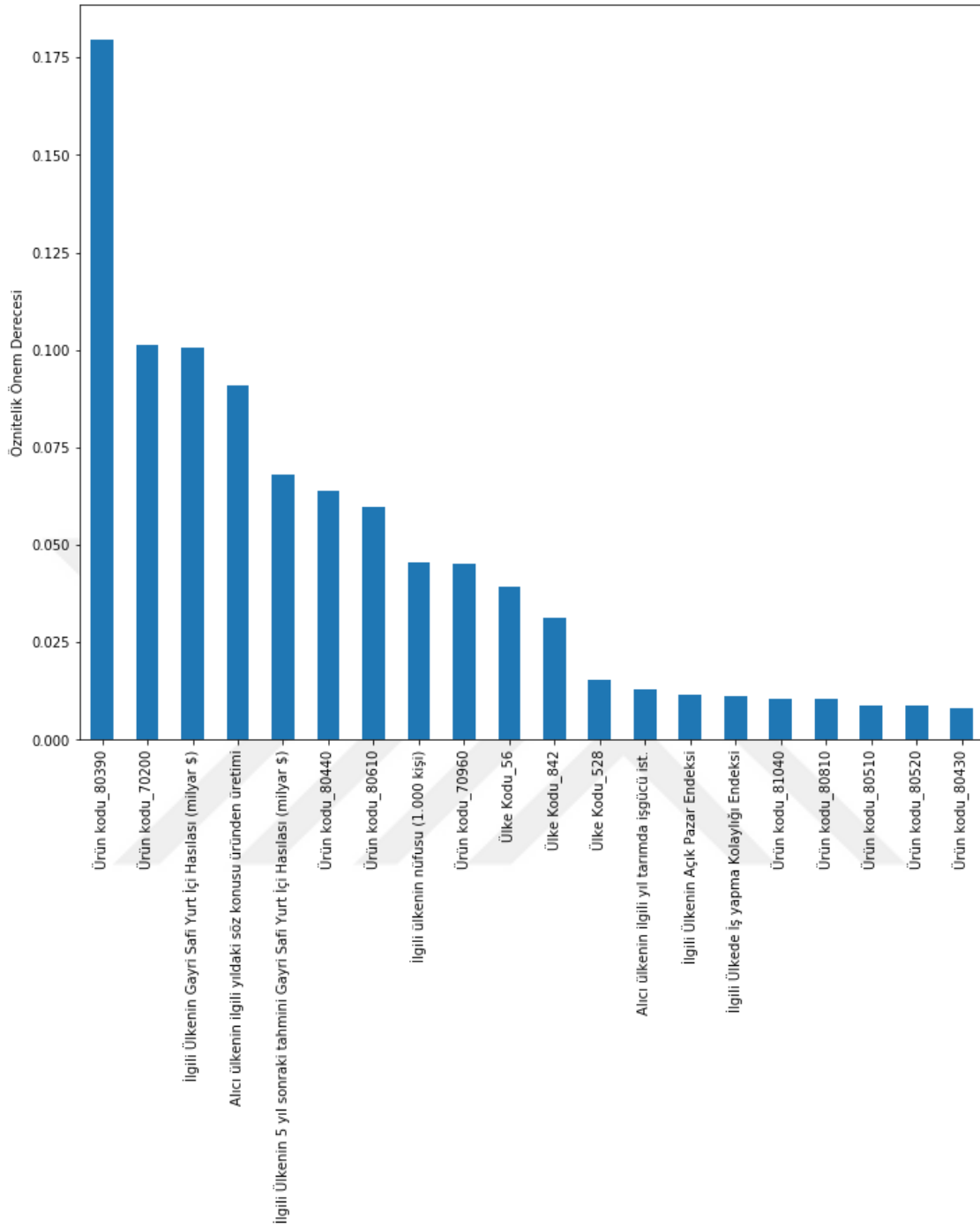
- Maksimum ağaç derinliği parametresi için: 15
- Bir yaprak düğümünde olması gereken minimum örnek sayısı parametresi için: 3
- Değerlendirmeye alınacak ağaç sayısı parametresi için: 250
- Bir iç yaprak düğümünü bölmek için gereken minimum örnek sayısı parametresi için: 5

Elde edilen en iyi parametre değerleri ile Rastgele Orman algoritmasının performansına yönelik olarak elde edilen performans değerleri Tablo 4.13’de yer almaktadır.

Tablo 4.13 Rastgele Orman Algoritması Performans Değerleri

	R^2	$RMSE$	MAE
Eğitim Veri Seti	0,92	38.547.372,05	16.274.259,76
Test Veri Seti	0,82	67.676.987,28	25.802.798,63
Süre (sn.)	10,75		

Rastgele Orman Algoritması da XGBoost algoritması gibi öznitelikleri önem düzeylerine göre sıralamaktadır. Tüm diğer öznitelikler sabitken değerlendirilen öznitelik için test verisinin değiştirilmesi durumunda tahminleme hatasındaki değişikliğin büyüklüğü, değerlendirilen öznitelikğin önemini ifade etmektedir (Liaw ve Wiener, 2002: 18). Rastgele Orman Algoritmasında özniteliklerin ithalat değerini etkileme konusundaki önem düzeyleri üzerine gerçekleştirilen incelemenin sonuçları Şekil 4.3’de sunulmaktadır.



Şekil 4.3 Rastgele Orman Algoritmasında Özniteliklerin Önem Düzeyleri

Şekil 4.3'ün incelenmesinden, ithalat değerini etkileyen en önemli özneliğin “080390 GTİP’li ürün” yani muz ürünü olduğu tespit edilmektedir. Muzu “070200 GTİP’li ürün” yani domates ürünü takip etmekte olup, diğer önemli öznelikler ise “İlgili ülkenin GSYİH”, “Alıcı ülkenin söz konusu üründen üretimi” ve “İlgili ülkenin 5 yıl sonraki GSYİH” olarak sıralanmaktadır. Hem XGBoost Algoritmasında hem de Rastgele Orman Algoritmasında da GSYİH değerleri ile üretim miktarının ithalatı etkileyen öznelikler olarak belirlenmesi bu özneliğin çalışmada kesinlikle bulunması gerektiğini ifade etmektedir.

Rastgele Orman Algoritması tarafından önemli bulunan özniteliklerin gerekçeleri şu şekilde açıklanabilmektedir:

- Ülkeler açısından “56” (Belçika), “842” (ABD) ve “528” (Hollanda) ülkelerinin modelin veri setini açıklama gücüne daha fazla katkı sağlamaktadır. Bölüm 4.5.1’de belirtildiği üzere OECD ülkeleri içinde en fazla yaş meyve sebze ithalatı gerçekleştiren ülkeler sıralamasında ABD birinci, Hollanda üçüncü ve Belçika yedinci sırada yer almaktadır. Ancak, oluşturulan veri setinde Belçika , Hollanda ve ABD en fazla gözlem sayısına sahip ülkelerdendir. Hollanda, XGBoost algoritması tarafından da önemli bir öznitelik olarak belirlenmektedir. İthalat değerini etkileyen önemli öznitelikler arasında söz konusu ülkelerin önemli bulunmasında bu durumların da etkisi olduğu düşünülmektedir.
- Ürünler açısından ise “080390-Muz”, “070200-Domates”, “080440-Avokado”, “080610-Taze üzüm.”, “070960-Biber”, “081040-Yaban mersini”, “080810-Elma”, “080510-Portakal”, “080520-Mandarin, tangerin” ve “080430-Ananas” diğer ürünlere kıyasla modele daha fazla katkı sağladığı anlaşılmaktadır. OECD ülkelerinin 2019 yılında yaş meyve sebze ithalatında muz (plantainler hariç) birinci, domates ikinci, avokado üçüncü, taze üzüm dördüncü ve biber beşinci sıradadır. Söz konusu ürünlerin en fazla ithal edilen ürünler olması dolayısıyla Rastgele Orman Algoritması tarafından da çıktı değerini etkileyen önemli ürün öznitelikleri olarak belirtilmektedir.

Rastgele Orman Algoritmasının önemli bulduğu öznitelikleri ile modelin çalıştırılması mümkün olmakla birlikte, öznitelik elemesi yapılmamış mevcut özniteliklerle sürece devam edilmiştir.

Tezin analiz sürecinde her bir algoritmanın sonuçlarının bir örnek ürün üzerinden kıyaslanabilmesi için Bölüm 4.5.1’de “080520- Mandarin, tangerin” ürünü seçilmişti. Söz konusu ürün için Rastgele Orman Algoritmasının tahminleme sonuçları Tablo 4.14’de sunulmaktadır.

Tablo 4.14’de görüleceği üzere, Rastgele Orman Algoritması mandarin ürünü için gerçek değere en yakın tahminlemeyi 2018 yılında İrlanda’nın gerçekleştirdiği ithalat için yapmıştır. 2018 yılında İrlanda’nın mandarin ithalatı 31.643.189 USD iken %0,14’lik bir sapma ile Rastgele Orman Algoritması 31.687.189 USD olarak tahmin etmiştir.

Tablo 4.14 Rastgele Orman Algoritması ile OECD Ülkelerinin Mandarin İthalatı Tahminleri

Gözlem No	Dönem	Ülke Kodu	Ülke	Ürün kodu	Gerçek Değer (USD)	Tahmin Değeri (USD)	% Fark
14435	2018	36	Avustralya	080520	8.927.412	13.234.561	48,25%
14467	2019	36	Avustralya	080520	7.706.000	13.150.804	70,66%
14436	2018	40	Avusturya	080520	49.588.066	35.254.619	-28,91%
14468	2019	40	Avusturya	080520	51.039.000	35.336.558	-30,77%
14437	2018	56	Belçika	080520	96.941.885	71.835.713	-25,90%
14469	2019	56	Belçika	080520	89.734.000	71.835.713	-19,95%
14438	2018	124	Kanada	080520	221.983.819	206.194.492	-7,11%
14470	2019	124	Kanada	080520	223.383.000	210.176.699	-5,91%
14439	2018	152	Şili	080520	1.810.113	3.619.833	99,98%
14471	2019	152	Şili	080520	1.713.000	3.640.646	112,53%
14440	2018	203	Çekya	080520	46.985.481	37.566.021	-20,05%
14472	2019	203	Çekya	080520	45.129.000	36.497.171	-19,13%
14441	2018	208	Danimarka	080520	27.126.349	32.669.226	20,43%
14473	2019	208	Danimarka	080520	22.588.000	32.841.396	45,39%
14442	2018	233	Estonya	080520	9.573.491	7.989.129	-16,55%
14474	2019	233	Estonya	080520	8.698.000	8.970.384	3,13%
14443	2018	246	Finlandiya	080520	45.980.469	36.864.711	-19,83%
14475	2019	246	Finlandiya	080520	43.036.000	36.483.890	-15,22%
14444	2018	251	Fransa	080520	427.329.000	285.803.007	-33,12%
14476	2019	251	Fransa	080520	417.071.000	286.387.145	-31,33%
14445	2018	276	Almanya	080520	439.411.424	424.661.258	-3,36%
14477	2019	276	Almanya	080520	392.490.000	406.795.410	3,64%
14446	2018	348	Macaristan	080520	13.482.455	18.654.443	38,36%
14478	2019	348	Macaristan	080520	14.970.000	19.111.717	27,67%
14447	2018	352	İzlanda	080520	2.062.297	3.869.402	87,63%
14479	2019	352	İzlanda	080520	1.802.000	3.925.618	117,85%
14448	2018	372	İrlanda	080520	31.643.189	31.687.189	0,14%
14480	2019	372	İrlanda	080520	29.949.000	36.005.748	20,22%
14481	2019	376	İsrail	080520	8.000	35.126.183	438.977,28%
14449	2018	381	İtalya	080520	95.371.752	59.296.072	-37,83%
14482	2019	381	İtalya	080520	89.480.000	60.240.762	-32,68%
14450	2018	392	Japonya	080520	33.395.241	44.707.717	33,87%
14483	2019	392	Japonya	080520	31.576.000	43.040.288	36,31%
14451	2018	428	Letonya	080520	12.831.300	12.102.196	-5,68%
14484	2019	428	Letonya	080520	13.742.000	12.157.344	-11,53%
14452	2018	440	Litvanya	080520	24.121.574	32.255.523	33,72%
14485	2019	440	Litvanya	080520	25.345.000	31.973.989	26,16%
14453	2018	442	Lüksemburg	080520	4.170.578	5.519.456	32,34%
14486	2019	442	Lüksemburg	080520	5.110.000	5.501.278	7,66%
13620	2019	484	Meksika	080520	3.211.000	12.010.577	274,04%
14454	2018	484	Meksika	080520	3.801.898	12.025.252	216,30%
14487	2019	484	Meksika	080520	3.211.000	12.010.577	274,04%
14455	2018	528	Hollanda	080520	217.230.085	176.772.164	-18,62%
14488	2019	528	Hollanda	080520	209.359.000	176.985.408	-15,46%
14456	2018	554	Yeni Zelanda	080520	11.751.453	17.141.184	45,86%
14489	2019	554	Yeni Zelanda	080520	12.097.000	20.576.778	70,10%
14457	2018	579	Norveç	080520	49.857.128	40.962.056	-17,84%
14490	2019	579	Norveç	080520	51.583.000	41.227.501	-20,08%
14458	2018	616	Polonya	080520	113.628.615	63.141.173	-44,43%
14491	2019	616	Polonya	080520	0	62.791.976	

Tablo 4.14 Rastgele Orman Algoritması ile OECD Ülkelerinin Mandarin İthalatı Tahminleri (...devamı)

Gözlem No	Dönem	Ülke Kodu	Ülke	Ürün kodu	Gerçek Değer (USD)	Tahmin Değeri (USD)	% Fark
14459	2018	620	Portekiz	080520	27.413.441	17.541.383	-36,01%
14492	2019	620	Portekiz	080520	27.312.000	16.746.465	-38,68%
14460	2018	410	Güney Kore	080520	20.000	38.225.570	191.027,85%
14493	2019	410	Güney Kore	080520	452.000	40.009.282	8751,61%
14461	2018	703	Slovakya	080520	24.902.420	28.003.709	12,45%
14494	2019	703	Slovakya	080520	26.208.000	29.228.989	11,53%
14462	2018	724	İspanya	080520	30.430.142	22.188.039	-27,09%
14495	2019	724	İspanya	080520	26.548.000	22.188.039	-16,42%
14463	2018	752	İsveç	080520	53.602.317	41.561.105	-22,46%
14496	2019	752	İsveç	080520	54.273.000	41.395.356	-23,73%
14464	2018	757	İsviçre	080520	64.525.578	43.016.710	-33,33%
14497	2019	757	İsviçre	080520	64.209.000	42.643.620	-33,59%
14465	2018	826	Birleşik Krallık	080520	379.305.137	388.952.872	2,54%
14498	2019	826	Birleşik Krallık	080520	372.879.000	389.768.277	4,53%
14466	2018	842	ABD	080520	555.970.191	197.678.191	-64,44%
14499	2019	842	ABD	080520	488.391.000	317.834.454	-34,92%

4.5.3. Yapay Sinir Ağları ile Modelleme

Oluşturulan veri seti üzerinde kullanılan bir diğer yöntem ise Yapay Sinir Ağlarıdır (YSA). OECD ülkelerinin meyve sebze ürünleri ithalatını tahmin ederken daha önce belirlenen 110 bağımsız değişken YSA modelinin girdi nöronlarını oluştururken tahmini yapılacak değişken olan ithalat değerleri çıktı nöronunu oluşturmaktadır.

Tang ve Chi (2005: 248) YSA'lar ile ilgili literatürdeki daha önceki çalışmalarda tek gizli katmanlı ağların kompleks sistemleri istenen doğrulukta modellemek için yeterli olduğunu belirtmişlerdir. Ne kadar çok gizli nöron varsa eğitim aşamasında o kadar fazla iterasyona ihtiyaç vardır. Ancak, ağın veriyi ezberlemesi mümkün olabileceğinden ağa gereğinden fazla gizli birim yüklenmemesi önemlidir.

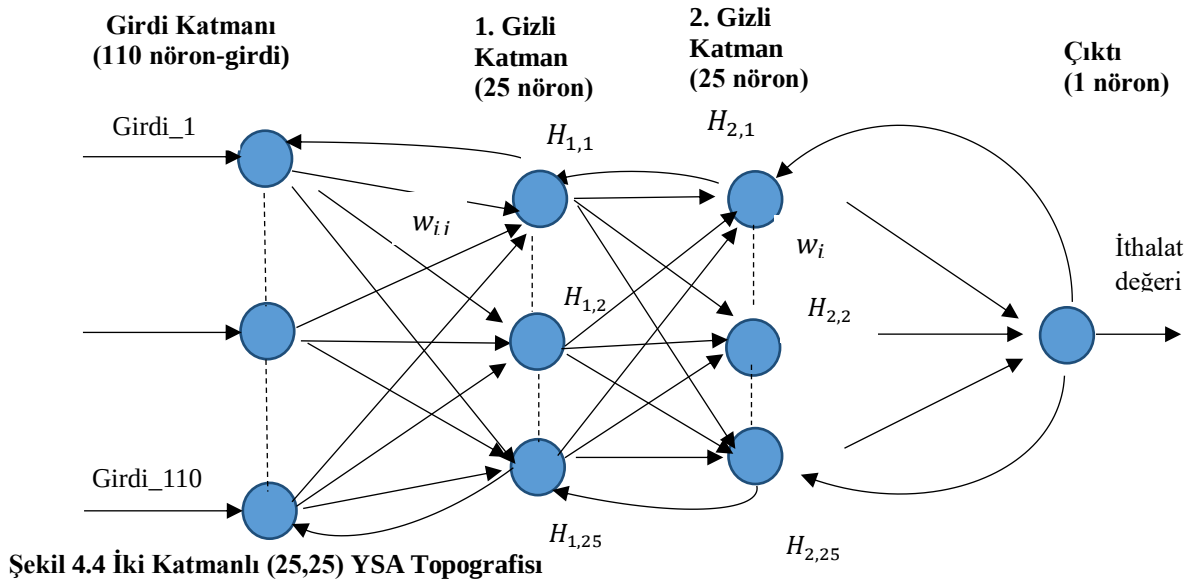
İleri beslemeli geri yayımlı ReLu Aktivasyon fonksiyonu kullanılarak YSA üzerinde çalışılmış ve normalizasyon işleminin sonrasında öncelikle tek gizli katmanlı YSA modeli üzerinde çalışılmıştır. Ancak, Python programında sekiz saatin üzerinde çalışmasına rağmen model veriyi öğrenememiştir.

Bu nedenle, gizli katman sayısı iki olarak ele alınarak modelin performansı değerlendirilmiştir. Gizli katmandaki nöron sayılarının belirlenmesi için [5, 10, 15, 20, 25] nöron sayıları sırasıyla denenmiştir. Her bir nöron sayısı için 10 farklı deneme gerçekleştirilmiş olup, denemelere ilişkin eğitim ve test veri setlerinin tahmin hataları Ek-3'te yer almaktadır. Yapılan deneyler sonucu her bir olasılık için elde edilen en iyi değerler ise Tablo 4.15'de gösterilmektedir.

Tablo 4.15 Yapay Sinir Ağları Denemelerinin Performans Değerleri

Ağ Yapısı	R^2		RMSE		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
(5, 5)	0,925	0,883	36.473.732,13	53.959.908,83	14.383.338,13	20.888.083,99	1.371,12
(5, 10)	0,922	0,884	37.214.441,89	53.933.178,54	14.716.412,36	21.343.475,22	1.220,14
(5, 15)	0,922	0,882	37.044.563,43	54.248.175,72	14.775.576,19	21.851.903,46	1.552,78
(5, 20)	0,927	0,884	35.929.843,34	53.881.827,65	14.557.593,26	21.664.902,06	1.317,56
(5, 25)	0,924	0,884	36.693.575,89	53.956.105,65	14.864.820,09	21.400.697,65	1.107,56
(10, 5)	0,939	0,897	32.880.734,80	50.736.846,50	11.916.249,26	18.689.883,40	1.307,41
(10, 10)	0,944	0,902	31.448.699,55	49.554.035,26	11.625.098,18	18.878.873,14	1.049,13
(10, 15)	0,939	0,899	32.827.722,89	50.288.586,16	11.920.920,29	18.620.067,78	967,72
(10, 20)	0,940	0,904	32.438.661,54	48.979.886,16	12.285.014,03	18.864.688,68	1.005,40
(10, 25)	0,940	0,899	32.473.195,83	50.113.623,86	11.609.383,33	18.139.023,42	942,54
(15, 5)	0,946	0,907	30.995.158,11	48.351.939,27	10.414.524,97	17.177.519,70	1.294,36
(15, 10)	0,949	0,911	29.967.918,87	47.044.806,51	10.458.626,75	17.350.498,61	999,19
(15, 15)	0,948	0,906	30.353.393,21	48.412.840,19	10.412.128,18	17.853.791,22	998,30
(15, 20)	0,949	0,906	30.156.936,10	48.354.748,63	10.495.907,48	17.583.327,51	942,28
(15, 25)	0,948	0,908	30.336.842,76	48.076.061,82	10.479.476,64	17.509.758,19	983,54
(20, 5)	0,950	0,909	29.578.672,85	47.593.581,17	9.799.251,93	17.037.151,90	1.055,91
(20, 10)	0,952	0,908	28.991.626,21	47.937.497,00	9.463.332,20	17.105.801,58	969,04
(20, 15)	0,952	0,909	29.263.863,52	47.686.480,66	9.666.152,49	17.126.390,22	956,13
(20, 20)	0,955	0,910	28.338.161,41	47.570.384,28	9.125.835,86	16.881.626,21	1.010,22
(20, 25)	0,953	0,912	28.879.994,85	47.015.626,95	9.376.638,21	16.557.976,19	832,99
(25, 5)	0,952	0,911	29.183.770,42	47.299.820,08	9.143.168,83	16.413.344,14	1.073,23
(25, 10)	0,953	0,911	28.760.218,02	47.060.100,13	8.764.828,47	16.235.689,72	1.163,14
(25, 15)	0,953	0,909	28.787.735,17	47.803.596,02	8.828.467,78	16.618.389,34	954,92
(25, 20)	0,956	0,912	27.832.919,97	46.903.288,15	8.501.121,04	16.261.904,61	1.003,89
(25, 25)	0,958	0,912	27.204.982,63	46.942.779,21	8.455.818,39	16.593.889,61	929,04

Eğitim ve test verileri üzerindeki R^2 değerinin yüksek olması, $RMSE$ ve MAE performans kriterleri açısından da göreceli olarak düşük sonuçlar vermesi nedeniyle her iki katmanda da 25 nöron sayısına sahip Yapay Sinir Ağı ile çalışılmasına karar verilmiştir. Seçilen ağ topolojisinin (her iki katmanda da 25 nöron sayısına sahip Yapay Sinir Ağı) görsel ifadesi Şekil 4.4’de yer almaktadır.



İki gizli katmanlı YSA (25,25) ile elde edilen tahminleme performansı sonuçları Tablo 4.16’de sunulmaktadır.

Tablo 4.16 Seçilen Yapay Sinir Ağının Performans Değerleri

	R^2	$RMSE$	MAE
Eğitim Veri Seti	0,958	27.204.982,63	8.455.818,39
Test Veri Seti	0,912	46.942.779,21	16.593.889,61
Süre (sn.)	929,04		

YSA’nın (25,25) “080520- Mandarin, tangerin” ürünü ithalatı açısından tahminleme sonuçları Tablo 4.17’de sunulmaktadır. Tablo 4.17’de görüleceği üzere YSA (25,25) Algoritması mandarin ürünü için gerçek değere en yakın tahminlemeyi 2019 yılında Çekya’nın gerçekleştirdiği ithalat için yapmıştır. 2019 yılında İrlanda’nın mandarin ithalatı 45.129.000 USD iken %0,11’lik bir sapma ile YSA 45.179.668 USD olarak tahmin etmiştir.

Tablo 4.17 YSA (25, 25) ile OECD Ülkelerinin Mandarin İthalatı Tahminleri

Gözlem No	Dönem	Ülke Kodu	Ülke	Ürün kodu	Gerçek Değer	Tahmin Değeri	% Fark
14435	2018	36	Avustralya	80520	8.927.412	3.563.407	-60,08%
14467	2019	36	Avustralya	80520	7.706.000	3.715.901	-51,78%
14436	2018	40	Avusturya	80520	49.588.066	68.380.346	37,90%
14468	2019	40	Avusturya	80520	51.039.000	67.718.446	32,68%
14437	2018	56	Belçika	80520	96.941.885	143.843.784	48,38%
14469	2019	56	Belçika	80520	89.734.000	141.554.720	57,75%
14438	2018	124	Kanada	80520	221.983.819	200.000.025	-9,90%
14470	2019	124	Kanada	80520	223.383.000	215.147.158	-3,69%
14439	2018	152	Şili	80520	1.810.113	676.930	-62,60%
14471	2019	152	Şili	80520	1.713.000	676.930	-60,48%
14440	2018	203	Çekya	80520	46.985.481	47.364.710	0,81%
14472	2019	203	Çekya	80520	45.129.000	45.179.668	0,11%
14441	2018	208	Danimarka	80520	27.126.349	42.680.398	57,34%
14473	2019	208	Danimarka	80520	22.588.000	40.169.721	77,84%
14442	2018	233	Estonya	80520	9.573.491	1.582.348	-83,47%
14474	2019	233	Estonya	80520	8.698.000	1.562.676	-82,03%
14443	2018	246	Finlandiya	80520	45.980.469	24.431.012	-46,87%
14475	2019	246	Finlandiya	80520	43.036.000	24.479.480	-43,12%
14444	2018	251	Fransa	80520	427.329.000	438.134.363	2,53%
14476	2019	251	Fransa	80520	417.071.000	457.113.100	9,60%
14445	2018	276	Almanya	80520	439.411.424	441.024.790	0,37%
14477	2019	276	Almanya	80520	392.490.000	448.285.572	14,22%
14446	2018	348	Macaristan	80520	13.482.455	5.986.569	-55,60%
14478	2019	348	Macaristan	80520	14.970.000	5.322.911	-64,44%
14447	2018	352	İzlanda	80520	2.062.297	676.930	-67,18%
14479	2019	352	İzlanda	80520	1.802.000	676.930	-62,43%
14448	2018	372	İrlanda	80520	31.643.189	22.387.695	-29,25%
14480	2019	372	İrlanda	80520	29.949.000	22.405.835	-25,19%

Tablo 4.17 YSA (25, 25) ile OECD Ülkelerinin Mandarin İthalatı Tahminleri (...devamı)

Gözlem No	Dönem	Ülke Kodu	Ülke	Ürün kodu	Gerçek Değer	Tahmin Değeri	% Fark
14481	2019	376	İsrail	80520	8.000	582.741	7184,26%
14449	2018	381	İtalya	80520	95.371.752	120.094.172	25,92%
14482	2019	381	İtalya	80520	89.480.000	121.062.374	35,30%
14450	2018	392	Japonya	80520	33.395.241	23.079.656	-30,89%
14483	2019	392	Japonya	80520	31.576.000	28.453.311	-9,89%
14451	2018	428	Letonya	80520	12.831.300	11.923.071	-7,08%
14484	2019	428	Letonya	80520	13.742.000	8.309.841	-39,53%
14452	2018	440	Litvanya	80520	24.121.574	64.288.430	166,52%
14485	2019	440	Litvanya	80520	25.345.000	56.741.402	123,88%
14453	2018	442	Lüksemburg	80520	4.170.578	3.129.556	-24,96%
14486	2019	442	Lüksemburg	80520	5.110.000	3.308.779	-35,25%
13620	2019	484	Meksika	80520	3.211.000	31.179.479	871,02%
14454	2018	484	Meksika	80520	3.801.898	41.814.811	999,84%
14487	2019	484	Meksika	80520	3.211.000	31.179.479	871,02%
14455	2018	528	Hollanda	80520	217.230.085	236.518.890	8,88%
14488	2019	528	Hollanda	80520	209.359.000	243.221.358	16,17%
14456	2018	554	Yeni Zelanda	80520	11.751.453	676.930	-94,24%
14489	2019	554	Yeni Zelanda	80520	12.097.000	676.930	-94,40%
14457	2018	579	Norveç	80520	49.857.128	31.850.021	-36,12%
14490	2019	579	Norveç	80520	51.583.000	41.127.473	-20,27%
14458	2018	616	Polonya	80520	113.628.615	99.154.408	-12,74%
14491	2019	616	Polonya	80520	0	97.489.450	#DIV/0!
14459	2018	620	Portekiz	80520	27.413.441	21.756.054	-20,64%
14492	2019	620	Portekiz	80520	27.312.000	18.509.875	-32,23%
14460	2018	410	Güney Kore	80520	20.000	676.930	3284,65%
14493	2019	410	Güney Kore	80520	452.000	676.930	49,76%
14461	2018	703	Slovakya	80520	24.902.420	16.914.957	-32,08%
14494	2019	703	Slovakya	80520	26.208.000	15.519.443	-40,78%
14462	2018	724	İspanya	80520	30.430.142	33.907.584	11,43%
14495	2019	724	İspanya	80520	26.548.000	31.416.416	18,34%
14463	2018	752	İsveç	80520	53.602.317	72.068.760	34,45%
14496	2019	752	İsveç	80520	54.273.000	73.656.720	35,72%
14464	2018	757	İsviçre	80520	64.525.578	80.196.133	24,29%
14497	2019	757	İsviçre	80520	64.209.000	83.617.083	30,23%
14465	2018	826	Birleşik Krallık	80520	379.305.137	384.724.744	1,43%
14498	2019	826	Birleşik Krallık	80520	372.879.000	383.875.848	2,95%
14466	2018	842	ABD	80520	555.970.191	520.901.440	-6,31%
14499	2019	842	ABD	80520	488.391.000	585.320.258	19,85%

4.6. Yöntemlerin Kıyaslanması

Yöntemlerin performans ölçümleri değerleri Tablo 4.18 ve Tablo 4.19’de yer almaktadır.

Tablo 4.18’de görüleceği üzere, R^2 değerleri açısından XGBoost Algoritmasının eğitim verisi sonuçları diğer algoritmalara göre daha iyi sonuçlar vermiş, test verisi sonuçları ise YSA’ya kıyasla daha düşük, Rastgele Orman Algoritmasına göre daha yüksek bir performans göstermiştir. $RMSE$ kriteri açısından değerlendirilecek olursa eğitim verisi

üzerinde XGBoost algoritması YSA ile Rastgele Orman Algoritmasına göre daha iyi sonuç gösterirken, test verisi üzerinde Rastgele Orman Algoritmasından daha iyi, YSA'ya kıyasla daha kötü sonuçlar vermiştir.

Tablo 4.18 Algoritmaların Performanslarının Kıyaslanması (1)

Algoritma	R^2		$RMSE$		Süre (sn.)
	Eğitim	Test	Eğitim	Test	
XGBoost	0,99	0,89	920.046,74	52.375.909,39	7,09
Rastgele Orman	0,92	0,82	38.547.372,05	67.676.987,28	10,75
YSA (25, 25)	0,958	0,912	27.204.982,63	46.942.779,21	929,04

Diğer taraftan, aynı anda bir çok çalışmanın gerçekleştirilmesi gereken durumlarda ve büyük ölçekli verilerde algoritmanın çalışma süresi de diğer kriterlere ilave olarak oldukça önemli bir değerlendirme kriteri olarak karşımıza çıkabilmektedir. Söz konusu durum göz önünde bulunduracak olursa Tablo 4.18’de yer alan çalışma süreleri açısından en hızlı çalışan algoritma XGBoost’un en iyi performansı verdiğini ifade edilebilmektedir.

Tablo 4.19 Algoritmaların Performanslarının Kıyaslanması (2)

Algoritma	MAE		MAE 'nin ithalat değerleri ortalamasına oranı (%)	
	Eğitim	Test	Eğitim	Test
XGBoost	1.371.282,62	15.700.976,44	% 3,23	% 37,02
Rastgele Orman	16.274.259,76	25.802.798,63	% 38,37	% 60,84
YSA (25, 25)	8.455.818,39	16.593.889,61	% 19,94	% 39,13

Çıktı olan, ithalat değerleri arasında “0” olan değerler olduğu için Ortalama Mutlak Yüzde Hata (MAPE) değerleri hesaplanamamaktadır. Bu nedenle, MAE değerleri veri setindeki çıktının ortalama ve standart sapmaları ile birlikte yorumlanmaktadır.

Oluşturulan veri setinin tanımlayıcı istatistiklerinin sunulduğu Tablo 4.4’de veri setinde “İthalatçı ülkenin ilgili üründen bir yıl sonraki ithalat değeri (\$)” ortalamasının “42.411.690” ve standart sapmasının “140.450.800” olduğu görülmektedir. Tablo 4.19’da MAE değerinin oransal büyüklüğünü gözlemleyebilmek adına MAE 'nin ortalama ithalat değerine oranı yüzdesel olarak hesaplanarak kıyaslanmıştır. Söz konusu değerler açısından hem eğitim hem de test veri seti üzerinde MAE performans kriteri açısından XGBoost Algoritması oldukça iyi sonuçlar sergilemiştir.

Analizde kullanılan her üç tekniğin test veri setinde en iyi gerçekleştirdiği tahminler örnek olması açısından Tablo 4.20’de yer almaktadır.

XGBoost algoritmasının en yakın tahminde bulunduğu değer veri setinin 13.613 numaralı satırında yer alan gözlem aittir. Söz konusu gözlem 2019 yılında Portekiz’in ithal

ettiği “portakal” ürünü değerine ilişkin olup, gerçek değeri 21.227.000 USD iken XG Boost algoritması 21.227.310 USD olarak tahmin etmiştir.

Tablo 4.20 Algoritmaların En İyi Performans Gösterdiği Tahminleme Değerleri

Yöntem	Gözlem No	Yıl	Ülke	Ürün	Gerçek Değer (USD)	Tahmin Değeri (USD)	% Fark
XGBoost	13.613	2019	Portekiz	Portakal	21.227.000	21.227.310	0,0015%
Rastgele Orman	11.228	2018	İzlanda	İspanak	1.114.325	1.114.985,28	0,0593%
YSA	13.045	2019	Estonya	Yaprak Kereviz	677.000,00	676.929,88	-0,0104%

Rastgele Orman Algoritmasının en yakın tahminde bulunduğu değer ise veri setinin 11.228 numaralı satırında yer alan gözleme ait olan 2018 yılında İzlanda’nın “ıspanak” ithalat değerine karşılık gelmektedir. İzlanda’nın 2018 yılında ıspanak ithalatının gerçek değeri 1.114.325,00 USD iken Rastgele Orman Algoritması 1.114.985,28 USD olarak tahmin etmiştir.

Her iki gizli katmanında 25 nöron kullanılan YSA modelinin en yakın tahminde bulunduğu değer ise veri setinin 13.045 numaralı satırında yer alan gözleme ait olan Estonya’nın 2019 yılında ithal ettiği “yaprak kerevizler - taze / soğutulmuş (kök kerevizleri hariç)” ürününün ithalat değeri ile ilgili gözleme karşılık gelmektedir. Estonya’nın 2019 yılındaki yaprak kereviz ithalatının gerçek değeri 677.000,00 USD iken YSA modeli söz konusu ürüne ilişkin ithalat değerini 676.929,88 USD olarak tahmin etmiştir.

Tüm test verisi için tahminleme sonuçları çok fazla alan kapladığı için en fazla ithalat yapılan ürünler açısından örnek değerler için sonuçlar özetlenmiştir. OECD ülkelerinin 2019 yılında en fazla ithal ettiği ilk beş yaş meyve sebze ürünü sırasıyla muz (plantainler hariç), domates, avokado, taze üzüm ve biber olmuştur. Söz konusu bu beş ürün için test veri seti üzerinde (2018 ve 2019 yılları ithalatı için) XGBoost, Rastgele Orman Algoritmaları ile YSA’nın gerçekleştirmiş olduğu tahmin değerleri Ek-7’de sunulmaktadır.

SONUÇ

Bu çalışmada makine öğrenmesi algoritmalarından XGBoost ve Rastgele Orman Algoritması ile Yapay Sinir Ağları üzerinde durulmuş, ülkelerin ithalat potansiyelini tahminlemek amacıyla oluşturulan bir veri seti üzerinde söz konusu algoritmalar çalıştırılarak performansları karşılaştırılmıştır.

Antalya ilinin Türkiye yaş meyve sebze üretim ve ihracatında en fazla paya sahip olan illerden olması nedeniyle, yaş meyve sebze ihracatının geliştirilmesine yönelik bir analiz çalışmasına odaklanılmıştır.

Türkiye yaş meyve sebze üretiminde ve ihracatında dünya genelinde önemli bir konumdadır. Üretim miktarı açısından üstünlüğümüz bulunan birkaç ürün dışında, ülkemiz üretimde gösterdiği performansı dünya ihracat sıralamasında sergileyememektedir.

Ülkemiz üretim potansiyelinin ihracata da yansıtılabilmesi için doğru pazara yönelebilmek ve bu amaçla kapsamlı bir hedef pazar çalışmasının gerçekleştirilmesi oldukça önemlidir. Pazara ilişkin talep tahmini çalışmaları, gelecek dönemde üretilmesi planlanan ürünlerin tür ve çeşitlerinin, ayrıca bu üretim planı için gerekli olan hammadde, işgücü, sermaye gibi kaynakların belirlenmesine yönelik stratejik kararlar alınabilmesi için gereklidir.

Ülkemizin yaş meyve sebze ihracatını artırmak ve dünya ihracatından daha fazla pay alabilmek amacıyla yürütülen talep tahmini çalışmasında incelenecek pazar olarak, GSYİH değerlerinin yüksek olması ve toplamda dünya yaş meyve sebze ithalatının %68'ini tedarik etmesi nedeniyle OECD ülkeleri seçilmiştir.

Dış ticaret verilerinin tahmin edilmesine yönelik literatür çalışmalarının incelenmesi neticesinde çok kriterli karar verme yöntemleri ile regresyon, zaman serisi analizleri gibi ekonometrik yöntemlerin yoğun bir şekilde kullanıldığı tespit edilmiştir. Veri madenciliği yöntemleri açısından ise dış ticaret verilerine yönelik çalışmalarda Yapay Sinir Ağları ile kümeleme analizlerine sıklıkla başvurulduğu görülmektedir.

Diğer taraftan, XGBoost Algoritmaları ile Rastgele Orman Algoritmalarının son dönemlerde veri madenciliği alanında tahminleme çalışmalarında daha fazla tercih edilmeye başlandığı gözlenmiştir.

Bu nedenle, bu tez çalışması ile Türkiye'nin yaş meyve sebze ihracatında önemli bir potansiyele sahip olan OECD ülkelerinin ürün ve ülke bazında gerçekleştireceği yaş meyve sebze ithalatının tahmin edilmesi için makine öğrenmesi algoritmalarından XGBoost, Rastgele Orman Algoritması ile Yapay Sinir Ağları kullanılmıştır. En iyi performansı

sergileyen algoritma kullanılarak ülkelerin ithalat değerlerinin öngörülmesi ve bu doğrultuda yürütülecek pazarlama çalışmalarına katkı sağlanması hedeflenmiştir.

Ayrıca, dış ticaret verilerinin tahminlenmesi çalışmalarında daha önce kullanılmayan XGBoost ile Rastgele Orman Algoritmaları ile analiz sürecini gerçekleştirip hem bu yöntemlerin tahminleme alanındaki başarısını dış ticaret alanına da yansıtmak hem de akademik anlamda bir farklılık yaratmak amaçlanmıştır.

Bunun için öncelikle, yaş meyve sebze sektöründe dış ticarete konu olan 66 ürün grubu açısından Türkiye dışında kalan 35 OECD ülkesinin 2012-2019 yılları arasındaki ithalat verileri, ikincil veri kaynağı olan Birleşmiş Milletler veri kaynağı “UN COMTRADE DATA BASE”den (<https://comtrade.un.org/data>, erişim tarihi: 20.05.2020) derlenmiştir. Değerlendirmeye alınan “Dönem”, “Alıcı ülke kodu”, “Ürün kodu”, “Alıcı ülkenin ilgili yıl enflasyon oranı”, “İlgili ülkenin nüfusu (1.000 kişi)”, “İlgili ülkenin GSYİH (milyar \$)”, “İlgili ülkenin 5 yıl sonraki tahmini GSYİH (milyar \$)”, “İlgili ülkede iş yapma kolaylığı endeksi”, “İlgili ülkenin açık pazar endeksi”, “Alıcı ülkenin ilgili yıldaki söz konusu üründen üretimi”, “Alıcı ülkenin ilgili yıl tarımda işgücü istatistiği”, “Alıcı ülkenin ilgili yıl para biriminin \$ cinsinden değeri” ile “İthalatçı ülkenin ilgili üründen ithalatı değeri (\$)” kriterleri açısından 14.499 gözlem elde edilmiştir. “İthalatçı ülkenin ilgili üründen ithalatı değeri (\$)” çıktı değerinin, girdi olarak alınan diğer değişkenler ile tahmin edilmesi amaçlanmıştır.

Yaş meyve sebze ürünleri GTİP’lerinde 2012 yılında uluslararası bir revizyon gerçekleştirilmiş ve bazı ürünlerin sadece GTİP kodu değişmiş iken bazı ürünlere ise ayrı GTİP’ler oluşturulmuştur. Bu durum, analiz için temin edilen verilerde 2012 yılı öncesinin incelenmesine bir engel oluşturmuş ve daha fazla gözlem ile analiz yapılabilmesi konusunda çalışmanın kısıtını oluşturmuştur. Bu doğrultuda, veri seti oluşturulurken 2012-2019 yılları arasında OECD ülkelerinin yaş meyve sebze ürünleri ithalatı verileri temin edilmiş, 2012 yılı öncesi değerlendirilmeye alınamamıştır.

Çalışmanın bir diğer kısıtı ise FAO’nun web sayfasından temin edilen üretim ile ilgili verilerin en güncel değerlerinin 2018 yılına ait olması durumudur. Bu durum nedeniyle, her period yılındaki girdi özniteliklerine karşılık bir sonraki yıla ait ithalat değerleri çıktı özniteliği olarak esas alınmıştır. “İthalatçı ülkenin ilgili üründen bir yıl sonraki ithalatı Değer (\$)” tahmini yapılacak çıktı olmuştur. Böylece ilgili yıl verileri kullanılarak bir sonraki yılın ithalat verilerinin tahmin edilmesi amaçlanmıştır.

Veri seti 14.499 satır ve 13 sütundan (12 girdi, bir çıktı (ithalat değeri)) oluşmaktadır. Kategorik özelliğe sahip öznitelikler “0-1 ikili değer alabilen özniteliklere dönüştürülmüş

böylece, 111 girdi ve 1 çıktı olmak üzere toplam 112 öznitelik ve 14.499 gözlemden oluşan bir veri seti elde edilmiştir.

Tezin analiz sürecinde Python programlama dili kullanılmıştır. Çalışılan yöntemlerden XGBoost ile Rastgele Orman Algoritmaları normalizasyona ihtiyaç duymamakla birlikte, YSA’larda normalizasyon işlemi önemlidir. YSA analizi için girdi özniteliklerine standardizasyon işlemi uygulanmıştır.

Analiz aşamasında ilk denemeler sonucunda modellerin başarı düzeyine katkı sunmayan “dönem” değişkeni özniteliklerden çıkarılmış, bu durum neticesinde 110 girdi ve 1 çıktı olmak üzere toplam 111 öznitelik ve 14.499 gözlem üzerinden analizler gerçekleştirilmiştir. Veri setinde 2012-2016 dönemine ilişkin veriler eğitim; 2017 ve 2018 yılına ilişkin veriler ise test veri seti olarak ayrılmıştır. Bu durumda eğitim veri seti 10.428 adet gözlemden; test veri seti ise 4.071 adet gözlemden oluşmuştur.

Makine öğrenmesi modelinin performansının doğrudan doğruya seçilen kontrol parametrelerine bağlı olması nedeniyle öncelikle doğru parametrenin seçilmesine yönelik çalışma gerçekleştirilmiştir. XGBoost ve Rastgele Orman Algoritmaları için 10-Kat Çapraz Doğrulama tekniği kullanılmış en iyi performansı veren parametreler üzerinden algoritmalar çalıştırılmıştır. Benzer şekilde Yapay Sinir Ağları da her ağ yapısı için 10 defa çalıştırılmış, söz konusu 10 deneme içerisinde en iyi performansı veren değerler açısından kıyaslama tablosu oluşturulmuş ve en iyi sonucu veren nöron sayıları açısından çalışılacak Yapay Sinir Ağı topolojisi oluşturulmuştur.

Söz konusu üç makine öğrenmesi yönteminin değerlendirilme ölçütleri olarak için R^2 (determinasyon katsayısı), $RMSE$ (Root Mean Squared Error-Hata Kareleri Ortalaması Kökü) ve MAE (Mean Absolute Error-Ortalama Mutlak Hata) ile tahmin değerlerinin ortalaması ve standart sapması performans kriterleri olarak kullanılmıştır.

R^2 değerleri açısından XGBoost Algoritmasının eğitim verisi sonuçları diğer algoritmalara göre daha iyi sonuçlar vermiş, test verisi sonuçları ise YSA’ya kıyasla daha düşük, Rastgele Orman Algoritmasına göre daha yüksek bir performans göstermiştir. $RMSE$ kriteri açısından değerlendirilecek olursa eğitim verisi üzerinde XGBoost algoritması YSA ile Rastgele Orman Algoritmasına göre daha iyi sonuç gösterirken, test verisi üzerinde Rastgele Orman Algoritmasından daha iyi, YSA’ya kıyasla daha kötü sonuçlar vermiştir. MAE değerinin oransal büyüklüğünü gözlemleyebilmek adına MAE ’nin ortalama ithalat değerine oranı yüzdesel olarak hesaplanarak kıyaslanmıştır. Söz konusu değerler açısından hem eğitim hem de test veri seti üzerinde MAE performans kriteri açısından XGBoost Algoritması oldukça iyi sonuçlar sergilemiştir.

Özellikle büyük ölçekli verilerde çalışma süresi oldukça önemli bir değerlendirme performansı niteliğindedir. Yöntemlerin çalışma süreleri ölçütüne göre kıyaslamasında XGBoost en kısa sürede sonuç veren algoritma olmuştur.

Böylece, en iyi performansı sergileyen yöntem olarak XGBoost algoritması seçilmiş olup, girdi özniteliklerinin 2019 yılı değerleri kullanılarak OECD ülkelerinin 2020 yılında hangi yaş meyve sebze ürününden ne kadarlık bir ithalat yapacağını tahmin edilmesi mümkün bulunmaktadır.

Tez çalışmasına katkı sağlamak amacıyla bundan sonraki süreçlerde gerçekleştirilebilecek çalışmalar ile ilgili öneriler ise aşağıda sunulmaktadır:

- XGBoost ile Rastgele Orman Algoritmaları öznitelikleri önem düzeylerine göre sıralayabilmektedir. Bu çalışmada “dönem” özniteliği dışındaki tüm öznitelikler değerlendirilmeye alınmıştır. Sonraki çalışmalarda sadece XGBoost ve Rastgele Orman Algoritmalarının önemli bulduğu öznitelikler ile analiz yapılarak çalışma yürütülebilir.
- Farklı öznitelikler eklenerek çalışma yinelenabilir.
- Bu çalışmayı takip eden süreçte hedef ülkelerin tercih ettiği ürünlerde aradığı tat ve lezzet ile pazarda en fazla talep gören ürün çeşitlerine yönelik gerekli araştırmalar yapılarak, pazardaki rakip ülke firmaları ve ürün fiyatları yakından incelenerek, üretim planı ve ürün deseni oluşturabilir ve pazarlama programı bu doğrultuda yapılabilir.
- OECD ülkelerinin ithal etmiş olduğu yaş meyve sebze ürünlerini hangi ülkelere ve ne kadar ithal ettiği ile ilgili geçmiş dönem verileri kullanılarak gelecek dönemdeki olası rakipler üzerinde durulabilir, bu süreçte ürün tedarik edilen ülke ile ithalatçı ülke arasındaki kültürel etkileşim, coğrafi mesafe, gümrük vergi oranları gibi öznitelikler de analiz sürecine dahil edilebilir.
- T.C. Ticaret Bakanlığı tarafından iki yılda bir hedef ve öncelikli ülkeler belirlenmekte ve bu ülkelere yönelik faaliyetlerde devlet destekleri oranlarına ilave olarak %10 destek sağlanmaktadır. Tez çalışmasında geliştirilen tahminleme mekanizması ile ülkelerin gelecekte oluşabilecek talepleri tahmin edilebilir ve belirlenecek hedef pazarlardan ülkemizin mümkün olduğunca fazla pay alabilmesi için devlet desteği sistemi de bu doğrultuda düzenlenebilir.

İhracat, ülkemiz ekonomisinin gelişimine katkı sağlayan en önemli yapı taşlarından biridir. İhracatın gelişimi için hedef pazarların gelecek dönem ithalat potansiyellerinin yakından izlenmesi ve ona uygun bir şekilde gerekli tedbirlerin alınması gerekmektedir. Bu çalışma ile

hedef pazarların talep tahmini çalışmalarında bir farklılık yaratılması amaçlanmıştır. Bu çalışmalar farklı özniteliklerin veri setine eklenmesi ve farklı algoritmaların kullanılması ile geliştirilebilecek olup önemli olan ülkemiz ihracatının gelişimine yönelik alınacak stratejik kararlara bilimsel veriler sunulabilmesidir.

Bu doğrultuda, bu çalışmanın dış ticarete ilişkin politikaların belirlenmesi sürecinde veri madenciliği yöntemlerinin kullanımına yönelik bir örnek teşkil etmesi sonraki çalışmalara ışık tutması beklenmektedir.



KAYNAKÇA

- Akar, Ö., Güngör, O. ve Akar, A., (2010). “Rastgele orman sınıflandırıcısı ile arazi kullanım alanlarının belirlenmesi.” *III. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu Bildiriler Kitabı*, 11-13 Ekim 2010, Gebze, Kocaeli, 142-152.
- Akar, Ö. ve Güngör, O. (2012). “Rastgele Orman Algoritması Kullanılarak Çok Bantlı Görüntülerin Sınıflandırılması.” *Journal of Geodesy and Geoinformation*, 1(2): 139-146.
- Akar, Ö. (2013). *Rastgele Orman Sınıflandırıcısına Doku Özellikleri Entegre Edilerek Benzer Spektral Özellikteki Tarımsal Ürünlerin Sınıflandırılması*. Yayımlanmamış Doktora Tezi. Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Trabzon.
- Akın, Y.K., (2008) *Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi*. Yayımlanmamış Doktora Tezi. Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Akgöbek, Ö., ve Çakır, F. (2009). “Veri madenciliğinde bir uzman sistem tasarımı”. *Akademik XI. Bilişim Konferansı Bildirileri*, 11-13 Şubat 2009, Şanlıurfa, 801-806.
- Akgün, A, (2012), *Seyahat Acentalarında Veri Madenciliği: Antalya Bölgesinde Bir Uygulama*. Yayımlanmamış Yüksek Lisans Tezi. Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Antalya.
- Akkurt, A. (2005). *Yapay Sinir Ağları ve Türkiye Elektrik Tüketimi Tahmin Modeli*. Yayımlanmamış Yüksek Lisans Tezi. İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Akküçük, U. (2011). *Veri Madenciliği: Kümeleme ve Sınıflama Algoritmaları*. Yalın Yayıncılık, İstanbul.
- Akman M., Genç Y. ve Ankaralı H. (2011). “Random Forests Yöntemi ve Sağlık Alanında Bir Uygulama”. *Türkiye Klinikleri Biyoistatistik Dergisi, Biostat*, 3(1): 36-48.
- Akpınar, H. (2000). “Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği”, *İ.Ü. İşletme Fakültesi Dergisi* , 29 (1): 1-22.
- Akpınar, H. (2014). *DATA -Veri Madenciliği Veri Analizi*. Papatya Bilim Üniversite Yayıncılığı, İstanbul.
- Alam, T. (2019). “Forecasting Exports and Imports Through Artificial Neural Network And Autoregressive Integrated Moving Average”. *Decision Science Letters* , 8(3): 249-260.

- Albayrak, A. S. ve Yilmaz, S. K. (2009). “Veri Madenciliği: Karar Ağacı Algoritmaları ve İmkb Verileri Üzerine Bir Uygulama”. *Süleyman Demirel University Journal of Faculty of Economics & Administrative Sciences*, 14(1): 31-52.
- Ata, N., Özkök, E. ve Karabey, U. (2008), “Survival Data Mining: an Application To Credit Card Holders”, *Sigma Mühendislik ve Fen Bilimleri Dergisi*, 26(1): 33-42.
- Atalay, G. (2012). *Türkiye Orman Ürünleri Dış Ticaretinde Hedef Pazar Yönelim Analizi*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Avuçlu, E. ve Ekmekci, D. (ed.). (2020). *Geleceğin Dünyasında Bilimsel ve Mesleki Çalışmalar*. Ekin Basın Yayın Dağıtım, Bursa.
- Aydemir, E., Aktürk, C., ve Yalçınkaya, M. A. (2020). “Yapay Zekâ ile Konut Fiyatlarının Tahmin Edilmesi”. *Turkish Studies-Information Technologies and Applied Sciences*, 15(2): 183-194.
- Ayık, Y. Z., Özdemir A. ve Yavuz U. (2007), “Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi”. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10 (2): 441-454.
- Babyak, M.A. (2004). “What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models”. *Statistical Corner*. 66 (3): 411-421.
- Bai, X., Liu, H., Zhang, F., Ning, Z., Kong, X., Lee, I., ve Xia, F. (2017). “An Overview on Evaluating and Predicting Scholarly Article Impact”. *Information*, 8(3), 73: 1-14.
- Balaban, M.E. ve Kartal, E. (2015), *Veri Madenciliği ve Makine Öğrenmesi, Temel Algoritmaları ve R Dili ile Uygulamaları*. Çağlayan Kitabevi, İstanbul.
- Baş, N. (2006). *Yapay Sinir Ağları Yaklaşımı ve Bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi. Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Bayır, F. (2006). *Yapay Sinir Ağları ve Tahmin Modellemesi Üzerine bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.
- Baynal, K., ve Çalış, A. (2016). “Veri Madenciliğinde Kümeleme Analizi ile Bankacılık Sektöründe Bir Uygulama”. *Beşkent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 9(1): 13-41.

- Benito Garzón, M., Sánchez de Dios, R. ve Sainz Ollero, H. (2008). "Effects of Climate Change on the Distribution of Iberian Tree Species". *Applied Vegetation Science*, 11(2): 169-178.
- Bezek Güre, Ö. (2019). *Öğrencilerin Matematik Başarılarını Etkileyen Faktörlerin Rastgele Orman, Çok Katmanlı Algılayıcı ve Radyal Tabanlı Fonksiyon Yapay Sinir Ağları Yöntemleri İle Tahminleme Yeteneği Açısından Karşılaştırılması: Türkiye Örneği*. Yayımlanmamış Doktora Tezi. Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van.
- Binici, K. (2019). "Makine Öğrenmesi Yaklaşımıyla e-Belgelere Standart Dosya Plan Numaralarının Otomatik Olarak Atanması Üzerine Bir Çalışma". *Bilgi Yönetimi Dergisi*, 2(2): 116-126.
- Bircan, H., Zontul, M. ve Yüksel, A. G. (2010). "SOM Tipinde Yapay Sinir Ağlarını Kullanarak Türkiye'nin İhracat Yaptığı Ülkelerin Kümelenmesi Üzerine Bir Çalışma". *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 20 (2): 219-239.
- Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. (2012). "Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6): 493-507.
- Bowles, M. (2015). *Machine Learning In Python*. John Wiley & Sons Inc., New Jersey.
- Breiman, L., Friedman, J. H., Olshen, R.A. ve Stone, C. J. (1984). *Classification and Regression Trees*. Monterey, CA.
- Breiman, L. (1996). "Bagging Predictors". *Machine Learning*, 24(2): 123-140.
- Breiman L. (2001). "Random Forests". *Machine Learning*, 45(1): 5-32.
- Bühlmann, P., ve Yu, B. (2002). "Analyzing Bagging". *The Annals of Statistics*, 30(4): 927-961.
- 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.
- Chen, T. ve Guestrin, C. (2016). "XGBoost: A scalable tree boosting system". *In Proceedings Of The 22nd ACM SigKDD International Conference On Knowledge Discovery and Data Mining*. 13-17 Ağustos 2016, San Francisco, 785-794.
- Christian, A.V., Zhang, Y. ve Salifou, C.K. (2016). "Country Selection for International Expansion: TOPSIS Method Analysis". *Modern Economy*, 7: 470-476.

- Chryssolouris G ve Guillot M (1990). “A Comparison of Statistical and AI Approaches to the Selection of Process Parameters in Intelligent Machining”. *ASME J Eng Ind*, 112 (2): 122–131.
- Chung, J. Y. ve Lee, S. (2019). “Dropout Early Warning Systems for High School Students Using Machine Learning”. *Children and Youth Services Review*, 96: 346-353
- Co, H.C. ve Boosarawongse, R. (2007). “Forecasting Thailand’s Rice Export: Statistical Techniques vs. Artificial Neural Networks”. *Computers & Industrial Engineering*, 53: 610-627.
- Coşlu, E. (2013). “Veri madenciliği”. *Akademik Bilişim 2013 –XV. Akademik Bilişim Konferansı Bildirileri*. 23-25 Ocak 2013, Antalya, 615-619.
- Cutler, A., Cutler, D.R. ve Stevens, J.R. (2012). “Random Forests”. C., Zhang ve Y. Ma (Ed.), *Ensemble Machine Learning: Methods and Applications*. Springer Science and Business Media LLC, Heidelberg, 157-176.
- Czinkota, M.R., (2002). “Export Promotion: A Framework for Finding Opportunity in Change”. *Thunderbird International Business Review*, 44(3): 315-324.
- Çakmak, T., Tekin A.T., Şenel Ç., Çoban, T., Uran, Z.E. ve Sakar C.O. (2019). “Accurate Prediction of Advertisement Clicks Based on Impression and Click-Through Rate Using Extreme Gradient Boosting”. In *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2019)*: 621-629.
- Çavuşgil, S.T. (1985). “Guidelines for Export Market Research”. *Business Horizons*, 28: 27-33.
- Çavuşgil, S. T. (1997). “Measuring the Potential of Emerging Markets: An Indexing Approach”. *Business Horizons*, 40(1):87–91.
- Çavuşgil, S.T., T. Kiyak ve S. Yenyurt (2004), “Complementary Approaches to Preliminary Foreign Market Opportunity Assessment: Country Clustering and Country Ranking.” *Industrial Marketing Management*, 33(7): 607-617.
- Çeşmeli, M.Ş., Bozkurt, Ö. Ç., Kalkan, A. ve Pençe, İ, (2015). “Yönetim Bilişim Sistemleri Bölümü Öğrenci-lerinin Yönetim Ve Bilişim Derslerindeki Başarılarının Veri Madenciliği Yöntemleri İle İncelenmesi”. *Yönetim Bilişim Sistemleri Dergisi*. 1(2): 36-47.
- Çuhadar, M. (2006). *Turizm Sektöründe Talep Tahmini için Yapay Sinir Ağları Kullanımı ve Diğer Yöntemlerle Karşılaştırmalı Analizi*. Yayımlanmamış Doktora Tezi. Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü, Isparta.

- Danacı, M., Çelik, M. ve Akkaya, A.E. (2010). “Veri Madenciliği Yöntemleri Kullanılarak Meme Kanseri Hücrelerinin Tahmin ve Teşhisi”, *Akıllı Sistemlerde Yenilikler ve Uygulama Sempozyumu*, 21-24 Haziran 2010, Kayseri, 9-12.
- Das, S., Dey, A., Pal, A. ve Roy, N. (2015). “Applications of Artificial Intelligence in Machine Learning: Review and Prospect.” *International Journal of Computer Applications*, 115(9): 31-41.
- Das, K. ve Behera, R. N. (2017). “A Survey on Machine Learning: Concept, Algorithms and Applications”. *International Journal of Innovative Research in Computer and Communication Engineering*, 5(2): 1301-1309.
- Deligiannis, A., Argyriou C. ve Kourtesis, D. (2020). “Building a Cloud-Based Regression Model to Predict Click-Through Rate in Business Messaging Campaigns”. *International Journal of Modeling and Optimization*, 10(1): 26-31.
- Demir, H. ve Gümüšoğlu, Ş. (2003). Üretim Yönetimi-İşlemler Yönetimi. *Beta Basım Yayım*, İstanbul
- Demir, L. ve Akkaş, S. (2018). “A Comparison of Sales Forecasting Methods For a Feed Company: A Case Study”. *Pamukkale University Journal of Engineering Sciences*, 24(4): 705-712.
- Dhaliwal, S. S., Nahid, A. A. ve Abbas, R. (2018). “Effective Intrusion Detection System Using XGBoost”. *Information*, 9(7), 149: 1-24.
- Dış Ekonomik İlişkiler Kurulu (DEİK) (2016). *Asya Yüzyılında Ejder & Hilâl Türkiye-Çin Ekonomik İlişkilerinin Geliştirilmesi İçin Bir Yol Haritası*. DEİK, İstanbul.
- Doğan, Ş. ve Türkoğlu, İ. (2008). “Iron-Deficiency Anemia Detection From Hematology Parameters By Using Decision Trees”, *International Journal of Science & Technology*, 3 (1): 85-92.
- Dolgun, M.Ö. (2014). *Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının; Bağımlı Değişken Prevelansı, Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası İlişki Yapısına Göre Karşılaştırılması*. Yayımlanmamış Doktora Tezi. Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
- Erdoğan, B.Z., Önder, G., Sayın, H.C., Tiltay, M.A., Kurulgan M., Sayılır, Ö., Suvacı, B., Besler, S. (2019). *İşletme İlkeleri*. T.C. Anadolu Üniversitesi Yayını No: 3914, Eskişehir.
- Ersen, N. (2016). *Türkiye'nin Ağaç ve Orman Ürünleri İhracat ve İthalat Değerlerinin Box-Jenkins ve Yapay Sinir Ağları Yöntemleri ile Tahmin Edilmesi ve Karşılaştırılması*.

- Yayımlanmamış Doktora Tezi. Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Trabzon.
- Ersöz, T., Düğenci, M., Ünver, M. ve Eyiöl, B. (2015), “Demir Çelik Sektörüne Genel Bir Bakış ve Beş Milyon Ton Üstü Demir Çelik İhracatı Yapan Ülkelerin Kümeleme Analizi ile İncelenmesi”. *Nevşehir Bilim ve Teknoloji Dergisi*, 4(2): 75-90.
- Everingham, Y., Sexton, J., Skocaj, D. ve Inman-Bamber, G. (2016). “Accurate Prediction of Sugarcane Yield Using a Random Forest Algorithm”. *Agronomy for Sustainable Development*, 36(2), 27:1-9.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). “The KDD Process for Extracting Useful Knowledge from Volumes of Data”. *Communication of ACM*, 39(11): 27-34.
- Feng, C. X. ve Wang, X. F. (2003). “Surface Roughness Predictive Modeling: Neural Networks Versus Regression”. *IIE Transactions*, 35(1):11-27.
- Ferreira, A.J. ve Figueiredo, M.A.T. (2012). “Boosting Algorithm: A Review of Methods, Theory and Applications”. C., Zhang ve Y. Ma (Ed.), *Ensemble Machine Learning: Methods and Applications*. Springer Science and Business Media LLC, Heidelberg, 35-86.
- Gao, J., Xie, Y., Cui, X., Yu, H. ve Gu, F. (2018). “Chinese Automobile Sales Forecasting Using Economic Indicators and Typical Domestic Brand Automobile Sales Data: A Method Based on Econometric Model”. *Advances in Mechanical Engineering*, 10(2): 1-11.
- Gardner M.W. ve Sorling S.R., (1998). “Artificial Neural Networks (The Multilayer Perceptron)—A Review Of Applications In The Atmospheric Sciences”. *Atmospheric Environment*, 32(14/15): 2627-2636.
- Gemici, B. (2012). *Veri Madenciliği ve Bir Uygulaması*. Yayımlanmamış Yüksek Lisans Tezi. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.
- Gislason, P. O., Benediktsson, J. A., ve Sveinsson, J. R. (2004). “Random Forest Classification of Multisource Remote Sensing and Geographic Data”. *2004 IEEE International Geoscience and Remote Sensing Symposium*, 20-24 Eylül 2004, 1049-1052.
- Gislason, P. O., Benediktsson, J. A., ve Sveinsson, J. R. (2006). “Random Forests for Land Cover Classification”. *Pattern Recognition Letters*, 27(4): 294-300.
- Gnananandarao, T., Dutta, R.K. ve Khatri, V.N. (2019). “Application of Artificial Neural Networks to Predict the Settlement of Shallow Foundations on Cohesionless Soils”.

- I.V., Anirudhan ve V.B, Mai (Ed.), *Geotechnical Applications*. Springer Nature Singapore Pte Ltd., Singapore, 51-58.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. Springer-Verlag, Berlin.
- Gupta, G.K. (2016). *Introduction to Data Mining with Case Studies*. PHI Learning Private Limited, Delhi.
- Güldal, H. ve Çakıcı, Y. (2017). “Ders Yönetim Sistemi Yazılımı Kullanıcı Etkileşimlerinin Sınıflandırma Algoritmaları ile Analizi”. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 21(4):1355-1367.
- Güresen, E. ve Kayakutlu, G. (2011). “Definition Of Artificial Neural Networks with Comparison to Other Networks”. *Procedia Computer Science*, 3: 426–433.
- Haglin, D., Roiger, R., Hakkila, J. ve Giblin, T. (2005). “A Tool for Public Analysis of Scientific Data”. *Data Science Journal*, 4(30): 39–53.
- Han, J. ve Kamber, M. (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, Massachusetts.
- Hand, D. , Mannila, H. ve Smyth, P. (2001). *Principles of Data Mining*. MIT Press, Cambridge, England.
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice Hall, New Jersey.
- Hazır, E., Koç, K. H. ve Esnaf, Ş. (2015). “Türkiye mobilya satış değerlerinin örnek bir yapay zekâ uygulaması ile tahmini”, 3. *Ulusal Mobilya Kongresi (UMK-2015)*, 10-12 Nisan 2015 Konya: 1172-1182.
- Helmets, C. ve Pasteels, J. M. (2006). “Assessing Bilateral Trade Potential at the Commodity Level: An Operational Approach”. *International Trade Centre, ITC Working Paper*.
- Horning N., (2010). “Random forests : an algorithm for image classification and generation of continuous fields data sets”. *International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences (GIS-IDEAS)*. 9-11 Aralık 2010, Hanoi, 1-6.
- Humphris, R.M.V. (2020). *Testing Algorithm Fairness Metrics for Binary Classification Problems by Supervised Machine Learning Algorithms*. Yayımlanmamış Yüksek Lisans Tezi. Vrije Universiteit, Faculty of Science, Amsterdam.
- Irmak, S. (2009). *Veri Madenciliği Yöntemleri ile Sağlık Sektörü Veritabanlarında Bilgi Keşfi: Tanımlayıcı ve Kestirimci Model Uygulamaları*. Yayımlanmamış Doktora Tezi. Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Antalya.

- Ji, W. ve Lin, S. (2019). "The Risk Prediction of Type 2 Diabetes based on XGBoost". *2019 2nd International Conference on Mechanical, Electronic and Engineering Technology*, 19-20 Ocak 2019. Xi'an, 145-150.
- Kalaycı, S. (2018). *Makine Öğrenmesi Yöntemleri İle Kredi Risk Analizi*. Yayımlanmamış Yüksek Lisans Tezi. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- 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.
- Karahan, M. (2015). "Yapay Sinir Ağları Metodu ile İhracat Miktarlarının Tahmini: ARIMA ve YSA Metodunun Karşılaştırmalı Analizi". *Ege Academic Review*, 15 (2): 165-172.
- Kobu, B. (2013). *Üretim Yönetimi*. Beta Yayıncılık, İstanbul.
- Korkem, E. (2003). *Mikroarray Gen Ekspresyon Veri Setlerinde Random Forest ve Naive Bayes Sınıflama Yöntemleri Yaklaşımı*. Yayımlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
- Kotsiantis, S.B., Zaharakis, I. ve Pintelas, P (2007). "Supervised Machine Learning: A Review of Classification Techniques". *Emerging Artificial Intelligence Applications in Computer Engineering*. 160: 3-24.
- Krajewski, L.J., Ritzman, L.P. ve Malhotra, M.K. (2013). *Üretim Yönetimi, Süreçler ve Tedarik Zincirleri*. (Çev. S. Birgün), Nobel Akademik Yayıncılık, Ankara.
- Kurt, R., Karayılmazlar, S., İmren, E. ve Çabuk, Y. (2017), "Yapay Sinir Ağları İle Öngörü Modellemesi: Türkiye Kâğıt-Karton Sanayi Örneği". *Bartın Orman Fakültesi Dergisi*, 19(2): 99-106.
- Kuzey, C. (2012). *Veri Madenciliğinde Destek Vektör Makinaları ve Karar Ağaçları Yöntemlerini Kullanarak Bilgi Çalışanlarının Kurum Performansı Üzerine Etkisinin Ölçülmesi ve Bir Uygulama*. Yayımlanmamış Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Larose, D. T. ve Larose C.D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons Inc., New Jersey.
- Le Cun, Y., Simard, P. Y. ve Pearlmutter, B. (1993). "Automatic Learning Rate Maximisation by Online Estimation of the Hessian's Eigen Vectors". *Advances in Neural Information Processing Systems*, 5: 156-163.
- Lee, W. (2019). *Python Machine Learning*. John Wiley & Sons Inc., Indianapolis.
- Lei, T., Chen, F., Liu, H., Sun, H., Kang, Y., Li, D., Li, Y. ve Hou, T. (2017). "ADMET Evaluation in Drug Discovery. Part17: Development of Quantitative and Qualitative

- Prediction Models for Chemical-induced Respiratory Toxicity". *Molecular Pharmaceutics*, 14(7): 2407-2421.
- Li, C., Shao, H., Zhang, T. ve Yu, G. (2019)."Analysis of Macro Factors of Welfare Lottery Marketing Based on Big Data". W., Ni, X., Wang, W. Song ve Y., Li. (Ed.) Web Information Systems and Applications. 16th International Conference, WISA 2019, 20-22 Eylül 2019, Qingdao, 189-200.
- Liaw, A. and Wiener, M. (2002). "Classification and Regression by Random Forest". *R News*, 2, 18-22.
- Mat Isaa, C.M., Samanb H.M. ve Nasirc S.R.M (2013). "Specific-factors Influencing Market Selection Decision by Malaysian Construction Firms into International Market". *Procedia - Social and Behavioral Sciences*, 129 : 4-10
- Merembayev, T., Yunussov, R. ve Yedilkhan, A. (2019). "Machine Learning Algorithms for Stratigraphy Classification on Uranium Deposits". *Procedia Computer Science*, 150: 46-52.
- Mieciński A., Stasyte V. ve Kazlauskaitė J., (2014). "Reasoning of Export Market Selection, Business". *Procedia Management and Education*. 110: 1166-1175
- Mobin, M., Dehghanimohammadabadi, M. ve Salmon, C. (2014). "Food Product Target Market Prioritization Using MCDM Approaches". In *Proceedings of the 2014 Industrial and Systems Engineering Research Conference*, Montreal, Canada, 1-10.
- Müller C. ve Guido S. (2017), *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media Inc., CA.
- Namlı, Ö. H., ve Özcan, T. (2017). "Makine Öğrenmesi Algoritmaları Kullanarak Gişe Hasılatının Tahmini". *Yönetim Bilişim Sistemleri Dergisi*, 3(2): 130-143.
- Nas, S. (2020). *Acil Servislerde Yapay Sinir Ağları ile Tahminleme ve Kaynak Planlaması*. Yayımlanmamış Doktora Tezi. Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.
- Nisbet, R., Miner, G. ve Yale, K. (2018). *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press, Burlington.
- Oğuzlar, A. (2003). "Veri Ön İşleme". *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 21: 67-76
- Onur, G. (2012), "Dış Pazar Araştırması Nasıl Yapılır?". *T.C. Ekonomi Bakanlığı İhracat Genel Müdürlüğü Ülke Masaları - I Daire Başkanlığı*, Ankara, 1-51.
- Özbay, Ö., ve Ersoy, H. (2017). "Öğrenme Yönetim Sistemi Üzerindeki Öğrenci Hareketliliğinin Veri Madenciliği Yöntemleriyle Analizi". *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 37(2): 523-558.

- Özbek, A. (2009). *Türk Hazır Giyim Sanayinin Örnek Ürün Bazında (Denim Pantolon) Gelecekteki İhracat Performansının İncelenmesi*. Yayınlanmamış Doktora Tezi. Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Özcan, C. (2014). *Veri Madenciliğinin Güvenlik Uygulama Alanları ve Veri Madenciliği İle Sahtekarlık Analizi*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul Bilgi Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Özdarıcı Ok A., Akar Ö. ve Güngör O. (2011). “Rastgele orman sınıflandırma yöntemi yardımıyla tarım alanlarındaki ürün çeşitliliğinin sınıflandırılması”. *TUFUAB VI. Teknik Sempozyumu*, 21-25 Şubat 2011, Antalya, 1-7.
- Özdemir, S. (2018). “Random Forest Yöntemi Kullanılarak Potansiyel Dağılım Modellemesi ve Haritalaması: Yukarıgökdere Yöresi Örneği”. *Turkish Journal of Forestry*, 19(1): 51-56.
- Özkan, Y. (2016). *Veri Madenciliği Yöntemleri*. Papatya Bilim Üniversite Yayıncılığı, İstanbul.
- Öztemel, E. (2012). *Yapay Sinir Ağları*. Papatya Yayıncılık, İstanbul.
- Öztürk, A., Joiner, E. ve Çavuşgil S.T. (2015). “Delineating Foreign Market Potential: A Tool for International Market Selection”. *Thunderbird International Business Review*, 57: 119-141.
- Pal M., (2005), “Random Forest Classifier for Remote Sensing Classification”. *International Journal of Remote Sensing*, 26(1): 217-222.
- Pan, B. (2018). “Application of XGBoost Algorithm in Hourly PM2. 5 Concentration Prediction”. In *IOP Conference Series: Earth and Environmental Science*, Vol. 113:1-7.
- Pekel, E. (2018). *Farklı Makine Öğrenmesi Algoritmalarının Karşılaştırılması*. Yayınlanmamış Yüksek Lisans Tezi. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Samsun.
- Polat, K. (2008). *Biyomedikal Sinyallerde Veri Ön-İşleme Tekniklerinin Medikal Teşhiste Sınıflama Doğruluğuna Etkisinin İncelenmesi*. Yayınlanmamış Doktora Tezi. Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Prasad A.M., Iverson L.R. ve Liaw A. (2006). “Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction”. *Ecosystems*, 9: 181-199.

- Priya, P. ve Venkatesh, A. (2012). "Integration of Analytic Hierarchy Process with Regression Analysis to Identify Attractive Locations for Market Expansion". *Journal of Multi-Criteria Decision Analysis*, 19(3-4): 143-153.
- Quintana, D., Sáez, Y. ve Isasi, P. (2017). "Random Forest Prediction of IPO Underpricing". *Applied Sciences*, 7(6) 636:1-16
- Rahmati, O., Moghaddam, D. D., Moosavi, V., Kalantari, Z., Samadi, M., Lee, S., ve Tien Bui, D. (2019). "An Automated Python Language-Based Tool For Creating Absence Samples In Groundwater Potential Mapping". *Remote Sensing*, 11(11): 1375.
- Raschka, S. ve Mirjalili V. (2019), *Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-learn, and Tensor Flow 2*. Packt Publishing, Birmingham.
- Røine, T. B. ve Holter, E. K. (2018). *Properties of the Gold Price: An Investigation Using Fractional Brownian Motion and Supervised Machine Learning Techniques*. Yayınlanmamış Yüksek Lisans Tezi. Norwegian School of Economics, Bergen.
- Rojas, R. (1996). *Neural Networks: A Systematic Introduction*. Springer-Verlag, Berlin Heidelberg.
- Sabancı, D. (2019). *Rastgele Orman Yaklaşımı Kullanılarak Çok Değişkenli Uyumlu Regresyon Şeritlerinde Model Seçimi*. Yayınlanmamış Doktora Tezi. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Samsun.
- Sarı, M. (2016). *Yapay Sinir Ağları ve Bir Otomotiv Firmasında Talep Tahmini*. Yayınlanmamış Yüksek Lisans Tezi. Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya.
- Schapire, R.E. (1990). "The Strength of Weak Learnability". *Machine Learning*, 5(2):197-227.
- Seyrek, İ. H., ve Ata, H. A. (2010). "Veri Zarflama Analizi ve Veri Madenciliği ile Mevduat Bankalarında Etkinlik Ölçümü". *Journal of BRSA Banking & Financial Markets*, 4(2): 67-84.
- Sheng, S.Y. ve Mullen M.R. (2006). "Extending and Comparing Cavusgil's Overall Market Opportunity Indexes". *Advances in International Marketing*, 17: 219-249.
- Sosyal, M. ve Ömürgönülşen, M. (2010). "Türk Turizm Sektöründe Talep Tahmini Üzerine Bir Uygulama". *Turizm Araştırma Dergisi*, 21(1): 129-136.
- Sürek, Ş. (2010). *İhracatçı Olmak, İhracat Pazarlamasında Yanlışlar, Doğrular*. Mercan Kitap, İstanbul.

- Tan P., Steinbach M. ve Kumar V. (2006). *Introduction to Data Mining*. Pearson Education Inc., Londra.
- Tang T. ve Chi L. (2005). “Neural Networks Analysis in Business Failure Prediction of Chinese importers: A Between Countries Approach”. *Expert Systems with Applications*, 29 (2): 244-255.
- Tek, B. (1999). *Pazarlama İlkeleri: Global Yönetimsel Yaklaşım Türkiye Uygulamaları*. Beta Basım Yayım, İstanbul.
- Temiz, M. (2018). *Doğal Taş Karolarının Görüntü İşleme ve Makine Öğrenmesi Teknikleri ile Sınıflandırılması*. Yayınlanmamış Yüksek Lisans Tezi. Sivas Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü, Sivas.
- Tosun, A. (2007). *Türkiye'nin Mermer İhracatını Artırma Olanaklarının Araştırılması*. Yayınlanmamış Yüksek Lisans Tezi. Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü, İzmir.
- Tozlu, İ. (2019). *Simplifying Balance Sheet Adjustment Process in Commercial Loan Applications Using Machine Learning Methods*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Tsai, F. ve Huang, L.J.W. (2017) “Using Artificial Neural Networks to Predict Container Flows Between The Major Ports Of Asia”, *International Journal of Production Research*, 55 (17): 5001-5010
- Tütek, H., Gümüsoğlu, Ş.(2000). *İşletme İstatistiği*. Barış Yayınları, İzmir.
- Ulutaş, A.E. (2018). *Protein Fragment Selection Using Machine Learning*, Yayınlanmamış Yüksek Lisans Tezi. Abdullah Gül Üniversitesi Fen Bilimleri Enstitüsü, Kayseri.
- Xu, Y., Ho, H. C., Wong, M. S., Deng, C., Shi, Y., Chan, T. C. ve Knudby, “A. (2018). Evaluation of Machine Learning Techniques with Multiple Remote Sensing Datasets in Estimating Monthly Concentrations of Ground-Level PM_{2.5}”. *Environmental Pollution*, 242: 1417-1426.
- Wang, H., Yin J., Pei, J., Yu, P.S. ve Yu J.X. (2006). “Suppressing Model Overfitting in Mining Concept-Drifting Data Streams”. *KDD '06: Proceedings of the 12th ACM SIGKDD International Conference On Knowledge Discovery And Data Mining*. Philadelphia, 736-741.
- Wang, G., Hao, J., Ma, J. ve Jiang, H. (2011). “A Comparative Assessment of Ensemble Learning for Credit Scoring”. *Expert Systems with Applications*, 38(1): 223-230.
- Witten, I.H., Frank, E. ve Hall, M.A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, Burlington.

- Yangın, G. (2019). *XGBoost ve Karar Ağacı Tabanlı Algoritmaların Diyabet Veri Setleri Üzerine Uygulaması*. Yayınlanmamış Doktora Tezi. Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Yılmazel, Ö., Afşar, A. ve Yılmazel, S. (2018). “Konut Fiyat Tahmininde Yapay Sinir Ağları Yönteminin Kullanılması”, *International Journal of Economic & Administrative Studies*, 20: 285-300.
- Yurtay, Y., Salman, Y., Salman, M. E., ve Gençali, F. (2013). “Kansızlık Tanısına İlişkin Bir Veri Madenciliği Uygulaması”. *1st International Symposium on Innovative Technologies in Engineering and Science*, 7-9 Temmuz 2013, Sakarya, 896-900.
- Yücel, A. (2006). *İhracat Pazarlaması Stratejilerinin Firma Performansı Üzerine Etkileri Hazır Giyim Firmaları Üzerine Bir Uygulama*. Yayınlanmamış Doktora Tezi. Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Yücesan, M. (2018). “YSA, ARIMA ve ARIMAX Yöntemleriyle Satış Tahmini: Beyaz Eşya Sektöründe bir Uygulama”. *İşletme Araştırmaları Dergisi*, 10(1): 689-706.
- Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B. ve Si Y. (2018). “A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGBoost”. *IEEE-Advancing Technology for Humanity*. 6: 21020- 21031.
- Zhang, R., Li, B. ve Jiao, B. (2019). “Application of XGboost Algorithm in Bearing Fault Diagnosis”. *In IOP Conference Series: Materials Science and Engineering*, 490 (7): 1-5.
- Zhao J., Li J., Li L. (2011). “An Analysis on the Target Market of China’s Textile and Garment Export Trade”. *Procedia Engineering*, 15: 4718-4722
- Zheng, H., Yuan, J. ve Chen, L. (2017). “Short-term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm For Feature Importance Evaluation”. *Energies*, 10(8), 1168: 1-20.
- Zontul, M. ve Yangın, A. (2017), “Yapay Sinir Ağı Teknikleri Kullanarak Eğitim Yayıncılığı Sektöründe Veri Madenciliği”. *AURUM-Mühendislik Sistemleri ve Mimarlık Dergisi*, 1(2): 1-15.
- Zou, S., Taylor, C.R. ve Osland, G.E, (1998). “The EXPERF Scale: A Cross-National Generalized Export Performance Measure”. *Journal of International Marketing*, 6(3) : 37-58.
- Zou, H. F., Xia, G. P., Yang, F. T. ve Wang, H. Y. (2007). “An Investigation and Comparison of Artificial Neural Network and Time Series Models for Chinese Food Grain Price Forecasting”. *Neurocomputing*, 70: 2913-2923.

İNTERNET KAYNAKLARI

- Adım Adım İhracat, TİM Akademi, 2014. <https://tr.scribd.com/doc/207549411/Adim-Adim-Ihracat-2014-Ilanli-Kitap3> (Erişim Tarihi: 25.05.2020).
- Apak, T., İhracat Bedellerinin Tahsili-Kapatılması Ve Kur Farkları (Gümrük-Kambiyo-Vergi-Muhasebe), <http://www.alomaliye.com/2003/06/30/ihracat-bedellerinin-tahsili-kapatilmasi-ve-kur-farklari-gumruk-kambiyo-vergi-muhasebe/> (erişim tarihi: 25.05.2020).
- <https://comtrade.un.org/>, (erişim tarihi: 16.07.2020)
- https://docs.oracle.com/cd/B19306_01/datamine.102/b14339/5dmtasks.htm , (erişim tarihi: 25.05.2020).
- <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG>, (erişim tarihi: 17.10.2019).
- <https://www.doingbusiness.org/en/custom-query>, (erişim tarihi: 19.10.2019).
- The Python Tutorial, <https://docs.python.org/3/tutorial/index.html#the-python-tutorial> , son (erişim tarihi: 04.08.2019.)
- <http://www.fao.org/statistics/en/>, (erişim tarihi: 19.10.2019).
- <https://www.gumruklem.com.tr/gumruklem-terimleri-sozlugu/gtip-nedir/>, (erişim tarihi: 14.11.2019).
- <https://iccwbo.org/publication/icc-open-markets-index-2017/>, (erişim tarihi: 19.10.2019).
- <https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD/CHN>, (erişim tarihi: 19.10.2019).
- İhracat Yönetmeliği, T.C. Resmi Gazete, 06.06.2006, <https://www.resmigazete.gov.tr/eskiler/2006/06/20060606-10.htm> , erişim tarihi: 20.05.2020.
- <https://population.un.org/wpp/Download/Standard/Population/>, (erişim tarihi: 17.10.2019).
- <https://xgboost.readthedocs.io/en/latest/parameter.html#parameters-for-tree-boost>, (erişim tarihi: 30.11.2019).
- What is Python?, <https://docs.python.org/3/faq/general.html#what-is-python>, (erişim tarihi: 04.08.2019).

EKLER

Ek 1-OECD Ülkeleri ve Ülke Kodları Listesi

Ülke Adı	Ülke Kodu	Ülke Adı	Ülke Kodu
ABD	842	İtalya	381
Almanya	276	İzlanda	352
Avustralya	36	Japonya	392
Avusturya	40	Kanada	124
Belçika	56	Letonya	428
Çekya	203	Litvanya	440
Danimarka	208	Lüksemburg	442
Estonya	233	Macaristan	348
Finlandiya	246	Meksika	484
Fransa	251	Norveç	579
Güney Kore	410	Polonya	616
Hollanda	528	Portekiz	620
İngiltere	826	Şili	152
İrlanda	372	Slovakya	703
İspanya	724	Slovenya	705
İsrail	376	Yeni Zelanda	554
İsveç	752	Yunanistan	300
İsviçre	757		

Ek 2-Yaş Meyve Sebze Ürünleri GTİP'leri ve Açıklamaları

GTİP	AÇIKLAMALARI
070110	Patates-tohumluk
070190	Patates - taze / soğutulmuş
070200	Domates
070310	Soğan-şalotlar
070320	Sarımsak
070390	Pırasa
070410	Karnabahar- brokoli
070420	Brüksel lahanası
070490	Beyaz lahana-kırmızı lahana- alabaşlar
070511	Baş marul
070519	Diğer marullar
070521	Başlı hindiba
070529	Diğer hindiba
070610	Havuç-şalgam
070690	Diğer turp. kırmızı pancar. teke sakalı. vb yenilen kökler - taze / soğutulmuş
070700	Hıyarlar ve kornişonlar
070810	Bezelye
070820	Fasulye -barbunya
070890	Kabuklu/ kabuksuz baklagiller
070920	Kuşkonmazlar - taze / soğutulmuş
070930	Patlıcanlar - taze / soğutulmuş
070940	Yaprak kerevizler - taze / soğutulmuş.(kök kerevizleri hariç)
070951	Agaricus cinsi mantarlar - taze / soğutulmuş
070959	Cüce kız mantarları -kuzu mantarlar -diğer mantarlar
070960	Biber
070970	Ispanak
070991	Enginar
070993	Kabak
070999	Salatalık sebzeler (marul (lactuca sativa) ve hindiba (cichorium spp.)/pazı ve yaban enginarı/ kebere rezene/ tatlı mısır/ bamya/ semizotu/ maydanoz/ asma yaprağı ve diğerleri
080241	Kabuklu-kestane
080242	Kabuksuz-kestane
080310	Plantain
080390	Muz (plantainler hariç)
080410	Hurma
080420	İncir
080430	Ananas
080440	Avokado
080450	Guavo, mango, mangosten
080510	Portakal
080520	Mandarin, tangerin

Ek 2-Yaş Meyve Sebze Ürünleri GTİP'leri ve Açıklamaları (...devamı)

GTİP	AÇIKLAMALARI
080521	Mandarin, tangerin (klemantinler hariç) : satsuma, yerli mandarin, okitsu, mandarin ve vilking, tanjerin, mineola mandarin
080522	Klemantinler, monreale, king, freumont, nova, marisol mandarin
080529	Mandalina diğerleri
080540	Greyfurt
080550	Limon
080590	Diğer narenciye ürünleri
080610	Taze üzüm
080711	Karpuzlar
080719	Kavunlar
080720	Papaya
080810	Elma
080830	Armut
080840	Ayva
080910	Kayısı (zerdali dahil)
080921	Vişne
080929	Kiraz
080930	Şeftali-nektarin
080940	Erik
081010	Çilek
081020	Ahududu
081030	Frenk üzümü
081040	Yaban mersini
081050	Kivi
081060	Durian
081070	Trabzon hurması
081090	Demirhindi, mahun elması, ekmek ağacı meyvesi, sapodillo meyvesi/ çarkıfelek meyvesi. karambola ve pitahaya /nar/ muşmula/ kuşburnu ve diğer meyveler

Ek 3-Yapay Sinir Ağları Sonuçları

Deneme	(5, 5)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,872	0,853	47.445.810,78	60.624.973,07	17.373.932,55	22.970.837,25	1.148
2	0,907	0,871	40.358.877,20	56.733.519,59	16.318.675,67	22.778.387,34	1.497
3	0,898	0,861	42.421.112,21	58.917.565,45	16.671.149,67	23.109.267,99	1.300
4	0,916	0,873	38.624.102,17	56.244.918,55	14.969.832,64	21.626.734,94	1.270
5	0,887	0,851	44.692.819,10	60.961.200,67	17.069.977,52	23.121.568,57	1.170
6	0,907	0,869	40.497.069,40	57.261.827,73	16.139.139,44	22.624.187,86	1.320
7	0,925	0,883	36.473.732,13	53.959.908,83	14.383.338,13	20.888.083,99	1.371
8	0,914	0,876	39.013.221,98	55.628.869,31	15.153.257,74	21.908.730,64	1.310
9	0,914	0,877	38.804.797,55	55.419.724,18	15.406.428,86	22.014.806,31	1.411
10	0,890	0,854	44.142.787,55	60.358.081,21	17.499.427,50	23.945.377,59	1.299
Ortalama	0,905	0,869	41.247.433,01	57.611.058,86	16.098.515,97	22.498.798,25	1.310

Deneme	(5, 10)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,895	0,863	43.021.820,52	58.485.891,49	16.542.887,19	22.875.293,60	1.004
2	0,880	0,849	45.987.379,84	61.458.113,02	17.339.820,43	23.411.005,82	1.028
3	0,922	0,884	37.214.441,89	53.933.178,54	14.716.412,36	21.343.475,22	1.220
4	0,888	0,854	44.545.068,31	60.404.545,40	16.887.119,85	23.006.029,25	1.207
5	0,920	0,879	37.509.485,59	54.884.936,90	15.117.123,55	21.607.189,55	1.280
6	0,878	0,844	46.358.167,79	62.469.016,07	17.709.458,09	23.682.787,39	1.394
7	0,809	0,782	58.117.119,93	73.862.930,78	20.543.485,38	26.388.019,72	890
8	0,915	0,874	38.836.447,32	56.218.806,94	15.853.469,12	22.886.408,04	1.333
9	0,908	0,877	40.379.686,38	55.355.705,73	16.527.772,17	22.297.870,67	1.424
10	0,919	0,881	37.781.299,02	54.634.966,29	14.897.609,26	21.314.944,36	1.270
Ortalama	0,893	0,859	42.975.091,66	59.170.809,12	16.613.515,74	22.881.302,36	1.205

Deneme	(5, 15)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,921	0,881	37.365.778,91	54.502.951,97	14.673.054,05	21.302.499,11	1.108
2	0,904	0,863	41.415.822,31	58.573.655,05	16.863.149,04	23.481.799,78	1.108
3	0,907	0,869	40.592.838,77	57.262.404,24	16.243.666,23	22.551.151,18	1.155
4	0,911	0,868	39.578.794,05	57.486.941,04	15.572.078,62	22.714.475,77	1.070
5	0,917	0,873	38.175.435,64	56.254.039,88	14.962.317,27	21.896.708,39	1.113
6	0,907	0,871	40.436.375,24	56.882.349,05	15.172.788,99	21.685.583,15	1.072
7	0,915	0,876	38.696.641,77	55.780.659,59	15.680.631,21	22.388.248,89	1.355
8	0,891	0,858	43.970.056,32	59.559.123,23	17.221.406,19	23.695.263,49	1.174
9	0,922	0,882	37.044.563,43	54.248.175,72	14.775.576,19	21.851.903,46	1.552
10	0,871	0,834	47.724.599,52	64.474.508,27	17.247.240,59	23.780.096,89	1.001
Ortalama	0,907	0,868	40.500.090,60	57.502.480,80	15.841.190,84	22.534.773,01	1.178

Ek 3-Yapay Sinir Ağları Sonuçları (...devamı)

Deneme	(5, 20)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,888	0,859	44.533.826,41	59.358.385,77	17.330.822,62	23.139.243,07	1.047
2	0,910	0,876	39.899.816,88	55.737.004,02	15.965.444,83	22.311.153,23	900
3	0,908	0,869	40.317.601,86	57.195.592,10	15.545.958,18	22.105.753,99	1.011
4	0,905	0,868	40.972.341,88	57.530.854,63	15.829.713,62	22.644.463,31	1.019
5	0,917	0,883	38.292.257,24	54.097.273,75	15.722.081,75	22.099.495,41	1.132
6	0,883	0,843	45.499.500,25	62.646.596,21	16.577.926,40	23.487.803,55	1.064
7	0,915	0,881	38.722.036,01	54.554.934,83	15.407.126,97	22.189.423,58	1.134
8	0,879	0,842	46.247.412,27	62.749.842,31	16.822.090,75	23.113.127,76	1.192
9	0,918	0,870	37.962.655,77	56.902.352,88	14.762.037,84	22.052.808,76	1.267
10	0,927	0,884	35.929.843,34	53.881.827,65	14.557.593,26	21.664.902,06	1.317
Ortalama	0,905	0,868	40.837.729,19	57.465.466,42	15.852.079,62	22.480.817,47	1.108

Deneme	(5, 25)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,920	0,879	37.655.419,86	54.860.306,91	14.699.392,17	21.220.181,57	960
2	0,904	0,868	41.139.985,75	57.369.447,28	16.427.589,74	23.128.085,02	1.078
3	0,908	0,870	40.291.201,42	57.118.165,21	15.689.600,86	21.831.687,67	853
4	0,915	0,878	38.711.088,95	55.331.007,50	15.525.957,02	21.901.797,41	1.240
5	0,919	0,882	37.846.331,71	54.303.063,59	15.632.978,92	21.795.162,95	977
6	0,920	0,881	37.602.462,40	54.508.966,60	14.548.734,92	21.342.762,79	1.013
7	0,924	0,881	36.677.847,66	54.490.028,47	14.963.608,32	21.552.262,69	900
8	0,888	0,852	44.475.099,63	60.819.476,68	16.953.567,72	23.085.367,66	1.017
9	0,924	0,884	36.693.575,89	53.956.105,65	14.864.820,09	21.400.697,65	1.107
10	0,924	0,881	36.601.238,95	54.445.607,76	14.630.545,55	21.215.710,06	1.003
Ortalama	0,905	0,868	40.837.729,19	57.465.466,42	15.852.079,62	22.480.817,47	1.108

Deneme	(10, 5)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,930	0,894	35.210.488,52	51.430.048,19	13.136.847,46	19.497.223,18	1.391
2	0,930	0,891	35.220.434,53	52.239.038,43	13.503.283,96	20.238.699,73	1.286
3	0,936	0,901	33.445.967,07	49.821.763,75	12.429.610,81	18.807.955,32	1.177
4	0,936	0,894	33.681.400,03	51.528.764,71	12.174.095,50	18.967.934,44	1.372
5	0,939	0,897	32.880.734,80	50.736.846,50	11.916.249,26	18.689.883,40	1.307
6	0,934	0,893	34.199.754,68	51.764.078,97	12.800.462,20	19.856.739,27	1.214
7	0,937	0,896	33.317.089,11	51.110.844,49	12.380.646,17	19.314.043,31	1.240
8	0,931	0,891	34.991.482,59	52.191.853,76	12.805.990,61	19.557.984,92	1.183
9	0,899	0,859	42.234.069,18	59.337.511,99	14.544.252,04	21.454.201,09	1.330
10	0,935	0,891	33.811.258,40	52.283.815,41	12.205.294,12	19.522.497,58	1.373
Ortalama	0,931	0,891	34.899.267,89	52.244.456,62	12.789.673,21	19.590.716,22	1.287

Ek 3-Yapay Sinir Ağları Sonuçları (...devamı)

Deneme	(10, 10)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,936	0,890	33.580.015,09	52.420.270,39	12.485.450,82	19.902.482,38	1.140
2	0,944	0,902	31.448.699,55	49.554.035,26	11.625.098,18	18.878.873,14	1.049
3	0,936	0,896	33.623.422,16	50.900.757,71	12.236.518,21	19.203.674,23	1.079
4	0,930	0,892	35.099.374,19	51.871.740,56	13.462.554,19	19.897.578,88	1.164
5	0,936	0,892	33.529.010,07	52.019.698,77	12.713.033,67	19.650.894,97	910
6	0,940	0,898	32.564.001,85	50.408.858,44	12.317.162,60	19.332.413,89	1.231
7	0,940	0,897	32.588.250,89	50.649.794,51	12.326.941,66	19.611.485,65	1.298
8	0,931	0,892	34.846.686,12	52.028.233,95	13.066.481,83	19.938.505,27	1.059
9	0,938	0,893	33.111.326,59	51.828.175,59	12.460.239,21	19.769.189,31	1.313
10	0,937	0,896	33.483.317,60	50.978.410,16	12.328.766,70	19.111.278,26	1.119
Ortalama	0,937	0,895	33.387.410,41	51.265.997,53	12.502.224,71	19.529.637,60	1.136

Deneme	(10, 15)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,937	0,896	33.266.504,80	50.866.153,67	12.018.205,48	18.855.850,42	1.257
2	0,939	0,899	32.726.380,59	50.186.603,59	12.198.534,18	19.027.767,71	1.152
3	0,935	0,891	33.826.940,10	52.134.909,99	12.611.792,07	19.388.760,32	962
4	0,936	0,896	33.505.626,14	50.982.980,81	12.253.495,35	18.797.733,77	1.069
5	0,939	0,899	32.827.722,89	50.288.586,16	11.920.920,29	18.620.067,78	967
6	0,936	0,897	33.755.645,19	50.813.397,18	12.513.792,47	18.786.490,84	1.087
7	0,937	0,899	33.369.112,46	50.253.935,86	11.979.390,12	18.746.345,88	1.011
8	0,932	0,889	34.642.032,79	52.610.058,68	12.978.643,24	19.752.680,09	1.006
9	0,939	0,898	32.763.008,51	50.602.905,85	11.417.660,44	18.440.197,80	1.061
10	0,938	0,895	33.064.736,31	51.265.951,43	12.515.308,99	19.716.030,26	1.072
Ortalama	0,937	0,895	33.387.410,41	51.265.997,53	12.502.224,71	19.529.637,60	1.136

Deneme	(10, 20)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,934	0,891	34.171.798,44	52.233.104,86	12.683.208,72	19.460.745,93	990
2	0,938	0,895	33.056.484,45	51.152.351,99	12.164.038,78	18.982.090,92	934
3	0,939	0,895	32.866.763,75	51.289.255,76	12.215.171,36	19.278.218,23	966
4	0,934	0,896	34.115.243,83	50.916.652,96	12.607.390,38	19.176.004,25	842
5	0,942	0,898	32.057.944,95	50.538.938,78	11.677.392,89	18.918.734,49	1.033
6	0,942	0,898	31.906.554,59	50.537.629,27	11.387.457,15	18.677.648,14	1.060
7	0,940	0,904	32.438.661,54	48.979.886,16	12.285.014,03	18.864.688,68	1.005
8	0,940	0,896	32.566.440,53	50.874.413,72	11.753.878,32	18.642.013,15	932
9	0,939	0,897	32.810.142,66	50.763.042,19	11.852.803,38	19.023.047,91	887
10	0,936	0,892	33.605.983,45	51.879.178,12	11.935.291,03	18.882.521,44	896
Ortalama	0,938	0,896	32.959.601,82	50.916.445,38	12.056.164,60	18.990.571,31	955

Ek 3-Yapay Sinir Ağları Sonuçları (...devamı)

Deneme	(10, 25)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,929	0,890	35.432.953,02	52.518.985,74	13.378.261,39	19.929.569,55	907
2	0,924	0,887	36.572.890,59	53.238.179,37	13.150.172,15	19.312.124,35	884
3	0,931	0,897	35.011.787,17	50.867.031,97	13.341.848,24	19.772.466,63	1.018
4	0,941	0,897	32.193.684,84	50.676.598,50	11.765.197,22	19.107.228,67	929
5	0,931	0,894	34.976.465,28	51.524.918,89	12.594.897,69	18.816.759,10	834
6	0,938	0,898	33.172.548,06	50.588.077,49	12.236.031,33	18.840.228,89	925
7	0,936	0,895	33.610.644,85	51.208.725,85	11.966.497,54	18.792.633,86	931
8	0,935	0,898	33.995.238,85	50.614.195,44	12.630.881,03	19.382.407,94	835
9	0,940	0,899	32.473.195,83	50.113.623,86	11.609.383,33	18.139.023,42	942
10	0,941	0,901	32.397.286,83	49.633.365,85	11.967.800,97	18.842.125,44	1.020
Ortalama	0,935	0,896	33.983.669,53	51.098.370,30	12.464.097,09	19.093.456,79	923

Deneme	(15, 5)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,943	0,899	31.730.797,98	50.266.521,80	10.874.298,14	18.195.608,77	1.128
2	0,945	0,904	30.738.089,15	49.067.096,93	10.554.157,88	17.676.188,21	1.064
3	0,941	0,902	32.166.141,67	49.538.440,24	10.945.352,11	17.668.671,17	1.090
4	0,942	0,898	32.092.416,36	50.433.901,10	11.080.656,35	18.077.400,82	1.464
5	0,941	0,900	32.173.807,71	49.937.414,24	11.200.717,65	18.094.442,61	1.277
6	0,942	0,901	31.947.782,16	49.814.791,85	11.058.710,05	18.026.647,53	1.084
7	0,944	0,903	31.458.632,43	49.185.102,59	10.920.271,53	17.871.381,21	1.681
8	0,946	0,903	30.770.800,13	49.274.819,82	10.463.388,72	17.736.080,03	1.410
9	0,946	0,907	30.995.158,11	48.351.939,27	10.414.524,97	17.177.519,70	1.294
10	0,945	0,901	31.210.126,74	49.681.535,36	11.182.805,74	18.198.595,11	1.195
Ortalama	0,944	0,902	31.528.375,24	49.512.683,22	10.869.488,31	17.949.446,16	1.269

Deneme	(15, 10)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,945	0,906	31.234.866,14	48.532.682,36	11.239.666,35	18.071.183,29	1.085
2	0,948	0,908	30.302.102,95	48.088.207,99	10.549.193,76	17.501.192,11	1.201
3	0,940	0,898	32.461.627,24	50.525.448,97	10.944.688,09	18.044.605,34	917
4	0,946	0,905	30.769.996,70	48.687.764,36	10.490.603,18	17.686.538,77	1.073
5	0,944	0,902	31.411.689,44	49.448.658,53	11.035.711,24	18.262.504,67	923
6	0,949	0,911	29.967.918,87	47.044.806,51	10.458.626,75	17.350.498,61	999
7	0,942	0,901	32.076.791,57	49.761.660,33	11.274.654,51	18.182.787,60	963
8	0,945	0,901	31.255.382,37	49.786.051,48	10.828.974,14	18.268.107,88	1.020
9	0,949	0,907	29.994.210,81	48.172.261,76	10.261.537,53	17.645.372,59	952
10	0,945	0,901	31.238.901,97	49.799.299,24	10.819.538,72	18.164.748,24	839
Ortalama	0,945	0,904	31.071.348,81	48.984.684,15	10.790.319,43	17.917.753,91	997

Ek 3-Yapay Sinir Ağları Sonuçları (...devamı)

Deneme	(15, 15)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,945	0,902	31.093.341,83	49.549.469,17	10.870.444,46	18.122.496,95	1.085
2	0,946	0,904	30.857.904,09	48.887.296,51	10.538.786,32	17.476.398,16	988
3	0,940	0,901	32.514.337,71	49.746.603,57	11.251.139,86	18.028.824,32	948
4	0,946	0,903	30.804.777,72	49.131.256,92	10.326.486,14	17.735.508,62	1.082
5	0,946	0,901	30.960.686,63	49.789.868,67	10.697.817,09	18.099.733,95	968
6	0,945	0,902	31.040.185,42	49.592.481,12	10.415.414,89	17.469.508,75	968
7	0,947	0,905	30.667.434,28	48.715.951,52	10.290.834,68	17.340.414,32	967
8	0,944	0,905	31.334.602,96	48.837.300,79	10.944.082,89	18.065.834,64	1.005
9	0,946	0,903	30.927.990,44	49.303.845,14	10.912.534,10	17.889.182,56	972
10	0,948	0,906	30.353.393,21	48.412.840,19	10.412.128,18	17.853.791,22	998
Ortalama	0,945	0,903	31.055.465,43	49.196.691,36	10.665.966,86	17.845.032,81	998

Deneme	(15, 20)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,945	0,904	31.252.620,37	48.958.166,48	11.025.380,59	17.827.058,01	927
2	0,945	0,905	31.144.558,55	48.815.623,82	10.840.917,69	17.810.117,16	857
3	0,944	0,903	31.490.196,85	49.356.020,78	10.607.397,20	17.700.100,96	936
4	0,946	0,903	30.910.838,23	49.214.674,85	11.063.177,70	18.092.340,18	848
5	0,947	0,903	30.517.566,08	49.178.014,37	10.197.510,09	17.326.710,21	1.012
6	0,948	0,903	30.348.320,75	49.130.001,72	10.244.519,74	17.771.037,68	956
7	0,946	0,902	31.011.310,92	49.542.659,00	10.490.275,15	18.224.576,44	963
8	0,945	0,898	31.170.529,90	50.515.781,58	10.518.192,88	17.860.322,41	820
9	0,949	0,906	30.156.936,10	48.354.748,63	10.495.907,48	17.583.327,51	942
10	0,947	0,902	30.654.074,52	49.432.617,56	10.470.477,22	18.204.503,26	938
Ortalama	0,946	0,903	30.865.695,23	49.249.830,88	10.595.375,57	17.840.009,38	920

Deneme	(15, 25)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,945	0,901	31.258.308,41	49.693.500,55	11.345.016,27	18.394.203,66	955
2	0,948	0,903	30.332.508,02	49.142.223,59	10.205.543,59	17.802.489,70	848
3	0,949	0,907	29.914.427,43	48.217.632,02	10.291.796,56	17.578.871,41	960
4	0,939	0,901	32.889.108,67	49.711.086,99	10.970.155,63	17.625.161,38	702
5	0,943	0,901	31.738.456,64	49.634.065,42	11.095.441,49	18.405.154,44	902
6	0,946	0,907	30.938.488,72	48.157.757,18	10.496.726,02	17.464.169,69	880
7	0,948	0,908	30.336.842,76	48.076.061,82	10.479.476,64	17.509.758,19	983
8	0,943	0,902	31.863.754,78	49.607.328,22	10.840.799,78	17.696.622,70	809
9	0,948	0,906	30.307.111,37	48.377.948,34	10.116.331,20	17.024.649,49	827
10	0,941	0,897	32.181.035,64	50.650.731,94	10.730.969,02	17.870.032,77	837
Ortalama	0,945	0,903	31.176.004,24	49.126.833,61	10.657.225,62	17.737.111,34	870

Ek 3-Yapay Sinir Ağları Sonuçları (...devamı)

Deneme	(20, 5)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,941	0,902	32.392.243,59	49.380.069,97	10.710.992,17	17.207.724,41	1.208
2	0,948	0,908	30.225.619,02	48.055.971,95	9.867.309,91	16.871.511,99	1.449
3	0,946	0,903	30.848.312,98	49.209.120,48	9.976.775,68	17.308.285,38	1.171
4	0,949	0,905	30.099.857,83	48.776.667,15	9.842.601,60	17.073.375,71	938
5	0,948	0,908	30.259.594,92	47.852.356,60	10.060.278,40	17.065.100,46	1.106
6	0,950	0,909	29.578.672,85	47.593.581,17	9.799.251,93	17.037.151,90	1.055
7	0,951	0,908	29.408.469,33	47.995.033,15	9.831.413,25	17.174.454,38	1.214
8	0,950	0,909	29.664.632,89	47.707.518,09	9.863.390,16	17.217.380,94	1.186
9	0,950	0,907	29.710.690,75	48.232.127,53	9.746.281,43	17.377.735,70	1.005
10	0,954	0,914	28.537.223,05	46.358.606,64	9.405.720,95	16.629.530,04	1.566
Ortalama	0,949	0,907	30.072.531,72	48.116.105,27	9.910.401,55	17.099.683,38	1.190

Deneme	(20, 10)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,946	0,904	30.973.928,44	48.935.549,67	10.195.814,58	17.277.901,17	979
2	0,949	0,908	29.927.723,24	48.015.226,89	9.858.951,56	16.892.008,66	1.026
3	0,950	0,908	29.586.921,49	48.067.523,49	9.477.923,18	16.686.061,96	1.022
4	0,945	0,903	31.158.200,36	49.174.129,68	10.386.893,30	17.498.868,11	991
5	0,947	0,905	30.640.352,49	48.740.481,15	9.909.593,54	17.085.849,64	1.109
6	0,952	0,905	29.237.720,58	48.746.993,63	9.489.588,20	17.332.763,80	1.054
7	0,952	0,908	29.122.097,46	47.960.348,47	9.691.324,03	17.635.182,99	1.152
8	0,948	0,906	30.186.034,38	48.453.896,62	9.824.086,54	17.091.796,46	1.115
9	0,951	0,908	29.388.185,74	47.984.969,62	9.569.625,20	17.410.552,42	1.074
10	0,952	0,908	28.991.626,21	47.937.497,00	9.463.332,20	17.105.801,58	969
Ortalama	0,949	0,906	29.921.279,04	48.401.661,62	9.786.713,23	17.201.678,68	1.049

Deneme	(20, 15)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,947	0,901	30.579.437,50	49.733.529,95	10.079.366,30	17.511.214,47	884
2	0,945	0,899	31.225.384,09	50.239.880,59	10.194.207,82	17.743.275,37	856
3	0,948	0,908	30.177.982,85	48.048.343,01	9.575.088,98	16.627.204,56	928
4	0,951	0,905	29.526.002,85	48.717.896,67	9.543.318,24	17.108.374,93	961
5	0,951	0,908	29.505.229,07	48.053.811,92	9.426.835,70	16.552.877,53	1.120
6	0,949	0,907	30.038.099,42	48.321.639,16	9.649.810,25	16.896.312,48	784
7	0,949	0,907	29.907.461,51	48.350.074,40	9.717.089,48	17.162.707,15	849
8	0,951	0,909	29.444.738,06	47.620.639,51	9.693.883,04	17.064.024,61	947
9	0,952	0,909	29.263.863,52	47.686.480,66	9.666.152,49	17.126.390,22	956
10	0,946	0,903	30.927.725,26	49.313.384,12	10.056.655,09	17.313.883,43	915
Ortalama	0,949	0,906	30.059.592,41	48.608.568,00	9.760.240,74	17.110.626,48	920

Ek 3-Yapay Sinir Ağları Sonuçları (...devamı)

Deneme	(20, 20)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,950	0,906	29.583.320,56	48.379.944,77	9.665.217,84	17.067.478,38	948
2	0,951	0,906	29.417.117,01	48.450.747,55	9.907.898,35	17.799.567,81	932
3	0,945	0,905	31.092.275,22	48.621.974,22	10.093.103,01	17.102.718,95	802
4	0,948	0,906	30.336.339,65	48.427.900,31	9.860.253,94	16.972.429,27	815
5	0,955	0,910	28.338.161,41	47.570.384,28	9.125.835,86	16.881.626,21	1.010
6	0,952	0,907	29.033.987,43	48.208.579,95	9.167.332,68	16.679.613,47	940
7	0,951	0,908	29.313.434,63	48.066.622,88	9.595.052,57	16.941.239,03	941
8	0,952	0,909	29.096.743,85	47.780.195,74	9.352.262,87	17.036.726,28	993
9	0,951	0,908	29.563.944,54	48.054.321,47	9.632.200,45	17.121.280,04	861
10	0,950	0,908	29.838.008,15	47.849.722,30	9.449.498,76	16.453.895,66	827
Ortalama	0,951	0,907	29.561.333,25	48.141.039,35	9.584.865,63	17.005.657,51	895

Deneme	(20, 25)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,952	0,910	29.234.221,01	47.456.220,39	9.631.628,84	16.970.281,87	848
2	0,952	0,909	29.119.643,13	47.718.791,53	9.275.449,61	16.770.523,99	910
3	0,953	0,908	28.741.613,19	47.858.659,09	9.408.303,16	17.247.711,04	889
4	0,948	0,902	30.457.143,09	49.426.134,13	9.869.075,62	17.252.497,65	805
5	0,952	0,907	29.191.414,24	48.228.340,51	9.289.228,07	16.905.170,16	853
6	0,952	0,905	29.106.030,54	48.667.311,01	9.526.893,03	17.373.360,91	848
7	0,954	0,911	28.503.340,08	47.291.079,88	9.146.495,42	16.904.926,89	916
8	0,946	0,904	30.864.917,23	48.968.534,19	10.213.064,68	17.182.703,98	762
9	0,953	0,912	28.879.994,85	47.015.626,95	9.376.638,21	16.557.976,19	832
10	0,953	0,911	28.806.911,68	47.245.081,81	9.453.165,80	16.758.051,37	841
Ortalama	0,952	0,908	29.290.522,90	47.987.577,95	9.518.994,24	16.992.320,41	850

Deneme	(25, 5)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,952	0,906	29.242.695,53	48.423.534,90	9.573.817,82	17.409.519,75	1.301
2	0,945	0,906	31.094.751,53	48.597.039,01	10.157.323,29	17.073.652,50	928
3	0,952	0,911	29.183.770,42	47.299.820,08	9.143.168,83	16.413.344,14	1.073
4	0,950	0,907	29.581.681,29	48.162.371,22	9.560.282,91	16.856.273,58	1.125
5	0,950	0,908	29.840.704,21	48.019.941,66	9.381.458,22	16.571.987,38	979
6	0,952	0,909	29.264.247,61	47.812.122,83	9.345.523,80	16.903.277,66	976
7	0,951	0,907	29.416.542,06	48.167.161,41	9.258.064,29	16.546.318,91	1.031
8	0,950	0,909	29.803.771,09	47.791.450,14	9.509.779,41	16.561.351,55	1.259
9	0,953	0,909	28.743.517,95	47.696.611,46	9.194.861,69	16.846.870,93	1.114
10	0,950	0,907	29.723.534,29	48.180.855,98	9.254.634,65	16.730.045,29	904
Ortalama	0,951	0,908	29.589.521,60	48.015.090,87	9.437.891,49	16.791.264,17	1.063

Ek 3-Yapay Sinir Ağları Sonuçları (...devamı)

Deneme	(25, 10)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,953	0,911	28.760.218,02	47.060.100,13	8.764.828,47	16.235.689,72	1.163
2	0,948	0,904	30.199.589,42	49.063.249,05	9.614.599,59	16.961.761,59	896
3	0,951	0,908	29.298.019,06	48.065.953,47	9.079.145,84	16.332.801,15	982
4	0,951	0,908	29.417.325,53	47.948.981,51	9.334.153,65	16.657.744,46	955
5	0,954	0,910	28.659.145,57	47.494.211,05	8.957.529,37	16.441.586,52	994
6	0,948	0,904	30.191.707,11	48.950.023,73	9.596.275,01	16.891.544,00	978
7	0,955	0,910	28.181.147,72	47.514.922,41	8.968.590,40	16.885.604,00	959
8	0,952	0,909	29.134.234,71	47.750.086,39	9.303.943,32	16.673.944,73	1.002
9	0,954	0,910	28.619.234,53	47.360.820,22	8.875.579,59	16.590.273,22	915
10	0,953	0,910	28.891.892,81	47.435.253,25	9.149.981,93	16.596.483,91	957
Ortalama	0,952	0,908	29.135.251,45	47.864.360,12	9.145.608,17	16.626.743,33	980

Deneme	(25, 15)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,952	0,907	29.134.004,77	48.337.641,52	9.066.689,77	16.782.392,22	855
2	0,948	0,904	30.256.599,06	49.022.253,84	9.581.715,09	17.060.257,02	803
3	0,954	0,910	28.416.988,95	47.393.852,36	9.058.454,22	16.697.035,91	886
4	0,950	0,906	29.579.511,15	48.574.103,22	9.439.090,26	16.657.379,99	992
5	0,953	0,905	28.685.052,85	48.699.488,84	8.874.532,74	17.214.256,61	995
6	0,949	0,906	30.085.136,52	48.570.901,02	9.484.898,13	16.743.863,96	788
7	0,949	0,906	29.995.081,95	48.569.137,46	9.766.507,28	16.996.391,93	936
8	0,953	0,909	28.787.735,17	47.803.596,02	8.828.467,78	16.618.389,34	954
9	0,954	0,911	28.623.623,53	47.162.666,65	9.175.026,36	16.720.424,31	987
10	0,954	0,910	28.646.262,89	47.529.846,58	9.049.879,06	16.814.839,78	1.016
Ortalama	0,952	0,907	29.287.374,81	48.166.348,75	9.238.914,93	16.830.523,11	921

Deneme	(25, 20)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,950	0,907	29.627.862,58	48.138.488,36	9.380.564,57	16.728.618,95	957
2	0,953	0,909	28.915.226,46	47.639.973,25	9.093.492,07	16.524.708,77	856
3	0,949	0,908	29.976.356,49	47.889.403,89	9.433.911,20	16.606.125,11	742
4	0,954	0,911	28.416.780,31	47.097.757,62	8.797.676,64	16.645.394,02	936
5	0,951	0,907	29.306.231,72	48.252.315,41	9.264.299,31	17.060.500,84	787
6	0,953	0,909	28.879.338,62	47.654.249,54	8.998.484,62	16.683.094,05	894
7	0,953	0,910	28.967.484,84	47.545.141,79	8.992.636,38	16.613.617,97	1.047
8	0,951	0,907	29.516.453,69	48.099.615,81	9.187.822,53	16.662.396,57	857
9	0,955	0,908	28.159.780,34	47.881.264,24	8.802.406,16	16.900.033,77	960
10	0,956	0,912	27.832.919,97	46.903.288,15	8.501.121,04	16.261.904,61	1.003
Ortalama	0,953	0,909	28.959.843,50	47.710.149,81	9.045.241,45	16.668.639,47	904

Ek 3-Yapay Sinir Ağları Sonuçları (...devamı)

Deneme	(25, 25)						
	R^2		$RMSE$		MAE		Süre (sn.)
	Eğitim	Test	Eğitim	Test	Eğitim	Test	
1	0,955	0,911	28.143.376,79	47.302.358,61	8.739.936,24	16.839.024,59	883
2	0,952	0,910	29.122.184,88	47.326.122,78	8.953.511,07	16.233.394,25	898
3	0,948	0,905	30.257.670,43	48.709.933,97	9.402.387,77	16.670.333,74	779
4	0,953	0,913	28.728.400,15	46.528.752,51	9.099.449,51	16.178.717,91	1.027
5	0,951	0,909	29.562.222,75	47.599.517,76	9.319.291,28	16.334.684,53	864
6	0,954	0,909	28.578.983,67	47.684.095,39	9.028.066,05	16.956.087,41	793
7	0,953	0,911	28.666.995,69	47.242.506,69	9.008.939,21	16.630.281,46	875
8	0,958	0,912	27.204.982,63	46.942.779,21	8.455.818,39	16.593.889,61	929
9	0,952	0,911	29.134.052,45	47.211.330,44	9.227.703,31	16.424.720,37	802
10	0,953	0,913	28.674.463,51	46.713.446,34	8.781.878,49	16.066.051,18	831
Ortalama	0,953	0,910	28.807.333,30	47.326.084,37	9.001.698,13	16.492.718,51	868

Ek 4- Algoritmaların Test Veri Seti Üzerinde Ürün Bazında R^2 Kriteri Açısından Ortalama Performansları

Ürün kodu	Ürün	XG_Boost	Rastgele Orman	YSA	Maksimum R^2
070110	Patates-tohumluk	0,797	0,452	0,678	0,797
070190	Patates - taze / soğutulmuş	0,900	0,895	0,883	0,900
070200	Domates	0,940	0,961	0,968	0,968
070310	Soğan-şalotlar	0,959	0,905	0,945	0,959
070320	Sarımsak	0,918	0,846	0,942	0,942
070390	Pırasa	0,843	0,354	0,624	0,843
070410	Karnabahar- brokoli	0,923	0,792	0,940	0,940
070420	Brüksel lahanası	0,934	0,858	0,949	0,949
070490	Beyaz lahana-kırmızı lahana- alabaşlar	0,959	0,857	0,963	0,963
070511	Baş marul	0,797	0,791	0,847	0,847
070519	Diğer marullar	0,920	0,903	0,964	0,964
070521	Başlı hindiba	0,258	0,305	0,438	0,438
070529	Diğer hindiba	0,280	0,568	0,407	0,568
070610	Havuç-şalgam	0,836	0,749	0,860	0,860
070690	Diğer turp. kırmızı pancar. teke sakalı. vb yenilen kökler - taze / soğutulmuş	0,678	0,647	0,754	0,754
070700	Hıyarlar ve kornişonlar	0,914	0,937	0,979	0,979
070810	Bezelye	0,921	0,850	0,857	0,921
070820	Fasulye -barbunya	0,917	0,737	0,872	0,917
070890	Kabuklu/ kabuksuz baklagiller	0,710	0,514	0,630	0,710
070920	Kuşkonmazlar - taze / soğutulmuş	0,995	0,986	0,997	0,997
070930	Patlıcanlar - taze / soğutulmuş	0,844	0,776	0,866	0,866
070940	Yaprak kerevizler - taze / soğutulmuş.(kök kerevizleri hariç)	0,924	0,566	0,864	0,924
070951	Agaricus cinsi mantarlar - taze / soğutulmuş	0,944	0,801	0,986	0,986
070959	Cüce kız mantarları -kuzu mantarlar -diğer mantarlar	0,792	0,638	0,737	0,792
070960	Biber	0,967	0,973	0,984	0,984
070970	Ispanak	0,915	0,681	0,802	0,915
070991	Enginar	0,662	0,371	0,296	0,662
070993	Kabak	0,959	0,924	0,965	0,965
070999	Salatalık sebzeler (marul (lactuca sativa) ve hindiba (cichorium spp.)/pazı ve yaban enginarı/ kebere rezene/ tatlı mısır/ bamya/ semizotu/ maydanoz/ asma yaprağı ve diğerleri	0,882	0,815	0,902	0,902
080241	Kabuklu-kestane	0,974	0,247	0,503	0,974
080242	Kabuksuz-kestane	0,694	0,290	0,307	0,694
080310	Plantain	0,989	0,924	0,978	0,989
080390	Muz (plantainler hariç)	0,954	0,948	0,970	0,970
080410	Hurma	0,791	0,834	0,764	0,834
080420	İncir	0,745	0,715	0,728	0,745
080430	Ananas	0,983	0,977	0,985	0,985
080440	Avokado	0,967	0,979	0,979	0,979

Ek 4-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında R^2 Kriteri Açısından Ortalama Performansları (...devamı)

Ürün kodu	Ürün	XG_Boost	Rastgele Orman	YSA	Maksimum R^2
080450	Guavo, mango, mangosten	0,961	0,953	0,961	0,961
080510	Portakal	0,809	0,845	0,855	0,855
080520	Mandarin, tangerin	0,977	0,929	0,986	0,986
080521	Mandarin, tangerin (klemantinler hariç) : satsuma, yerli mandarin, okitsu,mandarin ve vilking, tanjerin, mineola mandarin	0,507	0,508	0,896	0,896
080522	Klemantinler, monreale, king, freumont, nova, marisol mandarin	0,423	0,667	0,546	0,667
080529	Mandalina diğerleri	0,703	0,617	0,646	0,703
080540	Greyfurt	0,861	0,347	0,642	0,861
080550	Limon	0,933	0,938	0,952	0,952
080590	Diğer narenciye ürünleri	0,211	0,463	0,106	0,463
080610	Taze üzüm	0,968	0,980	0,984	0,984
080711	Karpuzlar	0,889	0,827	0,914	0,914
080719	Kavunlar	0,948	0,931	0,946	0,948
080720	Papaya	0,978	0,868	0,962	0,978
080810	Elma	0,949	0,937	0,927	0,949
080830	Armut	0,822	0,701	0,905	0,905
080840	Ayva	0,570	0,534	0,155	0,570
080910	Kayısı (zerdali dahil)	0,773	0,335	0,635	0,773
080921	Vişne	0,197	0,242	0,276	0,276
080929	Kiraz	0,743	0,640	0,734	0,743
080930	Şeftali-nektarin	0,752	0,597	0,859	0,859
080940	Erik	0,889	0,817	0,849	0,889
081010	Çilek	0,927	0,920	0,966	0,966
081020	Ahududu	0,980	0,964	0,981	0,981
081030	Frenk üzümü	0,193	0,799	0,231	0,799
081040	Yaban mersini	0,961	0,959	0,974	0,974
081050	Kivi	0,941	0,873	0,819	0,941
081060	Durian	0,332	0,177	0,891	0,891
081070	Trabzon hurması	0,675	0,482	0,508	0,675
081090	Demirhindi, mahun elması, ekmek ağacı meyvesi, sapodillo meyvesi/ çarkıfelek meyvesi. karambola ve pitahaya /nar/ muşmula/ kuşburnu ve diğer meyveler	0,954	0,876	0,910	0,954

Ek 5-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında *RMSE* Kriteri Açısından Ortalama Performansları

Ürün kodu	Ürün	XG_Boost	Rastgele Orman	YSA	Minimum RMSE değeri	Ürün bazında ortalama ithalat değeri	Minimum RMSE'nin ortalama ithalata oranı
070110	Patates-tohumluk	12.278.995	37.751.928	18.829.042	12.278.995	6.835.157	1,796
070190	Patates - taze / soğutulmuş	60.631.804	74.483.389	63.788.153	60.631.804	31.564.154	1,921
070200	Domates	167.035.996	164.910.165	128.494.448	128.494.448	54.556.047	2,355
070310	Soğan-şalotlar	29.102.034	57.031.281	34.128.776	29.102.034	14.805.942	1,966
070320	Sarımsak	15.519.635	18.750.876	29.032.606	15.519.635	7.469.740	2,078
070390	Pırasa	9.115.704	25.082.723	18.818.384	9.115.704	4.522.779	2,016
070410	Karnabahar-brokoli	19.371.273	29.629.720	17.264.564	17.264.564	9.010.175	1,916
070420	Brüksel lahanası	4.900.969	28.922.884	12.209.946	4.900.969	2.567.817	1,909
070490	Beyaz lahana-kırmızı lahana-alabaşlar	23.433.268	53.165.709	23.412.271	23.412.271	10.163.670	2,304
070511	Baş marul	23.596.753	19.182.832	26.751.233	19.182.832	10.700.808	1,793
070519	Diğer marullar	34.627.943	52.745.810	23.767.546	23.767.546	11.590.154	2,051
070521	Başlı hindiba	17.803.204	35.470.000	11.467.563	11.467.563	4.851.913	2,364
070529	Diğer hindiba	18.955.692	32.506.891	13.824.445	13.824.445	5.174.821	2,671
070610	Havuç-şalgam	23.457.308	27.765.783	23.321.205	23.321.205	11.418.371	2,042
070690	Diğer turp. kırmızı pancar. teke sakalı. vb yenilen kökler - taze / soğutulmuş	13.878.247	24.985.628	18.675.652	13.878.247	6.860.731	2,023
070700	Hıyarlar ve kornişonlar	78.225.259	96.225.085	39.057.802	39.057.802	24.213.128	1,613
070810	Bezelye	9.195.962	23.101.265	18.576.245	9.195.962	4.906.017	1,874
070820	Fasulye -barbunya	20.807.703	36.050.106	30.467.916	20.807.703	8.504.634	2,447
070890	Kabuklu/kabuksuz baklagiller	2.348.863	24.976.074	13.382.908	2.348.863	1.472.939	1,595
070920	Kuşkonmazlar - taze / soğutulmuş	19.980.276	38.664.502	27.092.928	19.980.276	8.769.881	2,278
070930	Pathıcanlar - taze / soğutulmuş	11.060.402	17.788.267	18.050.353	11.060.402	3.880.240	2,850
070940	Yaprak kerevizler - taze / soğutulmuş.(kök kerevizleri hariç)	7.757.837	17.782.513	12.746.705	7.757.837	3.536.421	2,194
070951	Agaricus cinsi mantarlar - taze / soğutulmuş	27.659.613	42.798.670	13.539.870	13.539.870	10.821.461	1,251
070959	Cüce kız mantarları -kuzu mantarlar -diğer mantarlar	16.270.998	22.279.768	22.490.916	16.270.998	8.524.032	1,909
070960	Biber	87.543.392	123.285.935	56.954.344	56.954.344	34.433.403	1,654
070970	Ispanak	10.897.383	19.765.765	16.735.493	10.897.383	5.477.433	1,990

Ek 5-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında RMSE Kriteri Açısından Ortalama Performansları (...devamı)

Ürün kodu	Ürün	XG_Boost	Rastgele Orman	YSA	Minimum RMSE değeri	Ürün bazında ortalama ithalat değeri	Minimum RMSE'nin ortalama ithalata oranı
070991	Enginar	3.678.614	17.902.370	14.118.389	3.678.614	2.103.167	1,749
070993	Kabak	22.224.879	41.772.554	27.034.004	22.224.879	8.246.191	2,695
070999	Salatalık sebzeler (marul (lactuca sativa) ve hindiba (cichorium spp.)/pazı ve yaban enginarı/ kebere rezene/ tatlı mısır/ bamya/ semizotu/ maydanoz/ asma yaprağı ve diğerleri	43.360.023	59.603.570	39.028.422	39.028.422	20.815.703	1,875
080241	Kabuklu-kestane	4.230.376	21.380.177	19.159.387	4.230.376	2.447.609	1,728
080242	Kabuksuz-kestane	4.239.180	17.882.219	14.756.255	4.239.180	2.434.836	1,741
080310	Plantain	6.204.069	20.864.853	15.571.593	6.204.069	4.335.151	1,431
080390	Muz (plantainler hariç)	155.983.538	172.102.692	127.656.668	127.656.668	74.198.610	1,720
080410	Hurma	16.141.869	14.340.168	20.653.141	14.340.168	7.841.458	1,829
080420	İncir	10.365.650	13.664.796	18.029.300	10.365.650	4.726.230	2,193
080430	Ananas	24.013.792	40.308.106	36.774.917	24.013.792	11.431.366	2,101
080440	Avokado	184.041.630	240.743.407	164.041.318	164.041.318	78.519.773	2,089
080450	Guavo, mango, mangosten	36.509.156	53.907.520	38.063.874	36.509.156	17.536.316	2,082
080510	Portakal	69.007.381	61.361.263	64.539.734	61.361.263	30.528.819	2,010
080520	Mandarin, tangerin	31.203.633	57.180.893	25.950.377	25.950.377	15.172.609	1,710
080521	Mandarin, tangerin (klemantinler hariç) : satsuma, yerli mandarin, okitsu,mandarin ve vilking, tanjerin, mineola mandarin	62.625.575	63.357.782	74.099.822	62.625.575	28.220.206	2,219
080522	Klemantinler, monreale, king, freumont, nova, marisol mandarin	73.702.505	71.254.980	106.841.129	71.254.980	39.908.630	1,785
080529	Mandalina diğerleri	13.893.453	14.733.959	122.216.634	13.893.453	10.472.260	1,327
080540	Greyfurt	21.870.913	46.490.280	32.933.065	21.870.913	9.418.388	2,322
080550	Limon	57.418.397	72.495.230	42.654.419	42.654.419	28.106.913	1,518
080590	Diğer narenciye ürünleri	2.878.027	23.176.851	13.301.950	2.878.027	1.930.320	1,491
080610	Taze üzüm	92.868.962	118.290.571	65.758.433	65.758.433	37.439.408	1,756

Ek 5-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında RMSE Kriteri Açısından Ortalama Performansları (...devamı)

Ürün kodu	Ürün	XG_Boost	Rastgele Orman	YSA	Minimum RMSE değeri	Ürün bazında ortalama ithalat değeri	Minimum RMSE'nin ortalama ithalata oranı
080711	Karpuzlar	42.690.542	61.842.967	37.049.072	37.049.072	16.652.030	2,225
080719	Kavunlar	27.629.910	38.344.828	40.581.916	27.629.910	11.610.318	2,380
080720	Papaya	4.784.103	32.321.478	18.850.373	4.784.103	2.281.931	2,097
080810	Elma	46.286.472	51.120.247	59.529.128	46.286.472	27.641.150	1,675
080830	Armut	29.918.207	38.507.604	33.347.758	29.918.207	13.015.093	2,299
080840	Ayva	2.198.756	19.333.000	15.331.842	2.198.756	1.067.967	2,059
080910	Kayısı (zerdali dahil)	12.425.770	25.900.408	19.384.668	12.425.770	5.354.752	2,321
080921	Vişne	10.743.300	37.549.970	13.747.622	10.743.300	4.333.400	2,479
080929	Kiraz	27.004.214	30.464.816	29.859.046	27.004.214	10.915.572	2,474
080930	Şeftali-nektarin	49.275.260	60.241.177	42.296.329	42.296.329	20.526.712	2,061
080940	Erik	9.110.508	16.266.217	20.592.996	9.110.508	4.772.955	1,909
081010	Çilek	66.855.970	81.357.290	38.233.189	38.233.189	25.478.388	1,501
081020	Ahududu	76.459.848	132.302.247	65.096.686	65.096.686	33.922.564	1,919
081030	Frenk üzümü	3.552.874	25.355.393	14.051.299	3.552.874	2.048.469	1,734
081040	Yaban mersini	104.973.430	144.385.292	81.933.944	81.933.944	45.058.411	1,818
081050	Kivi	36.651.125	57.985.277	54.990.337	36.651.125	18.785.096	1,951
081060	Durian	2.425.972	20.689.186	13.982.539	2.425.972	1.471.744	1,648
081070	Trabzon hurması	9.449.390	15.303.688	15.321.107	9.449.390	3.529.233	2,677
081090	Demirhindi, mahun elması, ekmek ağacı meyvesi, sapodillo meyvesi/ çarkıfelek meyvesi. karambola ve pitahaya /nar/ muşmula/ kuşburnu ve diğer meyveler	22.780.836	33.211.253	26.485.172	22.780.836	11.296.643	2,017

Ek 6-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında MAE Kriteri Açısından Ortalama Performansları

Ürün kodu	Ürün	XG_Boost	Rastgele Orman	YSA	Minimum MAE	Ürün bazında ortalama ithalat değeri	Minimum MAE'nin ortalama ithalata oranı
070110	Patates-tohumluk	6.835.157	20.200.232	9.425.223	6.835.157	13.648.496	0,501
070190	Patates - taze / soğutulmuş	31.564.154	38.122.460	31.465.908	31.465.908	81.078.577	0,388
070200	Domates	54.556.047	76.175.321	57.073.731	54.556.047	219.419.471	0,249
070310	Soğan-şalotlar	14.805.942	32.499.839	19.520.433	14.805.942	58.644.637	0,252
070320	Sarımsak	7.469.740	11.606.161	12.902.188	7.469.740	22.167.079	0,337
070390	Pırasa	4.522.779	15.084.829	7.862.874	4.522.779	10.345.666	0,437
070410	Karnabahar-brokoli	9.010.175	16.124.816	12.333.322	9.010.175	29.177.739	0,309
070420	Brüksel lahanası	2.567.817	19.936.642	3.309.643	2.567.817	6.059.295	0,424
070490	Beyaz lahana-kırmızı lahana-alabaşlar	10.163.670	25.269.775	14.481.579	10.163.670	39.283.633	0,259
070511	Baş marul	10.700.808	12.222.804	13.445.648	10.700.808	23.136.585	0,463
070519	Diğer marullar	11.590.154	25.256.132	15.255.272	11.590.154	47.794.461	0,242
070521	Başlı hindiba	4.851.913	21.912.390	3.561.046	3.561.046	2.920.335	1,219
070529	Diğer hindiba	5.174.821	19.881.023	5.045.123	5.045.123	4.536.793	1,112
070610	Havuç-şalgam	11.418.371	15.080.245	14.127.962	11.418.371	26.856.736	0,425
070690	Diğer turp. kırmızı pancar. teke sakalı. vb yenilen kökler - taze / soğutulmuş	6.860.731	15.808.724	9.374.622	6.860.731	13.846.658	0,495
070700	Hıyarlar ve kornişonlar	24.213.128	40.434.247	16.828.972	16.828.972	75.738.840	0,222
070810	Bezelye	4.906.017	13.866.763	6.576.363	4.906.017	12.482.117	0,393
070820	Fasulye - barbunya	8.504.634	19.502.004	12.855.717	8.504.634	28.700.180	0,296
070890	Kabuklu/ kabuksuz baklagiller	1.472.939	16.734.733	3.645.579	1.472.939	2.257.505	0,652
070920	Kuşkonmazlar - taze / soğutulmuş	8.769.881	16.350.216	10.048.597	8.769.881	44.072.196	0,199
070930	Patlıcanlar - taze / soğutulmuş	3.880.240	10.760.707	6.739.049	3.880.240	12.267.079	0,316
070940	Yaprak kerevizler - taze /soğutulmuş.(kök kerevizleri hariç)	3.536.421	12.351.165	4.552.415	3.536.421	8.860.618	0,399
070951	Agaricus cinsi mantarlar - taze / soğutulmuş	10.821.461	20.159.312	8.095.331	8.095.331	28.429.014	0,285
070959	Cüce kız mantarları -kuzu mantarlar -diğer mantarlar	8.524.032	14.132.089	11.556.947	8.524.032	19.023.556	0,448

Ek 6-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında MAE Kriteri Açısından Ortalama Performansları (...devamı)

Ürün kodu	Ürün	XG_Boost	Rastgele Orman	YSA	Minimum MAE	Ürün bazında ortalama ithalat değeri	Minimum MAE'nin ortalama ithalata oranı
070960	Biber	34.433.403	48.801.474	26.166.081	26.166.081	142.956.351	0,183
070970	Ispanak	5.477.433	12.858.624	6.506.342	5.477.433	10.805.587	0,507
070991	Enginar	2.103.167	13.378.609	3.826.789	2.103.167	1.776.197	1,184
070993	Kabak	8.246.191	18.891.581	11.138.535	8.246.191	34.913.241	0,236
070999	Salatalık sebzeler (marul (lactuca sativa) ve hindiba (cichorium spp.)/pazı ve yaban enginarı/ kebere rezene/ tatlı mısır/ bamya/ semizotu/ maydanoz/ asma yaprağı ve diğerleri	20.815.703	32.228.928	18.566.276	18.566.276	57.433.348	0,323
080241	Kabuklu-kestane	2.447.609	13.587.759	7.302.515	2.447.609	6.632.700	0,369
080242	Kabuksuz-kestane	2.434.836	13.232.573	3.844.241	2.434.836	1.834.495	1,327
080310	Plantain	4.335.151	12.004.993	6.018.186	4.335.151	15.411.728	0,281
080390	Muz (plantainler hariç)	74.198.610	85.312.332	61.638.678	61.638.678	314.051.687	0,196
080410	Hurma	7.841.458	8.936.833	10.185.436	7.841.458	16.915.261	0,464
080420	İncir	4.726.230	8.849.497	7.302.325	4.726.230	10.994.062	0,430
080430	Ananas	11.431.366	23.513.458	16.405.058	11.431.366	65.662.818	0,174
080440	Avokado	78.519.773	97.598.736	65.843.634	65.843.634	182.088.647	0,362
080450	Guavo, mango, mangosten	17.536.316	30.125.104	19.890.921	17.536.316	63.844.624	0,275
080510	Portakal	30.528.819	33.031.413	26.574.438	26.574.438	88.916.903	0,299
080520	Mandarin, tangerin	15.172.609	24.468.057	17.185.395	15.172.609	90.598.126	0,167
080521	Mandarin, tangerin (klemantinler hariç) : satsuma, yerli mandarin, okitsu, mandarin ve vilking, tanjerin, mineola mandarin	28.220.206	27.759.891	23.709.232	23.709.232	36.269.742	0,654
080522	Klemantinler, monreale, king, freumont, nova, marisol mandarin	39.908.630	39.168.123	41.509.588	39.168.123	49.241.733	0,795

Ek 6-Algoritmaların Test Veri Seti Üzerinde Ürün Bazında MAE Kriteri Açısından Ortalama Performansları (...devamı)

080529	Mandalina diğerleri	10.472.260	11.385.252	39.730.732	10.472.260	7.779.129	1,346
080540	Greyfurt	9.418.388	18.919.735	16.423.870	9.418.388	20.541.431	0,459
080550	Limon	28.106.913	38.132.121	20.534.898	20.534.898	83.316.572	0,246
080590	Diğer narenciye ürünleri	1.930.320	15.507.424	3.149.663	1.930.320	1.399.958	1,379
080610	Taze üzüm	37.439.408	52.001.512	32.919.559	32.919.559	187.602.229	0,175
080711	Karpuzlar	16.652.030	31.097.186	19.497.980	16.652.030	48.328.042	0,345
080719	Kavunlar	11.610.318	22.951.887	16.105.291	11.610.318	48.162.180	0,241
080720	Papaya	2.281.931	14.635.301	5.666.977	2.281.931	9.328.561	0,245
080810	Elma	27.641.150	29.726.904	36.020.204	27.641.150	94.331.102	0,293
080830	Armut	13.015.093	21.348.634	19.096.619	13.015.093	39.072.880	0,333
080840	Ayva	1.067.967	14.627.236	3.458.901	1.067.967	495.719	2,154
080910	Kayısı (zerdali dahil)	5.354.752	15.242.532	9.822.605	5.354.752	9.573.912	0,559
080921	Vişne	4.333.400	20.754.281	3.252.590	3.252.590	1.267.881	2,565
080929	Kiraz	10.915.572	15.723.093	14.476.286	10.915.572	25.004.812	0,437
080930	Şeftali-nektarin	20.526.712	32.091.317	24.128.631	20.526.712	48.229.675	0,426
080940	Erik	4.772.955	10.003.347	8.722.687	4.772.955	13.352.041	0,357
081010	Çilek	25.478.388	37.666.570	19.968.912	19.968.912	79.559.840	0,251
081020	Ahududu	33.922.564	55.731.368	32.061.284	32.061.284	89.498.214	0,358
081030	Frenk üzümü	2.048.469	16.550.540	3.810.086	2.048.469	2.015.107	1,017
081040	Yaban mersini	45.058.411	62.875.923	41.506.751	41.506.751	100.443.401	0,413
081050	Kivi	18.785.096	31.544.803	28.434.009	18.785.096	60.620.958	0,310
081060	Durian	1.471.744	16.368.194	3.256.621	1.471.744	782.398	1,881
081070	Trabzon hurması	3.529.233	10.811.985	6.049.848	3.529.233	7.192.055	0,491
081090	Demirhindi, mahun elması, ekmek ağacı meyvesi, sapodillo meyvesi/ çarkıfelek meyvesi. karambola ve pitahaya /nar/ muşmula/ kuşburnu ve diğer meyveler	11.296.643	17.353.603	14.640.435	11.296.643	35.830.693	0,315

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri

						XG_Boost		Rastgele Orman		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	% Fark	Tahmin Değeri	% Fark	Tahmin Değeri	% Fark
2018	36	Avustralya	080440	Avokado	56.438.973	81.838.824	45%	33.263.828	-41%	37.465.064	-34%
2019	36	Avustralya	080440	Avokado	50.774.000	80.542.860	59%	32.664.347	-36%	38.172.376	-25%
2018	40	Avusturya	080440	Avokado	37.716.187	19.335.490	-49%	30.160.946	-20%	31.555.643	-16%
2019	40	Avusturya	080440	Avokado	40.903.000	22.103.090	-46%	30.499.450	-25%	32.130.294	-21%
2018	56	Belçika	080440	Avokado	78.452.822	60.748.372	-23%	32.595.705	-58%	163.833.696	109%
2019	56	Belçika	080440	Avokado	97.092.000	57.141.244	-41%	32.595.705	-66%	166.276.563	71%
2018	124	Kanada	080440	Avokado	223.249.331	202.954.660	-9%	142.197.161	-36%	212.225.217	-5%
2019	124	Kanada	080440	Avokado	237.830.000	203.787.970	-14%	137.867.323	-42%	212.157.291	-11%
2018	152	Şili	080440	Avokado	41.074.958	10.281.420	-75%	5.103.755	-88%	13.837.194	-66%
2019	152	Şili	080440	Avokado	28.127.000	34.255.504	22%	5.103.755	-82%	13.191.110	-53%
2018	203	Çekya	080440	Avokado	18.431.649	6.929.424	-62%	7.252.440	-61%	2.851.917	-85%
2019	203	Çekya	080440	Avokado	18.184.000	9.938.070	-45%	7.201.406	-60%	4.339.965	-76%
2018	208	Danimarka	080440	Avokado	62.977.889	42.157.670	-33%	13.222.714	-79%	27.878.512	-56%
2019	208	Danimarka	080440	Avokado	65.119.000	43.403.172	-33%	16.178.944	-75%	29.870.315	-54%
2018	233	Estonya	080440	Avokado	4.452.599	2.130.807	-52%	1.415.321	-68%	676.930	-85%
2019	233	Estonya	080440	Avokado	5.317.000	1.835.880	-65%	1.798.990	-66%	676.930	-87%
2018	246	Finlandiya	080440	Avokado	28.855.717	24.754.838	-14%	7.769.070	-73%	19.657.818	-32%
2019	246	Finlandiya	080440	Avokado	28.423.000	24.195.874	-15%	7.638.760	-73%	18.659.717	-34%
2018	251	Fransa	080440	Avokado	0	355.940.030		177.557.102		322.292.469	
2019	251	Fransa	080440	Avokado	520.249.000	220.539.460	-58%	166.484.344	-68%	320.878.040	-38%
2018	276	Almanya	080440	Avokado	317.687.295	140.842.480	-56%	135.872.759	-57%	279.215.403	-12%
2019	276	Almanya	080440	Avokado	329.357.000	43.695.256	-87%	115.826.146	-65%	291.241.956	-12%
2018	348	Macaristan	080440	Avokado	5.541.093	4.016.461	-28%	3.594.294	-35%	676.930	-88%
2019	348	Macaristan	080440	Avokado	9.205.000	4.757.690	-48%	3.594.294	-61%	676.930	-93%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2018	352	İzlanda	080440	Avokado	4.823.134	3.562.858	-26%	1.291.858	-73%	610.362	-87%
2019	352	İzlanda	080440	Avokado	4.853.000	4.033.097	-17%	1.274.947	-74%	676.930	-86%
2018	372	İrlanda	080440	Avokado	24.241.522	45.956.484	90%	14.840.339	-39%	16.677.523	-31%
2019	372	İrlanda	080440	Avokado	22.609.000	55.051.460	143%	30.032.433	33%	16.160.910	-29%
2018	376	İsrail	080440	Avokado	2.000	3.221.803	160990%	17.414.200	870610%	5.494.346	274617%
2019	376	İsrail	080440	Avokado	0	11.764.976		26.574.957		3.354.630	
2018	381	İtalya	080440	Avokado	54.128.102	15.563.814	-71%	47.630.594	-12%	85.058.023	57%
2019	381	İtalya	080440	Avokado	81.505.000	15.796.626	-81%	47.630.594	-42%	86.826.527	7%
2018	392	Japonya	080440	Avokado	217.866.183	256.757.440	18%	161.614.566	-26%	187.945.836	-14%
2019	392	Japonya	080440	Avokado	239.702.000	214.832.940	-10%	159.512.198	-33%	188.860.795	-21%
2018	428	Letonya	080440	Avokado	8.340.215	2.463.528	-70%	1.849.042	-78%	676.930	-92%
2019	428	Letonya	080440	Avokado	9.913.000	3.304.321	-67%	1.849.042	-81%	676.930	-93%
2018	440	Litvanya	080440	Avokado	12.279.457	10.014.810	-18%	6.049.917	-51%	7.199.244	-41%
2019	440	Litvanya	080440	Avokado	14.376.000	8.576.512	-40%	6.080.741	-58%	6.237.658	-57%
2018	442	Lüksemburg	080440	Avokado	4.709.253	3.361.835	-29%	1.808.817	-62%	3.653.140	-22%
2019	442	Lüksemburg	080440	Avokado	0	3.812.752		1.808.817		4.628.392	
2019	484	Meksika	080440	Avokado	0	15.437.760		17.196.659		2.520.964	
2018	528	Hollanda	080440	Avokado	540.225.597	354.157.100	-34%	296.435.158	-45%	394.311.970	-27%
2019	528	Hollanda	080440	Avokado	1.043.851.000	342.319.040	-67%	296.564.811	-72%	392.591.194	-62%
2019	554	Yeni Zelanda	080440	Avokado	62.000	0	-100%	7.781.430	12451%	676.930	992%
2018	579	Norveç	080440	Avokado	59.472.386	45.544.028	-23%	31.828.592	-46%	36.577.153	-38%
2019	579	Norveç	080440	Avokado	64.982.000	42.412.304	-35%	31.731.256	-51%	33.844.828	-48%
2018	616	Polonya	080440	Avokado	52.384.220	21.074.950	-60%	23.233.104	-56%	19.746.987	-62%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2019	616	Polonya	080440	Avokado	72.045.000	21.758.560	-70%	23.386.427	-68%	20.904.288	-71%
2018	620	Portekiz	080440	Avokado	11.209.286	7.731.445	-31%	7.212.870	-36%	16.747.522	49%
2019	620	Portekiz	080440	Avokado	18.048.000	7.143.756	-60%	7.174.697	-60%	18.681.707	4%
2018	410	Güney Kore	080440	Avokado	0	25.152.732		69.558.198		38.330.992	
2019	410	Güney Kore	080440	Avokado	31.785.000	29.709.340	-7%	71.556.026	125%	40.902.376	29%
2018	703	Slovakya	080440	Avokado	5.839.164	5.650.103	-3%	5.461.084	-6%	642.022	-89%
2019	703	Slovakya	080440	Avokado	7.565.000	0	-100%	5.490.393	-27%	676.930	-91%
2018	724	İspanya	080440	Avokado	280.256.209	81.446.270	-71%	86.888.758	-69%	148.078.142	-47%
2019	724	İspanya	080440	Avokado	347.039.000	74.843.990	-78%	86.888.758	-75%	151.181.053	-56%
2018	752	İsveç	080440	Avokado	71.501.906	59.259.892	-17%	31.820.710	-55%	36.520.563	-49%
2019	752	İsveç	080440	Avokado	0	60.220.944		31.820.710		36.570.772	
2018	757	İsviçre	080440	Avokado	57.409.077	53.177.256	-7%	32.108.958	-44%	47.778.248	-17%
2019	757	İsviçre	080440	Avokado	62.526.000	42.015.988	-33%	32.011.126	-49%	46.042.452	-26%
2018	826	Birleşik Krallık	080440	Avokado	267.404.441	247.453.040	-7%	145.975.858	-45%	205.398.978	-23%
2019	826	Birleşik Krallık	080440	Avokado	342.698.000	121.513.200	-65%	134.810.432	-61%	211.154.161	-38%
2018	842	ABD	080440	Avokado	2.448.438.747	1.937.459.200	-21%	1.506.543.934	-38%	1.923.900.581	-21%
2019	842	ABD	080440	Avokado	2.864.125.000	1.936.515.000	-32%	1.542.837.774	-46%	2.000.546.989	-30%
2018	36	Avustralya	070960	Biber	3.073.158	0	-100%	12.144.484	295%	2.043.254	-34%
2019	36	Avustralya	070960	Biber	1.720.000	0	-100%	11.304.215	557%	2.481.332	44%
2018	40	Avusturya	070960	Biber	73.934.644	87.788.570	19%	59.561.219	-19%	52.446.439	-29%
2019	40	Avusturya	070960	Biber	115.686.000	92.025.016	-20%	59.844.148	-48%	52.999.391	-54%
2018	56	Belçika	070960	Biber	83.874.875	85.327.800	2%	50.011.104	-40%	136.035.048	62%
2019	56	Belçika	070960	Biber	86.140.000	84.130.360	-2%	50.585.876	-41%	134.775.912	56%
2018	124	Kanada	070960	Biber	266.165.923	238.505.310	-10%	266.288.069	0%	251.944.114	-5%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2019	124	Kanada	070960	Biber	286.488.000	253.616.080	-11%	272.000.755	-5%	260.594.866	-9%
2018	152	Şili	070960	Biber	19.171	3.101.585	16079%	3.148.893	16325%	1.092.990	5601%
2019	152	Şili	070960	Biber	13.000	1.496.183	11409%	3.826.741	29336%	1.766.272	13487%
2018	203	Çekya	070960	Biber	79.793.162	59.779.736	-25%	57.323.268	-28%	75.654.707	-5%
2019	203	Çekya	070960	Biber	76.874.000	63.302.036	-18%	56.352.291	-27%	75.023.877	-2%
2018	208	Danimarka	070960	Biber	60.741.913	50.035.800	-18%	50.568.369	-17%	49.716.163	-18%
2019	208	Danimarka	070960	Biber	53.684.000	52.441.070	-2%	51.081.157	-5%	49.972.031	-7%
2018	233	Estonya	070960	Biber	7.222.403	13.084.164	81%	6.833.290	-5%	8.085.316	12%
2019	233	Estonya	070960	Biber	7.317.000	13.807.608	89%	7.017.828	-4%	8.141.511	11%
2018	246	Finlandiya	070960	Biber	42.537.752	43.027.456	1%	27.816.113	-35%	38.894.388	-9%
2019	246	Finlandiya	070960	Biber	42.077.000	39.233.356	-7%	27.675.572	-34%	40.076.754	-5%
2018	251	Fransa	070960	Biber	0	236.968.700		210.674.211		331.129.509	
2019	251	Fransa	070960	Biber	252.542.000	241.620.300	-4%	230.381.365	-9%	353.458.332	40%
2018	276	Almanya	070960	Biber	831.781.097	892.065.900	7%	590.353.082	-29%	648.562.625	-22%
2019	276	Almanya	070960	Biber	800.075.000	242.718.750	-70%	454.000.504	-43%	662.734.301	-17%
2018	348	Macaristan	070960	Biber	16.460.244	13.141.426	-20%	9.699.760	-41%	14.865.485	-10%
2019	348	Macaristan	070960	Biber	18.730.000	12.991.004	-31%	8.938.515	-52%	15.120.299	-19%
2018	352	İzlanda	070960	Biber	3.951.313	5.407.582	37%	4.396.834	11%	2.633.422	-33%
2019	352	İzlanda	070960	Biber	4.146.000	8.952.896	116%	4.354.002	5%	3.564.151	-14%
2018	372	İrlanda	070960	Biber	42.893.413	44.485.850	4%	47.649.244	11%	34.406.020	-20%
2019	372	İrlanda	070960	Biber	41.247.000	56.806.536	38%	58.772.267	42%	34.654.718	-16%
2018	376	İsrail	070960	Biber	444.000	10.656.222	2300%	25.622.414	5671%	2.165.607	388%
2019	376	İsrail	070960	Biber	2.204.000	21.379.972	870%	41.555.986	1785%	3.222.712	46%
2018	381	İtalya	070960	Biber	114.641.291	132.215.520	15%	97.076.275	-15%	90.143.164	-21%
2019	381	İtalya	070960	Biber	126.386.000	109.136.170	-14%	101.255.620	-20%	91.143.641	-28%
2018	392	Japonya	070960	Biber	135.540.459	130.782.940	-4%	144.311.717	6%	118.070.694	-13%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2019	392	Japonya	070960	Biber	128.712.000	134.834.420	5%	148.129.635	15%	122.376.629	-5%
2018	428	Letonya	070960	Biber	9.184.372	22.721.074	147%	9.623.725	5%	10.847.260	18%
2019	428	Letonya	070960	Biber	10.015.000	26.249.420	162%	9.373.344	-6%	10.495.307	5%
2018	440	Litvanya	070960	Biber	21.821.407	77.159.864	254%	48.233.037	121%	79.594.003	265%
2019	440	Litvanya	070960	Biber	16.748.000	74.402.090	344%	47.439.090	183%	66.102.308	295%
2018	442	Lüksemburg	070960	Biber	4.964.490	5.050.090	2%	4.838.136	-3%	7.071.455	42%
2019	442	Lüksemburg	070960	Biber	0	4.961.204		4.980.620		7.063.182	
2018	484	Meksika	070960	Biber	891.512	1.686.546	89%	11.695.391	1212%	36.099.863	3949%
2019	484	Meksika	070960	Biber	1.549.000	1.169.730	-24%	10.384.612	570%	33.671.535	2074%
2018	528	Hollanda	070960	Biber	196.885.284	214.942.450	9%	170.204.120	-14%	200.600.630	2%
2019	528	Hollanda	070960	Biber	187.427.000	205.102.660	9%	171.858.504	-8%	203.769.352	9%
2018	554	Yeni Zelanda	070960	Biber	2.160.400	18.503.350	756%	13.956.542	546%	6.362.764	195%
2019	554	Yeni Zelanda	070960	Biber	1.403.000	14.156.860	909%	16.089.718	1047%	6.082.529	334%
2018	579	Norveç	070960	Biber	58.394.576	55.852.404	-4%	58.321.856	0%	49.751.875	-15%
2019	579	Norveç	070960	Biber	59.162.000	50.844.852	-14%	58.110.631	-2%	54.941.362	-7%
2018	616	Polonya	070960	Biber	128.196.025	111.910.080	-13%	81.343.684	-37%	114.765.682	-10%
2019	616	Polonya	070960	Biber	143.671.000	112.558.830	-22%	81.411.057	-43%	112.633.923	-22%
2018	620	Portekiz	070960	Biber	21.369.791	27.147.542	27%	20.449.497	-4%	22.332.714	5%
2019	620	Portekiz	070960	Biber	21.291.000	24.202.752	14%	28.864.834	36%	21.913.961	3%
2018	410	Güney Kore	070960	Biber	0	88.445.020		123.475.263		9.780.838	
2019	410	Güney Kore	070960	Biber	0	88.051.384		129.776.634		10.310.522	
2018	703	Slovakya	070960	Biber	30.777.480	30.181.052	-2%	24.027.954	-22%	25.639.536	-17%
2019	703	Slovakya	070960	Biber	33.299.000	37.589.784	13%	23.691.351	-29%	28.980.215	-13%
2018	724	İspanya	070960	Biber	81.117.306	65.259.600	-20%	37.098.516	-54%	63.188.050	-22%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2019	724	İspanya	070960	Biber	84.267.000	66.148.708	-22%	37.130.956	-56%	63.186.821	-25%
2018	752	İsveç	070960	Biber	74.459.583	71.206.470	-4%	66.328.667	-11%	82.614.432	11%
2019	752	İsveç	070960	Biber	0	72.167.544		66.261.294		83.957.923	
2018	757	İsviçre	070960	Biber	77.088.597	80.742.420	5%	67.705.462	-12%	67.499.218	-12%
2019	757	İsviçre	070960	Biber	79.280.000	76.969.976	-3%	67.707.344	-15%	67.366.653	-15%
2018	826	Birleşik Krallık	070960	Biber	435.215.465	383.000.030	-12%	472.768.911	9%	453.005.960	4%
2019	826	Birleşik Krallık	070960	Biber	464.433.000	379.689.100	-18%	479.081.559	3%	451.620.190	-3%
2018	842	ABD	070960	Biber	1.588.501.053	1.429.625.300	-10%	1.094.128.806	-31%	1.610.234.918	1%
2019	842	ABD	070960	Biber	1.794.431.000	1.540.352.500	-14%	1.107.272.612	-38%	1.694.443.518	-6%
2018	36	Avustralya	070200	Domates	1.070.244	8.387.453	684%	11.782.110	1001%	26.654.083	2390%
2019	36	Avustralya	070200	Domates	1.441.000	13.054.414	806%	11.221.088	679%	27.030.681	1776%
2018	40	Avusturya	070200	Domates	95.151.237	135.308.780	42%	102.674.977	8%	99.097.564	4%
2019	40	Avusturya	070200	Domates	97.197.000	134.643.980	39%	104.821.594	8%	98.365.650	1%
2018	56	Belçika	070200	Domates	101.880.389	167.765.710	65%	153.526.507	51%	281.684.524	176%
2019	56	Belçika	070200	Domates	107.956.000	166.979.650	55%	153.858.143	43%	279.282.518	159%
2018	124	Kanada	070200	Domates	320.982.736	298.917.440	-7%	367.283.685	14%	353.149.967	10%
2019	124	Kanada	070200	Domates	305.569.000	298.768.600	-2%	376.336.457	23%	366.016.125	20%
2018	152	Şili	070200	Domates	789.359	0	-100%	7.853.354	895%	668.095	-15%
2019	152	Şili	070200	Domates	622.000	5.833.746	838%	9.225.236	1383%	670.260	8%
2018	203	Çekya	070200	Domates	117.998.962	103.271.660	-12%	110.645.563	-6%	112.678.209	-5%
2019	203	Çekya	070200	Domates	113.612.000	108.749.810	-4%	110.544.888	-3%	111.962.520	-1%
2018	208	Danimarka	070200	Domates	76.451.082	78.753.304	3%	63.518.312	-17%	75.779.360	-1%
2019	208	Danimarka	070200	Domates	75.455.000	78.968.480	5%	65.528.236	-13%	75.334.202	0%
2018	233	Estonya	070200	Domates	21.531.163	31.650.726	47%	19.072.313	-11%	23.075.881	7%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2019	233	Estonya	070200	Domates	21.122.000	21.537.960	2%	19.574.658	-7%	23.120.328	9%
2018	246	Finlandiya	070200	Domates	63.326.093	60.002.532	-5%	59.225.974	-6%	67.786.546	7%
2019	246	Finlandiya	070200	Domates	64.285.000	59.492.930	-7%	59.604.306	-7%	68.926.613	7%
2018	251	Fransa	070200	Domates	0	349.690.140		479.451.618		631.017.611	
2019	251	Fransa	070200	Domates	704.362.000	350.714.660	-50%	411.426.126	-42%	649.984.900	-8%
2018	276	Almanya	070200	Domates	1.474.206.920	1.468.654.300	0%	905.693.957	-39%	964.123.784	-35%
2019	276	Almanya	070200	Domates	1.398.469.000	202.897.340	-85%	743.982.605	-47%	979.700.354	-30%
2018	348	Macaristan	070200	Domates	28.607.036	12.668.862	-56%	18.363.795	-36%	24.212.689	-15%
2019	348	Macaristan	070200	Domates	23.654.000	29.161.272	23%	16.392.216	-31%	25.135.703	6%
2018	352	İzlanda	070200	Domates	3.203.654	3.330.535	4%	5.111.693	60%	676.930	-79%
2019	352	İzlanda	070200	Domates	3.121.000	3.836.597	23%	5.113.801	64%	676.930	-78%
2018	372	İrlanda	070200	Domates	54.317.852	54.906.140	1%	69.248.830	27%	63.157.423	16%
2019	372	İrlanda	070200	Domates	48.103.000	50.416.724	5%	86.232.050	79%	63.726.434	32%
2018	376	İsrail	070200	Domates	21.301.000	43.252.836	103%	53.760.952	152%	9.845.817	-54%
2019	376	İsrail	070200	Domates	27.743.000	41.077.148	48%	79.728.060	187%	9.130.223	-67%
2018	381	İtalya	070200	Domates	146.083.933	162.844.620	11%	138.856.878	-5%	169.185.337	16%
2019	381	İtalya	070200	Domates	174.179.000	160.059.570	-8%	139.940.162	-20%	170.273.460	-2%
2018	392	Japonya	070200	Domates	34.129.487	27.162.626	-20%	71.559.149	110%	76.900.293	125%
2019	392	Japonya	070200	Domates	34.838.000	27.371.196	-21%	69.664.233	100%	81.436.214	134%
2018	428	Letonya	070200	Domates	28.757.058	37.525.596	30%	21.651.594	-25%	29.678.139	3%
2019	428	Letonya	070200	Domates	31.551.000	24.137.610	-23%	21.319.276	-32%	29.108.086	-8%
2018	440	Litvanya	070200	Domates	39.716.459	165.651.970	317%	89.511.529	125%	137.584.089	246%
2019	440	Litvanya	070200	Domates	40.381.000	165.015.020	309%	88.755.273	120%	123.438.264	206%
2018	442	Lüksemburg	070200	Domates	16.265.790	15.570.026	-4%	13.939.419	-14%	14.079.853	-13%
2019	442	Lüksemburg	070200	Domates	0	15.855.202		13.857.633		13.770.201	
2018	484	Meksika	070200	Domates	1.225.961	6.901.334	463%	14.918.678	1117%	37.003.360	2918%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2019	484	Meksika	070200	Domates	557.000	40.241.300	7125%	14.125.037	2436%	29.054.858	5116%
2018	528	Hollanda	070200	Domates	317.415.651	321.201.900	1%	285.982.215	-10%	399.318.032	26%
2019	528	Hollanda	070200	Domates	320.647.000	317.463.740	-1%	284.007.978	-11%	404.251.920	26%
2018	554	Yeni Zelanda	070200	Domates	1.116.900	48.440.460	4237%	29.182.243	2513%	4.456.123	299%
2019	554	Yeni Zelanda	070200	Domates	431.000	4.079.206	846%	12.957.862	2906%	3.470.441	705%
2018	579	Norveç	070200	Domates	63.662.759	51.884.572	-19%	76.674.625	20%	89.841.067	41%
2019	579	Norveç	070200	Domates	58.704.000	52.035.340	-11%	77.344.234	32%	98.137.209	67%
2018	616	Polonya	070200	Domates	242.254.812	191.644.940	-21%	132.378.137	-45%	211.722.193	-13%
2019	616	Polonya	070200	Domates	244.767.000	180.641.010	-26%	128.866.804	-47%	208.019.876	-15%
2018	620	Portekiz	070200	Domates	48.381.421	41.391.570	-14%	35.005.577	-28%	38.704.136	-20%
2019	620	Portekiz	070200	Domates	48.958.000	38.358.892	-22%	34.058.197	-30%	29.143.525	-40%
2019	410	Güney Kore	070200	Domates	0	55.073.796		140.245.808		44.567.337	
2018	703	Slovakya	070200	Domates	37.452.716	38.150.320	2%	42.464.007	13%	45.524.940	22%
2019	703	Slovakya	070200	Domates	45.224.000	38.588.884	-15%	40.412.048	-11%	47.628.236	5%
2018	724	İspanya	070200	Domates	145.848.484	118.579.304	-19%	79.888.015	-45%	114.554.289	-21%
2019	724	İspanya	070200	Domates	173.675.000	119.834.750	-31%	79.607.777	-54%	108.875.147	-37%
2018	752	İsveç	070200	Domates	170.197.756	141.696.180	-17%	148.234.844	-13%	178.469.695	5%
2019	752	İsveç	070200	Domates	0	140.910.980		148.140.944		180.927.773	
2018	757	İsviçre	070200	Domates	85.797.830	88.320.776	3%	84.233.345	-2%	104.947.381	22%
2019	757	İsviçre	070200	Domates	83.505.000	128.936.776	54%	82.448.111	-1%	108.935.029	30%
2018	826	Birleşik Krallık	070200	Domates	700.165.785	694.604.200	-1%	783.501.361	12%	688.063.390	-2%
2019	826	Birleşik Krallık	070200	Domates	646.517.000	689.997.200	7%	880.687.872	36%	689.292.416	7%
2018	842	ABD	070200	Domates	2.486.067.108	2.279.987.700	-8%	1.913.949.806	-23%	2.208.877.947	-11%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2019	842	ABD	070200	Domates	2.420.263.000	2.338.118.100	-3%	1.974.644.176	-18%	2.342.855.057	-3%
2018	36	Avustralya	080390	Muz (plantainler hariç)	269.036	16.988.168	6214%	78.266.219	28991%	12.496.160	4545%
2019	36	Avustralya	080390	Muz (plantainler hariç)	234.000	96.384.700	41090%	151.598.581	64686%	13.323.630	5594%
2018	40	Avusturya	080390	Muz (plantainler hariç)	124.633.504	186.173.520	49%	167.229.453	34%	134.529.428	8%
2019	40	Avusturya	080390	Muz (plantainler hariç)	132.937.000	186.594.340	40%	170.278.912	28%	134.358.593	1%
2018	56	Belçika	080390	Muz (plantainler hariç)	1.317.283.107	1.379.766.000	5%	1.103.409.937	-16%	1.189.885.750	-10%
2019	56	Belçika	080390	Muz (plantainler hariç)	1.073.867.000	1.425.939.500	33%	1.102.600.578	3%	1.191.407.681	11%
2018	124	Kanada	080390	Muz (plantainler hariç)	407.315.489	434.839.360	7%	374.351.162	-8%	443.767.871	9%
2019	124	Kanada	080390	Muz (plantainler hariç)	410.226.000	433.791.360	6%	403.364.290	-2%	455.488.171	11%
2018	152	Şili	080390	Muz (plantainler hariç)	79.122.874	57.123.176	-28%	72.136.006	-9%	75.463.015	-5%
2019	152	Şili	080390	Muz (plantainler hariç)	83.400.000	58.372.000	-30%	72.616.556	-13%	72.644.989	-13%
2018	203	Çekya	080390	Muz (plantainler hariç)	148.264.698	109.846.020	-26%	116.440.872	-21%	129.824.020	-12%
2019	203	Çekya	080390	Muz (plantainler hariç)	142.718.000	119.932.590	-16%	111.983.973	-22%	128.699.759	-10%
2018	208	Danimarka	080390	Muz (plantainler hariç)	89.538.947	83.391.256	-7%	81.660.926	-9%	78.727.117	-12%
2019	208	Danimarka	080390	Muz (plantainler hariç)	68.891.000	81.990.670	19%	83.613.888	21%	78.495.248	14%
2018	233	Estonya	080390	Muz (plantainler hariç)	17.123.411	7.727.306	-55%	15.801.436	-8%	27.352.242	60%
2019	233	Estonya	080390	Muz (plantainler	16.281.000	13.266.510	-19%	16.166.255	-1%	27.662.509	70%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
				hariç)							
2018	246	Finlandiya	080390	Muz (plantainler hariç)	98.909.026	96.922.410	-2%	89.368.579	-10%	82.515.811	-17%
2019	246	Finlandiya	080390	Muz (plantainler hariç)	91.838.000	98.568.280	7%	89.907.579	-2%	83.719.213	-9%
2018	251	Fransa	080390	Muz (plantainler hariç)	0	642.171.140		663.525.221		494.498.860	
2019	251	Fransa	080390	Muz (plantainler hariç)	536.981.000	658.696.800	23%	707.285.641	32%	509.669.698	-5%
2018	276	Almanya	080390	Muz (plantainler hariç)	940.347.228	1.137.093.200	21%	901.317.676	-4%	887.267.533	-6%
2019	276	Almanya	080390	Muz (plantainler hariç)	939.406.000	523.152.000	-44%	871.936.075	-7%	903.140.078	-4%
2018	348	Macaristan	080390	Muz (plantainler hariç)	35.890.583	36.231.176	1%	47.260.979	32%	31.135.547	-13%
2019	348	Macaristan	080390	Muz (plantainler hariç)	34.902.000	41.211.350	18%	83.722.689	140%	33.343.838	-4%
2018	352	İzlanda	080390	Muz (plantainler hariç)	3.258.116	4.002.598	23%	4.242.434	30%	3.771.553	16%
2019	352	İzlanda	080390	Muz (plantainler hariç)	2.899.000	8.139.705	181%	4.251.505	47%	1.717.911	-41%
2018	372	İrlanda	080390	Muz (plantainler hariç)	58.276.397	78.550.290	35%	118.622.638	104%	60.256.783	3%
2019	372	İrlanda	080390	Muz (plantainler hariç)	63.003.000	65.015.560	3%	122.463.182	94%	61.165.260	-3%
2018	376	İsrail	080390	Muz (plantainler hariç)	109.000	0	-100%	42.669.247	39046%	1.098.288	908%
2019	376	İsrail	080390	Muz (plantainler hariç)	61.000	3.952.281	6379%	77.028.165	126176%	508.813	734%
2018	381	İtalya	080390	Muz (plantainler hariç)	547.280.616	512.291.100	-6%	490.831.141	-10%	497.698.300	-9%
2019	381	İtalya	080390	Muz (plantainler	509.260.000	503.718.140	-1%	490.365.828	-4%	498.537.975	-2%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
				hariç)							
2018	392	Japonya	080390	Muz (plantainler hariç)	911.520.352	852.833.400	-6%	873.842.822	-4%	870.673.274	-4%
2019	392	Japonya	080390	Muz (plantainler hariç)	958.477.000	866.071.700	-10%	876.526.252	-9%	883.922.056	-8%
2018	428	Letonya	080390	Muz (plantainler hariç)	28.822.489	34.678.010	20%	26.159.276	-9%	36.152.127	25%
2019	428	Letonya	080390	Muz (plantainler hariç)	32.901.000	32.136.838	-2%	26.092.334	-21%	32.143.492	-2%
2018	440	Litvanya	080390	Muz (plantainler hariç)	35.630.861	47.723.580	34%	41.579.564	17%	87.176.411	145%
2019	440	Litvanya	080390	Muz (plantainler hariç)	32.540.000	45.101.932	39%	44.652.932	37%	69.728.499	114%
2018	442	Lüksemburg	080390	Muz (plantainler hariç)	8.762.344	1.462.817	-83%	44.612.216	409%	7.420.982	-15%
2019	442	Lüksemburg	080390	Muz (plantainler hariç)	0	1.373.931		44.581.023		5.874.409	
2018	484	Meksika	080390	Muz (plantainler hariç)	0	0		54.282.936		59.709.843	
2019	484	Meksika	080390	Muz (plantainler hariç)	0	0		73.526.360		52.770.007	
2018	528	Hollanda	080390	Muz (plantainler hariç)	743.379.258	388.941.400	-48%	284.333.201	-62%	400.601.225	-46%
2019	528	Hollanda	080390	Muz (plantainler hariç)	1.029.232.000	381.099.600	-63%	283.415.761	-72%	399.999.842	-61%
2018	554	Yeni Zelanda	080390	Muz (plantainler hariç)	68.940.530	73.505.740	7%	72.198.288	5%	66.344.174	-4%
2019	554	Yeni Zelanda	080390	Muz (plantainler hariç)	71.640.000	73.620.080	3%	72.659.903	1%	65.743.821	-8%
2018	579	Norveç	080390	Muz (plantainler hariç)	66.008.801	81.703.650	24%	82.072.845	24%	96.478.326	46%
2019	579	Norveç	080390	Muz (plantainler	65.831.000	86.295.470	31%	88.650.148	35%	99.688.244	51%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
				hariç)							
2018	616	Polonya	080390	Muz (plantainler hariç)	301.830.951	220.965.280	-27%	218.745.506	-28%	221.418.343	-27%
2019	616	Polonya	080390	Muz (plantainler hariç)	280.593.000	216.900.140	-23%	219.239.265	-22%	218.345.693	-22%
2018	620	Portekiz	080390	Muz (plantainler hariç)	138.991.014	107.490.070	-23%	114.486.070	-18%	106.540.897	-23%
2019	620	Portekiz	080390	Muz (plantainler hariç)	135.536.000	103.233.950	-24%	112.166.380	-17%	104.851.452	-23%
2018	410	Güney Kore	080390	Muz (plantainler hariç)	0	415.654.980		400.859.434		314.325.599	
2019	410	Güney Kore	080390	Muz (plantainler hariç)	301.550.000	429.890.500	43%	419.700.266	39%	322.406.164	7%
2018	703	Slovakya	080390	Muz (plantainler hariç)	71.913.444	56.640.772	-21%	58.082.523	-19%	44.551.879	-38%
2019	703	Slovakya	080390	Muz (plantainler hariç)	72.350.000	45.178.770	-38%	57.188.841	-21%	46.824.526	-35%
2018	724	İspanya	080390	Muz (plantainler hariç)	225.610.741	246.136.640	9%	211.984.739	-6%	203.167.590	-10%
2019	724	İspanya	080390	Muz (plantainler hariç)	226.185.000	262.226.820	16%	237.380.772	5%	202.841.585	-10%
2018	752	İsveç	080390	Muz (plantainler hariç)	161.968.278	151.578.940	-6%	192.139.103	19%	197.004.867	22%
2019	752	İsveç	080390	Muz (plantainler hariç)	0	160.609.300		192.667.949		199.059.281	
2018	757	İsviçre	080390	Muz (plantainler hariç)	107.342.208	146.470.140	36%	139.423.592	30%	149.635.186	39%
2019	757	İsviçre	080390	Muz (plantainler hariç)	102.860.000	144.405.040	40%	142.271.766	38%	152.949.149	49%
2018	826	Birleşik Krallık	080390	Muz (plantainler hariç)	740.223.877	788.383.740	7%	775.787.913	5%	826.272.988	12%
2019	826	Birleşik	080390	Muz (plantainler	711.041.000	834.313.600	17%	778.261.240	9%	828.779.800	17%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
		Krallık		hariç)							
2018	842	ABD	080390	Muz (plantainler hariç)	2.592.172.138	2.514.920.400	-3%	2.139.426.524	-17%	2.702.445.140	4%
2019	842	ABD	080390	Muz (plantainler hariç)	2.529.032.000	2.518.621.700	0%	2.141.946.139	-15%	2.778.825.398	10%
2018	36	Avustralya	080610	Taze üzüm	71.080.877	58.493.184	-18%	59.539.422	-16%	55.981.485	-21%
2019	36	Avustralya	080610	Taze üzüm	58.257.000	91.774.720	58%	62.033.726	6%	57.048.212	-2%
2018	40	Avusturya	080610	Taze üzüm	75.586.956	81.392.080	8%	74.502.790	-1%	77.707.014	3%
2019	40	Avusturya	080610	Taze üzüm	77.386.000	82.337.950	6%	74.194.919	-4%	75.351.570	-3%
2018	56	Belçika	080610	Taze üzüm	115.048.889	112.311.680	-2%	98.951.485	-14%	190.331.754	65%
2019	56	Belçika	080610	Taze üzüm	105.943.000	98.218.400	-7%	98.884.209	-7%	189.021.395	78%
2018	124	Kanada	080610	Taze üzüm	420.919.795	419.464.060	0%	366.998.760	-13%	425.557.916	1%
2019	124	Kanada	080610	Taze üzüm	432.654.000	411.281.630	-5%	367.913.952	-15%	436.611.567	1%
2018	152	Şili	080610	Taze üzüm	1.359.497	0	-100%	3.193.178	135%	700.959	-48%
2019	152	Şili	080610	Taze üzüm	1.672.000	248.040	-85%	3.071.518	84%	1.033.286	-38%
2018	203	Çekya	080610	Taze üzüm	89.712.841	77.638.030	-13%	60.815.547	-32%	66.558.196	-26%
2019	203	Çekya	080610	Taze üzüm	75.940.000	66.826.760	-12%	59.901.741	-21%	66.166.850	-13%
2018	208	Danimarka	080610	Taze üzüm	88.278.557	55.396.870	-37%	50.387.612	-43%	53.779.444	-39%
2019	208	Danimarka	080610	Taze üzüm	53.712.000	55.448.364	3%	53.971.754	0%	53.812.641	0%
2018	233	Estonya	080610	Taze üzüm	10.692.296	12.392.646	16%	10.860.667	2%	10.788.583	1%
2019	233	Estonya	080610	Taze üzüm	9.032.000	8.239.964	-9%	11.134.700	23%	10.836.801	20%
2018	246	Finlandiya	080610	Taze üzüm	53.807.529	46.137.916	-14%	46.639.235	-13%	50.069.674	-7%
2019	246	Finlandiya	080610	Taze üzüm	50.260.000	45.578.948	-9%	46.109.929	-8%	50.327.095	0%
2018	251	Fransa	080610	Taze üzüm	0	144.670.110		183.350.646		229.107.817	
2019	251	Fransa	080610	Taze üzüm	184.075.000	157.547.300	-14%	225.237.123	22%	236.370.258	28%
2018	276	Almanya	080610	Taze üzüm	753.619.196	721.184.600	-4%	572.456.256	-24%	604.559.862	-20%
2019	276	Almanya	080610	Taze üzüm	655.586.000	59.125.816	-91%	320.755.797	-51%	581.380.747	-11%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	XG_Boost		Random Forest		YSA	
						Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2018	348	Macaristan	080610	Taze üzüm	14.702.419	0	-100%	12.346.612	-16%	6.793.669	-54%
2019	348	Macaristan	080610	Taze üzüm	21.050.000	9.424.986	-55%	13.110.168	-38%	4.394.041	-79%
2018	352	İzlanda	080610	Taze üzüm	5.790.291	3.850.359	-34%	5.797.171	0%	4.396.621	-24%
2019	352	İzlanda	080610	Taze üzüm	5.426.000	4.267.061	-21%	5.866.018	8%	4.746.641	-13%
2018	372	İrlanda	080610	Taze üzüm	43.504.719	44.384.880	2%	50.944.212	17%	39.410.833	-9%
2019	372	İrlanda	080610	Taze üzüm	38.333.000	80.196.610	109%	77.076.554	101%	39.843.323	4%
2018	376	İsrail	080610	Taze üzüm	8.301.000	2.239.857	-73%	44.273.792	433%	4.697.018	-43%
2019	376	İsrail	080610	Taze üzüm	12.107.000	39.385.810	225%	69.454.447	474%	5.152.044	-57%
2018	381	İtalya	080610	Taze üzüm	47.027.542	72.485.170	54%	54.111.042	15%	56.284.939	20%
2019	381	İtalya	080610	Taze üzüm	48.077.000	68.599.160	43%	53.701.241	12%	62.592.396	30%
2018	392	Japonya	080610	Taze üzüm	106.235.790	89.114.560	-16%	82.849.608	-22%	78.314.744	-26%
2019	392	Japonya	080610	Taze üzüm	131.012.000	92.500.370	-29%	87.836.540	-33%	80.716.559	-38%
2018	428	Letonya	080610	Taze üzüm	13.875.574	18.949.114	37%	11.524.025	-17%	12.976.624	-6%
2019	428	Letonya	080610	Taze üzüm	13.205.000	13.089.850	-1%	11.492.825	-13%	12.552.800	-5%
2018	440	Litvanya	080610	Taze üzüm	21.243.073	35.271.396	66%	29.401.524	38%	82.701.481	289%
2019	440	Litvanya	080610	Taze üzüm	19.521.000	29.885.720	53%	29.812.354	53%	73.632.037	277%
2018	442	Lüksemburg	080610	Taze üzüm	7.950.076	6.581.933	-17%	7.570.414	-5%	11.049.692	39%
2019	442	Lüksemburg	080610	Taze üzüm	0	7.398.896		7.826.237		10.806.515	
2018	484	Meksika	080610	Taze üzüm	156.051.230	133.316.160	-15%	140.697.044	-10%	166.639.061	7%
2019	484	Meksika	080610	Taze üzüm	130.998.000	134.142.940	2%	140.994.790	8%	155.940.618	19%
2018	528	Hollanda	080610	Taze üzüm	823.531.640	690.864.260	-16%	573.012.302	-30%	686.157.084	-17%
2019	528	Hollanda	080610	Taze üzüm	1.023.923.000	702.185.300	-31%	573.566.932	-44%	692.484.399	-32%
2018	554	Yeni Zelanda	080610	Taze üzüm	38.674.357	29.734.810	-23%	40.461.833	5%	33.968.174	-12%
2019	554	Yeni Zelanda	080610	Taze üzüm	38.688.000	32.294.028	-17%	40.053.089	4%	32.621.867	-16%

Ek-7 OECD Ülkelerinin en fazla ithal ettiği ürünler (Muz, Domates, Avokado, Üzüm ve Biber) için XGBoost, Rastgele Orman Algoritmaları ile YSA Yönteminin Test Veri Seti Üzerindeki Tahmin Değerleri (...devamı)

						XG_Boost		Random Forest		YSA	
Dönem	Ülke Kodu	Ülke	Ürün kodu	Ürün	Gerçek Değer	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark	Tahmin Değeri	%Fark
2018	579	Norveç	080610	Taze üzüm	82.418.821	83.268.710	1%	84.556.950	3%	81.987.722	-1%
2019	579	Norveç	080610	Taze üzüm	78.650.000	88.031.080	12%	84.642.929	8%	87.566.357	11%
2018	616	Polonya	080610	Taze üzüm	187.355.831	169.315.790	-10%	167.687.522	-10%	202.590.619	8%
2019	616	Polonya	080610	Taze üzüm	182.795.000	156.074.140	-15%	167.924.527	-8%	199.443.029	9%
2018	620	Portekiz	080610	Taze üzüm	60.941.807	46.220.620	-24%	48.719.401	-20%	41.901.971	-31%
2019	620	Portekiz	080610	Taze üzüm	51.846.000	45.100.980	-13%	48.206.622	-7%	43.891.433	-15%
2018	410	Güney Kore	080610	Taze üzüm	0	107.306.720		194.785.596		161.480.541	
2019	410	Güney Kore	080610	Taze üzüm	202.311.000	134.702.750	-33%	197.364.949	-2%	172.023.286	-15%
2018	703	Slovakya	080610	Taze üzüm	31.028.513	43.888.316	41%	34.755.441	12%	26.699.810	-14%
2019	703	Slovakya	080610	Taze üzüm	28.416.000	43.389.210	53%	34.998.314	23%	25.190.793	-11%
2018	724	İspanya	080610	Taze üzüm	99.587.101	128.376.980	29%	74.437.923	-25%	67.441.829	-32%
2019	724	İspanya	080610	Taze üzüm	108.363.000	125.576.690	16%	74.980.820	-31%	63.751.581	-41%
2018	752	İsveç	080610	Taze üzüm	71.465.539	69.394.760	-3%	65.736.758	-8%	106.037.683	48%
2019	752	İsveç	080610	Taze üzüm	0	70.355.820		65.934.532		107.916.486	
2018	757	İsviçre	080610	Taze üzüm	74.132.788	83.465.920	13%	83.433.753	13%	93.940.990	27%
2019	757	İsviçre	080610	Taze üzüm	66.105.000	82.459.820	25%	83.419.064	26%	97.481.243	47%
2018	826	Birleşik Krallık	080610	Taze üzüm	687.344.947	562.231.100	-18%	594.543.287	-14%	685.044.132	0%
2019	826	Birleşik Krallık	080610	Taze üzüm	664.563.000	744.321.200	12%	639.754.283	-4%	683.803.205	3%
2018	842	ABD	080610	Taze üzüm	1.746.250.616	1.720.715.600	-1%	1.316.857.731	-25%	1.708.912.074	-2%
2019	842	ABD	080610	Taze üzüm	1.804.321.000	1.691.819.000	-6%	1.336.045.780	-26%	1.786.037.233	-1%

ÖZGEÇMİŞ

Adı ve SOYADI	Nedret TOSUN
EĞİTİM DURUMU	
Mezun Olduğu Lise	Nazilli Atatürk Lisesi
Lisans Diploması	Dokuz Eylül Üniversitesi Endüstri Mühendisliği Bölümü
Yüksek Lisans Diploması	Orta Doğu Teknik Üniversitesi Endüstri Mühendisliği Bölümü
Tez Konusu	A Linear Programming Approach to Quality Improvement Project and Product Mix Selection Under Inspection Error and Rework
Doktora Diploması	
Yabancı Dil / Diller	İngilizce
BİLİMSEL FAALİYETLER	
Daldır, I., Tosun N. ve Tosun Ö. (2019). Flight Delay Prediction. <i>17th International Logistics and Supply Chain Congress</i>, 17-18 Ekim 2019, İstanbul, Türkiye. Daldır, I., Tosun N. ve Tosun Ö. (2019). Supplier Evaluation using machine learning techniques. <i>17th International Logistics and Supply Chain Congress</i>, 17-18 Ekim 2019, İstanbul, Türkiye. Tosun, N., Arslanhan H. Ve Tosun Ö. (2019). Türkiye Mermer İhracat Değerinin Makine Öğrenmesi Algoritmaları ile Tahmin Edilmesi, <i>8. Ulusal Lojistik ve Tedarik Zinciri Kongresi</i>, 25-27 Nisan 2019, Niğde, Türkiye. Ustalı Koç, N. ve Tosun N. (2019). Bulanık AHP ve bulanık WASPAS yöntemleri ile yeni ürün seçimi, <i>Pazarlama İlgörüsü Üzerine Çalışmalar</i>, 3 (2), 25-34. Aktan Erdoğan, H. ve Tosun N. (2018). A hybrid MCDM methodology for an ERP system selection. <i>Economic and Administrative Sciences, Vol II.</i> ed. Azmi Yalçın, Akademisyen Kitapevi, Türkiye. Daldır, I., Tosun N. ve Tosun Ö. (2018). Supplier Selection Based on Artificial Neural Networks. <i>16th International Logistics and Supply Chain Congress</i>, 18-20 Ekim 2018, Denizli, Türkiye. Tosun, N., (2017). Target market selection in fresh fruit-vegetable sector using Fuzzy Vikor method. <i>Journal of Management, Marketing and Logistics</i>, 4 (4), 465-471. Aktan Erdoğan, H. ve Tosun N. (2017). Akıllı fabrika uygulanması kararını etkileyen kriterlerin ISM & MICMAC yöntemleri ile değerlendirilmesi. <i>3rd SCF International Conference on Economic and Social Impacts of Globalization</i>, 5-7 Ekim 2017, Antalya,	

Türkiye.	
İŞ DENEYİMİ	
Projeler	
Çalıştığı Kurumlar	Batı Akdeniz İhracatçılar Birliği Genel Sekreterliği: 2006-2020 Orta Anadolu İhracatçı Birlikleri Genel Sekreterliği: 2005-2006 Uğur Soğutma Sanayi A.Ş.: 2002-2003