

Yapı Kimyasalları Sektöründe Makine Öğrenmesi Destekli Çimento Fiyat Tahminlemesi

Mehmet Erol Kara

YÜKSEK LİSANS TEZİ

Endüstri Mühendisliği Anabilim Dalı

Temmuz 2025

Machine Learning-Supported Cement Price Forecasting in the Construction Chemicals
Industry

Mehmet Erol Kara

MASTER OF SCIENCE THESIS

Department of Industrial Engineering

July 2025

Yapı Kimyasalları Sektöründe Makine Öğrenmesi Destekli Çimento Fiyat
Tahminlemesi

Mehmet Erol Kara

Eskişehir Osmangazi Üniversitesi

Fen Bilimleri Enstitüsü

Lisansüstü Eğitim ve Öğretim Yönetmeliği Uyarınca

Endüstri Mühendisliği Anabilim Dalı

Yöneylem Araştırması Bilim Dalında

YÜKSEK LİSANS TEZİ

olarak hazırlanmıştır.

Danışman: Prof. Dr. Aydın Sipahioğlu

İkinci Danışman: Prof. Dr. Ezgi Aktar Demirtaş

Temmuz 2025

ONAY

Endüstri Mühendisliği Anabilim Dalı Yüksek Lisans öğrencisi **Mehmet Erol Kara**'nın YÜKSEK LİSANS tezi olarak hazırladığı “Yapı Kimyasalları Sektöründe Makine Öğrenmesi Destekli Çimento Fiyat Tahminlemesi” başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğin ilgili maddeleri uyarınca değerlendirilerek oybirliği ile kabul edilmiştir.

Danışman : Prof. Dr. Aydın Sipahioğlu

İkinci Danışman : Prof. Dr. Ezgi Aktar Demirtaş

Yüksek Lisans Tez Savunma Jürisi:

Üye: Prof. Dr. Aydın Sipahioğlu

Üye: Prof. Dr. Tuğba Saraç

Üye: Dr. Öğr. Üyesi Zeliha Ergül Aydın

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
tarih ve sayılı kararıyla onaylanmıştır.

Prof. Dr.

Enstitü Müdürü

ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Prof. Dr. Aydın SİPAHİOĞLU ve Prof. Dr. Ezgi AKTAR DEMİRTAŞ danışmanlığında hazırlamış olduğum “Yapı Kimyasalları Sektöründe Makine Öğrenmesi Destekli Çimento Fiyat Tahminlemesi” başlıklı Yüksek Lisans tezimin özgün bir çalışma olduğunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallarına uygun davrandığımı; tezimde verdiğim bilgileri, verileri akademik ve bilimsel ilke ve kurallara uygun olarak elde ettiğimi; tez çalışmamda yararlandığım eserlerin tümüne atıf yaptığımı ve kaynak gösterdiğimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduğumu beyan ederim. 08/07/2025

Mehmet Erol Kara

İmza

ÖZET

Bu tez çalışmasında, yapı kimyasalı üretiminde faaliyet gösteren bir firmanın Eskişehir fabrikasında çimento hammadde satın alma fiyatlarının tahminlenmesi amaçlanmıştır. Firma yetkilileri ile yapılan görüşmeler doğrultusunda geliştirilen bu çalışma, stratejik ve operasyonel karar mekanizmalarına veri temelli bir katkı sunmayı hedeflemektedir. Yapı kimyasalı üretiminde çimento, selüloz, kalsit, kraft kâğıdı ve polimerler olmak üzere beş temel hammadde yer almakta olup bu araştırma pilot bir uygulama olarak yalnızca çimento hammaddesine ve Eskişehir fabrikasına odaklanmaktadır. Araştırmada Ocak 2013 – Aralık 2024 dönemini kapsayan aylık olarak derlenmiş veri seti kullanılmış; arz-talep dengesi, enerji ve nakliye maliyetleri, ekonomik ve demografik göstergeler gibi farklı kategoriler altında toplam 30 aday öznitelik değerlendirmeye alınmıştır. İlk aşamada, öznitelik seçimlerinin önünü açmak amacıyla yüzeysel eksik veri doldurma işlemleri gerçekleştirilmiştir. Sonrasında, sektör uzmanlarının görüşlerine dayalı (Çalışma 1) ve akademik yöntemlerle desteklenmiş istatistiksel yaklaşımlara dayalı (Çalışma 2) iki farklı öznitelik seçim stratejisi benimsenmiş; Seçilen özniteliklerle oluşturulan veri setlerine ileri eksik veri doldurma teknikleri uygulanmış; devamında ise öznitelik kümelerinin birleştirilmesi (Çalışma 3), kapasite kullanım oranı değişkeninin modelden çıkarılması (Çalışma 4) ve korelasyon temelli öznitelik indirgeme (Çalışma 5) gibi çeşitli modelleme senaryoları üzerinden çalışmalar çeşitlendirilmiştir. Modelleme sürecinde sekiz farklı makine öğrenmesi algoritması hem varsayılan hem de eniyilenmiş hiperparametrelerle test edilmiştir. Modeller, çeşitli performans metrikleri ile değerlendirilmiştir. Sonuçlar, sektörel bilgiye dayalı öznitelik seçiminin özellikle makine öğrenmesi modelleri açısından yüksek tahmin başarısı sağladığını göstermiştir. Tez kapsamında geliştirilen modelleme yaklaşımının, sadece çimento değil, diğer hammaddelere yönelik tahmin sistemlerine de uyarlanarak yapısal bir karar destek aracı olarak yaygınlaştırılması önerilmektedir.

Anahtar Kelimeler: Yapı kimyasalları, makine öğrenmesi, fiyat tahminlemesi

SUMMARY

This thesis aims to forecast the procurement prices of cement—a key raw material—for the Eskişehir factory of a company operating in the field of construction chemicals. Developed in line with consultations held with company representatives, the study seeks to provide a data-driven contribution to strategic and operational decision-making processes. Among the five fundamental raw materials used in the production of construction chemicals—namely cement, cellulose, calcite, kraft paper, and polymers—this research focuses on cement within the scope of a pilot application, specifically targeting the Eskişehir facility. The analysis is based on a monthly dataset compiled for the period between January 2013 and December 2024. A total of 30 candidate features were evaluated across different thematic categories such as supply-demand balance, energy and transportation costs, economic indicators, and demographic factors. In the initial phase, superficial missing value imputation was conducted to enable feature selection. Subsequently, two distinct feature selection strategies were adopted: one based on expert opinions (Study 1) and the other on academic and statistical methodologies supported by machine learning techniques (Study 2). Advanced missing value imputation techniques were then applied to the resulting datasets. The study was further diversified through additional modeling scenarios, including feature set merging (Study 3), the exclusion of the capacity utilization variable from the model (Study 4), and correlation-based feature reduction (Study 5). Eight different machine learning algorithms were tested during the modeling process, using both default and optimized hyperparameters. The models were evaluated using various performance metrics. The results reveal that feature selection based on domain knowledge yields particularly high prediction accuracy in machine learning models. The modeling framework developed within the scope of this thesis is recommended to be extended to other raw materials beyond cement, thereby enabling the construction of a more comprehensive decision support system.

Keywords: Construction chemicals, machine learning, price forecasting

İÇİNDEKİLER

Sayfa

ÖZET	vi
SUMMARY	vii
İÇİNDEKİLER.....	viii
ŞEKİLLER DİZİNİ	xi
ÇİZELGELER DİZİNİ	xiii
KISALTMALAR DİZİNİ.....	xv
1. GİRİŞ VE AMAÇ.....	1
2. LİTERATÜR ARAŞTIRMASI.....	3
2.1. Yapı Kimyasalı Üretimi ve Makine Öğrenmesi Algoritmalarının Birlikte Kullanıldığı Çalışmalar	3
2.2. Makine Öğrenmesi Yöntemleri ile Fiyat Tahmini Çalışmaları	8
2.3. Makine Öğrenmesi ile Çimento Fiyat Tahminine Yönelik Akademik Çalışmalar ..	13
3. MATERYAL VE YÖNTEM	16
3.1. Öznitelik Seçim Süreci ve Özniteliklerin Sektörel Önemi	16
3.2. Veri Setinin Genel Analizi ve Veri Ön İşleme Süreci	20
3.3. Öznitelik Mühendisliği ve Çalışma Kapsamında Kullanılacak Makine Öğrenmesi Metotları.....	22
3.3.1. Çalışma Kapsamında Kullanılan Eksik Veri Doldurma Yöntemleri.....	23
3.3.1.1. <u>Lineer İnterpolasyon Yöntemi</u>	24
3.3.1.2. <u>Spline İnterpolasyon Yöntemi</u>	25
3.3.1.3. <u>Sabit Değer Ataması Yöntemi</u>	25
3.3.1.4. <u>Parazit Eklenmiş Ortalama Yöntemi</u>	26
3.3.1.5. <u>Parçalı Orantılı Büyüme Tahmini Yöntemi</u>	27
3.3.1.6. <u>Yıl İçi Sabit Değer Tahmini Yöntemi</u>	28
3.3.1.7. <u>Büyüme/Küçülme Katsayısı ile Zincirleme Tahmin Yöntemi</u>	28
3.3.1.8. <u>İlımlı Büyüme Oranı Destekli Trend Analizi Yöntemi</u>	29
3.3.1.9. <u>Pro-Rata (Temsili Ay Oranlama) yöntemi</u>	30

İÇİNDEKİLER

Sayfa

3.3.1.10. <u>STL Yöntemi</u>	31
3.3.1.11. <u>Benchmarking Yöntemi</u>	33
3.3.2. Öznitelik Seçim Kriterleri ve Araçları.....	34
3.3.2.1. <u>Uzman Görüşüne Dayalı Öznitelik Seçimi</u>	35
3.3.2.2. <u>Korelasyon Analizi</u>	37
3.3.3. Çalışma Kapsamında Kullanılacak Makine Öğrenmesi Modellerinin Seçimi ve Modellere İlişkin Detaylar	39
3.3.3.1. <u>Rastgele Orman - RF Algoritması</u>	40
3.3.3.2. <u>GB Algoritması</u>	42
3.3.3.3. <u>XGBoost Algoritması</u>	44
3.3.3.4. <u>LightGBM Algoritması</u>	45
3.3.3.5. <u>Ridge Regresyon Algoritması</u>	47
3.3.3.6. <u>Lasso Regresyon Algoritması</u>	49
3.3.3.7. <u>SVR Algoritması</u>	50
3.3.3.8. <u>MLPR Algoritması</u>	51
3.4. Çalışma Kapsamında Kullanılan Performans Metrikleri.....	53
3.4.1. Ortalama Mutlak Hata (MAE).....	54
3.4.2. Ortalama Karesel Hata (MSE).....	55
3.4.3. Hata Kareleri Ortalamasının Kare Kökü (RMSE).....	56
3.4.4. Belirlilik Katsayısı (R^2).....	57
3.5. GridSearchCV ile Hiperparametre Eniyilemesi	58
4. UYGULAMA VE MODELLEME ADIMLARI	60
4.1. Ham Veri Setinin Tanıtımı ve Genel Özellikleri.....	61
4.2. Ön Değerlendirme Amaçlı Yüzeysel Eksik Veri Doldurma İşlemleri	64
4.3. Çalışma 1: Uzman Görüşü Temelli Öznitelik Seçimi	85
4.4. Çalışma 2: Korelasyon Analizi Temelli Öznitelik Seçimi	93
4.5. Gelişmiş Eksik Veri Doldurma ve Nihai Veri Setlerinin Oluşturulması.....	98

İÇİNDEKİLER

Sayfa

4.5.1. Çalışma 1'de Kullanılacak Veri Seti için Eksik Verilerin İleri Yöntemler ile Doldurulması.....	99
4.5.2. Çalışma 2'de Kullanılacak Veri Seti için Eksik Verilerin İleri Yöntemler ile Doldurulması.....	99
4.6. Çalışma 1 için Modelleme Süreci.....	105
4.7. Çalışma 2 için Modelleme Süreci.....	112
4.8. Çalışma 3: Öznitelik Birleştirme ve Modelleme	119
4.9. Çalışma 4: Kapasite Kullanım Oranı Özniteliğinin Modelden Çıkarılması.....	124
4.10. Çalışma 5: Korelasyon Tabanlı Nihai Öznitelik İndirgeme	126
5. BULGULAR VE TARTIŞMA	129
5.1. Model Bazlı Karşılaştırmalı Performans Analizi	129
5.2. Çalışma Bazlı Karşılaştırmalı Performans Analizi.....	142
5.3. Genel (Bütünleşik) Değerlendirme ve Stratejik Sonuçlar	150
6. SONUÇ VE ÖNERİLER	153
KAYNAKLAR DİZİNİ.....	155
EK AÇIKLAMALAR	161
Ek Açıklama-A: Korelasyon Analizi Temelli Öznitelik Seçimi için Korelasyon Matrisi Isı Haritası	162
Ek Açıklama-B: Korelasyon Analizi Temelli Öznitelik Seçimi için Korelasyon Matrisi	163
Ek Açıklama-C: R^2 Değerine Göre Büyükten Küçüğe Sıralanmış Çalışma-Model İkili Karşılaştırma Tablosu.....	164

ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
3.1. Tematik Başlıklar Temelinde Potansiyel Öznitelik Seti.....	17
3.2. Kullanılan Makine Öğrenmesi Algoritmalarının Sınıflandırılması.....	40
3.3. Performans Metrikleri.....	53
4.1. Çimento Sanayi İstihdam Sayısı (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği.....	64
4.2. Çimento İhracatı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği.....	65
4.3. Klinker İhracatı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	66
4.4. Çimento Üretim Miktarı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	67
4.5. Klinker Üretim Miktarı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	68
4.6. Yapı Ruhsatına Göre Bina Sayısı (adet) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği.....	69
4.7. Petrokok Birim Fiyatı (USD) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	70
4.8. Petrokok İthalatı (kg) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	71
4.9. Türkiye Nüfusu (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	72
4.10. Türkiye Kent Nüfusu (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	73
4.11. Türkiye Kırsal Nüfusu (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	74
4.12. Brent Petrol (USD) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	75
4.13. Kalker Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	76
4.14. Kalker Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	77
4.15. Alçıtaşı Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	78
4.16. Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	79
4.17. Çimento Kili Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	80
4.18. Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	81
4.19. Demir Cevheri Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	82
4.20. Demir Cevheri Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği	83
4.21. Eskişehir Fabrikası Gri Çimento Satın Alım Fiyatı Hedef Değişkeni için Yüzeysel Eksik Veri Doldurma Grafiği.....	84

ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
4.22. Çimento Üretim Miktarı (Ton) ve Yapı Kimyasalı Üreticisi Satış Endeksi Karşılaştırması.....	88
4.23. Uzman Görüşü Temelli Öznitelik Seçim Sonuçları	93
4.24. Korelasyon Analizi Temelli Öznitelik Seçim Sonuçları	98
4.25. Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Aylık Ortalama Yöntemi ile Eksik Verilerin Doldurulması Grafiği.....	100
4.26. Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Aylık Ortalama + Parazit Yöntemi ile Eksik Verilerin Doldurulması Grafiği.....	100
4.27. Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Regresyon Yöntemi ile Eksik Verilerin Doldurulması Grafiği.....	101
4.28. Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Aylık Ortalama Yöntemi ile Eksik Verilerin Doldurulması Grafiği.....	102
4.29. Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Aylık Ortalama + Parazit Yöntemi ile Eksik Verilerin Doldurulması Grafiği.....	102
4.30. Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Regresyon Yöntemi ile Eksik Verilerin Doldurulması Grafiği.....	103
4.31. Klinker İhracatı (ton) Özniteliğinin Çeşitli Yöntemler ile Eksik Verilerin Doldurulması Grafiği	104

ÇİZELGELER DİZİNİ

Çizelge

Sayfa

2.1. Yapı Kimyasalı Üretimi ve Makine Öğrenmesi Algoritmalarının Birlikte Kullanıldığı Çalışmalar için Literatür Araştırması Özet Tablosu.....	7
2.2. Makine Öğrenmesi Yöntemleri ile Fiyat Tahmini Çalışmaları için Literatür Araştırması Özet Tablosu	12
2.3. Makine Öğrenmesi Yöntemleri ile Çimento Fiyat Tahmini Çalışmaları için Literatür Araştırması Özet Tablosu	15
3.1. Ham Veri Setindeki Eksik Veri Gözlemleri	21
4.1. Çimento Üretim Miktarı (Ton) ve Yapı Kimyasalı Üreticisi Satış Endeksi için Betimsel İstatistikler	87
4.2. Hedef Değişken ile En Yüksek Pozitif Korelasyona Sahip Öznitelikler.....	94
4.3. Hedef Değişken ile En Yüksek Negatif Korelasyona Sahip Öznitelikler	95
4.4. Çalışma 1 için RF Modeli Performans Metrikleri	106
4.5. Çalışma 1 için GB Modeli Performans Metrikleri	106
4.6. Çalışma 1 için XGBoost Modeli Performans Metrikleri.....	107
4.7. Çalışma 1 için LightGBM Modeli Performans Metrikleri	108
4.8. Çalışma 1 için Ridge Regresyon Modeli Performans Metrikleri	108
4.9. Çalışma 1 için Lasso Regresyon Modeli Performans Metrikleri	109
4.10. Çalışma 1 için SVR Modeli Performans Metrikleri	110
4.11. Çalışma 1 için MLPR Modeli Performans Metrikleri	111
4.12. Çalışma 2 için RF Modeli Performans Metrikleri	113
4.13. Çalışma 2 için GB Modeli Performans Metrikleri	114
4.14. Çalışma 2 için XGBoost Modeli Performans Metrikleri.....	115
4.15. Çalışma 2 için LightGBM Modeli Performans Metrikleri	115
4.16. Çalışma 2 için Ridge Regresyon Modeli Performans Metrikleri	116
4.17. Çalışma 2 için Lasso Regresyon Modeli Performans Metrikleri	116
4.18. Çalışma 2 için SVR Modeli Performans Metrikleri	117
4.19. Çalışma 2 için MLPR Modeli Performans Metrikleri	118
4.20. Çalışma 3 için RF Modeli Performans Metrikleri	119
4.21. Çalışma 3 için GB Modeli Performans Metrikleri	120
4.22. Çalışma 3 için XGBoost Modeli Performans Metrikleri.....	120
4.23. Çalışma 3 için LightGBM Modeli Performans Metrikleri	121
4.24. Çalışma 3 için Ridge Regresyon Modeli Performans Metrikleri	121
4.25. Çalışma 3 için Lasso Regresyon Modeli Performans Metrikleri	122
4.26. Çalışma 3 için SVR Modeli Performans Metrikleri	123
4.27. Çalışma 3 için MLPR Modeli Performans Metrikleri	123

ÇİZELGELER DİZİNİ

<u>Çizelge</u>	<u>Sayfa</u>
4.28. Çalışma 4 için Makine Öğrenmesi Modellerinin Çalıştırılması Sonucu Elde Edilen Performans Metrikleri	125
4.29. Çalışma 5 için Makine Öğrenmesi Modellerinin Çalıştırılması Sonucu Elde Edilen Performans Metrikleri	127
5.1. RF Modeline İlişkin Performans Metrikleri	130
5.2. GB Modeline İlişkin Performans Metrikleri	131
5.3. XGBoost Modeline İlişkin Performans Metrikleri	132
5.4. LightGBM Modeline İlişkin Performans Metrikleri	134
5.5. Ridge Regresyon Modeline İlişkin Performans Metrikleri	135
5.6. Lasso Regresyon Modeline İlişkin Performans Metrikleri.....	137
5.7. SVR Modeline İlişkin Performans Metrikleri	138
5.8. MLPR Modeline İlişkin Performans Metrikleri	140
5.9. Çalışma 1 Sonucunda Elde Edilen Performans Metrikleri	142
5.10. Çalışma 2 Sonucunda Elde Edilen Performans Metrikleri	144
5.11. Çalışma 3 Sonucunda Elde Edilen Performans Metrikleri	145
5.12. Çalışma 4 Sonucunda Elde Edilen Performans Metrikleri	147
5.13. Çalışma 5 Sonucunda Elde Edilen Performans Metrikleri	148
5.14. Bütünleşik Karşılaştırma Tablosu (En Yüksek R^2 Değerine Sahip İlk 10 Çalışma-Model İkili).....	150

KISALTMALAR DİZİNİ

Kısaltmalar	Açıklama
GB	Gradient Boosting
GridSearchCV	GridSearch with Cross-Validation
LightGBM	Light Gradient Boosting Machine
MAE	Ortalama Mutlak Hata (Mean Absolute Error)
MLP	Multi-Layer Perceptron
MLPR	Multi-Layer Perceptron Regressor
MSE	Ortalama Karesel Hata (Mean Squared Error)
R^2	Belirlilik Katsayısı (Coefficient of Determination)
RF	Random Forest (Rastgele Orman)
RMSE	Ortalama Hata Karelerinin Kökü (Root Mean Squared Error)
SVR	Support Vector Regression (Destek Vektör Regresyonu)
XGBoost	Extreme Gradient Boosting

1. GİRİŞ VE AMAÇ

Yapı Kimyasalı sektörü, inşaat sektörünün gelişimine paralel olarak sürekli büyüyen ve teknolojik yeniliklerle şekillenen stratejik bir sektördür. Bu sektördeki üretim faaliyetleri, çimento, selüloz, kalsit, kraft kâğıdı ve polimer gibi beş temel hammaddeye dayanmaktadır. Türkiye genelinde dört farklı lokasyonda üretim yapan bir yapı kimyasalı üreticisinin Eskişehir Fabrikası özelinde gerçekleştirilen bu tez çalışması, sektörün en temel girdilerinden biri olan çimentonun aylık satın alma fiyatlarını öngörmeye yönelik bir karar destek mekanizması geliştirmeyi hedeflemektedir. Çalışma, söz konusu yapı kimyasalı üreticisinin yetkililerinden gelen talep doğrultusunda pilot bir uygulama niteliği taşımakta ve çimento hammaddesi ile Eskişehir lokasyonuna odaklanmaktadır. Nihai amaç, firma içi satın alma ve planlama süreçlerinde veri temelli kararlar alınmasını sağlayarak kurumsal fayda elde etmektir.

Tez çalışması boyunca, modelleme sürecinin güvenilirliğini ve doğruluğunu artırmak adına veri ön işleme aşamasına büyük önem verilmiştir. İlk aşamada öznitelik seçimine olanak tanıyacak şekilde yüzeysel eksik veri doldurma yöntemleri uygulanmıştır. Daha sonra seçilen özniteliklere ilişkin gelişmiş ve öznitelik türüne özgü olarak belirlenen eksik veri doldurma yöntemleri ile nihai modelleme için güçlü ve anlamlı veri setleri oluşturulmuştur. Söz konusu veri seti; ekonomik göstergeler, enerji maliyetleri, ulaşım giderleri, sektörel talep sinyalleri, politik ve demografik etkenler gibi çeşitli temalarda gruplanmış, çok boyutlu bir yapıya sahip 30 öznitelikten oluşmaktadır. Bu veri setine 11 farklı gelişmiş eksik veri tamamlama yöntemi uygulanmış; böylece hem zamansal bütünlük hem de istatistiksel tutarlılık sağlanmıştır.

Veri ön işleme sürecinin ardından modelleme aşamasında, sektörel uzman görüşlerine (Çalışma 1) ve korelasyon analizine dayalı (Çalışma 2) olarak oluşturulan iki farklı öznitelik kümesi kullanılarak bağımsız çalışmalar gerçekleştirilmiş, ardından bu iki kümenin birleşimiyle oluşturulan bütünlük öznitelik kümesiyle (Çalışma 3) yeni bir modelleme yapılmıştır. Ayrıca tahmin doğruluğuna etki edebileceği düşünülen ve sonradan öznitelik setine eklenen kapasite kullanım oranı özniteliğinin modelleme üzerindeki etkileri analiz edilerek özellik çıkarımı (Çalışma 4) ve bütünlük öznitelik kümesinin indirgeme

alıřmaları sonrası elde edilen veri setiyle (alıřma 5) de analizler yrtlmřtr. Tm bu srelerde, Rastgele Orman (Random Forest - RF), Gradient Boosting (GB), Exreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Ridge Regresyon, Lasso Regresyon, Destek Vektr Regresyonu (Support Vector Regression - SVR) ve Multi-Layer Perceptron Regressor (MLPR) olmak zere sekiz farklı makine ğrenmesi algoritması hem varsayılan hiperparametrelerle hem de GridSearch with Cross-Validation (GridSearchCV) yntemi ile eniyilenmiř hiperparametrelerle alıřtırılmıřtır. Elde edilen sonular Ortalama Mutlak Hata (Mean Absolute Error – MAE), Ortalama Karesel Hata (Mean Squared Error – MSE), Ortalama Hata Karelerinin Kare Kk (Root Mean Squared Error – RMSE) ve Belirlilik Katsayısı (R^2) performans metrikleri aracılıėıyla karřılařtırılmal olarak deėerlendirilmiřtir.

Bu alıřma, yalnızca yksek doėruluk saėlayan tahmin modelleri geliřtirmeyi deėil, aynı zamanda farklı veri yapıları ve model kombinasyonlarının ngr bařarısına olan katkılarını analiz ederek sektrel karar destek sistemlerine bilimsel zemin kazandırmayı da amalamaktadır. Geliřtirilen modellerin sektrel baėlamda ne lde uygulanabilir olduėu ve hangi kořullarda hangi modelin tercih edilmesi gerektiėi gibi stratejik sorulara da yanıt aranmaktadır.

Tezin ilerleyen blmleri řu řekilde yapılandırılmıřtır: İkinci blmde, imento fiyat tahminlemesine ynelik alıřmalara ve makine ğrenmesi uygulamalarına odaklanan bir literatr taraması sunulmaktadır. nc blmde, veri setinin yapısı, znitelik mhendisliėi sreci, kullanılan eksik veri doldurma yntemleri, znitelik seim stratejileri ve tercih edilen makine ğrenmesi algoritmalarının detayları yer almaktadır. Drdnc blmde uygulama adımları ayrıntılı olarak aıklanmakta; farklı znitelik kmeleri ile oluřturulan modeller ve performans analizleri yer almaktadır. Beřinci blmde ise elde edilen bulgular deėerlendirilmekte ve karřılařtırılmal sonulara dayalı ıkarımlar yapılmaktadır. Altıncı ve son blmde, genel sonulara yer verilerek neriler sunulmakta ve gelecekte yapılabilecek alıřmalara ıřık tutulmaktadır.

2. LİTERATÜR ARAŞTIRMASI

Günümüzde yapı kimyasalları üretimi, veri odaklı karar destek sistemleri ve makine öğrenmesi yöntemleri ile yeniden şekillenmekte; hammadde temininden üretim eniyilemesine, maliyet öngörüsünden piyasa stratejilerine kadar geniş bir yelpazede akademik ve sektörel araştırmalara konu olmaktadır. Bu bağlamda literatür çalışmaları, hem disiplinler arası gelişmeleri takip etmek hem de mevcut yöntemsel boşlukları belirlemek adına büyük önem taşımaktadır.

Bu bölümde, yapı kimyasalı üretimi ve makine öğrenmesi algoritmalarının birlikte kullanıldığı çalışmalar, farklı sektörlerde gerçekleştirilen fiyat tahmini uygulamaları ve özel olarak çimento fiyatlarının öngörülmesine odaklanan akademik yayınlar sistematik bir şekilde incelenmiştir. Literatür değerlendirmeleri kapsamında makine öğrenmesinin sektörel bağlamdaki potansiyelini, kullanılan yöntemlerin çeşitliliğini ve uygulama başarılarını ortaya koymak hedeflenmiş; nihayetinde bu tez çalışmasının literatürdeki boşlukları doldurma potansiyeli gerek nicel gerekse nitel yönleriyle vurgulanmıştır.

2.1. Yapı Kimyasalı Üretimi ve Makine Öğrenmesi Algoritmalarının Birlikte Kullanıldığı Çalışmalar

Yapı kimyasalı üretim süreci, birçok fiziksel, kimyasal ve operasyonel değişkenin birlikte etkili olduğu karmaşık sistemlerdir. Bu sistemlerin anlaşılması ve veriye dayalı şekilde modellenmesi hem kalite kontrolünü iyileştirmek hem de maliyet ve üretim planlaması süreçlerini eniyilemek açısından giderek daha önemli hale gelmektedir. Bu kapsamda son yıllarda makine öğrenmesi algoritmaları, yapı kimyasallarına ilişkin veri yapılarının analizinde ve üretim/performans öngörüsünde etkin bir şekilde kullanılmaya başlanmıştır. Bu başlık altında, yapı kimyasalı üretime doğrudan ya da dolaylı olarak odaklanan ve makine öğrenmesi tabanlı yaklaşımlar içeren yedi farklı akademik yayın incelenmiştir.

Ozcan vd. (2024) tarafından yayımlanan bu çalışmada, geleneksel çimentonun yerine zeolit ve diatomit gibi doğal puzolanların kullanıldığı beton karışımlarının basınç dayanımlarının tahmini için farklı makine öğrenmesi algoritmaları uygulanmıştır. Deneysel

çalışmada, 7 farklı beton karışımı ile 63 numune üretilmiş ve bu numunelerin 28, 56 ve 90. günlerdeki basınç dayanımı ölçülmüştür. Bu veriler kullanılarak, RF, Support Vector Machines, XGBoost, LightGBM ve CatBoost algoritmaları eğitilmiş ve hiperparametre eniyilemesi ile eniyi tahmin modelleri belirlenmiştir. Karşılaştırmalı sonuçlara göre, CatBoost algoritması eniyilendiğinde %98 üzerinde R^2 başarısı göstermiştir. Bu da makine öğrenmesi modellerinin fiziksel dayanım testlerine kıyasla çok daha düşük maliyetli ve hızlı tahmin imkânı sunduğunu göstermektedir. Çalışma, yapı kimyasallarına yönelik beton katkılarının etkilerini veri temelli analiz eden önemli bir örnek teşkil etmekte ve bu yönüyle yapı kimyasalı üretiminde makine öğrenmesinin potansiyel kullanım alanlarına güçlü bir örnek sunmaktadır.

Parhi vd. (2024) tarafından yürütülen bu çalışmada, çevre dostu ve yüksek performanslı bir alternatif olan alkali aktivasyonlu cüruf betonlarının üretimi ve bu betonların mekanik özelliklerinin tahmini üzerine odaklanılmıştır. Deneysel olarak geliştirilen bu kendiliğinden yerleşen betonlar, granüle yüksek fırın cürufu ve farklı oranlarda alkali aktivatör çözeltisi kullanılarak elde edilmiştir. Araştırmanın ikinci aşamasında, elde edilen 135 gözlemden oluşan veri seti ile XGBoost, RF, SVR, Karar Ağacı, GB gibi makine öğrenmesi algoritmaları kullanılarak basınç ve çekme dayanımı tahminleri yapılmıştır. Tüm modellerin hiperparametreleri, doğadan esinlenilmiş bir algoritma olan Spotted Hyena Optimization yöntemi ile eniyilenmiştir. En başarılı sonuçları veren model, %97 R^2 değerine ulaşan eniyilenmiş XGBoost modeli olmuştur. Çalışma, hem alternatif bağlayıcı malzemelerle beton üretimini hem de dayanım tahmininde makine öğrenmesi uygulamasını bir araya getirerek yapı kimyasalı üretiminde veri temelli analizlerin gücünü göstermektedir.

Childs (2021) tarafından Carnegie Mellon Üniversitesi'nde hazırlanan bu doktora tezinde, karmaşık çimentolu sistemlerin modellenmesi ve eniyilenmesi amacıyla makine öğrenmesi teknikleri ile malzeme bilimi bilgisinin bütünleştirilmesi üzerine odaklanılmıştır. Tezde, özellikle beton ve harç sistemlerinin performansını etkileyen onlarca bileşenin (çimento, katkı, su/çimento oranı, katkı maddesi türü vb.) etkilerini öğrenebilecek, genelleme yeteneği yüksek makine öğrenmesi modelleri geliştirilmiştir. Yazar, klasik makine öğrenmesi modellerine doğrudan veri vermek yerine, malzeme bilimine ait fiziksel kısıtları ve bağlayıcı bilgileri de modele entegre ederek daha yorumlanabilir ve öngörü gücü

yüksek modeller oluşturmuştur. Tezde Gaussian Process Regression, Bayesian Optimization, Kernel Ridge Regression, ve Graph-Based Featurization gibi yöntemler kullanılmış; çok bileşenli sistemlerde farklı eniyileme hedefleriyle modelleme yapılmıştır. Çalışma, sadece yapı kimyasalları üretimi değil, aynı zamanda malzeme tasarımı sürecinde makine öğrenmesinin nasıl fiziksel bilgiyle desteklenebileceğine dair öncü bir akademik katkı niteliğindedir.

Tran vd. (2022) tarafından gerçekleştirilen çalışmada, kendiliğinden yerleşen betonların basınç dayanımı, karışım oranları ve kür süresi gibi temel değişkenler dikkate alınarak makine öğrenmesi teknikleri ile modellenmiştir. Çalışmada, başta XGBoost olmak üzere, Lineer Regresyon, K-Nearest Neighbors, Support Vector Machine, Karar Ağaçları, RF, GB ve iki farklı eğitim algoritmasıyla uygulanan Yapay Sinir Ağı modelleri karşılaştırılmıştır. Bununla birlikte, sekiz makine öğrenmesi yönteminin her biri, Parçacık Sürü Optimizasyonu ile hibrit hale getirilmiş ve sonuçlar performans açısından analiz edilmiştir. Elde edilen bulgulara göre, XGBoost'un Parçacık Sürü Optimizasyonu hibrit modeli tüm yöntemler arasında en iyi tahmin performansına ulaşmış ($R^2 = 0,9644$, RMSE = 4,78, MAE = 3,48) ve özellikle mühendislik uygulamalarında kullanılabilirlik açısından öne çıkmıştır. Bu kapsamlı kıyaslama, yapı kimyasalları alanında karışım tasarımı ve dayanım öngörüsüne yönelik makine öğrenmesi tabanlı karar destek sistemlerinin güvenilirliğini ortaya koymakta ve literatüre güçlü bir katkı sunmaktadır.

Kumar vd. (2024) tarafından gerçekleştirilen çalışmada, yüksek oranda uçucu kül ve silis dumanı içeren kendiliğinden yerleşen betonlarda dayanım ve dayanıklılık özelliklerinin tahmini amacıyla, yapay sinir ağları tabanlı meta-sezgisel eniyileme modelleri geliştirilmiştir. Özellikle, Rapid Chloride Penetration Test (RCPT) sonuçlarının, farklı bağlayıcı oranları ve yüksek sıcaklık kür koşullarında nasıl değiştiğini tahmin etmek amacıyla geliştirilen bu modellerde, teaching-learning-based optimization, ant colony optimization, imperialist competitive algorithm ve shuffled complex evolution gibi popüler meta-sezgisel algoritmalar, yapay sinir ağlarının öğrenme oranı, ağırlıkları ve öngerilme değerlerini eniyilemek için kullanılmıştır. Modellerin eğitiminde 7 giriş parametresine sahip 360 gözlemden oluşan deneysel bir veri seti bulunmaktadır. Model performansı 8 farklı başarı ölçütüyle değerlendirilmiş ve Yapay Sinir Ağı – Teaching-Learning-Based Optimization modeli, eğitimde $R^2 = 0,9962$ ve testte $R^2 = 0,9676$ değerleriyle en başarılı

tahmin performansını sergilemiştir. Elde edilen sonuçlar, RCPT gibi deneysel olarak zahmetli ve maliyetli testlerin yerini alabilecek veri temelli sistemlerin yapı kimyasalları üretimi ve kalite kontrolü açısından oldukça değerli olabileceğini göstermektedir.

Polat vd. (2024) tarafından yayımlanan çalışmada, Güneydoğu Anadolu Bölgesi'nde faaliyet gösteren bir çimento fabrikasına ait 5 yıllık üretim verisi kullanılarak, çimento üretim tahmininde makine öğrenmesi uygulaması gerçekleştirilmiştir. Özellikle, Şubat 2023 depreminin ardından hızla büyüyen çimento talebi ve artan üretim dinamikleri göz önüne alınarak, SVR algoritmasının farklı çekirdek fonksiyonlarıyla üretim tahmin başarısı karşılaştırılmıştır. Radial basis function (RBF), lineer, sigmoid ve polinomsal kernel türlerinin test edildiği deneylerde, en başarılı sonuçları SVR-RBF modeli vermiştir. Modelin başarısı şu metriklerle ifade edilmiştir: MSE = 0,002926, RMSE = 0,0541, MAE = 0,0486, MAPE = %5,27. Çalışma, yalnızca makine öğrenmesi algoritmalarının üretim planlamasına entegrasyonunu değil, aynı zamanda dinamik piyasa koşullarında stratejik karar destek sistemleri geliştirilmesini mümkün kılan veri temelli yaklaşımın önemini vurgulamaktadır. Bu yönüyle yapı kimyasalları sektöründe veri madenciliği temelli üretim tahmini literatürüne önemli bir katkı sunmaktadır.

Kittinaraporn vd. (2022) tarafından gerçekleştirilen bu çalışmada, geri dönüştürülmüş agrega içeren beton kullanılarak gerçekleştirilen inşaat faaliyetlerinin çeşitli çıktılarının (maliyet, beton hacmi ve toplam inşaat süresi) yapay sinir ağı modelleri ile tahmin edilmesi hedeflenmiştir. Gelişmekte olan ekonomilere sahip ülkelerde sürdürülebilir inşaat çözümlerinin yaygınlaştırılması amacıyla yapılan bu araştırma, toplam 909 gözlemden oluşan bir veri setiyle yürütülmüştür. Modelleme sonuçlarına göre; tahmin edilen çıktılar için sırasıyla agrega içeren beton maliyetinde $R^2 = 0,999$, beton hacminde $R^2 = 0,976$, ve toplam inşaat süresinde $R^2 = 0,968$ gibi oldukça yüksek başarı değerlerine ulaşılmıştır. Ayrıca MAPE değeri maliyet için %1,98 gibi düşük bir seviyede kalmıştır. Bu bulgular, yapay sinir ağı modellerinin hem yüksek doğruluk sunduğunu hem de zaman serisi analizi yoluyla sürdürülebilir beton üretimi ve planlamasında etkin şekilde kullanılabileceğini göstermektedir. Bu yönüyle çalışma, yapı kimyasalları alanında geri kazanılmış malzemelerin kullanımının ve yapay zekâ destekli tahmin mekanizmalarının entegrasyonu açısından dikkat çekici bir katkı sunmaktadır.

İncelenen çalışmaların karşılaştırmalı özeti Çizelge 2.1’de tablo olarak sunulmuştur.

Çizelge 2.1. Yapı Kimyasalı Üretimi ve Makine Öğrenmesi Algoritmalarının Birlikte Kullanıldığı Çalışmalar için Literatür Araştırması Özet Tablosu

Yazar(lar) – Yıl	Konu	Kullanılan Yöntemler
(Ozcan vd., 2024)	Zeolit ve diatomit katkılı betonların basınç dayanımının makine öğrenmesi ile tahmini	RF, Support Vector Machines, XGBoost, LightGBM, CatBoost
(Parhi vd., 2024)	Alkali aktivasyonlu kendiliğinden yerleşen betonlarda dayanım tahmini	XGBoost, RF, SVR, Karar Ağacı, GB
(Childs, 2021)	Çimentolu sistemlerde malzeme bilimi ve makine öğrenmesi ile modelleme ve eniyileme	Gaussian Process Regression, Bayesian Optimization, Kernel Ridge Regression, Graph-Based Featurization
(Tran vd., 2022)	Kendiliğinden yerleşen betonlarda dayanım tahmini ve model karşılaştırması	XGBoost, Lineer Regresyon, K-Nearest Neighbors, Support Vector Machine, Karar Ağaçları, RF, GB ve Parçacık Sürü Optimizasyonu ile hiperparametre eniyilemesi
(Kumar vd., 2024)	Kendiliğinden yerleşen betonlarda dayanım ve dayanıklılık özelliklerinin tahmini	Yapay Sinir Ağı & teaching-learning-based optimization, ant colony optimization, imperialist competitive algorithm ve shuffled complex evolution
(Polat vd., 2024)	Çimento üretim tahmini (5 yıllık veriye dayalı)	SVR & Radial basis function (RBF), Linear, Sigmoid, Polynomial çekirdek fonksiyonları
(Kittinaraporn vd., 2022)	Geri dönüştürülmüş agrega betonla inşaat çıktılarının tahmini	Yapay Sinir Ağı

Bu alt başlık altında özetlenen çalışmalar göstermektedir ki gerek klasik üretim süreçlerinde gerekse alternatif yapı kimyasalları ile geliştirilen yenilikçi beton sistemlerinde, makine öğrenmesi yöntemleri güçlü öngörü araçları olarak kullanılmaktadır. Algoritmaların performansı, veri yapısı ve problem türüne göre değişmekle birlikte, özellikle ağaç tabanlı modeller (XGBoost, RF) ve yapay sinir ağı tabanlı hibrit yaklaşımlar, yüksek doğruluk oranlarıyla öne çıkmaktadır. Ayrıca meta-sezgisel algoritmalarla yapılan eniyilemeler ve duyarlılık analizleri, model başarılarını artırmakta ve yorumlanabilirlik sağlamaktadır. Bu bölümde sunulan literatür, yapı kimyasalları üretimi ve performans öngörüsünde veri temelli

karar destek sistemlerinin gelişimine önemli katkılar sunmaktadır. İzleyen bölümde, makine öğrenmesi yöntemleri ile fiyat tahmini üzerine odaklanan çalışmalar incelenmektedir.

2.2. Makine Öğrenmesi Yöntemleri ile Fiyat Tahmini Çalışmaları

Fiyat tahmini, ekonomik karar destek sistemlerinde kritik bir rol oynamakta olup özellikle enerji, emtia, finans ve yapı sektörlerinde stratejik öngörüler geliştirilmesine olanak tanımaktadır. Geleneksel zaman serisi analizleri ve istatistiksel modelleme tekniklerinin, doğrusal varsayımlar ve sınırlı veri esnekliği nedeniyle karmaşık ekonomik ilişkileri yansıtmakta yetersiz kaldığı pek çok çalışmada ortaya konmuştur. Bu sınırlılıklara karşılık, son yıllarda makine öğrenmesi temelli yöntemler; doğrusal olmayan ilişkileri yakalayabilme kapasiteleri, otomatik öğrenme yetenekleri ve yüksek uyarlanabilirlikleri sayesinde literatürde öne çıkmaktadır. Bu bağlamda, bu bölümde son beş yıl içerisinde yayımlanmış, farklı sektörlerde makine öğrenmesi tabanlı fiyat tahminine odaklanan sekiz akademik yayın detaylı olarak incelenmiş ve kullanılan algoritmalar, veri yapıları ve öngörü başarısına ilişkin bulgular karşılaştırmalı olarak sunulmuştur.

Lu vd. (2025) tarafından gerçekleştirilen çalışmada, nadir toprak elementlerinin üretim süreçlerinin uzunluğu ve karmaşıklığı göz önünde bulundurularak, kısa vadeli fiyat tahminlerinin yetersizliği vurgulanmakta ve bu nedenle uzun vadeli zaman serisi tahminlerine (long sequence time-series forecasting - LSTF) odaklanan bir model önerilmektedir. Araştırmada geliştirilen variational mode decomposition - series random forest (VMD–SRF) hibrit modeli, ham fiyat verilerinin VMD ile ön işlenmesini ve ardından SRF algoritması ile öngörü yapılmasını içermektedir. SRF modeli, RF algoritmasının uzun zaman serilerinde yaşadığı doğruluk kaybını gidermek amacıyla, Dynamic Time Warping (DTW) mesafesini sezgisel bilgi olarak kullanmakta ve tahmin performansını artırmaktadır. Dört farklı nadir toprak elementi fiyat serisi üzerinde yapılan deneysel analizler, önerilen VMD–SRF yaklaşımının rakip modellere kıyasla daha yüksek doğruluk sağladığını ortaya koymuştur. Bu yönüyle çalışma, yalnızca nadir toprak ürünleri için değil, aynı zamanda uzun dönemli tahmin gereksinimi olan diğer sektörlerde de uygulanabilir bir model önerisi sunmaktadır.

Saeed vd. (2025) tarafından yayımlanan çalışma, emtia fiyat tahmini konusuna odaklanan sistematik bir literatür incelemesidir. Çalışmada, özellikle konut fiyatları, hisse senedi fiyatları ve doğal gaz fiyatları gibi farklı emtia türleri ayrı ayrı ele alınarak, bu alanlarda kullanılan makine öğrenmesi teknikleri kıyaslanmıştır. Kullanılan yöntemler arasında topluluk öğrenmesi modelleri, yapay sinir ağları, SVM, zaman serisi analizleri ve regresyon modelleri yer almaktadır. Araştırma, bu yöntemlerin avantajları, sınırlılıkları ve hangi durumlarda daha başarılı olduklarını karşılaştırmalı olarak değerlendirmiştir. Özellikle konut fiyat tahmininde makine öğrenmesinin güçlü potansiyeli vurgulanmakta, model başarısını etkileyen başlıca değişkenler tartışılmakta ve gelecek çalışmalar için yönlendirici bir çerçeve sunulmaktadır. Bu kapsamda çalışma, fiyat tahmini literatürüne teorik katkılar sağlamasının yanı sıra, karar vericiler ve politika geliştiriciler için de rehber niteliği taşımaktadır.

Kantanantha ve Pattaraumpornchai (2023) tarafından gerçekleştirilen çalışmada, Tayland'daki polipropilen granül fiyatlarının tahmini amacıyla üç farklı makine öğrenmesi algoritması kıyaslanmıştır. Çalışma kapsamında SVR, XGBoost ve Yapay Sinir Ağları yöntemleri kullanılmış; tahmin doğruluğu MAPE metriğiyle ölçülmüştür. 2011–2021 dönemini kapsayan 11 yıllık verinin aylık bazda yapılan analizinde; polipropilen granül fiyatı, üretim endeksleri, ithalat-ihracat verileri, petrol fiyatları ve ekonomik göstergeler gibi birçok açıklayıcı değişken kullanılmıştır. Sonuçlara göre, Yapay Sinir Ağları algoritması test setinde %5,89 MAPE değeri ile en yüksek doğruluğu sağlamış ve uygulamaya en uygun model olarak belirlenmiştir. Çalışma, plastik hammaddelerinin maliyet kontrolü açısından makine öğrenmesinin endüstriyel planlamaya katkı potansiyelini göstermesi açısından dikkat çekicidir.

Guennioui vd. (2024) tarafından yapılan bu çalışma, yüksek piyasa oynaklığı dönemlerinde hisse senedi fiyatlarının tahminine yönelik eniyilenmiş LightGBM modeli kullanımını konu almaktadır. Çalışmada, Covid-19 pandemisinin etkisi altındaki Fas borsası örnek olarak ele alınmış ve özellikle değişkenliği en yüksek sektörlerden biri olan kimya sektörüne odaklanılmıştır. Model, yalnızca sektör içindeki hisse senetlerini değil, aynı zamanda yüksek korelasyona sahip diğer sektörlerin hisse senetlerini de dikkate alarak tahmin gücünü artırmıştır. Elde edilen bulgular, eniyilenmiş LightGBM modelinin hem doğruluk ($R^2 = 0,99$) hem de işlem süresi açısından Long Short-Term Memory (LSTM),

Gated Recurrent Unit (GRU) ve convolutional neural network (CNN) gibi derin öğrenme modellerinden üstün olduğunu ortaya koymuştur. Bu sonuçlar, özellikle kriz dönemlerinde, farklı sektörlerden verilerin entegre edilmesinin makine öğrenmesi tabanlı fiyat tahmin performansını artırabileceğini göstermektedir.

Pavankumar ve Velmurugan (2023) tarafından gerçekleştirilen çalışmada, Bitcoin fiyat tahmini için farklı regresyon tabanlı makine öğrenmesi algoritmalarının karşılaştırması yapılmıştır. Araştırma kapsamında Ridge Lineer Regresyon ve Lasso Regresyon modelleri karşılaştırılmış; model başarıları çeşitli istatistiksel testlerle analiz edilmiştir. Çalışmada 5 dakikalık Bitcoin fiyat verileri, piyasa hacmi ve yatırımcı ilgisi gibi çeşitli değişkenler kullanılmıştır. Bulgulara göre, Ridge Lineer Regresyon modelinin ortalama doğruluğu %93,05, Lasso Regresyonun doğruluğu ise %91,14 olarak hesaplanmıştır. Ayrıca bağımsız örneklem T-testi sonuçları, Ridge Regresyon modelinin performansının istatistiksel olarak anlamlı derecede daha yüksek olduğunu göstermektedir ($p < 0,05$). Bu çalışma, özellikle kripto para piyasalarında regresyon tabanlı modellerin karşılaştırmalı performans analizlerine yönelik literatüre katkı sağlamaktadır.

Li (2021) tarafından yayımlanan çalışmada, Çin'deki kömür fiyatlarının tahmini için RF ve Lasso Regresyon modelleri birlikte değerlendirilmiştir. Araştırma kapsamında, Qinhuangdao Limanı güç kömürü fiyatları temel alınarak yedi farklı makroekonomik gösterge (hammadde üretimi, tüketici fiyat endeksi, ithalat-ihracat, petrol fiyatları vb.) modele dâhil edilmiştir. RF algoritması ile değişken önem düzeyleri sıralanmış, en belirleyici faktörün hammadde üretimi olduğu tespit edilmiştir. Ardından Lasso Regresyon modeli ile fazla parametreler elenmiş, 10 katlı çapraz doğrulama ile eniyi lambda değeri belirlenmiş ve modelin sadeleştirilmiş versiyonu geliştirilmiştir. Ayrıca çalışmada, COVID-19 gibi beklenmeyen dış faktörlerin etkisini modellemek amacıyla kukla değişkenler tanımlanarak regresyon modeli güncellenmiştir. Bulgulara göre Lasso Regresyon modeli düşük hata oranları ile yüksek doğruluk sağlamış ve haftalık/günlük kömür fiyatları için de çok terimli polinomlarla tahmin başarıyla genişletilmiştir. Bu çalışma, fiyat tahmininde hem algoritmik hem de politika odaklı çok katmanlı bir yaklaşım sunmasıyla dikkat çekicidir.

Cai, vd. (2024), karbon fiyat tahmininde karşılaşılan bilgi sızıntısı ve sınır etkisi gibi temel sorunları ele almak üzere, genişletilmiş kayan pencere ayrıştırması (extended sliding

window decomposition – ESWD) ve çok deęişkenli ampirik mod ayrıştırması (multivariate empirical mode decomposition - MEMD) temelli yenilikçi bir yaklaşım önermektedir. Model, MEMD ile zaman serisinin çok ölçekli bileşenlerini ayrıştırarak LSTM ve SVR algoritmalarını bir araya getirmektedir. Çin’in en büyük karbon piyasası olan Guangzhou Karbon Borsası’ndan elde edilen verilerle yapılan deneysel analizler, önerilen MEMD-ESWD-LSTM-SVR modelinin geleneksel LSTM ve SVR modellerine kıyasla RMSE açısından %66,6’ya varan iyileştirmeler sağladığını ortaya koymuştur. Ayrıca, yönsel doğruluk (Dstat) gibi göstergelerde de önemli gelişmeler kaydedilmiştir. Bu çalışma, karbon fiyatlarının doğruluğunu artırmaya yönelik melez modellerin önemini vurgularken, bilgi sızıntısı gibi uygulama problemlerine karşı ileri düzeyde çözümler sunmaktadır.

Asgari vd. (2024) tarafından gerçekleştirilen çalışmada, kripto para piyasasının öncü varlığı olan Bitcoin’in fiyat tahmini amacıyla, Differential Evolution Algorithm (DE) meta-sezgisel algoritması ile makine öğrenmesi modellerinin (MLPR, Lineer Regresyon, Bayesian Ridge Regresyon, RF, Lasso Regresyon, SVR) kullanıldığı melez bir yaklaşım önerilmiştir. Çalışmada, ilk kez DE algoritması kullanılarak istifleme modelindeki makine öğrenimi algoritmalarının parametreleri eniyilenmiş ve en uygun model kombinasyonu belirlenmiştir. OHLC (Open-High-Low-Close) verileri, teknik analiz göstergeleri (160’tan fazla) ve temel analiz göstergeleri (7 adet) olmak üzere üç farklı senaryoda model eğitimi gerçekleştirilmiş, her senaryoda farklı algoritma sıralamalarının öngörü doğruluğunu etkilediği raporlanmıştır. Bitcoin’in günlük USD bazlı fiyatları kullanılarak yapılan analizlerde, temel analiz verilerinin de model girdisine dâhil edilmesinin tahmin gücünü artırdığı gözlemlenmiştir. Çalışma, kripto para tahmin modellerinde kullanılan geleneksel teknik göstergelere alternatif olarak temel analiz temelli fiyatlama bileşenlerinin entegre edilmesini sağlayan ilk örneklerden biri olması bakımından literatüre özgün bir katkı sunmaktadır.

İncelenen çalışmaların karşılaştırmalı özeti Çizelge 2.2’de tablo olarak sunulmuştur.

Çizelge 2.2. Makine Öğrenmesi Yöntemleri ile Fiyat Tahmini Çalışmaları için Literatür Araştırması Özet Tablosu

Yazar(lar) – Yıl	Konu	Kullanılan Yöntemler
(Lu vd., 2025)	Nadir toprak elementlerinde uzun vadeli fiyat tahmini	VMD, SRF, DTW
(Saeed vd., 2025)	Konut, hisse ve doğal gaz fiyat tahmini	Topluluk Öğrenmesi Modelleri, Yapay Sinir Ağları, SVM, Zaman Serisi Analizleri, Regresyon Modelleri
(Kantanantha ve Pattaraumpornchai, 2023)	Polipropilen granül fiyat tahmini	SVR, XGBoost, Yapay Sinir Ağları
(Guennioui vd., 2024)	Hisse senedi fiyat tahmini (piyasa stresi)	LightGBM, LSTM, GRU, CNN
(Pavankumar ve Velmurugan, 2023)	Bitcoin fiyat tahmini	Ridge Lineer Regresyon, Lasso Regresyon
(Li, 2021)	Çin kömür fiyat tahmini	RF, Lasso Regresyon
(Cai vd., 2024)	Karbon fiyat tahmini	ESWD, MEMD & LSTM, SVR
(Asgari vd., 2024)	Bitcoin fiyat tahmini (melez modelleme)	DE & MLPR, Lineer Regresyon, Bayesian Ridge Regresyon, RF, Lasso Regresyon, SVR

İncelenen çalışmalar göstermektedir ki; regresyon, ağaç tabanlı yöntemler (RF, XGBoost, LightGBM), yapay sinir ağları (ANN, MLP), LSTM gibi derin öğrenme mimarileri ve melez modelleme yaklaşımları, farklı fiyat tahmin senaryolarında başarıyla uygulanabilmektedir. Bununla birlikte, kullanılan algoritmaların başarısı; veri setinin yapısı, özniteliklerin doğası, tahmin edilmek istenen zaman aralığı (kısa–orta–uzun vadeli) ve sektörel dinamikler gibi pek çok faktörden etkilenmektedir. Ayrıca, hiperparametre eniyilemesi, özellik seçimi ve zaman serisi ayrıştırma gibi ön işleme teknikleri model başarısını önemli ölçüde artırmaktadır. Bu nedenle fiyat tahmini için geliştirilecek makine öğrenmesi tabanlı modellerde, yalnızca algoritma seçimi değil, bütünsel bir veri işleme süreci de belirleyici olmaktadır.

Farklı sektörlerde makine öğrenmesiyle gerçekleştirilen fiyat tahmini uygulamalarına ilişkin literatür geniş bir yelpazeye yayılmaktadır. Bu çalışmaların sunduğu yöntemsel çeşitlilik, çimento sektörü gibi üretim süreci ve maliyet dinamikleri oldukça karmaşık olan alanlara da uyarlanabilir yaklaşımlar geliştirilmesine olanak sağlamaktadır. Bu doğrultuda, literatür araştırması kapsamında genelden özele doğru ilerleyen yapı gereği,

bir sonraki bölümde özellikle çimento fiyatlarının tahminine odaklanan akademik çalışmalara yer verilerek sektör özelinde geliştirilen modeller, kullanılan veri kaynakları ve yöntemsel tercihler karşılaştırmalı olarak ele alınacaktır.

2.3. Makine Öğrenmesi ile Çimento Fiyat Tahminine Yönelik Akademik Çalışmalar

Makine öğrenmesi yöntemleri son yıllarda farklı sektörlerde fiyat tahminine yönelik çalışmalarda sıklıkla kullanılmaya başlanmış olsa da çimento ve yapı kimyasalı üretimi sektörüne özgü uygulamaların sayısı literatürde oldukça sınırlıdır. Yapı Kimyasalı malzemelerinin hammaddeleri arasında stratejik bir öneme sahip olan çimentoya yönelik fiyat öngörülerinin doğruluğu, özellikle üretim planlaması ve maliyet kontrolü açısından büyük önem arz etmektedir. Bu bölümde, yalnızca makine öğrenmesi teknikleri kullanılarak çimento fiyatlarının tahmin edilmesini amaçlayan sınırlı sayıdaki akademik yayın detaylı olarak incelenmiştir. Elde edilen bulgular, bu alandaki çalışmaların sayıca az olmasının yanı sıra yöntemsel olarak da çeşitlilikten uzak olduğunu göstermekte; dolayısıyla konuya özgü kapsamlı, uygulamaya dönük ve karar destek potansiyeli yüksek çalışmalara duyulan ihtiyacı ortaya koymaktadır.

Afolabi ve Abimbola (2022) tarafından gerçekleştirilen bu çalışmada, Nijerya'daki inşaat sektörüne yönelik olarak geliştirilen web tabanlı bir sistem aracılığıyla çimento fiyatlarının tahmini hedeflenmiştir. Geliştirilen sistem, makine öğrenmesi algoritmalarını geçmiş veriler (çimento fiyatları, petrol ve dizel fiyatları, faiz oranı ve döviz kuru) üzerinde eğiterek gelecekteki çimento fiyatlarını öngörebilmektedir. Python kullanılarak geliştirilen lineer regresyon temelli model, Flask çatısı altında HTML, CSS, MySQL ve PHP ile entegre edilmiştir. Modelin doğruluk oranı yaklaşık %80 olarak raporlanmış ve değişkenlerin çimento fiyatı üzerindeki etkisi serpme diyagramları ile görselleştirilmiştir. Çalışma, büyük veri analitiği ve bilgi yönetimiyle desteklenen bu sistemin, ihaleye hazırlık ve maliyet planlaması süreçlerinde karar vericilere önemli katkılar sağlayabileceğini vurgulamaktadır. Bu yönüyle çalışma, makine öğrenmesi tekniklerinin çimento fiyat tahmini amacıyla sektörel bazda kullanılmasına dair önemli bir örnek teşkil etmektedir.

Liu vd. (2024) tarafından gerçekleştirilen bu derleme çalışması, inşaat projelerinin finansal başarısını etkileyen en kritik unsurlardan biri olan inşaat malzemeleri fiyatlarının

öngörülmesi üzerine veri odaklı yaklaşımları sistematik biçimde ele almaktadır. Çalışmada, makine öğrenmesi tekniklerinin fiyat tahmininde giderek daha fazla benimsendiği belirtilmiş ve yöntemler nedensel modelleme ve zaman serisi analizleri olarak iki ana kategoriye ayrılmıştır. Multiple-Linear Regression (MLR), Artificial Neural Networks (ANN) ve Least Square Support Vector Machine (LSSVM) gibi modellerin Tüketici Fiyat Endeksi (TÜFE), Üretici Fiyat Endeksi (ÜFE), gayri safi yurtiçi hasıla (GSYH) vb. ekonomik göstergeler üzerinden fiyatları tahmin edebildiği, zaman serisi tabanlı modellerin ise tarihsel fiyat eğilimlerini analiz ederek daha az veri gereksinimiyle kısa vadeli tahminler için avantaj sağladığı vurgulanmaktadır. Ayrıca Monte Carlo simülasyonunun belirsizlik analiziyle birlikte kullanıldığı örnekler incelenmiş, karma modellerin hem doğruluk hem de güvenilirlik açısından daha etkin sonuçlar verdiği ifade edilmiştir. Derleme çalışması, makine öğrenmesinin yapısal tahmin gücü ile zaman serisi analizlerinin zamansal dinamikleri birlikte ele alarak gelecekteki fiyat tahmini çalışmalarında melez yaklaşımların daha fazla öne çıkacağını önermektedir. Bu yönüyle çalışma, çimento fiyat tahmini araştırmalarında hem yöntemsel çeşitliliği hem de araştırma boşluklarını ortaya koyan kapsamlı bir kaynak niteliğindedir.

Ma vd. (2024) tarafından gerçekleştirilen bu sistematik derleme çalışması, inşaat sektöründe kullanılan fiyat tahmin yöntemlerini değerlendirmek amacıyla, 2012–2022 yılları arasında yayımlanmış literatürü kapsamlı şekilde incelemiştir. Üç aşamalı bir değerlendirme süreci sonucunda seçilen 34 makale, özellikle inşaat malzemeleri fiyatları, toplam inşaat maliyetleri ve inşaat maliyet endeksine (Construction Cost Index - CCI) odaklanmaktadır. Çalışma, araştırmacıların en sık tercih ettiği tahmin modellerinin Vektör Hata Düzeltme Modeli (Vector Error Correction Model - VECM), ANN ve Otoregresif Entegre Hareketli Ortalama (Autoregressive Integration Moving Average – ARIMA) olduğunu ortaya koymuştur. Ayrıca zaman serisi analizleri ve makine öğrenmesi tekniklerinin sektörde yaygın biçimde kullanıldığı tespit edilmiştir. Makalede, karma modellerin, farklı modelleme tekniklerinin güçlü yönlerini bir araya getirerek tahmin doğruluğunu artırabileceği vurgulanmakta ve bu yöntemlerin gelecekteki araştırmalar için güçlü bir alternatif oluşturabileceği belirtilmektedir. Ayrıca doğal afetler, pandemiler ve jeopolitik olayların inşaat sektöründeki fiyat dalgalanmalarına etkisi üzerine yapılacak çalışmaların, sektörel kırılganlığı azaltmada önemli katkılar sunabileceği ifade edilmiştir.

Bu çalışma hem araştırmacılara hem de sektördeki karar vericilere model seçimi konusunda rehberlik edebilecek geniş kapsamlı bir literatür sentezi sunmaktadır.

Makine Öğrenmesi Yöntemleri ile Çimento Fiyat Tahmini Çalışmaları için Literatür Araştırması Özet Tablosu Çizelge 2.3'te verilmiştir.

Çizelge 2.3. Makine Öğrenmesi Yöntemleri ile Çimento Fiyat Tahmini Çalışmaları için Literatür Araştırması Özet Tablosu

Yazar(lar) – Yıl	Konu	Kullanılan Yöntemler
(Afolabi ve Abimbola, 2022)	Web tabanlı sistem ile Nijerya'da çimento fiyat tahmini	Lineer Regresyon
(Liu vd., 2024)	İnşaat malzemesi fiyatlarının veri odaklı tahmini	MLR, ANN, LSSVM & zaman serisi analizi, Monte Carlo Simülasyonu
(Ma vd., 2024)	İnşaat sektörü için fiyat tahmini modellerinin sistematik derlemesi	VECM, ANN, ARIMA, zaman serisi analizi

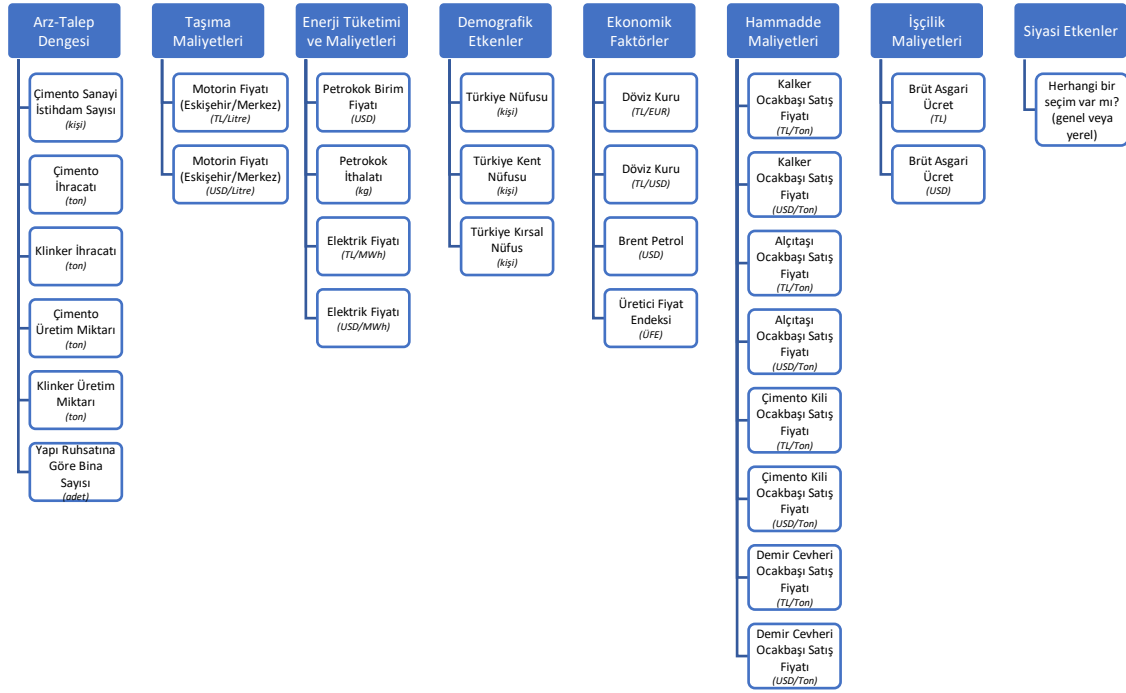
Bu bölümde ele alınan literatür taramaları sonucunda, makine öğrenmesi temelli modeller ile çimento hammaddesi fiyatlarının doğrudan tahmin edildiği yalnızca üç akademik yayına ulaşılabilmektedir. Üstelik bu üç yayının hiçbiri, bu tez çalışmasında olduğu gibi Türkiye özelinde, kapsamlı ekonomik ve sektörel özniteliklerle desteklenmiş bir fiyat tahmin sistemi sunmamaktadır. Bu durum, alana yönelik mevcut akademik literatürün hâlen gelişmekte olduğunu ve yöntemsel çeşitlilik ile veri derinliği açısından önemli bir boşluk barındırdığını göstermektedir. Dolayısıyla bu tez kapsamında yapılan çalışma, çimento fiyat tahmini konusunda makine öğrenmesinin uygulanabilirliğini sorgulayan ve bu alana yeni ve özgün bir katkı sağlayan önemli bir çalışma olarak değerlendirilebilir.

3. MATERYAL VE YÖNTEM

Araştırma süreci, veri toplama, ön işleme, öznitelik mühendisliği, modelleme ve model performanslarının değerlendirilmesi olmak üzere ardışık ve sistematik bir yapıya sahiptir. Bu bölümde; mevcut veri setinin genel analizi; öznitelik seçim araçları; geçmiş dönemlere ait çok değişkenli veri seti ile çimento satın alım fiyatlarını tahmin çalışması yapılırken kullanılan eksik veri doldurma metotları, makine öğrenmesi yöntemleri; sonuçların değerlendirilmesinde kullanılan performans metrikleri ve hiperparametre eniyilemesi yapmak için kullanılan algoritmanın açıklamalarına yer verilmiştir.

3.1. Öznitelik Seçim Süreci ve Özniteliklerin Sektörel Önemi

Bu çalışmada kullanılan veri seti, 2013 yılı Ocak ayından 2024 yılı Aralık ayına kadar olan dönemi kapsayan, aylık olarak derlenmiş çok değişkenli zaman serisi verilerinden oluşmaktadır. Veriler, yapı kimyasalı üreticisinin Eskişehir fabrikasında yürütülen çimento satın alım süreçlerinin tahminine yönelik karar destek modelleri geliştirmek amacıyla yapılandırılmıştır. Bununla birlikte veri seti yalnızca firma içi kayıtlarla sınırlı tutulmamış, Türkiye Çimento Müstahsilleri Birliği (TÇMB), Türkiye İstatistik Kurumu (TÜİK), Türkiye Cumhuriyet Merkez Bankası Elektronik Veri Dağıtım Sistemi (TCMB EVDS), Enerji Piyasası Düzenleme Kurumu (EPDK) ve Google Trends gibi güvenilir ulusal ve uluslararası kaynaklardan elde edilen verilerle desteklenmiştir. Böylece çimento fiyatlarını etkileyebilecek yapısal ve konjonktürel faktörlerin kapsamlı bir şekilde analiz edilmesi hedeflenmiştir. Eskişehir Fabrikası için kullanılacak öznitelikler Şekil 3.1’de sunulmuştur.



Şekil 3.1. Tematik Başlıklar Temelinde Potansiyel Öznitelik Seti

Şekil 3.1’de görüleceği üzere toplamda otuz değişkenden oluşan veri seti, sektörel dinamikleri ve makroekonomik çevresel faktörleri temsil eden sekiz tematik başlık altında toplanmıştır.

Arz-talep dengesi temalı öznitelikler, çimento üretim ve tüketim dengesini anlamlandırmaya yöneliktir. Literatürde, çimento fiyatlarını etkileyen başlıca faktörler arasında üretim hacmi, ulaşım maliyetleri ve talep elastikiyeti öne çıkmaktadır (Miller vd., 2025). Üretim ve ihracat verileri, piyasaya arz edilen çimento miktarını doğrudan etkilerken; yapı ruhsatı sayıları ve sanayi istihdamı gibi göstergeler sektörel talep potansiyelini yansıtmaktadır. Özellikle yapı ruhsatları, inşaat sektöründe geleceğe yönelik üretim talebinin öncü göstergeleri arasında yer almakta ve ekonomik aktiviteyle sıkı ilişkisi olduğu ifade edilmektedir (Mare ve Graham, 2013). Benzer şekilde, klinker üretim ve ihracat düzeyleri, çimento tedarik zincirinin temel bileşenleri olan yarı mamul hareketlerini anlamada kritik rol oynamaktadır. Özellikle üretim hacmi, ihracat kapasitesi ve lojistik maliyetlerin birlikte değerlendirilmesi, çimento fiyatlarının öngörülmesinde önemli bir bütünlük sunar (Miller vd., 2025).

Nakliye maliyetleri kapsamında yer alan motorin fiyatları, çimento gibi ağır ve hacimli malzemelerin taşınmasında lojistik giderlerin büyük bir bölümünü oluşturduğu için dikkate alınmıştır. Ulaştırma ekonomisi literatürüne göre enerji fiyatlarındaki artışlar doğrudan taşımacılık maliyetlerini yükseltmekte ve bu da tedarik zinciri maliyetlerine yansımaktadır (Rodrigue, 2020). Bu bağlamda, Eskişehir özelindeki motorin fiyatları, çimento hammaddesinin fabrikaya taşınmasında bölgesel bir maliyet göstergesi sunmaktadır.

Enerji tüketimi ve maliyetleri kategorisi, çimento üretim sürecinde yüksek oranda enerji tüketilmesi nedeniyle büyük öneme sahiptir. Özellikle döner fırınlar ve öğütme tesislerinde kullanılan elektrik enerjisi ile petrokok gibi fosil yakıtlar, çimento üretim maliyetlerinin önemli bileşenlerindendir (Benhelal vd., 2013). Bu nedenle hem ilgili kaynakların birim fiyatları hem de tedarik hacimleri veri setine dâhil edilmiştir. Enerji maliyetleri, yalnızca üretim süreçlerine değil aynı zamanda dolaylı olarak tedarik sürekliliğine ve kapasite kullanımına da etki etmektedir.

Demografik etkenler, özellikle yapı kimyasallarına olan talebin sosyoekonomik temellerini anlamaya yardımcı olur. Nüfus artışı, kentleşme oranındaki değişim ve kırsal-kentsel nüfus dengesi, inşaat sektörü talebini tetikleyen temel unsurlar arasında yer almaktadır. Türkiye’de 1960–1990 döneminde yaşanan hızlı kentleşme, çimento üretimi ve tüketimini artırmış; bu süreçte hem üretici sayısı hem de talep kayda değer şekilde yükselmiştir. Küresel ölçekte de kentleşme, çimento talebinin önemli belirleyicilerinden biri olarak kabul edilmekte ve şehirleşme oranının artışıyla beraber kişi başına düşen çimento tüketimi yükselmektedir (Beygo ve Cakmak, 1998).

Türkiye özelinde artan kent nüfusu, yapılaşma ihtiyacını beslemekte; bu da yapı kimyasalları ve çimentoya olan talebi yükseltmektedir. Dolayısıyla, demografik göstergeler hem dolaylı talep analizi hem de üretim planlaması açısından kritik bir bağlam sunmaktadır.

Ekonomik faktörler arasında döviz kuru, çimento üretim maliyetlerini etkileyen önemli bir unsur olarak öne çıkar. Türkiye’de enerji ve hammadde girdilerinin çoğu dövizle ithal edildiği için USD/TRY kurundaki oynaklık doğrudan maliyetlere yansımaktadır (Tatliyer ve Yigit, 2016). Ayrıca, Brent petrol fiyatları lojistik ve akaryakıt giderlerini

belirlediğinden, taşımacılık maliyetleri ve dolaylı çimento fiyatlarında önemli bir rol oynamaktadır (Başarır ve Erçakar, 2016). Üretici fiyat endeksi (ÜFE) ise genel ekonomik ortamın üretim maliyetleri üzerindeki baskısını yansıtarak fiyatlama stratejilerinde önemli bir referans oluşturur (Bilgili vd., 2024).

Hammadde maliyetleri başlığı altında yer alan kalker, çimento kili, alçıtaşı ve demir cevheri gibi malzemeler, çimento üretiminin doğrudan bileşenlerini oluşturur. Bu hammaddelerin ocakbaşı satış fiyatları, üreticinin karşılaştığı doğrudan maliyetlerin önemli bir göstergesidir. Ayrıca bu fiyatların hem TL hem de USD cinsinden takibi, döviz kurundaki değişimlerin satın alma maliyetlerine etkisini ayırtırmaya olanak tanır. Literatürde, yerli ve ithal hammadde kaynaklarının birlikte kullanıldığı durumlarda çift para birimli (dual-currency) analizlerin fiyat öngörülebilirliğini artırdığı vurgulanmaktadır (da Silva vd., 2020).

İşçilik maliyetleri, iş gücü yoğun üretim ve dolum süreçlerinin yerleşik olduğu çimento sektöründe dikkatle takip edilen bir değişkendir. Brüt asgari ücretin hem TL hem de USD bazında değerlendirilmesi, iş gücü maliyetlerinin zaman içindeki reel ve nominal değişimini gözlemlemeye olanak tanır. Özellikle enflasyonist dönemlerde işçilik maliyetlerinin üretim üzerindeki yükünün artması, satın alma ve fiyatlama stratejilerini doğrudan etkilemektedir. Örneğin, Çin gibi emek-yoğun büyüme modellerine sahip ekonomilerde, işgücü maliyeti artışları nihai ürün fiyatlarına tam olarak yansımamakta, bu da üretici kar marjlarını daraltmaktadır. Yine de yükselen nominal ücretler, özellikle üretici ülke döviz kurları sabit kalırsa, enflasyon üzerinde baskı oluşturabilir (Holz ve Mehrotra, 2014).

Son olarak, siyasi etkenler kapsamında ele alınan seçim dönemleri, kamusal harcamaların ve ekonomik beklentilerin yeniden şekillendiği zaman dilimlerini temsil etmektedir. Literatürde seçim dönemlerinin inşaat yatırımlarını artırıcı bir etkisi olabileceği, bunun da inşaat malzemelerine olan talebi geçici olarak artırabileceği belirtilmektedir (Shi ve Svensson, 2006). Bu bağlamda, seçim dönemlerinin varlığı, talep dinamikleri üzerinden çimento satın alım fiyatlarını dolaylı olarak etkileyebilir niteliktedir.

Belirtilen öznitelikler, analiz sürecinin ilk adımında herhangi bir eleme yapılmaksızın veri setine dâhil edilmiş ancak model performanslarını artırmak ve değişkenler arası ilişkileri yönetilebilir düzeye indirmek amacıyla daha sonra öznitelik seçimi süreçlerine başvurulmuştur. Bu kapsamda kullanılan yöntemler ve oluşturulan alt çalışmalara ilişkin detaylar, ilgili bölümlerde kapsamlı biçimde sunulmaktadır.

3.2. Veri Setinin Genel Analizi ve Veri Ön İşleme Süreci

Bu çalışma kapsamında ilk aşama kullanılan veri seti, Ocak 2013 ile Mayıs 2024 arasındaki dönemi kapsayan ve aylık olarak derlenmiş toplam 137 satırdan oluşmaktadır. Makine Öğrenmesi modellerinin çalıştırılma aşamasında tarihin ilerlemesine bağlı olarak veri seti Ocak 2013 ile Aralık 2024 arasındaki dönemi kapsayan yine aylık olarak derlenen toplam 144 satırlık veriyi içerecektir.

Veri seti, yapı kimyasal üreticisinin Eskişehir Fabrikasında çimento hammaddesine ilişkin satın alım fiyatlarının tahminine yönelik yapılandırılmıştır. Veri setinde başlangıçta 30 öznitelik yer almakta olup bu öznitelikler, sektörel bilgi ve akademik literatür ışığında belirlenmiştir. Fakat tahmin modellerinin güvenilirliğini ve genellenebilirliğini artırmak amacıyla, veri sayısı ile öznitelik sayısı arasındaki denge göz önünde bulundurularak öznitelik indirilmesi zorunlu hale gelmiştir.

Regresyon tabanlı makine öğrenimi yöntemlerinde her bir öznitelik başına düşen gözlem sayısının yeterliliği, modelin istikrarı açısından büyük önem arz etmektedir. Özellikle Babyak (2004); Austin ve Steyerberg (2015) gibi çalışmalar, parametrik tahmin modellerinde her bir öznitelik başına en az 10, tercihen 50 gözlem satırının bulunmasının modelin aşırı öğrenme riskini azaltmada kritik rol oynadığını ortaya koymuştur. Bu bağlamda, elde edilecek nihai veri setinde yalnızca 144 aylık gözlem bulunması, 30 öznitelikle doğrudan tahmin modelleri kurulmasının istatistiksel olarak anlamlı sonuçlar üretmesini engelleme riski taşımaktadır. Bu nedenle öznitelik seçim süreçleri yalnızca model performansını artırma amacıyla değil, aynı zamanda bilimsel uygunluk ve metodolojik doğruluk kriterleri çerçevesinde değerlendirilmiştir.

Ham veri setine ilişkin eksik veri gözlemleri Çizelge 3.1’de sunulmuştur.

Çizelge 3.1. Ham Veri Setindeki Eksik Veri Gözlemleri

Öznitelik Kategorisi	Öznitelik İsmi	Birim	Toplam Satır	Mevcut Veri Sayısı	Eksik Veri Sayısı	Eksik Veri Oranı
Arz-Talep Dengesi	Çimento Sanayi İstihdam Sayısı	(kişi)	137	11	126	92%
	Çimento İhracatı	(ton)	137	90	47	34%
	Klinker İhracatı	(ton)	137	90	47	34%
	Çimento Üretim Miktarı	(ton)	137	86	51	37%
	Klinker Üretim Miktarı	(ton)	137	86	51	37%
Nakliye Maliyetleri	Yapı Ruhsatına Göre Bina Sayısı	(adet)	137	135	2	1%
	Motorin Fiyatı (Eskişehir/Merkez)	(TL/Litre)	137	137	0	0%
	Motorin Fiyatı (Eskişehir/Merkez)	(USD/Litre)	137	137	0	0%
Enerji Tüketimi ve Maliyetleri	Petrokok Birim Fiyatı	(USD)	137	135	2	1%
	Petrokok İthalatı	(kg)	137	135	2	1%
	Elektrik Fiyatı	(TL/MWh)	137	137	0	0%
	Elektrik Fiyatı	(USD/MWh)	137	137	0	0%
Demografik Etkenler	Türkiye Nüfusu	(kişi)	137	132	5	4%
	Türkiye Kent Nüfusu	(kişi)	137	132	5	4%
	Türkiye Kırsal Nüfus	(kişi)	137	132	5	4%
Ekonomik Faktörler	Döviz Kuru	(TL/EUR)	137	137	0	0%
	Döviz Kuru	(TL/USD)	137	137	0	0%
	Brent Petrol	(USD)	137	134	3	2%
	Üretici Fiyat Endeksi	(ÜFE)	137	137	0	0%
Hammadde Maliyetleri	Kalker Ocakbaşı Satış Fiyatı	(TL/Ton)	137	108	29	21%
	Kalker Ocakbaşı Satış Fiyatı	(USD/Ton)	137	108	29	21%
	Alçıtaşı Ocakbaşı Satış Fiyatı	(TL/Ton)	137	108	29	21%
	Alçıtaşı Ocakbaşı Satış Fiyatı	(USD/Ton)	137	108	29	21%
	Çimento Kili Ocakbaşı Satış Fiyatı	(TL/Ton)	137	108	29	21%
	Çimento Kili Ocakbaşı Satış Fiyatı	(USD/Ton)	137	108	29	21%
	Demir Cevheri Ocakbaşı Satış Fiyatı	(TL/Ton)	137	72	65	47%
	Demir Cevheri Ocakbaşı Satış Fiyatı	(USD/Ton)	137	72	65	47%
İşçilik Maliyetleri	Brüt Asgari Ücret	(TL)	137	137	0	0%
	Brüt Asgari Ücret	(USD)	137	137	0	0%
Siyasi Etkenler	Herhangi bir seçim var mı? (genel veya yerel)	(Evet/Hayır)	137	137	0	0%
Hedef Değişken	Eskişehir Fabrika Gri Çimento Satın Alım Fiyatı	Endeks	137	134	3	2%

Çizelge 3.1’den görüleceği üzere ham veri seti incelendiğinde bazı özniteliklerde önemli oranlarda eksik veri bulunduğu tespit edilmiştir. Bu eksiklikler, özellikle kamu kurumları tarafından sağlanan ekonomik göstergelerde dönemsel raporlama gecikmeleri veya bazı özel sektör verilerinde erişim kısıtlamalarından kaynaklanmaktadır. Bu doğrultuda analizlerin sağlıklı yürütülebilmesi adına veri seti üzerinde aşamalı bir ön işleme süreci uygulanmıştır. İlk olarak, öznitelik seçim yöntemlerine geçilebilmesi amacıyla veri setine yüzeysel (geçici) doldurma işlemleri uygulanmış ve ilk öznitelik eleme analizleri bu set üzerinde gerçekleştirilmiştir. Ardından nihai analizler için daha ileri düzeyde eksik veri tamamlama teknikleri kullanılarak eksiksiz ve analizlere uygun veri setleri oluşturulmuştur. Böylece hem modellemeye uygun nitelikte eksiksiz veri setleri elde edilmiş hem de analiz öncesi öznitelik seçimlerinin sağlıklı biçimde yapılabilmesi sağlanmıştır.

Veri ön işleme sürecinde yalnızca eksik verilerle başa çıkmakla kalınmamış, aynı zamanda değişken türlerinin (sayısal/kategorik) kontrolü, tarih bilgileri ile etiketleme yapılan “Date” değişkeninin bağımsız değişkenler arasından çıkarılması, gerek duyulan durumlarda ölçü birliği sağlanması (örneğin USD ve TL cinsinden verilerin yorumlanabilirliği) gibi işlemler de gerçekleştirilmiştir. Ayrıca, aynı konsepti temsil eden

bazı değişkenlerin korelasyon analiziyle çoklu doğrusal bağlantı riski taşıyıp taşımadıkları sonraki bölümlerde ele alınmak üzere kayıt altına alınmıştır.

Tüm bu ön işleme süreci, yalnızca modelleme altyapısını hazırlamakla kalmamış, aynı zamanda çalışmanın metodolojik sağlamlığını ve tekrarlanabilirliğini güvence altına alan bir aşama olarak yapılandırılmıştır. Bundan sonraki bölümlerde eksik veri doldurma yöntemleri, öznitelik seçim stratejileri ve kullanılan makine öğrenimi algoritmaları detaylı biçimde açıklanacak; ardından bu yöntemlerin hangi çalışma kapsamında nasıl uygulandığı sistematik olarak sunulacaktır.

3.3. Öznitelik Mühendisliği ve Çalışma Kapsamında Kullanılacak Makine Öğrenmesi Metotları

Veri madenciliği ve makine öğrenmesi temelli modelleme süreçlerinde öznitelik mühendisliği ve doğru algoritma seçimi, tahmin başarısını doğrudan etkileyen en kritik adımlardır. Bu bölümde, yapı kimyasal üreticisinin çimento satın alım fiyatlarına ilişkin karar verme mekanizmasını güçlendirmek üzere geliştirilen tahmin modeli kapsamında izlenen öznitelik mühendisliği adımları ile çalışmada kullanılan makine öğrenmesi algoritmalarının seçimi ve uygulanma gerekçeleri sistematik bir şekilde açıklanacaktır.

Yüksek boyutlu veri yapılarında yalnızca mevcut verilerin doğrudan modele aktarılması yeterli olmayıp, modelleme için uygun nitelikte ve temsili gücü yüksek özniteliklerin seçilmesi, dönüştürülmesi ve gerekiyorsa yeni değişkenlerin türetilmesi büyük önem taşımaktadır. Bu süreç, veri kalitesini artırma ve aşırı öğrenme gibi olası modelleme risklerini azaltma açısından temel bir rol oynamaktadır (Guyon ve Elisseeff, 2003). Çalışmada başlangıçta 30 öznitelik analiz kapsamına alınmış; ancak her bir öznitelik için yeterli gözlem sayısına ulaşma zorunluluğu doğrultusunda indirgeme stratejileri uygulanmıştır.

Bu nedenle öznitelik seçim süreçleri yalnızca model başarısını değil, aynı zamanda metodolojik geçerliliği de güvence altına alacak biçimde yapılandırılmıştır. Öncelikle, sektörel bilgiye dayalı olarak uzman görüşüyle belirlenen değişkenler ile Çalışma 1 gerçekleştirilmiş; ardından korelasyon analizi değerlerine dayalı seçim stratejileri ile Çalışma 2 yapılandırılmıştır. Daha sonra, bu iki yaklaşımın harmanlandığı öznitelik

kümesiyle Çalışma 3 oluşturulmuş; bu veri setinden kapasite kullanım oranı özneteliğinin çıkarılmasıyla Çalışma 4 elde edilmiş; son olarak, hedef değişkenle anlamlı ilişkisi olan ve çoklu doğrusal bağlantı içermeyen değişkenlerin belirlendiği rafine veri seti ile Çalışma 5 gerçekleştirilmiştir.

Öznetelik mühendisliği sürecine paralel olarak, veri setinde yer alan eksik gözlemlerin giderilmesine yönelik olarak yüzeysel ve ileri düzeyde eksik veri doldurma stratejileri uygulanmıştır. Bu yöntemler, öznetelik seçimlerinin sağlıklı yürütülmesine zemin hazırlamış ve modelleme aşamasında eksiksiz, istikrarlı veri yapısının oluşturulmasını sağlamıştır. Aynı zamanda bazı değişkenlerde para birimi (TL/USD) farklılıklarının korunması ya da tarih etiketlerinin model dışı tutulması gibi kararlar da bu mühendislik sürecinin bir parçası olmuştur.

Öznetelik seçimi ve veri iyileştirme işlemlerini takiben, çalışma kapsamında kullanılacak olan makine öğrenmesi algoritmaları belirlenmiş ve her bir modelin çalışma mantığı, avantajları ve bu çalışmaya uygunluk gerekçeleri literatür desteğiyle açıklanmıştır. Bu bağlamda, takip eden alt başlıklarda sırasıyla eksik veri doldurma yöntemleri, öznetelik seçim kriterleri ve araçları ile seçilen makine öğrenmesi algoritmalarına ilişkin detaylar sunulacaktır.

3.3.1. Çalışma kapsamında kullanılan eksik veri doldurma yöntemleri

Veri madenciliği ve makine öğrenmesine dayalı modelleme çalışmalarında, eksik veri problemi araştırma kalitesini ve model başarısını doğrudan etkileyen temel bir sorundur. Eksik gözlemler, özneteliklerin temsili gücünü azaltmakta ve modelin genellenebilirliğini sınırlamaktadır. Bu nedenle, eksik verilerin sistematik ve doğru yöntemlerle tamamlanması, tahmin süreçlerinin sağlıklı yürütülebilmesi açısından kritik önemdedir. Literatürde yer alan farklı doldurma teknikleri; eksikliğin türüne, verinin zaman serisi yapısına, mevsimsellik içerip içermediğine ve istatistiksel dağılımına göre çeşitlilik göstermektedir (Little ve Rubin, 2019).

Bu çalışmada hem basit hem de ileri düzeyde interpolasyon yöntemleri, istatistiksel tahmin tabanlı doldurma stratejileri ve referans serilere dayalı dolaylı tahmin teknikleri kullanılarak eksik gözlemler tamamlanmıştır. Kullanılan yöntemler arasında lineer ve spline

interpolasyon gibi zaman serisi doğrultusunda sürekliliği koruyan yöntemlerin yanı sıra, sabit değer atamaları, parazit (noise) eklenmiş ortalama tahminleri, yıl içi sabit değer tahmini, zincirleme büyüme/küçülme katsayısına dayalı hesaplamalar, temsili ay oranlama (Pro-Rata Dağılım), Locally Estimated Scatterplot Smoothing (LOESS) Tabanlı Mevsimsel-Trend Ayırıştırması (STL) ve benchmarking gibi yöntemler yer almaktadır. Bu çeşitlilik, eksik verinin doğasına ve özniteliğin bağlamsal önemine uygun çözüm üretebilmek amacıyla benimsenmiştir.

İzleyen bölümde, çalışmada kullanılan her bir eksik veri doldurma yöntemi detaylı biçimde açıklanmış; teorik temelleri, matematiksel ifadeleri ve uygulanma gerekçeleriyle birlikte sunulmuştur.

3.3.1.1. Lineer interpolasyon yöntemi

Eksik veri doldurma sürecinde kullanılan yöntemlerden biri lineer interpolasyondur. Bu yöntem, eksik olan bir gözlem değerinin, kendisinden önce ve sonra gelen bilinen iki değer arasında doğrusal bir çizgi boyunca birbirine bağlanmasıyla tahmin edilmesine dayanır. Lineer interpolasyon, özellikle zaman serisi verilerinde ardışık gözlemler arasında belirgin bir sıçrama veya yapısal kırılma olmadığı varsayımı altında etkin şekilde kullanılabilir (Shumway vd., 2000).

Matematiksel olarak lineer interpolasyon yöntemi, x_t zamanında eksik olan değer, $x_{(t-1)}$ ve $x_{(t+1)}$ zamanındaki bilinen komşu gözlemler arasında doğrusal olarak tahmin edilmesiyle elde edilir. Bu yöntem Denklem (3.1) ile ifade edilir.

$$x_t = \frac{(x_{(t+1)} - x_{(t-1)})}{(t + 1 - (t - 1))} * (t - (t - 1)) = \frac{x_{(t-1)} + x_{(t+1)}}{2} \quad (3.1)$$

Özetle eksik değer, iki komşu değer aritmetik ortalaması olarak hesaplanır. Bu yaklaşım, veri kümesinin süreklilik özelliğini bozmadan eksik gözlemleri tamamlamada basit, hızlı ve düşük sapma riski taşıyan bir çözüm sunmaktadır. Yöntemin etkinliği, verideki eğilim veya mevsimsellik gibi yapısal bileşenlerin çok güçlü olmaması durumlarında daha yüksektir (Little ve Rubin, 2019).

3.3.1.2. Spline interpolasyon yöntemi

Spline interpolasyon yöntemi, eksik veri doldurma süreçlerinde doğrusal interpolasyona kıyasla daha esnek ve yüksek doğruluk sağlayabilen bir yaklaşımdır. Bu yöntemde eksik değerler, yalnızca komşu iki nokta arasındaki düz bir doğru ile değil, birden fazla veri noktasını kapsayan daha karmaşık eğriler yardımıyla tahmin edilir. Bu eğriler genellikle parçalı polinom fonksiyonları şeklinde tanımlanır ve her bir parça kendi tanım aralığında veri setine sürekli ve türevlenebilir şekilde uyum sağlar (De Boor, 2001).

Matematiksel olarak spline interpolasyon, eksik bir değer x_t zamanında tahmin edilmesini sağlayacak şekilde, veri noktalarının belirli düğüm noktalarına (knots) karşılık gelen üçüncü dereceden polinom eğrilerle temsil edilmesiyle gerçekleştirilir. En yaygın kullanılan tür olan kübik spline interpolasyonu, her bir veri aralığı için Denklem (3.2)'deki gibi üçüncü dereceden bir polinom tanımlar.

$$S_i(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3 \quad (3.2)$$

Bu fonksiyonlar parçalar arasındaki süreklilik (C^0), birinci türev sürekliliği (C^1) ve ikinci türev sürekliliği (C^2) koşulları sağlanacak şekilde belirlenir. Bu koşullar sayesinde spline interpolasyonu, özellikle veri setinin yumuşak geçişler içerdiği, trend ve eğilim barındıran zaman serisi yapılarında oldukça başarılı sonuçlar verebilir. Literatürde spline yöntemlerinin, eksik veri problemlerinde hem düşük MAE hem de düşük varyans üretme potansiyeli nedeniyle tercih edildiği belirtilmektedir (Hastie vd., 2009).

3.3.1.3. Sabit değer ataması yöntemi

Sabit değer ataması, eksik veri doldurma süreçlerinde uygulanan en basit fakat etkili olabilen yöntemlerden biridir. Bu yöntemde eksik gözlemler, önceden belirlenen tek bir sabit değer atanması yoluyla tamamlanır. Söz konusu sabit değer, genellikle gözlemlerin aritmetik ortalaması, medyanı veya modu gibi merkezi eğilim ölçüleri olabilir. Alternatif olarak, belirli bir bağlamda anlamlı görülen başka bir sabit değer (örneğin sıfır, geçmiş dönem değeri, sektörel eşik vs.) de atanabilir (Allison, 2002).

Matematiksel olarak, sabit değeri atamasıyla eksik gözlem x_t , veri kümesinin bilinen kısmından elde edilen $\bar{x}$ gibi sabit bir istatistiksel değere eşitlenir. Bu durum Denklem (3.3)'teki gibi ifade edilebilir.

$$x_t = c, \quad c \in \{\text{ortalama}(X), \text{medyan}(X), \text{mod}(X)\} \quad (3.3)$$

Sabit değeri ataması, eksik veri oranının düşük olduğu ve dağılım yapısının istikrarlı olduğu durumlarda kullanılabilir. Özellikle zaman serilerinde verinin durağan kabul edilebildiği veya yapay veri üretmenin yanıltıcı olabileceği durumlarda, bu tür sabit atamalar analizlerin güvenliğini korumak adına tercih edilebilmektedir ancak literatürde sabit değeri atamasının varyansı azaltma ve korelasyon yapısını bozma gibi potansiyel dezavantajlarına da dikkat çekilmektedir (Little ve Rubin, 2019).

3.3.1.4. Parazit eklenmiş ortalama yöntemi

Parazit eklenmiş ortalama yöntemi, eksik veri doldurma süreçlerinde sabit değeri atama yaklaşımının bir türeği olarak değerlendirilmekte olup, istatistiksel anlamda veri setinin varyans yapısını koruma amacı taşımaktadır. Geleneksel ortalama ile doldurma yönteminde tüm eksik değerlere aynı sabit sayı atanırken, bu yaklaşım verideki doğal varyasyonu göz ardı edebilir. Parazit eklenmiş ortalama yönteminde ise, her eksik gözlem yalnızca ortalama ile değil, aynı zamanda ortalama etrafında küçük rastgele sapmalar (parazitler) içeren değerlerle doldurulmaktadır (Rubin, 1987). Bu yöntem matematiksel olarak Denklem (3.4)'teki gibi ifade edilebilir.

$$x_t = \bar{x} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_{\text{parazit}}^2) \quad (3.4)$$

Denklem (3.4)'te x_t , eksik gözlemi; $\bar{x}$, değişkenin bilinen gözlemlerinden hesaplanan aritmetik ortalamayı; ε_t ise sıfır ortalamalı ve genellikle küçük bir varyansa sahip normal dağılımlı paraziti temsil eder. Bu sayede eksik değerler doldurulurken veri setine rastlantısallık kazandırılır ve doldurulan değerler sabit yapıda değil, değişken yapıda olur.

Parazit büyüklüğü dikkatli bir şekilde belirlenmelidir. Bu çalışma kapsamında parazit değerleri, ilgili özniteliklerin standart sapmasına oranla küçük kalacak şekilde yapılandırılmış; böylece aşırı uç değerler oluşturulmadan varyans korunmuştur. Bu yöntem özellikle eksik veri oranı görece düşük olan, ancak sabit doldurma işleminin varyans kaybı yaratabileceği özniteliklerde uygulanmıştır.

3.3.1.5. Parçalı orantılı büyüme tahmini yöntemi

Parçalı orantılı büyüme tahmini, özellikle nüfus, istihdam ve üretim gibi zaman içerisinde kümülatif olarak artan ve yapısal olarak monoton davranış sergileyen makro göstergelerde eksik veri tamamlamada kullanılan bir trend tabanlı yaklaşımdır. Bu yöntemde, eksik olan dönemler, bir önceki ve bir sonraki döneme ait bilinen yıllık toplamlar arasında varsayılan sabit bir büyüme oranı ile kademeli biçimde dağıtılır. Her bir eksik gözlem, bir önceki değere orantılı artış eklenerek tahmin edilir. Bu durum, sabit artış oranına dayalı Denklem (3.5)'teki genel ifade ile tanımlanabilir.

$$x_t = x_{(t-1)} * (1 + r), \quad r: \text{sabit artış oranı} \quad (3.5)$$

Denklem (3.5)'te $x_{(t-1)}$ bir önceki dönemin değeri, x_t ise tahmin edilen dönem değeridir. Artış oranı r , genellikle yıllık toplam farkın eksik dönem sayısına bölünmesiyle elde edilen sabit yüzdesel artıştır. Bu yapı literatürde “orantılı büyüme oranı tahmini” ya da “zamansal ayrıştırma” teknikleriyle ilişkilendirilmekte olup, özellikle ulusal ve uluslararası istatistiksel kurumların (örn. Eurostat, IMF) yıllık verileri alt dönem düzeylerine indirmek amacıyla kullandığı yöntemlerle benzerlik göstermektedir (Chow ve Lin, 1971; Denton, 1971).

Bu çalışmada söz konusu yöntem, Türkiye nüfus verisinde eksik gözlemlerle karşılaşılan sınırlı dönemlerde uygulanmış; nüfus serisinin yapısal monotonluğunu koruyarak eksiksiz hale getirilmesi sağlanmıştır.

3.3.1.6. Yıl içi sabit değer tahmini yöntemi

Yıl içi sabit değer tahmini, özellikle eksik aylık gözlemlerin bulunduğu ancak aynı yıla ait yıllık ortalama veya yıllık toplam değer bilindiği durumlarda tercih edilen basit ama etkili bir doldurma yöntemidir. Bu yaklaşımda, eksik olan tüm aylık değerler, ilgili yılın sabit bir değeriyle (genellikle yıllık ortalama veya yıllık toplamın 12'ye bölünmesiyle elde edilen aylık değer) doldurularak serinin sürekliliği korunur. Bu yöntem, özellikle yıl içindeki değerlerin önemli sapmalar göstermediği ve değişkenliğin düşük olduğu serilerde uygun sonuçlar verir. Yöntem matematiksel olarak Denklem (3.6)'daki gibi ifade edilebilir.

$$x_t = \frac{X_y}{n} \quad (3.6)$$

Denklem (3.6)'da x_t eksik olan her bir aylık değeri, X_y ilgili yılın toplam değerini ve n genellikle 12 olan ay sayısını ifade etmektedir. Eğer yıllık ortalama doğrudan biliniyorsa, tüm eksik aylar bu ortalama değere eşitlenebilir. Literatürde bu yaklaşım “yıllık ortalamaya dayalı sabit tahmin” ya da “yıl içi tekdüze dağılım” adıyla tanımlanmakta; özellikle yıllık düzeyde planlanan ekonomik göstergelerin alt dönemlerine dağıtılmasında kullanılmaktadır (Bloem vd., 2001).

Bu çalışmada yıl içi sabit değer tahmini yöntemi, bazı makroekonomik ve demografik özniteliklerde yıllık veri bulunmasına karşın eksik aylık verilerin olması durumunda uygulanmıştır. Böylece serideki süreksizlik giderilmiş ve modelleme için istikrarlı veri elde edilmiştir.

3.3.1.7. Büyüme/küçülme katsayısı ile zincirleme tahmin yöntemi

Büyüme/küçülme katsayısı ile zincirleme tahmin yöntemi, özellikle bir değişkenin zaman içerisindeki görece değişim oranlarının bilindiği durumlarda eksik verilerin ardışık biçimde tamamlanmasına imkân tanıyan bir yöntemdir. Bu yaklaşımda, eksik değerler, bir önceki bilinen ya da tahmin edilen değer, büyüme ya da küçülme katsayısı ile çarpılmasıyla elde edilir. Böylece her bir değer zincirleme şekilde bir öncekine bağlı olarak

oluşturulur ve bu yapı sayesinde serinin yapısal sürekliliği korunur. Yöntem matematiksel olarak Denklem (3.7)'deki gibi ifade edilebilir.

$$x_t = x_{(t-1)} * (1 + g_t) \quad (3.7)$$

Denklem (3.7)'de x_t , t . Döneme ait tahmini değeri; $x_{(t-1)}$ ise bir önceki döneme ait bilinen veya tahmin edilen değeri ifade eder. g_t , büyüme/küçülme katsayısını temsil eder ve bu katsayı, verinin doğal eğilimine, mevsimselliğe ya da bilinen geçmiş eğilimlere göre belirlenebilir. Katsayının sabit olması durumunda bu yöntem geometrik diziye dönüşürken, dönemsel olarak değişmesi durumunda dinamik bir zincirleme modelleme elde edilir.

Bu yöntem literatürde “zincirleme ekstrapolasyon” ya da “yinelemeli orantılı büyüme tahmini” olarak tanımlanmakta ve özellikle ekonomik zaman serilerinde veri sürekliliğini sağlamak amacıyla kullanılmaktadır (Eurostat, 2017; Lequiller ve Blades, 2014). Zincirleme yapısı nedeniyle, serinin hem kısa vadeli hem de uzun vadeli eğilimlerinin modellenmesine olanak sağlar.

Bu çalışmada ilgili yöntem, belirli makro değişkenlerin bazı eksik dönemlerinde, geçmiş verilerden türetilen büyüme/küçülme katsayılarıyla eksik aylık verilerin tahmini amacıyla uygulanmıştır. Böylece serinin tutarlılığı korunarak tahminleme modelleri için daha sağlıklı bir veri seti elde edilmiştir.

3.3.1.8. İlimli büyüme oranı destekli trend analizi yöntemi

Bu yöntemde, geçmiş dönemlerdeki gözlemlerden elde edilen genel eğilim doğrultusunda bir trend analizi gerçekleştirilmekte ve elde edilen eğilimin ileriye taşınmasıyla eksik veriler tahmin edilmektedir. Ancak tahmin sürecinde, trend doğrultusuna ılımlı (küçük) bir büyüme oranı eklenerek serinin doğal gelişimi muhafaza edilmekte ve aşırı sapmalardan kaçınılmaktadır. Bu yaklaşım, kısa dönemli ileri tahminlerde ve hedef değişkenin doğal değişim yapısının korunmasının istendiği durumlarda tercih edilmektedir. Bu yöntem matematiksel olarak Denklem (3.8)'deki gibi ifade edilebilir.

$$\hat{y}_t = y_{(t-1)} + \Delta\bar{y} + \varepsilon \quad (3.8)$$

Denklem (3.8)'de $\hat{y}_t$, eksik döneme ilişkin tahmini değeri; $y_{(t-1)}$, bir önceki döneme ait değeri; $\Delta\bar{y}$, geçmiş dönem ortalama değişim miktarını; ε ise küçük ve kontrollü artış katsayısını temsil etmektedir. Bu yapı, serinin genel eğilimine (trendine) sadık kalarak ileri yönlü tahmin yapmayı mümkün kılar. Eğer geçmiş yılların aynı aylarına ilişkin değişim oranları mevcutsa, bu oranlar referans alınarak geleceğe yönelik küçük artışlar kurgulanabilir.

Literatürde bu yöntem, kısa dönemli ekonomik tahminlerde ve işletme planlaması uygulamalarında önerilmektedir (Makridakis vd., 1997; Hyndman ve Athanasopoulos, 2018). Bu yöntemin en temel avantajı, geçmiş verinin trendini bozmadan, öngörülen küçük değişimlerle seriyi tamamlamasıdır.

Bu çalışmada söz konusu yöntem, hedef değişken olan *çimento satın alma fiyatı* değişkeninin veri setinin ham hâlinde 2024 yılının son üç ayında eksik kalan gözlemlerini tamamlamak üzere uygulanmıştır. Geçmiş yıllara ait benzer aylardaki eğilim göz önüne alınarak, mevcut verinin eğrisel yapısı bozulmadan küçük artış oranlarıyla seriye süreklilik kazandırılmıştır.

3.3.1.9. Pro-Rata (temsili ay oranlama) yöntemi

Bu yaklaşım, yıllık toplam değer bilindiği ancak yıl içindeki bazı ayların verisinin eksik olduğu durumlarda kullanılan, geçmiş dönemlere ait aylık dağılım yüzdeleri üzerinden tahminleme yapan bir orantı yöntemidir. “Pro Rata” esasına dayanan bu yöntem, yıllık toplamın geçmişteki aylık dağılım örüntüsüne göre ay bazlı orantılanarak eksik verilerin türetilmesini mümkün kılar. Bu özellikleriyle özellikle bütçe planlaması, kaynak tahsisi ve üretim hacmi gibi zamanla parçalanmış büyüklüklerin geri tahminlenmesinde yaygın olarak kullanılmaktadır. Yöntemin matematiksel ifadesi Denklem (3.9)'daki gibi gösterilebilir.

$$\hat{x}_t = X_{toplam} * \frac{r_t}{\sum r_t} \quad (3.9)$$

Denklem (3.9)'da;

- $\hat{x}_t$: Tahmin edilecek ay t 'ye ait değer
- $X_{toplaml}$: Bilinen yıllık toplam değer
- r_t : Geçmiş yıllar ortalaması olarak t . ayın yıllık toplam içindeki oranı (örnek vermek gerekirse Ocak ayı geçmişte ortalama %6 kadar katkı sağlıyorsa $r_{Ocak} = 0,06$)
- $\sum r_t$: Eksik ayların geçmişten gelen toplam katkı oranlarıdır (örneğin Ocak-Mart aylarının verileri eksikse bu üç ayın oranları toplanır).

Bu yöntem sayesinde, geçmiş yıllardaki benzer dönemlerin ay ay yapısı korunarak eksik ayların temsili değerleri üretilebilmekte ve yıllık bütünlük bozulmadan tahminleme gerçekleştirilebilmektedir.

Literatürde bu yöntem “orantılı zamansal dağıtım” ya da “mevsimsel orantılı dağıtım yöntemi” olarak adlandırılmakta olup, Eurostat, IMF ve çeşitli ulusal istatistik enstitüleri tarafından da önerilen resmi bir yöntemdir (Eurostat, 2018; Bloem vd., 2001). Özellikle aylık veri setlerinin tamamlanmasında analitik tutarlılık ve operasyonel kolaylık sağlamaktadır.

Bu çalışmada söz konusu yöntem, yalnızca yıl bazında toplam değeri bilinen ancak bazı aylarında eksiklikler bulunan değişkenlerin ay düzeyinde tamamlanması amacıyla uygulanmıştır. Böylece hem yıllık değerle tutarlı hem de geçmiş yılların ay temelinde dağılımını temsil eden bir yapı elde edilmiştir.

3.3.1.10. STL yöntemi

STL decomposition tabanlı dağılım yöntemi, zaman serilerindeki yapısal örüntüleri ayrı ayrı analiz ederek eksik verilerin daha doğru biçimde tahmin edilmesini sağlayan gelişmiş bir ayrıştırma tekniğidir. STL, seriyi üç bileşene ayırır: trend bileşeni (T_t), mevsimsel bileşen (S_t) ve artık bileşen (R_t). Bu yapı Denklem (3.10)'daki gibi ifade edilir.

$$X_t = T_t + S_t + R_t \quad (3.10)$$

Denklem (3.10)'da;

- X_t : t . döneme ait gözlemlenen veya tahmin edilmek istenen değer
- T_t : Uzun dönemli yapısal eğilimi temsil eden trend bileşeni
- S_t : Mevsimsel döngüyü temsil eden periyodik bileşen
- R_t : Modelde açıklanamayan, rastgele değişkenliği temsil eden hata (residual) bileşenidir.

STL yönteminin en önemli özelliği, mevsimsel bileşeni LOESS yaklaşımı ile esnek bir biçimde öğrenebilmesi ve hem sabit hem değişken mevsimselliğe sahip serilere uygulanabilir olmasıdır. Eksik değerlerin tahminlenmesinde genellikle, eldeki gözlemler yardımıyla trend ve mevsimsel bileşenler yeniden tahmin edilmekte ve bu bileşenlerin toplamı eksik döneme ait değeri oluşturmaktadır.

Eksik bir gözlem X_τ 'ya ilişkin tahmin Denklem (3.11)'deki gibi yapılır; genellikle artık bileşen sıfır olarak kabul edilir veya geçmişteki benzer dönem ortalamalarıyla belirlenebilir.

$$\hat{X}_\tau = \hat{T}_\tau + \hat{S}_\tau \quad (3.11)$$

STL tabanlı yaklaşımlar, özellikle güçlü mevsimsellik gösteren ekonomik ve sektörel verilerde, klasik interpolasyon yöntemlerinden çok daha başarılı tahminler sağlamaktadır (Cleveland vd., 1990). Ayrıca eksik veri yalnızca bir değer değil, ardışık birkaç dönem için söz konusu olduğunda da veri yapısına zarar vermeden doldurma imkânı sunar.

Bu çalışmada STL tabanlı dağılım yöntemi, özellikle yıl içi güçlü mevsimsel örüntüye sahip değişkenlerde tercih edilmiş; geçmiş dönem verilerinden çıkarılan mevsim

ve trend bileşenleri yardımıyla eksik gözlemler tahmin edilmiştir. Böylece hem doğal eğilim korunmuş hem de mevsimsellik etkisi modele dâhil edilmiştir.

3.3.1.11. Benchmarking yöntemi

Benchmarking yöntemi, eksik veri içeren bir serinin, eksiksiz veya daha güvenilir kabul edilen bir referans seriye (benchmark seriye) bağlanarak tahmin edilmesi esasına dayanan dolaylı bir doldurma tekniğidir. Özellikle doğrudan ölçümü mümkün olmayan ya da eksikliği bulunan ekonomik göstergelerde, yapısal benzerlik veya tarihsel korelasyon ilişkisi güçlü olan başka bir seriden faydalanılarak tahmin yapılması hem doğruluk hem de model tutarlılığı açısından tercih edilen bir yaklaşımdır.

Bu yöntemde, referans serinin ilgili döneme ait gözlemleri ile eksik gözleme sahip hedef serinin geçmişteki ilişki düzeyi kullanılarak eksik değer tahmin edilir. Temel varsayım, referans ve hedef seriler arasında doğrusal orantı veya sabit oran ilişkisi olduğudur. Bu yöntem matematiksel olarak Denklem (3.12)'deki gibi ifade edilebilir.

$$\hat{X}_t^{(hedef)} = \frac{X_{r,t}}{\bar{X}_r} * \bar{X}_h \quad (3.12)$$

Denklem (3.12)'de;

- $\hat{X}_t^{(hedef)}$: Hedef serideki t . döneme ait eksik değer tahmini
- $X_{r,t}$: Referans serinin t . dönemdeki değeri
- $\bar{X}_r$: Referans serinin geçmişteki ortalama değeri (örneğin son 12 ay)
- $\bar{X}_h$: Hedef serinin geçmişteki ortalama değeri (aynı dönem) olarak açıklanabilir.

Bazı uygulamalarda doğrudan sabit oran (θ) belirlenerek de tahmin yapılabilir. Sabit oran belirlenmesi hâlinde Denklem (3.12) sadeleşerek Denklem (3.13)'e dönüşecektir.

$$\hat{X}_t^{(hedef)} = \theta * X_{r,t} \quad (3.13)$$

Bu yöntem literatürde “orantılı kıyaslama”, “gösterge tabanlı tahmin” veya “ilgili seriler aracılığıyla dolaylı tahmin” gibi terimlerle de yer almaktadır (Bloem vd., 2001; Eurostat, 2013). Özellikle makroekonomik göstergelerde, sektörel endekslerde veya üretim-tüketim tahminlerinde yaygın olarak kullanılır.

Bu çalışmada benchmarking yöntemi, eksik veri barındıran bazı özniteliklerde doğrudan tahminin zor olduğu durumlarda uygulanmış; benzer davranış örüntüsü gösteren başka bir seriden elde edilen değerler yardımıyla eksik gözlemler dolaylı olarak tamamlanmıştır. Böylece istatistiksel tutarlılık korunmuş ve bağlamsal anlamlılık sağlanmıştır.

Bu bölümde genel hatlarıyla açıklanan eksik veri doldurma yöntemlerine ilişkin uygulama tercihleri, veri setinde karşılaşılan eksikliklerin niteliği ve çalışmanın ihtiyaçlarına göre hangi yöntemin daha uygun olduğu ise izleyen uygulama adımlarında detaylı şekilde ele alınmıştır.

3.3.2. Öznitelik seçim kriterleri ve araçları

Makine öğrenmesi uygulamalarında modelin başarısı, yalnızca kullanılan algoritmalarından değil, aynı zamanda bu algoritmalara sunulan veri yapısından da önemli ölçüde etkilenmektedir. Özellikle yüksek boyutlu veri setlerinde, gereksiz veya çoklu doğrusal bağlantı içeren özniteliklerin varlığı; modelin genellenebilirliğini düşürmekte, öğrenme sürecini karmaşıktırmakta ve yorumlanabilirliği azaltmaktadır (Guyon ve Elisseeff, 2003; Chandrashekar ve Sahin, 2014). Bu nedenle öznitelik seçim süreci, yalnızca boyut indirgeme amacıyla değil; aynı zamanda modelin öngörü gücünü artırmak ve veri temsiliyetini güçlendirmek amacıyla da kritik bir adımdır.

Bu çalışmada öznitelik seçimi süreci çok boyutlu bir perspektifle ele alınmış ve üç temel yaklaşıma dayalı olarak yapılandırılmıştır. İlk olarak, sektörel bilgiye dayalı uzman görüşleri doğrultusunda önemli olduğu değerlendirilen değişkenler belirlenmiştir. Bu yaklaşım, literatürde “uzman görüşüne dayalı öznitelik seçimi” olarak tanımlanmakta ve

saha deneyimini modelleme sürecine entegre etme fırsatı sunmaktadır (Liu ve Motoda, 2007). İkinci olarak, istatistiksel yöntemler kullanılarak korelasyon analizi hesaplamaları gerçekleştirilmiş; hedef değişkenle anlamlı ilişkisi olmayan ya da düşük seviyede ilişkisi olan öznitelikler elenmiştir. Son olarak, her iki yöntemin çıktıları bir araya getirilerek hibrit bir öznitelik kümesi oluşturulmuş, ardından bu küme üzerinden ileri analizler yapılmıştır.

Süreç boyunca yalnızca özniteliklerin hedef değişken üzerindeki etkileri değil, aynı zamanda öznitelikler arası ilişkiler de dikkate alınmış; gereksiz tekrar veya örtüşmelerin önüne geçilerek sadeleştirilmiş ama temsil gücü yüksek bir veri yapısı elde edilmiştir. Aşağıda, bu seçim süreçlerinde uygulanan yöntemler ve araçlar teorik çerçeveleriyle birlikte sunulacaktır.

3.3.2.1. Uzman görüşüne dayalı öznitelik seçimi

Veri madenciliği ve makine öğrenmesi tabanlı tahminleme modellerinde öznitelik seçim süreci, modelin hem genel başarısını hem de yorumlanabilirliğini doğrudan etkileyen kritik bir adımdır. Özellikle açıklayıcı değişken sayısının yüksek olduğu ve veri setinin sınırlı sayıda gözlem içerdiği durumlarda, anlamlı ve etkili özniteliklerin belirlenmesi, modelin aşırı öğrenme veya çoklu doğrusal bağlantı gibi risklerle karşılaşmasını engellemek açısından oldukça önemlidir. Bu bağlamda, öznitelik seçiminde sadece istatistiksel yaklaşımların değil, sektör tecrübesine dayalı uzman görüşlerinin de dâhil edilmesi, tahmin modellerinin hem güvenilirliğini hem de karar destek kapasitesini artırmaktadır.

Literatürde uzman görüşü temelli öznitelik seçimi, veri setinde yer alan değişkenler arasından, saha tecrübesine ve sektörel bilgiye dayalı olarak anlamlı görülenlerin belirlenmesi süreci olarak tanımlanmakta ve özellikle yüksek düzeyde alan bilgisi gerektiren disiplinlerde sıklıkla tercih edilmektedir (Liu ve Motoda, 2007; Saeys vd., 2007). Bu yöntem, örüntülerin yalnızca sayısal verilerle değil, gerçek yaşam pratikleri ve sektörel gözlemlerle de şekillendiği alanlarda oldukça etkin sonuçlar doğurabilmektedir.

Yapı kimyasalı üretim süreçleri, doğrudan inşaat sektörüne bağlılık, enerji ve nakliye maliyetlerindeki dalgalanmalar, hammaddeye erişim durumu, sezonluk talep değişimleri ve politik gelişmeler gibi birçok sektörel ya da dolaylı olarak sektörel olmayan faktörden etkilenmektedir. Bu faktörlerin büyük bir bölümü, istatistiksel korelasyon veya varyans

analizleri ile her zaman güçlü ilişkiler sergileyemeyebilir. Bu gibi durumlarda sektörde uzun yıllar görev yapmış uzmanların deneyimleri sayesinde, bazı değişkenlerin dolaylı etkileri, zamansal dinamikleri veya piyasa duyarlılıkları gibi boyutlar daha iyi analiz edilebilmektedir. Bu nedenle, bu çalışmada öznitelik seçim süreci yalnızca istatistiksel ilişkilerle değil, saha uzmanlarının bilgi birikimi ile şekillendirilmiştir.

Bu kapsamda yapı kimyasalı üreticisinin satın alma, planlama ve üretim departmanlarında görevli uzman personel ile yapılan yapılandırılmış görüşmeler sonucunda; sektörel dalgalanmalara duyarlı olduğu bilinen, piyasa takibiyle ilişkilendirilen ve operasyonel karar alma süreçlerinde sıklıkla kullanılan bazı değişkenlerin modelleme sürecine dâhil edilmesine karar verilmiştir. Örneğin, “Herhangi bir seçim var mı?” değişkeni, politik istikrarın ve piyasa belirsizliğinin tedarik zinciri üzerindeki etkisini yansıtabilecek bir gösterge olarak önerilmiştir. Benzer şekilde, motorin fiyatları, kalker ve alçıtaşı gibi hammadde satış fiyatları veya yapı ruhsatı verileri gibi sektöre özgü değişkenlerin modele katkı sağlayabileceği öngörülmüştür.

Söz konusu uzman görüşlerinin modelleme sürecine dâhil edilmesi yalnızca pratik bilgi aktarımıyla sınırlı kalmamış, aynı zamanda çalışmanın sektörle kurumsal bağı güçlendirmiş, model çıktılarının saha gerçekliğiyle uyumlu hâle gelmesini sağlamıştır. Bu durum, modelin yalnızca teknik başarısına değil, aynı zamanda karar destek süreçlerine entegre edilebilirliğine de katkı sunmuştur çünkü akademik doğruluk kadar, saha uygulamasında anlamlılık da bir tahmin modelinin başarısında temel unsurlardan biridir.

Ayrıca bu yaklaşımın metodolojik açıdan şeffaf ve tekrarlanabilir olmasına özen gösterilmiştir. Yapılan uzman görüşmeleri kayıt altına alınmış, hangi özniteliklerin neden önerildiği ve nasıl değerlendirildiği belgelenmiştir. Böylece öznitelik seçim süreci yalnızca sezgisel değil, belgelenmiş bir uzmanlık katkısı olarak ele alınmış ve çalışmanın bilimsel izlenebilirliği güçlendirilmiştir (Guyon ve Elisseeff, 2003).

Bu doğrultuda geliştirilen Çalışma 1, yalnızca uzman görüşü ile belirlenen özniteliklerin dâhil edildiği bir tahmin modeli olarak yapılandırılmıştır. Bu model ile sektörde aktif rol alan paydaşların bilgi birikimi, veri temelli karar destek mekanizmasına

entegre edilmiş ve elde edilen sonuçlar, sektörel öngörülerin doğrulanması açısından kıymetli çıktılar sağlamıştır.

Özetle, uzman görüşü temelli öznitelik seçimi bu tez çalışmasında yalnızca bir öznitelik azaltma tekniği değil; aynı zamanda sektörle kurulan anlam köprüsünün bir aracı olarak değerlendirilmiştir. Akademik yaklaşımlar ile sektörel sezgilerin birleştiği bu modelleme çerçevesi, ilerleyen aşamalarda geliştirilecek tahmin sistemleri için sürdürülebilir ve esnek bir temel sunmaktadır.

3.3.2.2. Korelasyon analizi

Korelasyon analizi, özellikle regresyon tabanlı tahminleme ve makine öğrenmesi modellerinde, öznitelikler ile hedef değişken arasındaki doğrusal ilişkinin yönünü ve şiddetini belirlemek amacıyla sıklıkla başvurulmuş temel istatistiksel araçlardan biridir. Bu analiz yöntemi, modelin açıklayıcı gücünü artıracak özniteliklerin belirlenmesinde ilk adım olarak değerlendirilmekte olup, yüksek boyutlu veri setlerinde öznitelik indirgeme stratejisinin de vazgeçilmez bir parçasını oluşturmaktadır. Özellikle çok sayıda bağımsız değişken içeren veri setlerinde, bu değişkenlerin tahmin sürecine olan katkısının değerlendirilmesi ve gereksiz özniteliklerin elenmesi açısından korelasyon analizi oldukça etkili bir ön eleme yöntemi olarak kullanılmaktadır (Dormann vd., 2013).

Korelasyon analizi kapsamında en yaygın kullanılan ölçüt, Pearson korelasyon katsayısıdır. Bu katsayı, iki sürekli değişken arasındaki doğrusal ilişkinin gücünü -1 ile $+1$ arasında bir değerle ifade eder. Pearson korelasyon katsayısı r , Denklem (3.14) ile hesaplanabilir.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.14)$$

Burada x_i ve y_i sırasıyla X ve Y değişkenlerinin i . gözlemine karşılık gelen değerleridir. $\bar{x}$ ve $\bar{y}$, ilgili değişkenlerin aritmetik ortalamalarını ifade eder. n , toplam gözlem sayısını belirtir.

Elde edilen r katsayısının değeri 1'e yaklaştıkça değişkenler arasında pozitif yönde güçlü bir doğrusal ilişki olduğu, -1 'e yaklaştıkça negatif yönde güçlü bir doğrusal ilişki bulunduğu, sifıra yakın olması durumunda ise doğrusal ilişkinin zayıf ya da yok denecek düzeyde olduğu yorumlanmaktadır (Mukaka, 2012). Ancak korelasyon analizi yalnızca hedef değişkenle olan ilişkilerin değil, aynı zamanda bağımsız değişkenlerin kendi aralarındaki ilişkilerin de analiz edilmesine olanak tanır. Bu olanak, özellikle çoklu doğrusal bağlantı gibi modelleme açısından sorun meydana getirebilecek yapıları erken aşamada fark edebilmek adına oldukça önemlidir. Zira birbirleriyle yüksek oranda ilişkili olan değişkenlerin birlikte modele dâhil edilmesi durumunda modelin kararlılığı düşebilir, tahmin gücü zayıflayabilir ve özniteliklerin bireysel etkileri anlamını yitirebilir (Graham, 2003).

Bu çerçevede, korelasyon analizinin tahminleme çalışmalarındaki rolü hedef değişkenle anlamlı ilişki gösteren değişkenlerin belirlenmesi ve açıklayıcı değişkenler arasında yüksek korelasyon gösteren çiftlerin tespitiyle çoklu bağlantı riskinin azaltılması olmak üzere iki yönlüdür. Yine de korelasyon katsayısının yalnızca doğrusal ilişkileri ölçtüğü, doğrusal olmayan fakat güçlü ilişkilere duyarsız kalabileceği unutulmamalıdır. Bu nedenle korelasyon analizi, genellikle daha derin istatistiksel analizlerin öncülü olarak değerlendirilir ve karar verme süreçlerinde tek başına yeterli bir kriter olarak ele alınmaz. Bununla birlikte, doğru uygulandığında öznitelik seçimi sürecine yön verici ve sadeleştirici katkısı tartışmasıdır.

Bu çalışmada da korelasyon analizine, veri setindeki 30 özniteliğin hedef değişken olan “Çimento Satın Alım Fiyatı” ile olan ilişkisini anlamak amacıyla başvurulmuştur. Korelasyon analizinin güçlü yönlerinden biri de veri ön işleme süreçleriyle bütünleşmiş çalışabilmesidir. Özellikle eksik veri doldurma işlemlerinden sonra yapılan korelasyon analizi, eksik doldurma yöntemlerinin veri yapısına olan etkisinin de kontrol edilmesine imkân tanımaktadır. Böylece yapay etkileşimlerin önüne geçilmiş ve öznitelik seçim süreci daha güvenilir hâle getirilmiştir.

Sonuç olarak korelasyon analizi, bu tez kapsamında hem hedef değişkenle anlamlı ilişkiler kurabilecek özniteliklerin seçilmesinde hem de olası çoklu bağlantıların önüne geçilmesinde yol gösterici bir rol üstlenmiştir.

3.3.3. Çalışma kapsamında kullanılacak makine öğrenmesi modellerinin seçimi ve modellere ilişkin detaylar

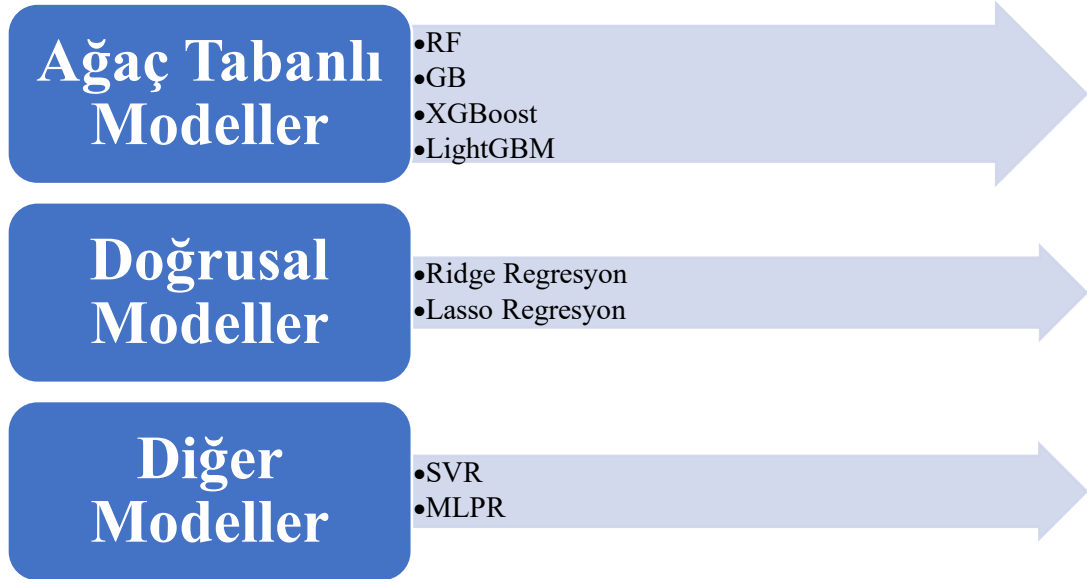
Bu çalışmada, yapı kimyasalı üreticisinin çimento satın alım fiyatlarını tahmin etmek amacıyla hem doğrusal hem doğrusal olmayan ilişkileri modelleyebilecek kapasitede çeşitli makine öğrenmesi algoritmaları uygulanmıştır. Model seçiminde, veri setinin yapısal özellikleri (örneğin gözlem sayısının sınırlı olması, öznitelik sayısının yüksekliği ve bazı değişkenler arasındaki olası korelasyonlar) ile hedef değişkenin sürekli nicel bir yapı taşıması gibi faktörler belirleyici olmuştur. Bu doğrultuda, modeller Şekil 3.2’de görülebileceği üzere üç ana kategoride ele alınmıştır: Ağaç Tabanlı Modeller, Doğrusal Modeller ve Diğer Modeller.

İlk olarak, ağaç tabanlı modeller kapsamında yer alan RF, GB, XGBoost ve LightGBM algoritmaları, veri içindeki karmaşık ve doğrusal olmayan desenleri öğrenme yetenekleri ve genelleme kabiliyetleri nedeniyle tercih edilmiştir (Pedregosa vd., 2011). Bu modellerin öne çıkan özellikleri, yüksek esneklik, görece düşük varsayım ihtiyacı ve güçlü performans sergileme potansiyelidir. Aynı zamanda bu algoritmaların topluluk (ensemble) öğrenmesine dayalı yapıları, önyargı-varyans dengesini iyileştirmeye olanak tanıyarak model doğruluğunu artırmaktadır.

İkinci olarak, doğrusal modeller başlığı altında Ridge ve Lasso regresyon yöntemleri değerlendirilmiştir. Bu algoritmalar, özellikle öznitelik sayısının fazla olduğu ve çoklu doğrusal bağlantı riskinin bulunduğu veri setlerinde hem aşırı öğrenmeyi engellemek hem de daha sade modeller üretmek amacıyla kullanılmaktadır. Ridge regresyon, parametre küçültme (shrinkage) yoluyla, Lasso ise değişken seçimi yoluyla model karmaşıklığını kontrol altına almakta ve daha açıklanabilir tahmin modelleri oluşturulmasına imkân tanımaktadır (Pedregosa vd., 2011).

Üçüncü grupta yer alan diğer modeller, doğrusal olmayan modelleme kapasitesi sunan alternatif yaklaşımları içermektedir. Özellikle SVR, yüksek boyutlu verilerde margin optimizasyonuna dayanan güçlü genelleme kabiliyetiyle öne çıkar; bu sayede sınırlı veri setlerinde bile etkinlik sağlar (Roozbeh vd., 2023). Buna karşılık MLPR, katmanlı yapısı sayesinde doğrusal olmayan desenleri öğrenmekte oldukça etkilidir (Zhou vd., 2024). Bu

modeller, karmaşık ilişkileri yakalama becerileri nedeniyle çimento fiyat tahmini gibi uygulamalarda sıkça tercih edilmektedir.



Şekil 3.2. Kullanılan Makine Öğrenmesi Algoritmalarının Sınıflandırılması

Her bir modelin uygulanması öncesinde varsayılan hiperparametreler ile ilk analizler gerçekleştirilmiş, ardından hiperparametre eniyilemeleri yapılarak modellerin nihai performansları değerlendirilmiştir. Bu yaklaşım, yalnızca farklı model yapılarını karşılaştırmakla kalmayıp, aynı zamanda her bir algoritmanın veri setine ne ölçüde uyum sağlayabildiğini de ortaya koymayı amaçlamaktadır.

Bu bölümde sırasıyla her bir modelin teorik temelleri, tahminleme mekanizmaları, varsa matematiksel formülasyonları ve parametrik yapıları açıklanarak, söz konusu algoritmaların neden tercih edildiği ayrıntılı biçimde sunulacaktır.

3.3.3.1. Rastgele Orman - RF Algoritması

RF algoritması, regresyon ve sınıflandırma problemlerinde yüksek doğruluk ve genellenebilirlik sağlayan bir topluluk öğrenmesi yöntemidir. Leo Breiman tarafından geliştirilen bu yöntem, temel olarak çok sayıda karar ağacının oluşturulması ve bu

ağaçlardan elde edilen tahminlerin birleştirilmesi prensibine dayanır. Regresyon problemleri bağlamında, bu tahminlerin aritmetik ortalaması alınarak nihai çıktı elde edilir. Nihai çıktının elde edilmesi Denklem (3.15)'teki gibi ifade edilebilir.

$$\bar{y} = \frac{1}{T} * \sum_{t=1}^T \hat{y}^{(t)} \quad (3.15)$$

Denklem (3.15)'te $\hat{y}^{(t)}$, t . karar ağacının tahminini; T ise toplam ağaç sayısını temsil etmektedir (Breiman, 2001).

Algoritmanın başarısının ardında iki temel mekanizma yer alır, bu iki temel mekanizma önyüklemeli birleştirme (bootstrap örnekleme, bagging) ve rastgele öznitelik alt kümesi seçimidir. Önyüklemeli birleştirme sayesinde, eğitim verisinin rastgele seçilen alt kümeleriyle her bir ağaç ayrı ayrı eğitilir. Bu durum, modeller arası çeşitliliği artırır ve varyansı düşürür. Rastgele öznitelik alt kümesi seçimi ise, her bir bölme kararında, tüm öznitelikler yerine sadece rastgele seçilen m öznitelik içerisinde en iyi bölünmenin yapılmasını sağlar. Bu süreç, aşırı öğrenme eğilimi yüksek olan karar ağacı algoritmalarının genellenabilirliğini artırır (Breiman, 2001).

RF algoritmasında öğrenme süreci sırasında yapılan öznitelik seçimleri, modelin değişken önem düzeylerini de ortaya koyar. Bu önem düzeyleri genellikle, her özneliğin modele olan katkısının, hata fonksiyonundaki iyileştirme (örneğin, karışıklık derecesini ölçen gini azaltımı veya MSE düşüşü) üzerinden hesaplanması ile elde edilir. Regresyon bağlamında kullanılan hata ölçütü MSE olduğundan, her ağacın inşa sürecinde yapılan bölmelerin MSE'yi ne kadar azalttığı değerlendirilerek öznitelik katkısı belirlenir (Scornet, 2020).

RF'nin hiperparametreleri arasında ağaç sayısı ($n_estimators$), her ağacın maksimum derinliği (max_depth), her bölme işleminde göz önüne alınacak öznitelik sayısı ($max_features$) gibi bileşenler yer alır. Bu hiperparametreler, modelin karmaşıklığını ve genel başarısını doğrudan etkilemektedir (Probst vd., 2019).

Bu çalışmada, RF algoritması kullanılarak; farklı öznitelik setlerine göre oluşturulan beş ayrı analiz yapısında model performansının istikrarlılığı ve doğruluğu test edilmiştir. Algoritmanın doğası gereği, öznitelik sayısının fazla olması durumunda bile aşırı öğrenmeye karşı dirençli yapısı sayesinde yüksek güvenilirlik sunması, tahminleme süreçlerinde bu yöntemin tercih edilme nedenlerinden biri olmuştur. Ayrıca, özniteliklerin hedef değişken üzerindeki göreceli öneminin anlaşılması ve değişken seçim sürecine katkı sağlaması açısından da RF, yalnızca bir tahmin aracı değil, aynı zamanda yorumlanabilirlik sağlayan bir analiz aracıdır.

3.3.3.2. GB Algoritması

GB, tahmin hatalarını enküçüklemeyi hedefleyen, ardışık olarak kurulan zayıf öğreniciler (genellikle karar ağaçları) üzerinden çalışan bir topluluk öğrenme yöntemidir. İlk olarak Friedman (2001) tarafından sistematikleştirilen bu yaklaşımda amaç, her bir yeni modelin, önceki modelin hatalarını azaltacak şekilde eğitilmesidir. Bu yapısıyla, doğruluk açısından oldukça güçlü, fakat dikkatli uygulanmadığında aşırı öğrenmeye açık bir yöntemdir.

Bu algoritma, temel olarak aşağıdaki adımları izler.

Adım 1: İlk model $F_0(x)$, Denklem (3.16)'daki gibi genellikle hedef değişkenin ortalaması alınarak başlatılır.

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma) \quad (3.16)$$

Denklem (3.16)'da $L(y_i, \gamma)$ kayıp fonksiyonudur (regresyon problemlerinde çoğunlukla kare hata fonksiyonu tercih edilir: $L = (y - \hat{y})^2$).

Adım 2: Ardından her iterasyonda $m = 1, 2, \dots, M$ için mevcut modelin hataları hedef alınarak gradyan yönünde yeni bir model eğitilir. Yani her adımda Denklem (3.17)'deki gradyan değerlerine karşılık gelen sözde-hatalar hesaplanır.

$$r_i^{(m)} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{(m-1)}(x)} \quad (3.17)$$

Adım 3: Bu sözde hataları en iyi açıklayan karar ağacı $h_m(x)$ eğitilir ve yeni model Denklem (3.18)'deki gibi güncellenir.

$$F_m(x) = F_{(m-1)}(x) + v * h_m(x) \quad (3.18)$$

Denklem (3.18)'de v (learning rate) modelin öğrenme hızını belirleyen hiperparametredir. Daha düşük bir v , daha yavaş ancak daha güvenilir bir öğrenme süreci sağlar (Friedman, 2001).

GB algoritmasının hiperparametreleri arasında öğrenme oranı (v), ağaç sayısı ($n_estimators$), her ağacın maksimum derinliği (max_depth), enküçük örnek sayısı gibi parametreler yer alır. Bu parametrelerin dikkatli şekilde ayarlanması, algoritmanın yüksek doğruluk kapasitesini aşırı öğrenme olmadan gerçekleştirmesi açısından kritiktir (scikit-learn, 2025).

Özellikle regresyon problemlerinde hata fonksiyonunun (örneğin MSE) enküçüklenmesine dayalı olarak çalıştığı için, GB yüksek derecede hataya duyarlı ve esnek bir yapıya sahiptir. Modelin her adımda hata eğrisini takip ederek ilerlemesi, doğruluk oranlarını artırsa da aşırı model karmaşıklığına karşı dikkatli olunmasını gerektirir. Bu nedenle sıklıkla erken durdurma gibi önlemlerle desteklenir (Raskutti vd., 2014).

Bu çalışmada GB algoritması, her bir veri seti (Çalışma 1–5) üzerinde uygulanarak farklı öznitelik yapılarına nasıl tepki verdiği gözlemlenmiş ve hiperparametre eniyilemesi sonrasında modelin istikrarlı ve tutarlı tahmin performansları elde ettiği belirlenmiştir. Özellikle küçük artışlarla ilerleyerek daha hassas tahminler yapabilme kapasitesi, bu yöntemin yapı kimyasalı üreticisinin satın alma süreçlerinde stratejik karar vermelerine katkı sağlayacak bir modelleme aracı olarak kullanılmasını desteklemektedir.

3.3.3.3. XGBoost Algoritması

XGBoost, GB algoritmasının eniyilenmiş ve ölçeklenebilir bir versiyonudur. Yüksek doğruluk sunmasının yanı sıra hız, paralel işlem desteği ve aşırı öğrenmeye karşı sunduğu gelişmiş önlemlerle öne çıkmaktadır. Özellikle büyük veri setlerinde ve karmaşık yapıli tahmin problemlerinde yüksek başarı sağlaması nedeniyle endüstriyel uygulamalarda da sıklıkla tercih edilmektedir (Chen ve Guestrin, 2016).

Bu algoritmanın temel felsefesi, önceki ağaçlardan öğrenilen hataları enküçükleyecek yeni bir karar ağacını, ikinci mertebeden gradyan bilgilerini (Hessian) de dikkate alarak öğrenmektir. Modelin t . iterasyondaki çıktısı Denklem (3.19)'daki gibi güncellenir.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3.19)$$

Denklem (3.19)'da $f_t(x_i)$, yeni eklenen karar ağacıdır. XGBoost'un eniyileme hedefi Denklem (3.20)'deki gibi tanımlanır.

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \quad (3.20)$$

Denklem (3.20)'de; $l(y_i, \hat{y}_i^{(t)})$, tahmin hatasını ölçen diferansiyellenebilir kayıp fonksiyonunu; $\Omega(f_k)$ ise ağaçların karmaşıklık cezasını içeren düzenleyici fonksiyondur.

XGBoost bu fonksiyonu ikinci dereceden Taylor açılımıyla yakınsar ve ardından her yaprağın toplam gradyan (g_i) ve Hessian (h_i) değerlerini kullanarak eniyileme problemini çözer. Bu yaklaşım yalnızca modelin öğrenmesini hızlandırmakla kalmaz, aynı zamanda aşırı öğrenmeyi kontrol altına alan güçlü bir düzennileştirme etkisi sağlar. Ağaç karmaşıklığını cezalandıran düzenleyici terim Denklem (3.21)'deki gibi ifade edilir.

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (3.21)$$

Denklem (3.21)'de; T , ağacın yaprak sayısı; w_j, j . Yaprığın skoru (ağırlığı); γ ve λ ise düzenleştirme parametrelerini ifade eder (Chen ve Guestrin, 2016).

XGBoost, aynı zamanda ön budama, sütun alt örnekleme ve eksik değerlerle doğrudan baş edebilme gibi modern mekanizmalarla donatılmıştır. Bu yönüyle yalnızca güçlü değil, aynı zamanda esnek ve pratik bir modelleme aracıdır.

XGBoost modelinin performansı, çeşitli hiperparametrelerle hassas biçimde ayarlanabilir. Modelin en temel hiperparametrelerinden bazıları şunlardır:

- Ağaç Sayısı ($n_estimators$): Aşamalı öğrenme iterasyonlarının sayısını belirtir. Yüksek değerlerde aşırı öğrenmeye sebep olabilir.
- Maksimum Derinlik (max_depth): Ağaçların derinliğini kontrol eder. Düşük değerlerde eksik öğrenme, yüksek değerlerde aşırı öğrenme riski taşır.
- Öğrenme Hızı ($learning_rate$): Her ağaçtan sonra yapılan tahmin güncellemesinin büyüklüğünü kontrol eder. Genellikle küçük değerler tercih edilip ağaç sayısı ($n_estimators$) artırılır.
- $subsample$: Her ağaç için kullanılan örnekleme oranıdır. Aşırı öğrenme ihtimalini azaltmak için < 1 değeri önerilir.

Bu parametrelerin dikkatli bir şekilde ayarlanması, modelin genellenebilirliğini ve doğruluk oranını önemli ölçüde artırabilir (Verma, 2024).

3.3.3.4. LightGBM Algoritması

LightGBM, Microsoft tarafından geliştirilen ve Gradient Boosting Decision Tree (GBDT) algoritmasının eniyilenmiş, hafif ve hızlı bir versiyonudur. Bu algoritma özellikle

büyük ölçekli veri setleri ve yüksek boyutlu öznitelik alanlarıyla çalışırken işlem süresi ve bellek kullanımında kayda değer avantajlar sağlamaktadır (Ke vd., 2017). LightGBM, geleneksel aşamalı öğrenme algoritmalarına göre daha verimli bir yapıya sahiptir ve bu nedenle tahminleme problemlerinde yaygın olarak kullanılmaktadır.

LightGBM'in başarısının arkasındaki temel fark, ağaç büyütme stratejisi ve veri örnekleme yaklaşımıdır. Geleneksel aşamalı öğrenme algoritmaları seviye bazlı ağaç büyütme stratejisi kullanırken, LightGBM yaprak bazlı büyütme stratejisini benimser. Bu yaklaşım, her iterasyonda kayıp fonksiyonunu en çok azaltan yaprağı derinleştirerek daha hızlı yakınsama sağlar (scikit-learn, 2025). Matematiksel olarak modelin t . iterasyonunda tahmin güncellemesi Denklem (3.22)'deki gibi ifade edilir.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3.22)$$

Aşamalı öğrenme sürecinde eniyileme hedefi Denklem (3.23)'teki gibi tanımlanır.

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \quad (3.23)$$

LightGBM, bu eniyileme problemini ikinci mertebeden Taylor genişlemesiyle yakınsar ve gradyan (g_i) ile Hessian (h_i) değerlerini kullanarak her bölünme adımında en iyi ayırım noktasını seçer. Bu işlem sırasında Gradient-based One-Side Sampling - GOSS (Gradyan Tabanlı Tek Taraflı Örnekleme) ve Exclusive Feature Bundling - EFB (Ayrıcalıklı Özellik Demetleme) gibi iki temel teknikten yararlanır. GOSS, gradyan büyüklüğü yüksek olan örnekleri tercih ederek bilgi açısından zengin alt küme oluşturur ve bu sayede eğitim süresi kısalırken modelin doğruluğu korunur. EFB ile, aralarında korelasyon bulunmayan seyrek öznitelikler bir araya getirilerek özellik boyutu azaltılır, bu da bellek kullanımını düşürür ve modeli hızlandırır (Ke vd., 2017).

LightGBM'in yaprak bazlı stratejisi daha derin ağaçlar üretebildiğinden, model karmaşık örüntüleri daha iyi yakalayabilir. Ancak bu durum aynı zamanda aşırı öğrenme

riskini de artırabileceği için, budama, enbüyük derinlik sınırlaması ve erken durdurma gibi önlemlerle denge sağlanır. LightGBM aynı zamanda eksik verileri otomatik olarak işleyebilir, kategorik verileri doğrudan modele dâhil edebilir ve paralel işlem destekleyerek büyük veri ortamlarında etkili sonuçlar üretebilir (scikit-learn, 2025).

LightGBM modeli, performans açısından oldukça duyarlı olduğu için model çalıştırılırken hiperparametre ayarlaması büyük önem taşır. LightGBM modelinin başlıca hiperparametreleri şunlardır:

- num leaves: Her karar ağacında oluşturulabilecek azami sayıdaki yaprak sayısını belirler. Algoritmanın doğası gereği $2^{\max_depth}$ ’ten küçük tutulmalıdır.
- max depth: Ağaçların azami derinliğini sınırlayarak aşırı uyum (overfitting) riskini kontrol eder.
- learning rate: Her iterasyonda modelin ne kadar öğrenme yapacağını belirler. Düşük değerler daha yavaş ancak istikrarlı öğrenme sağlar.
- min data in leaf: Her yaprakta bulunması gereken asgari örnek sayısıdır. Aşırı öğrenmeye karşı koruma sağlar.
- lambda l1 ve lambda l2: L1 ve L2 düzenleme parametreleridir. Aşırı öğrenmeyi baskılar.

LightGBM’in başarılı bir şekilde çalışabilmesi için bu hiperparametrelerin veri setine uygun şekilde belirlenmesi büyük önem arz etmektedir (scikit-learn, 2025).

3.3.3.5. Ridge Regresyon Algoritması

Ridge Regresyon, çoklu doğrusal bağlantı sorununun önüne geçebilmek ve modelin genellenebilirliğini artırmak amacıyla standart doğrusal regresyona ceza terimi ekleyen bir yöntemdir. Özellikle öznitelik sayısının gözlem sayısına yakın ya da daha fazla olduğu durumlarda klasik doğrusal regresyon modellerinin tahmin gücü düşmekte, katsayılar

istikrarsız hâle gelmekte ve varyansları artmaktadır (Hoerl ve Kennard, 1970). Ridge Regresyon bu problemi adresleyerek, daha istikrarlı ve güvenilir bir modelleme yapısı sunmaktadır.

Ridge Regresyon'da amaç, klasik hata kareler toplamının enküçüklenmesine ek olarak katsayıların büyüklüğünü de kontrol altına almak için L2 normu kullanılarak bir ceza terimi eklemektir. Matematiksel olarak Ridge Regresyonun eniyileme problemi Denklem (3.24)'teki gibi ifade edilir.

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (3.24)$$

Denklem (3.24)'te; $\lambda \geq 0$, ceza parametresini temsil eder ve modelin katsayılarının büyüklüğünü sınırlar (Hoerl ve Kennard, 1970).

λ arttıkça modelin katsayıları sıfıra doğru küçülür; bu da modelin varyansını azaltırken bir miktar yanlılık artışına neden olur ancak bu denge çoğu zaman genel tahmin hatasını azaltmak açısından faydalı bir etki yaratır (James vd., 2013).

Ridge Regresyon, tüm öznitelikleri modelde tutmaya devam eder; yani herhangi bir değişkenin katsayısını sıfıra indirmez. Bu yönüyle, öznitelik seçimi için değil, modelin düzenlenmesi ve genellenebilirliğin artırılması için tercih edilen bir tekniktir. Ayrıca Ridge Regresyon'un çözümü kapalı formda bulunabilir ve hesaplaması görece hızlıdır, bu da özellikle çok sayıda değişkene sahip veri setlerinde önemli bir avantajdır (Hastie vd., 2009).

Bu çalışmada Ridge Regresyon, özellikle çoklu bağlantı riski taşıyan öznitelik gruplarının yer aldığı veri setlerinde, modelin istikrarını ve tahmin doğruluğunu artırmak amacıyla uygulanmıştır. Ayrıca diğer yöntemlerle karşılaştırıldığında düşük varyanslı tahminler üretmesi sebebiyle model karşılaştırmalarında önemli bir referans niteliği taşımaktadır.

3.3.3.6. Lasso Regresyon Algoritması

Lasso Regresyon, doğrusal regresyon modellerinin genellenebilirliğini artırmak ve aşırı öğrenmeyi engellemek amacıyla geliştirilen, L1 normuna dayalı bir regülazirasyon yöntemidir (Tibshirani, 1996). Ridge Regresyona benzer biçimde bir ceza terimi eklese de Lasso'nun en belirgin özelliği bazı katsayıları tam olarak sıfır yapabilmesidir. Bu özelliği sayesinde, yalnızca tahmin başarısını artırmakla kalmaz, aynı zamanda değişken seçimi işlemini de doğrudan modelleme sürecine entegre eder. Lasso Regresyon algoritması Denklem (3.25)'teki gibi formüle edilebilir.

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (3.25)$$

Denklem (3.25)'te; λ , regülazirasyon (ceza) katsayısıdır. $\sum_{j=1}^p |\beta_j|$, L1 normudur ve modelin sadeleştirilmesini sağlar. λ değeri arttıkça, bazı katsayıların katkısı sıfıra zorlanır ve bu durum modelin sadece anlamlı özniteliklerle tahmin yapmasına olanak tanır.

Modelin en önemli hiperparametresi olan λ , ceza düzeyini belirler. Çok küçük bir λ değeri modelin klasik doğrusal regresyona yakın davranmasına yol açarken, çok yüksek bir λ değeri modelin fazla sadeleşmesine ve önemli özniteliklerin elenmesine neden olabilir. Bu nedenle doğru hiperparametre seçimi, modelin başarısı açısından oldukça kritiktir (Tibshirani, 1996).

Özellikle yüksek boyutlu veri setlerinde, öznitelik seçim süreciyle birlikte çalışması nedeniyle Lasso Regresyon hem yorumlanabilirliği yüksek hem de aşırı öğrenmeye karşı dayanıklı bir model olarak literatürde sıklıkla tercih edilmektedir. Bu çalışmada da Lasso'nun bu özelliğinden faydalanarak daha yalın ve etkili tahmin modelleri geliştirilmesi hedeflenmiştir.

3.3.3.7. SVR Algoritması

SVR, doğrusal olmayan ilişkilerin etkin şekilde modellenmesini sağlayan çekirdek (kernel) tabanlı güçlü bir makine öğrenmesi yöntemidir. SVR'nin temel felsefesi, veriler arasındaki karmaşık ilişkileri yüksek boyutlu özellik uzaylarına yansıtarak açıklayabilecek bir regresyon fonksiyonu öğrenmektir. SVR modeli, belirli bir tolerans aralığında (ε) sapmaları göz ardı ederek sadece bu eşiği aşan hatalar üzerinde eniyileme yapar (Smola ve Schölkopf, 2004). SVR, Denklem (3.26)'daki kısıtlar altında, Denklem (3.27)'deki eniyileme problemi çerçevesinde tanımlanır.

$$\begin{cases} y_i - (\mathbf{w}^T \phi(x_i) + b) \leq \varepsilon + \xi_i \\ (\mathbf{w}^T \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (3.26)$$

$$\min \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right\} \quad (3.27)$$

Denklem (3.26) ve Denklem (3.27)'de; $\mathbf{w}$, model ağırlık vektörü; $\phi(x_i)$, verilerin yüksek boyutlu uzaya dönüştürülmesi; b , sabit terim (bias); ξ_i ve ξ_i^* , tolerans dışı hata terimleri; ε , hata tüpü genişliği (tahminlerinde tolere edeceği azami hata payı); C , ceza parametresi olup modelin aşırı öğrenme / eksik öğrenme dengesini belirlemektedir (Smola ve Schölkopf, 2004).

SVR modellerinde doğrusal olmayan yapılar için çekirdek fonksiyonları kullanılır. En yaygın kullanılan çekirdekler; doğrusal, polinomsal ve radyal tabanlı fonksiyonlardır. Bu fonksiyonlar, girdilerin benzerliğini ölçerek daha yüksek doğrulukta tahmin yapılmasını sağlar (Cristianini ve Shawe-Taylor, 2000).

SVR'nin önemli hiperparametreleri aşağıdaki gibidir (Drucker vd., 1996):

- C : Hataların cezalandırma düzeyini belirler. Büyük C değerleri daha az toleranslı modeller üretmektedir.
- ε : Modelin hataları hangi sınır değerleri içinde tolere edeceğini belirler. Küçük ε değeri, daha hassas bir model anlamına gelir.
- Çekirdek tipi ve parametreleri: Çekirdek fonksiyonun türü ve buna ait parametreler tahmin performansını doğrudan etkiler.

Bu yönüyle SVR, özellikle örnek sayısının az olduğu, ancak açıklayıcı özneliliklerin fazla olduğu tahmin problemleri için güçlü bir çözümdür. Yapı kimyasal üreticisinin geçmiş dönem satın alma verileri kullanılarak çimento fiyatı tahminlemesi gibi sorunlarda SVR, istikrarlı ve açıklanabilir çıktılar üretme potansiyeli taşımaktadır.

3.3.3.8. MLPR Algoritması

ANN, biyolojik sinir sistemlerinden esinlenerek geliştirilmiş ve birçok karmaşık regresyon probleminde başarılı sonuçlar sunan güçlü öğrenme algoritmalarıdır (Haykin, 2009; Bishop ve Nasrabadi, 2006). Bu ağların en temel ve yaygın biçimlerinden biri olan Multi-Layer Perceptron (MLP), katmanlı bir yapıya sahip olup doğrusal olmayan ilişkileri yakalama kapasitesi oldukça yüksektir. Özellikle yapay sinir ağlarının tam bağlantılı ve ileri beslemeli versiyonu olan MLP modelleri, çok katmanlı yapıları sayesinde yüksek derecede esneklik ve genellenebilirlik sunar.

MLPR, en az bir gizli katman içeren çok katmanlı bir yapay sinir ağı modelidir. Her bir katmanda yer alan nöronlar, kendilerine gelen girdileri bir ağırlık ve sapma terimi ile işleyerek doğrusal olmayan aktivasyon fonksiyonları aracılığıyla bir sonraki katmana iletir. Modelin temel yapı taşları ağırlıklar (W^i), sapma terimleri ($b^{(i)}$) ve aktivasyon fonksiyonlarıdır ($\phi^{(i)}$). Genel olarak, üç katmanlı bir MLP modelinin ileri besleme süreci Denklem (3.28)'deki gibi ifade edilebilir.

$$\hat{y} = \phi^{(3)}(W^{(3)} * \phi^{(2)}(W^{(2)} * \phi^{(1)}(W^{(1)}x + b^{(1)}) + b^{(2)}) + b^{(3)}) \quad (3.28)$$

Denklem (3.28)'de; x , girdi vektörü; $\hat{y}$, modelin tahmini çıktısı; $W^{(i)}$, i . katmanın ağırlık matrisi; $b^{(i)}$, i . katmanın bias vektörü; $\phi^{(i)}$, i . katmanda uygulanan aktivasyon fonksiyonunu (ReLU, tanh, sigmoid gibi) ifade eder (Haykin, 2009).

Eğitim sürecinde, model çıktısı $\hat{y}$ ile gerçek değer y arasındaki farkı enküçükleyecek şekilde ağırlık ve sapma parametreleri güncellenir. Modelin öğrenme süreci, geriye yayılım algoritması ve gradyan inişi yöntemi ile ağırlıkların güncellenmesine dayanır. Her bir yapay nöronun çıktısı Denklem (3.29) ve Denklem (3.30)'daki gibi tanımlanabilir:

$$z_j = \sum_{i=1}^n w_{ij}x_i + b_j \quad (3.29)$$

$$a_j = f(z_j) \quad (3.30)$$

Denklem (3.29) ve Denklem (3.30)'da; x_i , girdi nöronların değerleri; w_{ij} , giriş ve nöron arasındaki ağırlık; b_j , sapma terimi; $f(\cdot)$, aktivasyon fonksiyonu; z_j , nöronun doğrusal birleşimi; a_j , aktivasyon fonksiyonundan çıkan çıktı değeridir. Regresyon görevinde MLP'nin amacı, eğitim verisi üzerinden tahmin hatasını enküçüklemektir. Bu amaçla sıklıkla MSE kayıp fonksiyonu kullanılır.

MLPR'de öğrenme sürecini etkileyen temel hiperparametreler şunlardır:

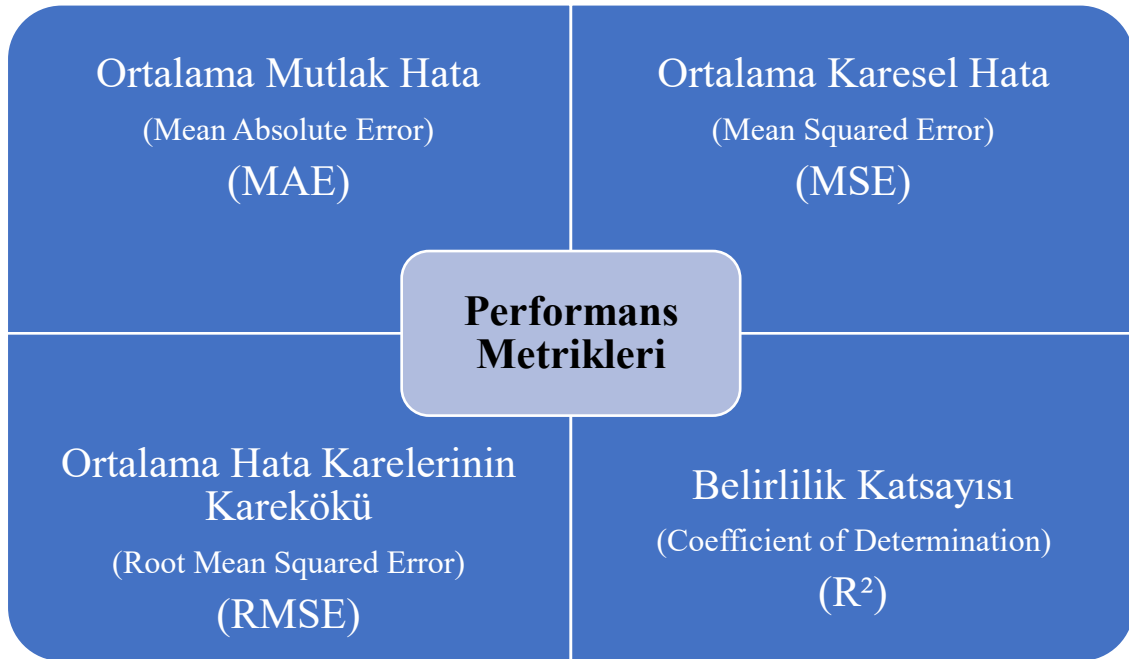
- hidden layer sizes: Gizli katmandaki nöron sayısı.
- activation: Aktivasyon fonksiyonu seçimi.
- learning rate init: Başlangıç öğrenme oranı.
- max iter: Azami iterasyon sayısı
- alpha: Ağırlıkların büyüklüğünü cezalandıran L2 regülarizasyon katsayısı.

MLPR, özellikle çok boyutlu, doğrusal olmayan ve etkileşimli veri yapılarında önemli avantajlar sunmaktadır. Ancak modelin başarıyla eğitilebilmesi, parametre seçimleri ve veri ön işleme kalitesiyle doğrudan ilişkilidir. Bu nedenle, standartlaştırılmış (z-score) verilerle çalışılması genellikle önerilmektedir (Goodfellow vd., 2016).

3.4. Çalışma Kapsamında Kullanılan Performans Metrikleri

Makine öğrenmesi tabanlı tahmin modellerinin değerlendirilmesinde, yalnızca modelin öğrenme başarısı değil, aynı zamanda genelleme yeteneği, hata yapısı ve öngörü gücü de dikkate alınmalıdır. Bu bağlamda, tahmin performansının sistematik bir şekilde ölçülmesi, model karşılaştırmalarının sağlıklı biçimde yapılabilmesi için zorunludur. Bu çalışmada, yapı kimyasal üreticisinin çimento satın alma fiyatlarını öngörmeyi amaçlayan makine öğrenmesi modellerinin başarısını değerlendirmek amacıyla dört temel performans metriği kullanılmıştır: MAE, MSE, RMSE ve R^2 .

Kullanılan performans metriklerinin şematize edilmiş hâli Şekil 3.3'te sunulmuştur.



Şekil 3.3. Performans Metrikleri

Bu metrikler farklı hata boyutlarına duyarlılıkları ve yorumlanabilirlik düzeyleri bakımından birbirini tamamlayıcı niteliktedir. MAE, mutlak hataların ortalamasını alarak tahminlerin ortalama sapmasını ölçerken; MSE, hataların karesini alarak daha büyük hatalara daha fazla ağırlık vermekte ve uç değerler karşısında duyarlılığı artırmaktadır. RMSE, MSE'nin karekökü alınarak elde edilmekte ve birimsel olarak hedef değişken ile aynı ölçekte yorumlanabilme avantajı sağlamaktadır. R^2 katsayısı ise modelin açıklayıcılık düzeyini gösterir; hedef değişkenin varyansının ne kadarının bağımsız değişkenler aracılığıyla açıklandığını ifade eder.

Bu bölümde, söz konusu dört metriğin teorik altyapısı, avantajları ve sınırlılıkları açıklanarak; çalışmada kullanılan tahminleme modellerinin hangi ölçütlerle değerlendirildiği ayrıntılı biçimde ortaya konacaktır.

3.4.1. Ortalama Mutlak Hata (MAE)

MAE, bir tahmin modelinin tahmin ettiği değerler ile gerçek değerler arasındaki farkların mutlak değerlerinin ortalamasını ifade eden temel bir hata metriğidir. MAE, tahmin hatalarının büyüklüğünü doğrudan gözlemleyebilme imkânı sunması açısından oldukça yorumlanabilir ve sezgisel bir metrik olarak literatürde yaygın şekilde kullanılmaktadır (Willmott ve Matsuura, 2005). MAE, Denklem (3.31)'deki gibi formüle edilir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.31)$$

Denklem (3.31)'de; n , gözlem sayısını; y_i , i . gözlemdeki gerçek değeri; $\hat{y}_i$ ise i . gözlemdeki model tahminini ifade etmektedir.

MAE'nin en önemli avantajlarından biri, tüm hata terimlerine eşit ağırlık vermesi ve ölçülen birimle aynı ölçekte ifade edilebilmesidir. Bu özelliği sayesinde, modelin ortalama ne kadar yanlış olduğunu doğrudan belirtir. Örneğin, bir modelin MAE değeri 3,5 USD/ton olarak hesaplandığında bu, modelin tahminlerinde ortalama olarak 3,5 USD/ton sapma yaptığını göstermektedir.

Bununla birlikte MAE, hataların mutlak değerlerini kullanmasından dolayı pozitif ve negatif sapmalar nötr duruma sebebiyet vermez. Ayrıca, diğer bazı metriklerin aksine, uç değerlerin etkisine karşı daha az duyarlıdır. Bu özelliği, aşırı uç tahminlerin bulunduğu veri setlerinde daha kararlı bir performans ölçütü olmasını sağlar (Chai ve Draxler, 2014).

Bu çalışmada MAE metriği, özellikle satın alma fiyat tahminlerinin mutlak sapmalarını anlamak ve modellerin aylık karar destek düzeyindeki etkinliğini değerlendirebilmek amacıyla kullanılmıştır.

3.4.2. Ortalama Karesel Hata (MSE)

MSE, bir tahmin modelinin başarısını değerlendirmek amacıyla kullanılan temel performans ölçütlerinden biridir. MSE, model tarafından yapılan tahmin hatalarının karesini alarak ortalama bir değer sunar ve bu yönüyle sapmaların büyüklüğünü daha hassas biçimde dikkate alır. Bu özellik, büyük hatalara daha fazla ceza verilmesini sağlar (Hyndman ve Athanasopoulos, 2018). MSE, Denklem (3.32)'deki gibi ifade edilir.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.32)$$

Denklem (3.32)'de; n , gözlem sayısını; y_i , i . gözlemdeki gerçek değeri; $\hat{y}_i$ ise i . gözlemdeki model tahminini ifade etmektedir.

MSE'nin kare alma işlemi, pozitif ve negatif sapmaların birbirini nötr duruma sebebiyet vermesini engellerken, aynı zamanda büyük hataların etkisini abartarak modele ceza uygular. Bu durum, özellikle sistematik sapmaların tespit edilmesi ve modelin aykırı değerlere duyarlılığının ölçülmesi açısından önemlidir. Ancak bu hassasiyet, veri setinde uç değerlerin varlığında MSE'nin oldukça büyük değerlere ulaşmasına da neden olabilir. Bu nedenle, MSE'nin yorumlanması sırasında veri dağılımındaki ekstrem değerler göz önünde bulundurulmalıdır (Makridakis et al., 1997).

MSE, özellikle parametre eniyileme süreçlerinde bir maliyet fonksiyonu olarak sıkça tercih edilmektedir. Örneğin, doğrusal regresyon ya da yapay sinir ağları gibi birçok

denetimli öğrenme algoritması, eğitim sürecinde bu fonksiyonun asgari seviyeye indirgenmesini hedefler.

Bu çalışmada MSE metriği, tahmin hatalarının dağılımı ve yoğunluğu hakkında genel bir fikir sunmak; farklı modellerin hata kareleri bakımından kıyaslanmasını sağlamak ve özellikle uç değer duyarlılığı bakımından değerlendirme yapmak amacıyla kullanılmıştır.

3.4.3. Hata Kareleri Ortalamasının Kare Kökü (RMSE)

RMSE, regresyon tabanlı model performans değerlendirmelerinde en yaygın kullanılan ölçütlerden biridir. RMSE, ortalama karesel hatanın karekökünün alınmasıyla elde edilir ve hata değerlerini orijinal ölçüm biriminde sunarak yorumlamayı kolaylaştırır (Chai ve Draxler, 2014). Bu yönüyle hem hata büyüklüğünü hem de modelin genel tahmin doğruluğunu yansıtan sezgisel bir ölçüttür. RMSE, Denklem (3.33) ile hesaplanır.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.33)$$

Denklem (3.33)'te; n , gözlem sayısını; y_i , i . gözlemdeki gerçek değeri; $\hat{y}_i$ ise i . gözlemdeki model tahminini ifade etmektedir.

RMSE hesaplanırken, tahmin hatalarının karesi alınıp ortalaması hesaplandıktan sonra karekök işlemi uygulanır. Böylece RMSE, MSE'nin büyüklüğüne duyarlı yapısını korurken, sonuçları hedef değişkenin ölçü birimiyle aynı düzlemde sunarak pratik bir değerlendirme imkânı sağlar. Özellikle farklı modellerin performanslarının doğrudan karşılaştırılması gerektiğinde RMSE, sezgisel anlamı yüksek bir metrik olarak öne çıkar (Willmott ve Matsuura, 2005).

RMSE, aynı zamanda hataların dağılımındaki büyük sapmalara karşı duyarlılığı ile bilinir çünkü kare alma işlemi, uç değerlerin etkisini büyütür ve bu durum RMSE'nin cezalandırıcı doğasını yansıtır. Bu özellik, modelin küçük hatalar üretmeye ne derece yatkın

olduğunu test etmek için avantajlıdır; ancak veri setinde uç değerlerin bulunması hâlinde RMSE'nin dikkatli yorumlanması gerekir.

Bu çalışma kapsamında RMSE, farklı makine öğrenmesi algoritmalarının performanslarının aynı ölçekte değerlendirilmesi ve özellikle MSE ile hata yapılarının ayrıntılı biçimde analiz edilmesi amacıyla kullanılmıştır. Böylece sadece hataların kare ortalaması değil, bu hataların tahmin edilen değerler bağlamında ne kadar anlamlı olduğu da ortaya konmuştur.

3.4.4. Belirlilik Katsayısı (R^2)

Tahmin modellerinin başarısını değerlendirmede en çok başvurulan istatistiksel ölçütlerden biri olan R^2 , modelin bağımlı değişkendeki toplam varyansın ne kadarını açıkladığını gösterir (Kutner vd., 2005). Bu değer, 0 ile 1 arasında değişir ve 1'e yaklaştıkça modelin tahmin performansının arttığını, yani gözlemlenen verinin model tarafından daha iyi açıklandığını ifade eder. R^2 hesabında kullanılan matematiksel ifade Denklem (3.34)'teki gibidir.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.34)$$

Denklem (3.34)'te; y_i , gerçek değer; $\hat{y}_i$, modelin tahmin ettiği değer; $\bar{y}$, gerçek değerlerin ortalaması; n , gözlem sayısıdır. Pay kısmında modelin tahmin hataları toplamı (Sum Squares of Error – SSE), payda kısmında ise ortalama modele göre yapılan hataların toplamı (Sum Squares of Total – SST) yer alır. Böylece R^2 , modelin rastgele bir tahmine göre ne derece isabetli olduğunu gösteren göreceli bir başarı ölçütü sunar.

R^2 , özellikle çoklu regresyon ve makine öğrenmesi temelli tahmin modellerinde modelin açıklayıcılık gücünü anlamak açısından kritik bir ölçüttür. $R^2 = 1$ ise kurulan tahmin modeli varyansın tamamını açıklamaktadır yani tahminler mükemmeldir. $R^2 = 0$ ise kurulan tahmin modeli hiçbir varyansı açıklayamamaktadır, yani model tahmin açısından işe yaramazdır. R^2 , genellikle 0 ile 1 arasında değişse de özellikle modelin veriye çok zayıf uyum sağladığı veya modelin standart varsayımlarının ihlal edildiği durumlarda negatif

değerler alabilir. Negatif bir R^2 değeri, kurulan tahmin modelinin veriyi açıklama yeteneğinin olmadığını, hatta ortalamadan bile daha kötü performans gösterdiğini gösterir.

R^2 metriği her ne kadar sezgisel ve yaygın olsa da bazı sınırlılıklara sahiptir. Örneğin, bağımsız değişken sayısı arttıkça R^2 değeri genellikle artma eğilimindedir; bu durum modelin aşırı öğrenme eğilimini maskeleyebilir. Bu sebeple, özellikle çok sayıda değişken içeren modellerde düzeltilmiş R^2 (Adjusted R^2) kullanılması önerilmektedir (Wooldridge, 2013).

Bu çalışma kapsamında R^2 metriği hem modelin açıklayıcılık düzeyini değerlendirmek hem de MAE, MSE ve RMSE gibi mutlak hata bazlı ölçütlerle birlikte kullanılarak çok boyutlu bir model başarı analizi sunmak amacıyla tercih edilmiştir.

3.5. GridSearchCV ile Hiperparametre Eniyilemesi

Makine öğrenmesi modellerinin başarısı, yalnızca seçilen algoritmaların yapısal özelliklerine değil, aynı zamanda bu algoritmaların yapılandırılmasında kullanılan hiperparametrelerin doğruluğuna da bağlıdır. Hiperparametreler, modelin öğrenme sürecini doğrudan yönlendiren ve genellikle kullanıcı tarafından belirlenen ayar değerleridir. Bu değerlerin uygun şekilde belirlenmemesi, modelin aşırı öğrenmesine veya eksik öğrenmesine neden olabilmektedir. Bu nedenle, hiperparametre eniyilemesi, makine öğrenmesi sürecinin ayrılmaz bir bileşeni hâline gelmiştir (Bergstra ve Bengio, 2012).

Bu çalışmada, model başarısını artırmak ve geçerli sonuçlara ulaşabilmek amacıyla hiperparametre ayarlamaları sistematik bir biçimde gerçekleştirilmiştir. Bu süreçte, yaygın olarak kullanılan GridSearchCV yöntemi tercih edilmiştir. GridSearchCV, kullanıcı tarafından belirlenen hiperparametre değerleri kümesinin her bir kombinasyonunu sistematik olarak dener ve çapraz doğrulama stratejisi ile her kombinasyonun model performansını değerlendirir. Böylece en uygun hiperparametre seti, sadece eğitim verisi üzerinde değil, aynı zamanda genel veri dağılımını temsil eden alt kümeler üzerinde test edilerek belirlenmiş olur.

Matematiksel olarak, hiperparametre kümesi $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ olsun. GridSearchCV yöntemi, bu küme üzerindeki tüm olası hiperparametre kombinasyonlarını içeren kartezyen çarpımı yani $\Theta_1 \times \Theta_2 \times \dots \times \Theta_k$ uzayını tarar. Her bir parametre kombinasyonu $\theta^{(i)}$ için bir çapraz doğrulama stratejisi uygulanır ve her bir katman için doğrulama hatası hesaplanır. En iyi kombinasyon Denklem (3.35)'teki gibi hesaplanır.

$$\theta^* = \arg \min_{\theta^{(i)} \in \Theta} \frac{1}{k} \sum_{j=1}^k \mathcal{L}(f_{\theta^{(i)}}, D_j^{\text{val}}) \quad (3.35)$$

Denklem (3.35)'te; $\mathcal{L}$, seçilen kayıp fonksiyonunu temsil ederken D_j^{val} her katmandaki doğrulama veri kümesidir. Bu yapı modelin genellenebilirliğini artırmak ve aşırı öğrenme riskini azaltmak açısından önemli katkılar sağlamaktadır.

GridSearchCV'nin önemli avantajlarından biri, sistematik tarama yoluyla hiperparametre kombinasyonlarının tamamının denenmesidir fakat bu yöntem, özellikle hiperparametre uzayının büyük olduğu durumlarda yüksek hesaplama maliyetine neden olabilir. Bu sebeple bazı kaynaklarda Rastgele Arama veya Bayesyen Eniyileme gibi alternatif yöntemler önerilmektedir (Bergstra vd., 2011). Bu çalışmada kullanılan model ve hiperparametre sayısı görece sınırlı olduğundan, GridSearchCV yöntemi hesaplama açısından uygulanabilir olmuş ve tercih edilmiştir.

Sonuç olarak, GridSearchCV yöntemi bu çalışmada model performanslarını artırmak ve modellerin hiperparametre yapılarını en uygun hâle getirmek amacıyla kullanılmıştır. Bu süreç sayesinde, elde edilen tahmin modelleri yalnızca doğruluğu yüksek değil, aynı zamanda genellenebilirliği yüksek modeller olarak kurgulanmıştır. İzleyen bölümlerde, hiperparametre eniyilemesi sonrası elde edilen modellerin analizlerine ve uygulama sürecine ilişkin detaylara yer verilecektir.

4. UYGULAMA VE MODELLEME ADIMLARI

Bu bölümde, yapı kimyasalı üreticisine ait çimento satın alma fiyatlarını öngörebilmek amacıyla yürütülen tüm veri hazırlama, öznitelik mühendisliği ve makine öğrenmesi modelleme süreçleri ayrıntılı olarak sunulmaktadır. Öncelikle veri setinin genel yapısı tanıtılmış, ardından yüzeysel ve gelişmiş eksik veri doldurma stratejileriyle verideki boşluklar giderilmiştir. Bu hazırlıkların ardından, farklı öznitelik seçim stratejilerine dayalı beş ayrı modelleme çalışması gerçekleştirilmiştir. Her bir çalışma, kendine özgü bir değişken yapısına ve kurguya sahiptir.

Modelleme sürecinde RF, GB, XGBoost, LightGBM, Ridge Regresyon, Lasso Regresyon, SVR ve MLPR olmak üzere sekiz farklı makine öğrenmesi algoritması kullanılmıştır. Her algoritma hem varsayılan hiperparametrelerle hem de GridSearchCV yöntemiyle eniyilenmiş parametrelerle çalıştırılmıştır. Böylece model başarısı sadece öznitelik yapısına değil, aynı zamanda hiperparametrelerin etkisine göre de değerlendirilmiştir. Tahmin başarısı; MAE, MSE, RMSE ve R^2 gibi yaygın performans metrikleri üzerinden analiz edilmiştir.

Bu bölüm, toplamda beş ayrı modelleme çalışması (Çalışma 1–5) ve bunlara ait veri hazırlık adımlarını içermektedir. Çalışmalar arasında uzman görüşü temelli öznitelik seçimi (Çalışma 1), korelasyon analizi yaklaşımıyla öznitelik seçimi (Çalışma 2), bu iki yapının birleşimiyle oluşturulan hibrit öznitelik kümesi (Çalışma 3), kapasite kullanım oranı değişkeninin çıkarılmasıyla yapılan hassasiyet analizi (Çalışma 4) ve korelasyon tabanlı nihai öznitelik indirgeme uygulaması (Çalışma 5) yer almaktadır. Bu kapsamlı yapı sayesinde, sadece en başarılı modelin değil, aynı zamanda en anlamlı öznitelik setinin de belirlenmesi amaçlanmıştır.

Bu bütüncül uygulama ve modelleme yaklaşımı, yapı kimyasalları sektöründe karar verme sistemleri için makine öğrenmesi tabanlı tahmin modellerinin nasıl geliştirilebileceğini göstermesi bakımından akademik ve sektörel açıdan önemli bir katkı sunmaktadır.

4.1. Ham Veri Setinin Tanıtımı ve Genel Özellikleri

Tez çalışmasında kullanılan ham veri seti, yapı kimyasalı üreticisi tarafından araştırma amacıyla sağlanmıştır. İlk olarak Ocak 2013 – Mayıs 2024 dönemini kapsayacak şekilde 137 satırdan oluşan veri seti, yapılan yüzeysel eksik veri doldurma işlemleri ve öznitelik seçim süreci sonrası güncellenerek Ocak 2013 – Aralık 2024 dönemini kapsayan ve toplamda 144 satırdan oluşan bir yapıya kavuşturulmuştur. Veri setinde yer alan “Tarih” sütunu yalnızca etiketleme amacıyla kullanılmış olup, modelleme sürecine dâhil edilmemiştir.

Veri seti, çimento sektöründeki satın alma fiyatlarının tahminine yönelik olarak oluşturulmuş olup, çok yönlü bir yaklaşımı yansıtan özniteliklerden meydana gelmektedir. Öznitelikler; arz-talep dengesi, nakliye maliyetleri, enerji tüketimi ve maliyetleri, demografik etkenler, ekonomik göstergeler, hammadde maliyetleri, işçilik giderleri ve siyasi etkenler başlıkları altında gruplanmıştır. Her bir grup, sektörel anlamlılığı göz önünde bulundurularak dikkatle seçilmiş ve farklı kaynaklardan temin edilmiştir. Ancak verilerin önemli bir kısmı sektörel gizlilik gerekçesiyle manipüle edilmiş olup, veri kaynaklarının paylaşılması firma tarafından uygun bulunmamıştır.

Arz-talep dengesi başlığı altındaki ilk öznitelik olan Çimento Sanayi İstihdam Sayısı (kişi), normalde yalnızca yıllık olarak yayımlanan bir veridir. Dolayısıyla veri setinde bu özniteliğe ilişkin yalnızca yılların Aralık aylarındaki değerler bulunmakta, kalan aylarda ise veri boş bırakılmaktadır. Bu durum, veri setinin aylık yapısına uyarlanması için dönüşüm yapılması gerektiğini ortaya koymuştur. Aynı grup altında yer alan Çimento İhracatı (ton) ve Klinker İhracatı (ton) değişkenleri için Ocak 2013 – Aralık 2016 aralığında yalnızca yıl sonlarına ilişkin (Aralık ayı) yıllık toplam değerler mevcuttur. Ancak Ocak 2017 – Şubat 2024 aralığında aylık bazda eksiksiz veriler sağlanmıştır. Mart 2024 – Mayıs 2024 dönemine ait değerler ise eksik olarak kalmıştır. Benzer bir şekilde, Çimento Üretim Miktarı (ton) ve Klinker Üretim Miktarı (ton) öznitelikleri de yalnızca Ocak 2017 – Şubat 2024 döneminde eksiksiz şekilde yer almakta olup, önceki ve sonraki aylarda veri eksiklikleri söz konusudur. Ayrıca Yapı Ruhsatına Göre Bina Sayısı (adet) özniteliği Ocak 2013 – Mart 2024 aralığında eksiksiz olarak mevcut iken, Nisan–Mayıs 2024 aylarında veri bulunmamaktadır. Bu

özniteliğin izleyen analiz adımlarında metrekare cinsinden yapı ruhsatı bilgisine dönüştürülmesi planlanmıştır.

Nakliye maliyetleri kategorisinde, Motorin Fiyatı (TL/Litre) ve Motorin Fiyatı (USD/Litre) olmak üzere iki ayrı gösterge kullanılmıştır. Her iki öznitelik de Ocak 2013 – Mayıs 2024 döneminde eksiksiz şekilde veri setine dâhil edilmiştir. Bu öznitelikler, özellikle taşıma ve lojistik maliyetlerinin fiyatlara etkisini gözlemlemek açısından kritik rol oynamaktadır.

Enerji tüketimi ve maliyetleri kapsamında yer alan Petrokok Birim Fiyatı (USD) ile Petrokok İthalatı (kg) değişkenleri yalnızca iki ayda (Mayıs 2013 ve Mayıs 2024) eksik değer içermektedir. Diğer aylar için veri tamdır. Elektrik Fiyatı (TL/MWh) ve Elektrik Fiyatı (USD/MWh) öznitelikleri ise Ocak 2013 – Mayıs 2024 aralığında eksiksiz şekilde kaydedilmiştir. Enerji bileşenleri, çimento üretim süreçlerinin maliyet yapısında önemli bir yer tuttuğundan modelin performansına doğrudan katkı sağlaması beklenmektedir.

Demografik etkenler başlığı altında; Türkiye Nüfusu, Türkiye Kent Nüfusu ve Türkiye Kırsal Nüfusu olmak üzere üç farklı nüfus göstergesi kullanılmıştır. Bu veriler yalnızca yıllık olarak yayımlandığı için her yılın tamamında aynı değerler tekrarlanmıştır. Ocak 2024 – Mayıs 2024 dönemi için ise henüz yıl tamamlanmadığından veri bulunmamaktadır. Bu nedenle ilgili aylar eksik değer olarak işaretlenmiştir. Ayrıca, özellikle Türkiye Kent Nüfusu değişkeni ilerleyen adımlarda nüfus artış hızına dönüştürülerek ordinal düzeyde modele dâhil edilmiştir.

Ekonomik göstergeler arasında yer alan Döviz Kuru (TL/USD) ve Döviz Kuru (TL/EUR) öznitelikleri Ocak 2013 – Mayıs 2024 dönemi boyunca eksiksiz şekilde sağlanmıştır. Brent Petrol (USD) değişkeni, Mart 2024 – Mayıs 2024 döneminde eksik olsa da önceki dönemlerde tam olarak mevcuttur. Ayrıca, Üretici Fiyat Endeksi (ÜFE) değişkeni de ilgili tüm dönemde eksiksiz veri sunmaktadır. Bu değişkenlerin ekonomik dalgalanmalara bağlı olarak satın alma fiyatlarına etkide bulunabileceği öngörülmektedir.

Hammadde maliyetleri özniteliklerinin çoğu, yıllık yayımlanan ve sabit değerler üzerinden hesaplanan verilerden oluşmaktadır. Kalker, Alçıtaşı, Çimento Kili ve Demir Cevheri gibi hammaddelere ilişkin ocakbaşı satış fiyatları hem TL hem de USD cinsinden

veri setine dâhil edilmiştir. Ocak 2013 – Aralık 2014 döneminde söz konusu hammaddelerin hiçbirine ait veri bulunmazken, Ocak 2015 – Aralık 2023 döneminde büyük ölçüde eksiksiz veriler sağlanmıştır. Ocak 2024 – Mayıs 2024 döneminde ise veri eksikliği yeniden ortaya çıkmıştır. Bu özniteliklerin çoğunun doğası gereği yıllık belirlenen fiyatlara sahip olması, tüm yıl boyunca aynı değerin atanmasını gerektirmiştir.

İşçilik giderleri kapsamında yalnızca Brüt Asgari Ücret (TL) ve Brüt Asgari Ücret (USD) değişkenlerine yer verilmiş, bu değişkenler Ocak 2013 – Mayıs 2024 dönemi boyunca eksiksiz şekilde sağlanmıştır. Asgari ücretin yıllık olarak belirlendiği göz önüne alındığında, bu öznitelikler için her yılın tüm aylarında aynı değer kullanılmıştır.

Siyasi etkenler başlığında yer alan Herhangi bir seçim var mı? (genel/yerel) özniteliği, Ocak 2013 – Mayıs 2024 aralığında eksiksiz şekilde derlenmiş ve Evet/Hayır formatında ikili bir değişken olarak veri setine dâhil edilmiştir. Bu değişkenin izleyen çalışmalarda, seçim söylentileri ve politik beklentilerin daha iyi modellenebilmesi adına Google Trends verileriyle yeniden yapılandırılarak sayısal forma dönüştürülmüştür.

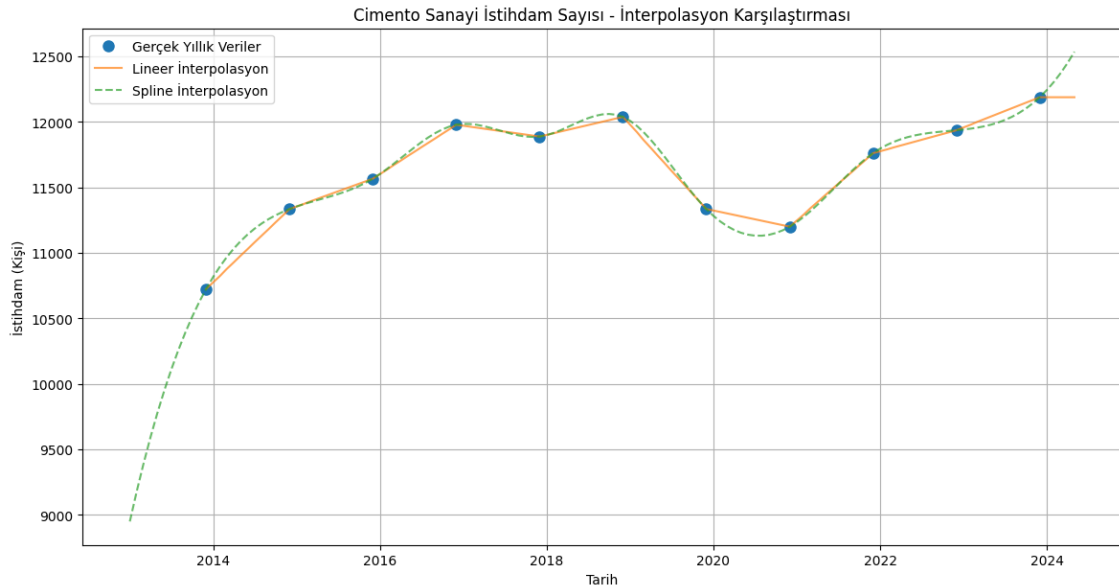
Tüm bu öznitelikler dışında, çalışmanın hedef değişkeni olarak Eskişehir Fabrikası Gri Çimento Satın Alım Fiyatı, yapı kimyasalı üreticisi tarafından sağlanmış ve sektörel gizlilik nedeniyle gerçek değerleri manipüle edilerek bir endeks biçiminde paylaşılmıştır. Ocak 2013 – Şubat 2024 aralığında eksiksiz olan bu veri, Mart 2024 – Mayıs 2024 döneminde eksik durumdadır. Modelleme sürecinde bu hedef değişkenin tahmin edilmesi esas alınmıştır.

Genel olarak değerlendirildiğinde, bu veri seti; sektörün dinamiklerini yansıtan çok boyutlu, zengin ve karmaşık bir yapıya sahiptir. Eksik veri durumları ve yapısal farklılıklar dikkate alınarak veri seti üzerinde çeşitli dönüşümler yapılması gerekmiş, bu dönüşümlere ilişkin detaylar çalışmanın izleyen bölümlerinde açıklanmıştır. Bu yönüyle veri seti, sektörel bilgiyle desteklenmiş, uygulama odaklı ve bilimsel araştırmalara uygun bir modelleme zemini sunmaktadır.

4.2. Ön Değerlendirme Amaçlı Yüzeysel Eksik Veri Doldurma İşlemleri

Makine öğrenmesi algoritmalarının etkin bir şekilde uygulanabilmesi için öncelikli gerekliliklerden biri, analizde kullanılacak veri setinde eksik değer bulunmamasıdır. Bu kapsamda, çalışmanın ilk aşamasında modelleme sürecine hazırlık amacıyla, öznitelik seçimlerinin sağlıklı bir biçimde gerçekleştirilebilmesi için geçici nitelikte yüzeysel eksik veri doldurma işlemleri uygulanmıştır. Bu aşamadaki temel hedef, veri setini tamamıyla analiz edilebilir hâle getirmek değil; ilerleyen aşamalarda uygulanacak öznitelik seçim yöntemlerine uygun bir veri ortamı sağlamaktır. Bu nedenle uygulanan doldurma yöntemleri, gelişmiş tahmin algoritmalarına dayanmaktan ziyade pratik ve kolay uygulanabilir yöntemler aracılığıyla gerçekleştirilmiştir. İlgili işlemlerin detayları ve kullanılan yaklaşımlar, bu bölümün izleyen kısımlarında adım adım açıklanacaktır.

Çimento Sanayi İstihdam Sayısı (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.1’de sunulmuştur. Öznitelige ilişkin toplam satır = 137, mevcut veri sayısı = 11, eksik veri sayısı = 126 şeklindedir.

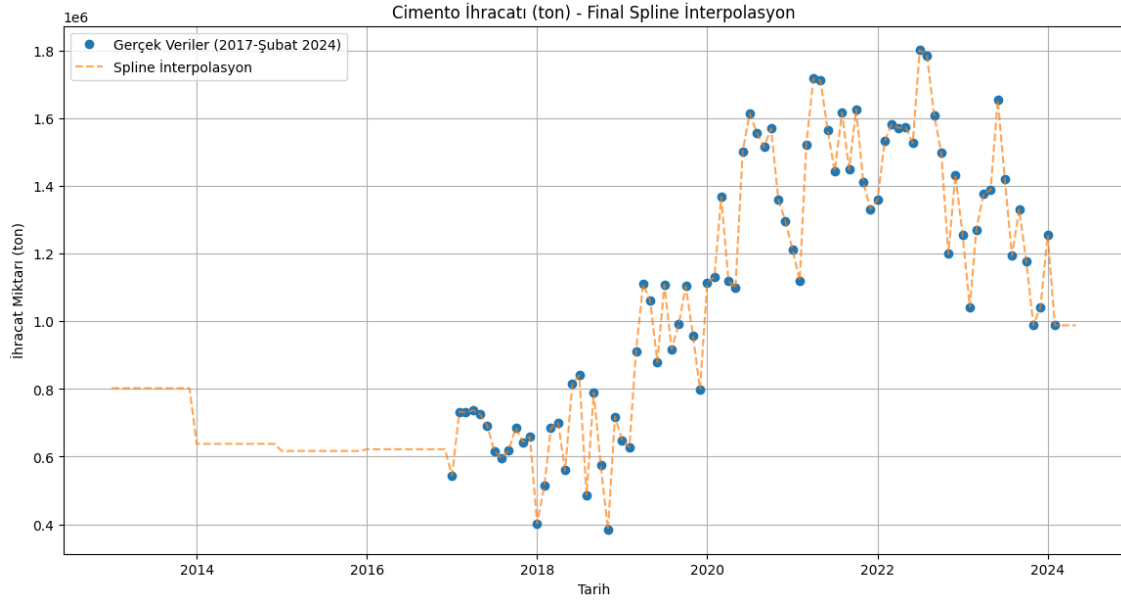


Şekil 4.1. Çimento Sanayi İstihdam Sayısı (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Veri setinde yer alan Çimento Sanayi İstihdam Sayısı (kişi) özniteliği, yalnızca her yılın Aralık ayına ait gözlemleri içermektedir. Bu nedenle, bu değişkenin aylık formata

uygun hâle getirilebilmesi amacıyla yüzeysel eksik veri doldurma işlemleri gerçekleştirilmiştir. Şekil 4.1’de görülebileceği üzere, bu öznitelik için öncelikle lineer interpolasyon ve spline interpolasyon yöntemleri uygulanmış ve sonuçlar görsel olarak değerlendirilmiştir. Yapılan incelemeler sonucunda, zaman içerisindeki istihdam değişiminin belirgin sıçramalar içermediği ve görece düzgün bir eğilim sergilediği dikkate alınarak, spline interpolasyon yöntemi tercih edilmiştir. Bu yöntemin, zaman serisinin doğal akışını ve sektörel büyüme eğilimini daha doğru yansıtacağı düşünülmüştür (De Boor, 2001; Hastie vd., 2009). Diğer yandan, lineer interpolasyonun özellikle uzun süreli eksikliklerde trend doğrusunu yapay biçimde kırabileceği ve zamana bağlı dalgalanmaları yeterince temsil edemeyeceği öngörülerek bu seçenek elenmiştir.

Çimento İhracatı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.2’de sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı= 90, eksik veri sayısı = 47 şeklindedir.

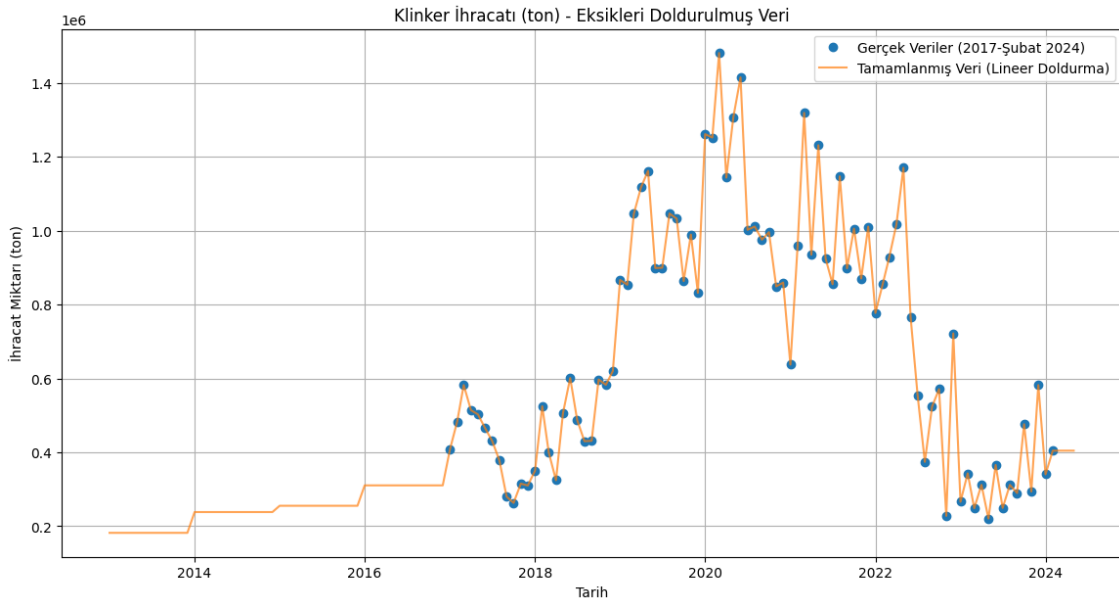


Şekil 4.2. Çimento İhracatı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Çimento İhracatı (ton) özniteliği, veri setinin 2013–2016 dönemine ait kısmında yalnızca yıl sonlarına ait (Aralık ayı) yıllık toplam değerleri içermektedir. Bu nedenle, ilgili dönemde eksik olan aylık gözlemler, her yılın toplam değerinin on iki aya eşit şekilde dağıtılması yöntemiyle doldurulmuştur. Bu yaklaşım, yıllık bazda sağlanan sınırlı veriden

yola çıkarak aylık bazda tahmin üretilebilmesini mümkün kılmış ve veri setinin yapısal bütünlüğünü korumayı hedeflemiştir. 2017 sonrasında ise daha ayrıntılı gözlemlerin mevcut olduğu ancak bazı aylarda eksikliklerin bulunduğu tespit edilmiştir. Şekil 4.2’de görülebileceği gibi bu eksik veriler için spline interpolasyon yöntemi uygulanmış, böylece mevcut trendin sürekliliği korunarak zaman serisinin yapay kırılmalar olmaksızın tamamlanması sağlanmıştır (De Boor, 2001; Hastie vd., 2009). Alternatif olarak değerlendirilen sabit değer atama yönteminden ise özellikle dönemsel dalgalanmaları temsil etme yetersizliği nedeniyle vazgeçilmiştir; zira bu tür bir yaklaşım, dış ticaret hacmindeki mevsimsel değişimleri göz ardı ederek modele yapay bir durağanlık kazandırabilirdi (Shumway ve Stoffer, 2000).

Klinker İhracatı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.3’te sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı= 90, eksik veri sayısı = 47 şeklindedir.

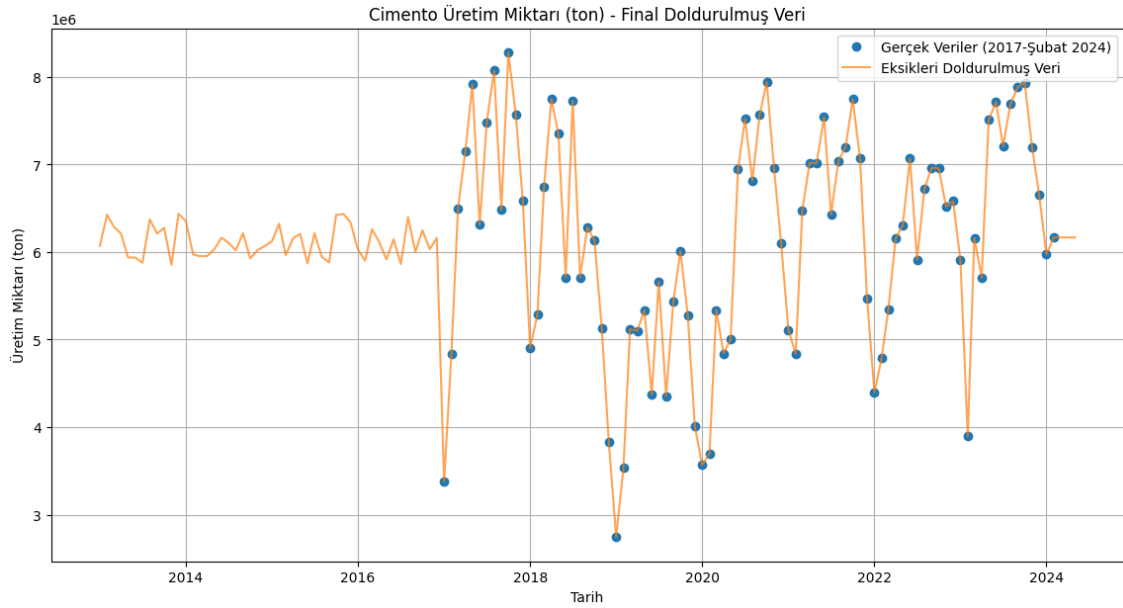


Şekil 4.3. Klinker İhracatı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Klinker İhracatı (ton) özniteliği, 2013–2016 yılları arasında yalnızca yıllık toplam ihracat değerlerine sahiptir. Bu döneme ait eksik aylık veriler, ilgili yılın toplam değerinin aylara eşit şekilde dağıtılması yoluyla doldurulmuştur. Ortalama dağıtım yöntemi, özellikle ihracat verilerinin yıllık bazda sunulduğu ve dönemsel dalgalanmaların detaylarının mevcut

olmadığı durumlarda analitik bir geçiş sağlamaktadır. 2017 yılı itibarıyla aylık bazda veri mevcudiyeti artmış olmakla birlikte, bazı aylarda eksiklikler gözlemlenmiştir. Bu eksik gözlemler, Şekil 4.3'te de görülebileceği üzere Çimento İhracatı (ton) özniteliği ile yapısal ve dönemsel olarak yüksek düzeyde paralellik göstermesi nedeniyle aynı yöntemle, yani spline interpolasyon tekniği ile tamamlanmıştır. Bu tercih, dış ticaret göstergelerindeki olası sürekliliği ve yumuşak geçişleri korumak amacıyla yapılmış olup, zamansal trendin doğallığını sürdürmeye yönelik bir karar olmuştur (De Boor, 2001; Hastie vd., 2009).

Çimento Üretim Miktarı (ton) Özniteliği içi Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.4'te sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı = 86, eksik veri sayısı = 51 şeklindedir.

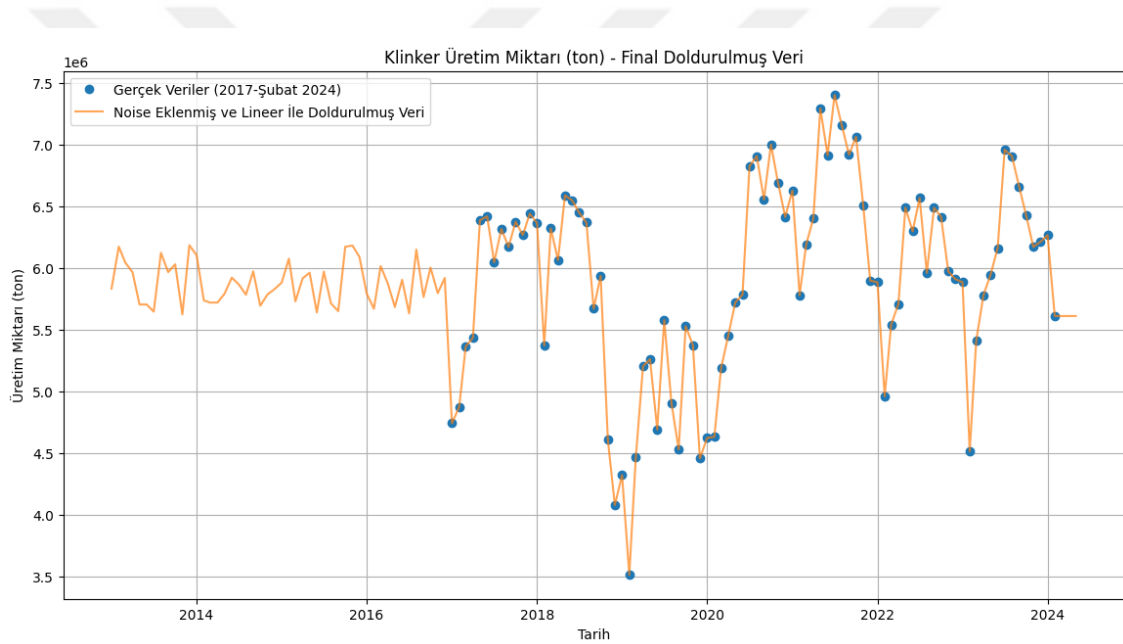


Şekil 4.4. Çimento Üretim Miktarı (ton) Özniteliği içi Yüzeysel Eksik Veri Doldurma Grafiği

Çimento Üretim Miktarı (ton) özniteliği için veri setinde 2013–2016 dönemine ait veriler tamamen eksik durumdaydı. Dört yıllık süre zarfında hiçbir gözlem mevcut olmadığından, doğrudan zaman serisi interpolasyonuna dayanan yöntemler yerine daha kontrollü ve gerçekçiliği artırmaya yönelik bir strateji benimsenmiştir. Bu doğrultuda, Şekil 4.4'te görüldüğü üzere söz konusu dönemdeki eksik aylık veriler için parazit eklenmiş ortalama yöntemi uygulanmıştır. Bu yöntemde, gözlemlerin mevcut olduğu dönemlerdeki

ortalama değeri esas alınmış, ardından bu ortalamaya düşük varyansa sahip rassal sapmalar eklenerek veri setine doğal bir varyasyon kazandırılmıştır. Böylece hem yapay bir sabitlikten kaçınılmış hem de zaman serisinin istatistiksel karakteristiği korunmuştur (Schafer, 1997). Özniteliğin tamamen boş olduğu bir dönemde spline ya da lineer interpolasyon gibi tekniklerin uygulanması, elde yeterli veri bulunmaması nedeniyle aşırı spekülasyon ve güvenilirmez sonuçlar doğurabileceğinden bu yöntemler bilinçli olarak tercih edilmemiştir.

Klinker Üretim Miktarı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.5'te sunulmuştur. Öznitelige ilişkin toplam satır = 137, mevcut veri sayısı = 86, eksik veri sayısı = 51 şeklindedir.

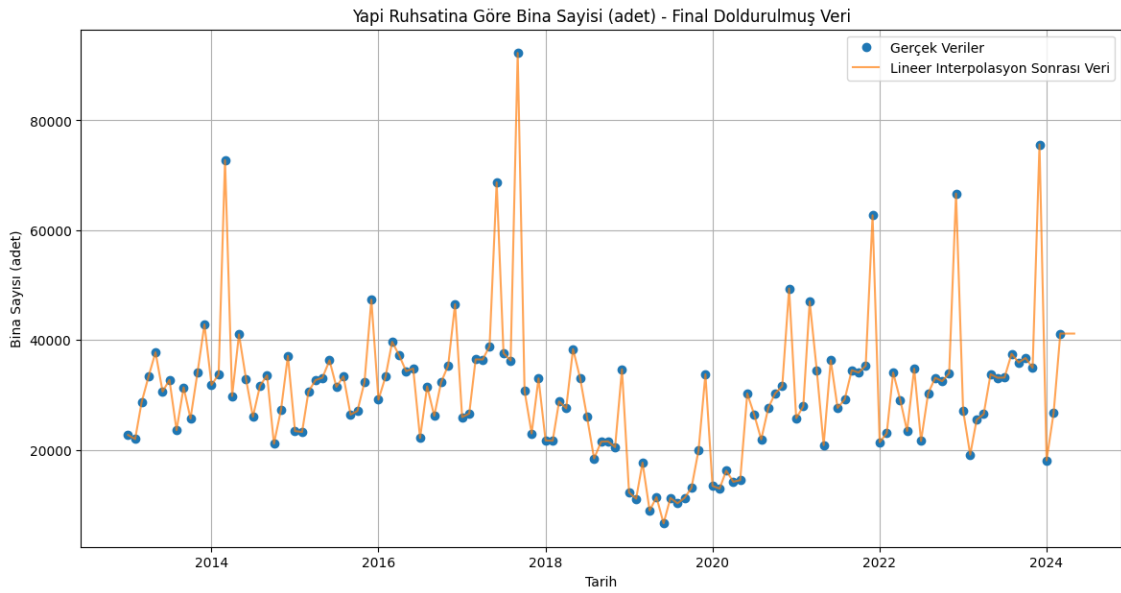


Şekil 4.5. Klinker Üretim Miktarı (ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Klinker Üretim Miktarı (ton) özniteliği, veri setinin 2013–2016 döneminde tamamen boş durumdadır. Bu eksiklik, ilgili dönemde herhangi bir gözlem bulunmaması nedeniyle, zaman serisi interpolasyon yöntemlerinin doğrudan uygulanmasına elverişli değildir. Bu nedenle Şekil 4.5'te görüldüğü üzere, ilgili öznitelik için parazit eklenmiş ortalama yöntemi tercih edilmiştir. Bu yöntem kapsamında, gözlemlerin mevcut olduğu 2017 ve sonrası dönemlerdeki klinker üretim verilerinin ortalaması esas alınmış, ardından bu ortalama etrafında düşük varyansa sahip rassal sapmalar oluşturularak eksik değerler üretilmiştir. Bu yaklaşım, verinin yapay bir sabitlikten uzak durarak zamansal doğal değişkenliğini

korumasını sağlamıştır (Schafer, 1997). Ayrıca, klinker üretimi ile çimento üretimi arasında gözlenen yapısal ve korelasyonel benzerlik dikkate alınarak, Çimento Üretim Miktarı (ton) özniteliğinde uygulanan yöntemin burada da geçerli olduğu değerlendirilmiştir. Bu sayede metodolojik tutarlılık korunmuştur.

Yapı Ruhsatına Göre Bina Sayısı (adet) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.6'da sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı= 135, eksik veri sayısı = 2 şeklindedir.

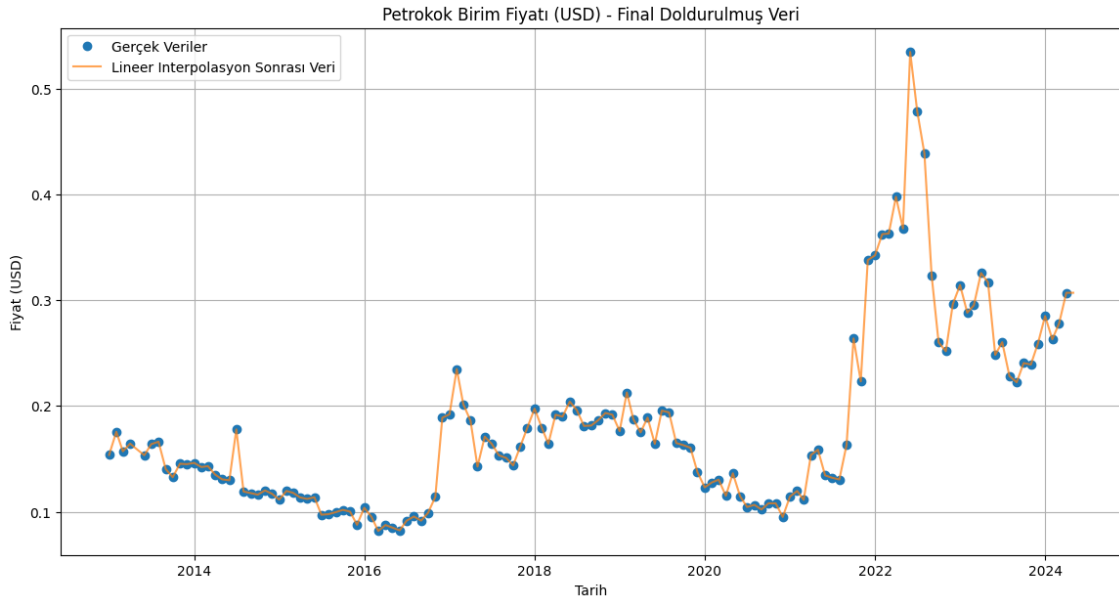


Şekil 4.6. Yapı Ruhsatına Göre Bina Sayısı (adet) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Yapı Ruhsatına Göre Bina Sayısı (adet) özniteliğinde yalnızca Nisan ve Mayıs 2024 aylarına ait olmak üzere iki eksik gözlem tespit edilmiştir. Bu sınırlı eksiklik karşısında, zaman serisinin genel eğilimini bozmayacak ve doğrudan gözlemler arasındaki geçişi temsil edebilecek nitelikte bir yöntem olarak lineer interpolasyon tekniği tercih edilmiştir. Söz konusu eksikliklerin zamansal bağlamda ardışık ve düşük hacimli olması, bu yöntemin yeterli doğrulukta sonuçlar sunacağına işaret etmiştir. Ayrıca eksik noktalar, istatistiksel olarak serinin genel trendinin parçası kabul edilebilecek pozisyonda yer aldığından, doğrusal geçişin güvenilirliği desteklenmiştir (Little ve Rubin, 2019). Şekil 4.6'da görüldüğü üzere, eksik gözlemler, çevreleyen verilerle uyumlu biçimde, doğallığı bozmadan doldurulmuştur. Daha karmaşık yöntemler, özellikle spline interpolasyon gibi teknikler, yalnızca iki veri

noktasının eksik olduğu bu gibi durumlarda gereksiz model karmaşıklığına neden olabileceğinden tercih edilmemiştir.

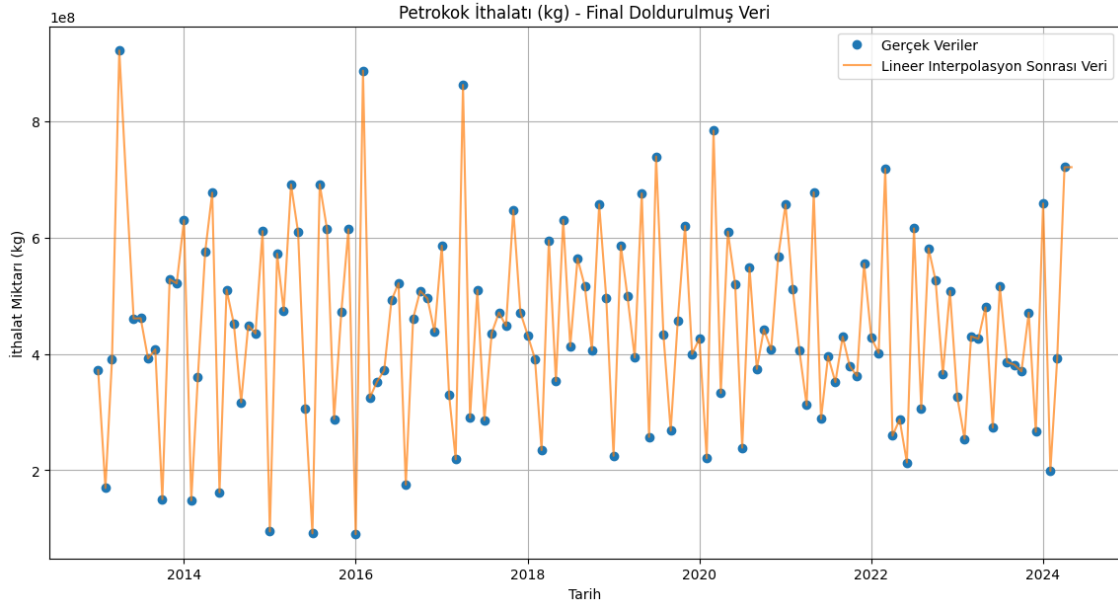
Petrokok Birim Fiyatı (USD) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.7’de sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı= 135, eksik veri sayısı = 2 şeklindedir.



Şekil 4.7. Petrokok Birim Fiyatı (USD) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Petrokok Birim Fiyatı (USD) özniteliğinde yalnızca iki veri noktasında (Mayıs 2013 ve Mayıs 2024) eksiklik tespit edilmiştir. Bu sınırlı eksikliğin giderilmesi amacıyla lineer interpolasyon yöntemi uygulanmıştır. Petrokok fiyatlarının zaman içinde görece istikrarlı ve düzenli bir artış eğilimi sergilemesi, doğrusal interpolasyonun bu bağlamda yeterli ve ekonomik bir çözüm sunabileceğini göstermiştir. Ayrıca, fiyat verilerinin doğası gereği ani sıçramalar yerine daha öngörülebilir dalgalanmalar barındırması, bu yöntemin uygunluğunu desteklemiştir. Şekil 4.7’de görüldüğü üzere, eksik değerler, çevreleyen gözlemlerle tutarlı ve eğilimsel sürekliliği koruyacak şekilde hesaplanmıştır. Alternatif olarak değerlendirilen spline interpolasyon yöntemi ise, özellikle kısa dönemli eksikliklerde gereğinden fazla oynaklık yaratma riski taşıdığından, bu çalışmada uygun bulunmamıştır.

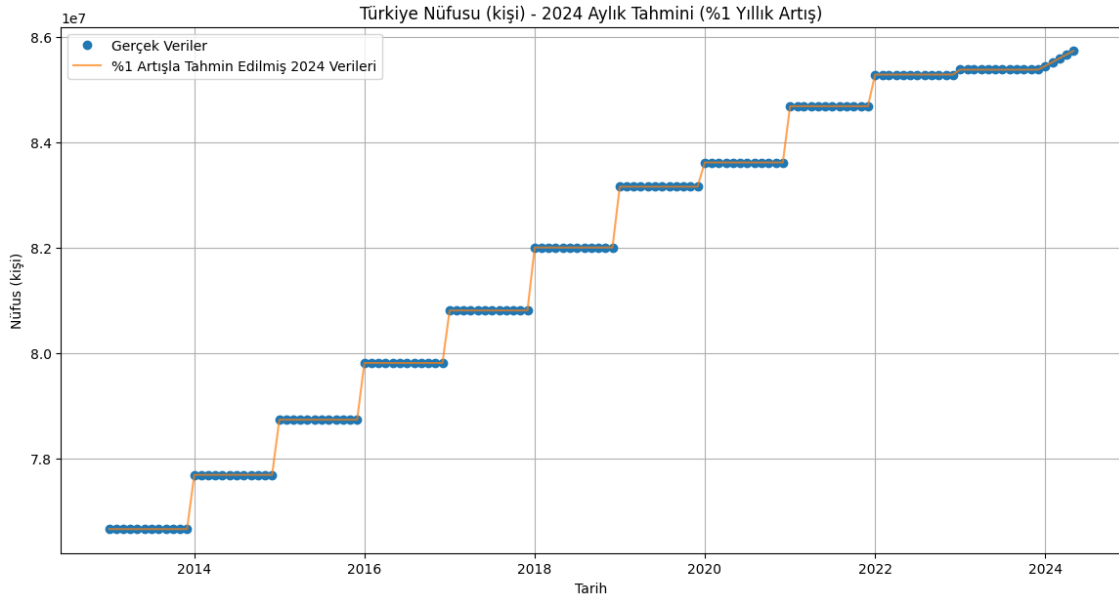
Petrokok İthalatı (kg) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.8’de sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı= 135, eksik veri sayısı = 2 şeklindedir.



Şekil 4.8. Petrokok İthalatı (kg) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Petrokok İthalatı (kg) özniteliğinde yalnızca iki gözlem noktasında eksiklik belirlenmiş ve bu eksikliklerin giderilmesi için lineer interpolasyon yöntemi tercih edilmiştir. İthalat verilerinin zaman içinde görece düzenli ve öngörülebilir bir eğilim sergilemesi, bu yöntemin güvenle uygulanmasına imkân tanımıştır. Ayrıca, eksik verilerin sınırlı sayıda olması ve serinin genel akışıyla yapısal uyum göstermesi, doğrusal interpolasyonun doğruluk açısından yeterli bir çözüm olduğunu ortaya koymuştur. Şekil 4.8’de yer verilen grafik, interpolasyon sonrası elde edilen veri yapısının, zaman serisinin doğal trendiyle bütünlük arz ettiğini açıkça ortaya koymaktadır. Daha karmaşık ve eğrisel yöntemlerin uygulanması bu ölçekte gereksiz karmaşıklık yaratacağından, hesaplama ekonomisi ve model sadeliği ilkesi doğrultusunda lineer interpolasyon yöntemi tercih edilmiştir.

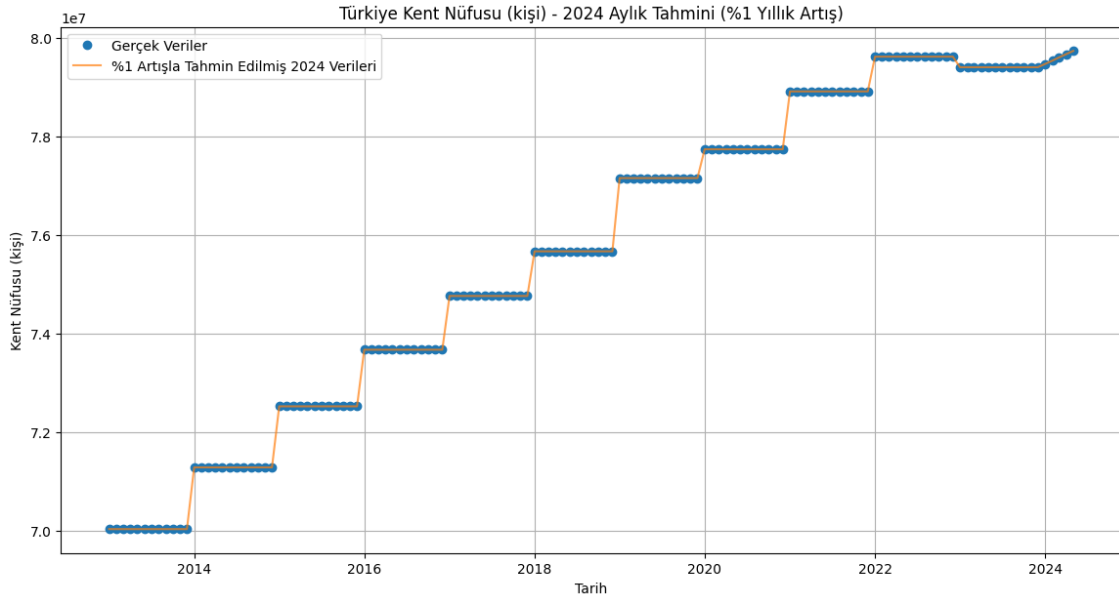
Türkiye Nüfusu (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.9’da sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı= 132, eksik veri sayısı = 5 şeklindedir.



Şekil 4.9. Türkiye Nüfusu (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Türkiye Nüfusu (kişi) özniteliği, veri setinde yalnızca yıllık bazda sağlanmış olup, her yılın tamamında aynı değerin tekrarlanması yoluyla aylık forma dönüştürülmüştür. Bununla birlikte, çalışmanın gerçekleştirildiği süre zarfında Ocak 2024 – Mayıs 2024 dönemine ilişkin herhangi bir veri yayımlanmadığından, söz konusu dönemde eksik olan aylık gözlemler, Türkiye'nin tarihsel nüfus verileri doğrultusunda yaklaşık %1 oranında yıllık büyüme varsayımıyla tahmin edilmiştir. Türkiye İstatistik Kurumu (2023) verilerine göre ülke nüfusu uzun yıllardır düzenli bir artış eğilimi göstermektedir. Bu eğilimi yansıtan yıllık büyüme oranı, aylık bazda kümülatif olarak dağıtılmış ve böylece demografik değişimin doğal yapısını daha doğru biçimde yansıtan bir veri üretimi gerçekleştirilmiştir. Şekil 4.9'da görüldüğü üzere, bu yöntemle elde edilen değerler, geçmiş yılların trendine paralel olarak tutarlı bir dağılım sergilemiştir. Lineer interpolasyon gibi yöntemler ise nüfus artışının doğasındaki kümülatif yapıyı ve yıllık büyüme etkisini temsil etme konusunda yetersiz kalacağından tercih edilmemiştir.

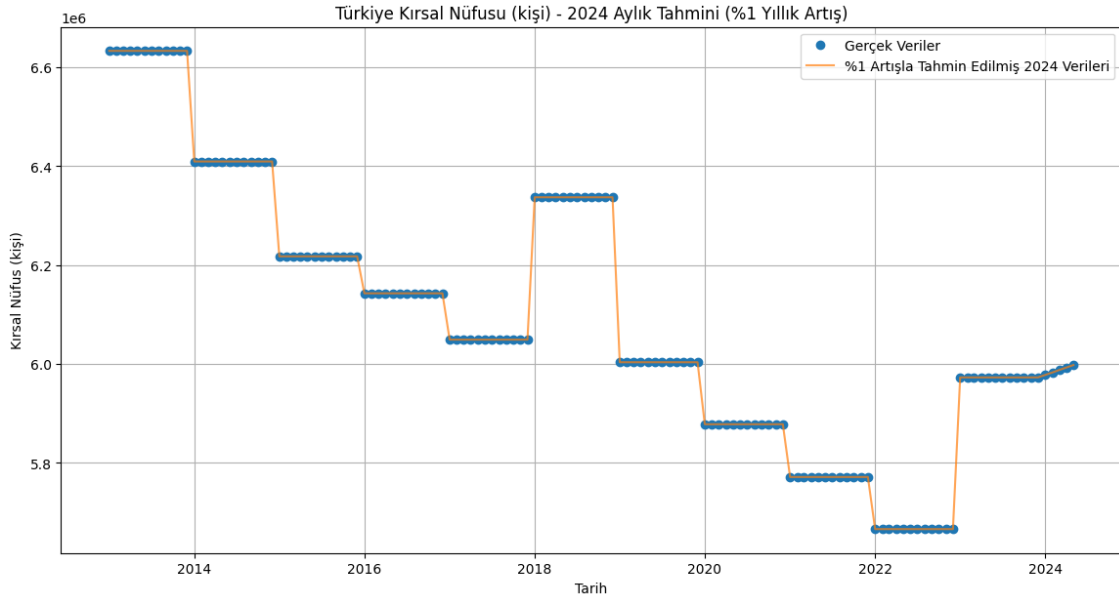
Türkiye Kent Nüfusu (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.10'da sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı = 132, eksik veri sayısı = 5 şeklindedir.



Şekil 4.10. Türkiye Kent Nüfusu (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Türkiye Kent Nüfusu (kişi) özniteliği, diğer demografik göstergelerde olduğu gibi yalnızca yıllık bazda sağlanmış ve ilgili yılların tamamı boyunca sabit değerlerin atanması suretiyle aylık düzleme uyarlanmıştır. Çalışmanın gerçekleştirildiği süre zarfında Ocak 2024 – Mayıs 2024 dönemine ilişkin resmi veriler henüz yayımlanmadığı için bu dönem eksik gözlemlerle temsil edilmiştir. Eksik aylık değerlerin tahmin edilmesinde, Türkiye’de son yıllarda süregelen kentleşme eğilimi ve nüfus dinamikleri dikkate alınarak, yıllık bazda yaklaşık %1 oranında bir artış varsayımıyla kümülatif bir büyüme tahmini yapılmıştır. Türkiye gibi gelişmekte olan ve şehirleşme oranı yüksek ülkelerde kent nüfusunun yıllık düzeyde artış gösterdiği bilinmektedir. Bu eğilim doğrultusunda oluşturulan tahmin değerleri, geçmiş yıllarla tutarlı bir eğilim sergilemiş ve Şekil 4.10’da görüldüğü üzere, zaman serisinin genel trendiyle uyum içinde olmuştur. Alternatif olarak değerlendirilen lineer interpolasyon yöntemi ise nüfus değişiminin doğal kümülatif yapısını yansıtmaktan uzak kaldığı ve eğilimsel artışı düz bir çizgiye indirgediği için uygun bulunmamıştır.

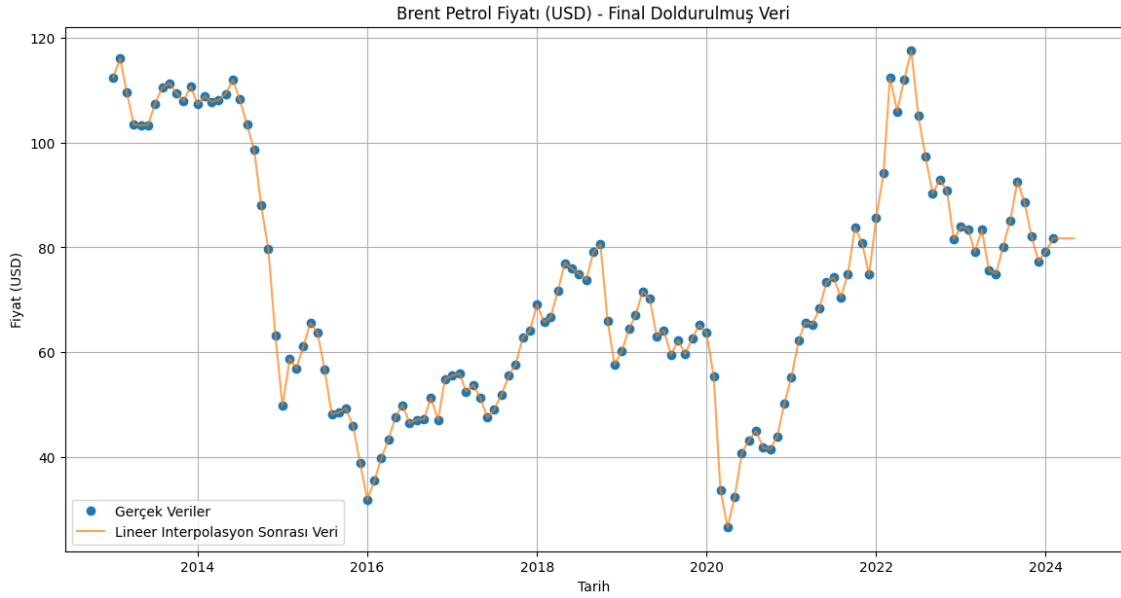
Türkiye Kırsal Nüfusu (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.11’de sunulmuştur. Öznitelige ilişkin toplam satır = 137, mevcut veri sayısı= 132, eksik veri sayısı = 5 şeklindedir.



Şekil 4.11. Türkiye Kırsal Nüfusu (kişi) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Türkiye Kırsal Nüfusu (kişi) özniteliği, yalnızca yıllık bazda yayımlanmakta olup, her yılın tüm aylarına sabit değerlerin atanmasıyla aylık veri formatına dönüştürülmüştür. Ocak 2024 – Mayıs 2024 dönemine ait eksik gözlemler içinse, yıllık yaklaşık %1 oranında bir artış varsayımıyla tahmin yapılmıştır. Her ne kadar kırsal nüfus, Türkiye özelinde uzun vadede durağan ya da azalan bir eğilim gösterebilse de bu çalışmada veri yapısal bütünlüğünü ve tahmin tutarlılığını koruyabilmek adına genel nüfus artış oranıyla uyumlu bir yaklaşım benimsenmiştir. Bu sayede hem metodolojik tutarlılık sağlanmış hem de eksik dönemlerdeki değerler, zaman serisinin genel eğilimine zarar vermeyecek biçimde tahmin edilmiştir. Şekil 4.11’de görüldüğü üzere, kümülatif artış varsayımıyla üretilen değerler, geçmiş dönemlerle yapısal uyum içerisinde bir dağılım sergilemiştir. Alternatif olarak düşünülebilecek daha karmaşık interpolasyon yöntemleri, kırsal nüfusun görece düşük değişkenliğine kıyasla fazla duyarlılık meydana getirebileceği için bu kapsamda tercih edilmemiştir.

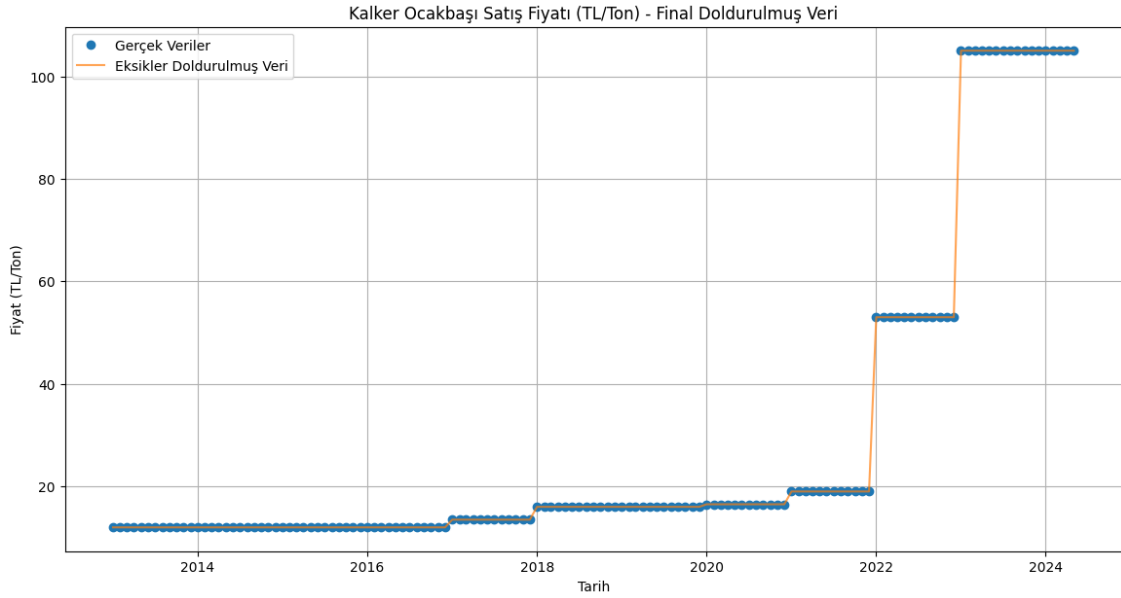
Brent Petrol (USD) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.12’de sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı= 134, eksik veri sayısı = 3 şeklindedir.



Şekil 4.12. Brent Petrol (USD) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Brent Petrol (USD) özniteliği, veri setinde yalnızca Mart 2024 – Mayıs 2024 dönemini kapsayan son üç ayda eksik gözlemler içermektedir. Bu kısa dönemli eksikliklerin giderilmesi amacıyla lineer interpolasyon yöntemi uygulanmıştır. Brent petrol fiyatlarının finansal piyasalarda günlük bazda yüksek oynaklık gösterebildiği bilinmekle birlikte, bu çalışmada kullanılan aylık ortalamalar, daha istikrarlı bir trend sunmaktadır. Bu nedenle, eksik verilerin, kendisinden önceki ve sonraki aylara ait ortalama fiyatlar esas alınarak doğrusal bir şekilde tahmin edilmesi, serinin genel yapısını bozmadan tutarlı bir tamamlamaya olanak tanımıştır. Şekil 4.12’de görüldüğü üzere, interpolasyonla elde edilen değerler, serinin genel eğilimiyle uyum içinde bir görünüm sunmuştur. Alternatif olarak değerlendirilen spline interpolasyon yöntemi, özellikle kısa dönemli finansal serilerde gereksiz dalgalanmalar meydana getirme riski taşıdığı için tercih edilmemiştir. Bu bağlamda, lineer interpolasyon hem metodolojik sadelik hem de doğruluk açısından uygun bulunmuştur.

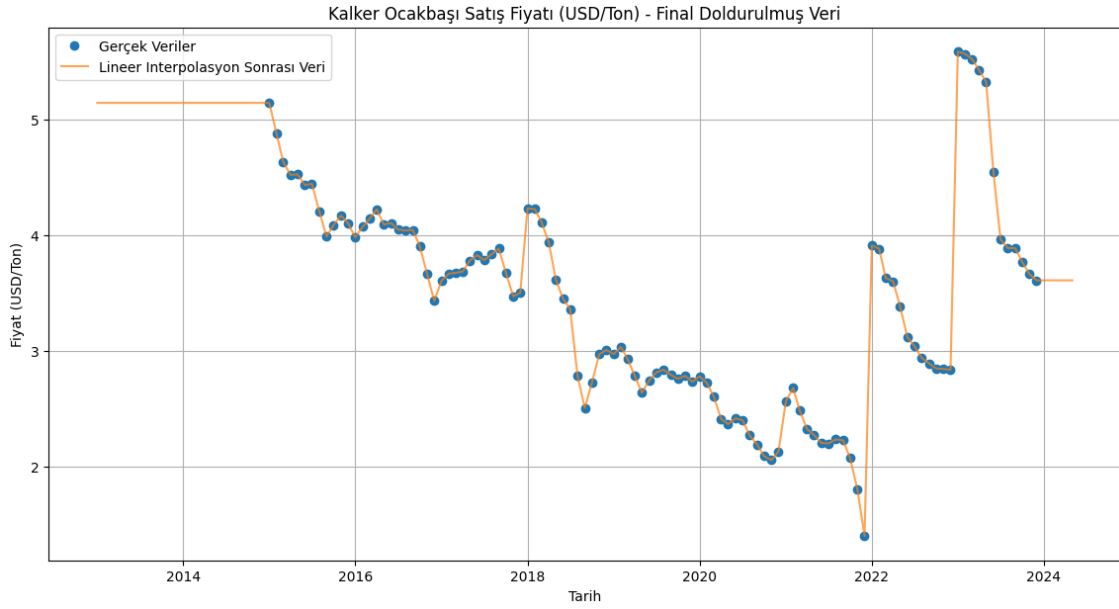
Kalker Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.13’te sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı = 108, eksik veri sayısı = 29 şeklindedir.



Şekil 4.13. Kalker Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Kalker Ocakbaşı Satış Fiyatı (TL/Ton) özniteliği, veri setinde bazı yıllarda eksik gözlemler içermekle birlikte, verinin mevcut olduğu dönemlerde fiyatların yıl boyunca sabit kaldığı gözlemlenmiştir. Bu yapısal özelliği dikkate alarak, eksik gözlemlerin bulunduğu yıllarda da yıl içi sabit değer tahmini yöntemi uygulanmıştır. Bu yöntemde, her bir eksik yıl için tek bir fiyat tahmini üretilmiş ve yılın tüm aylarına bu değer yansıtılmıştır. Bu yaklaşım, fiyatların yıllık bazda belirlendiği ve yıl içinde güncellenmediği düşünüldüğünde, veri yapısının doğrallığını koruyan tutarlı bir çözüm sunmuştur. Şekil 4.13'te yer alan grafik de bu yöntemle elde edilen değerlerin, geçmiş yılların fiyat davranışıyla bütünlük arz ettiğini göstermektedir. Alternatif olarak değerlendirilen lineer veya spline interpolasyon yöntemleri, yıl içinde sabit kalan fiyat yapısını bozma ve yapay dalgalanmalar üretme riski taşıdığından bu çalışmada uygun bulunmamıştır.

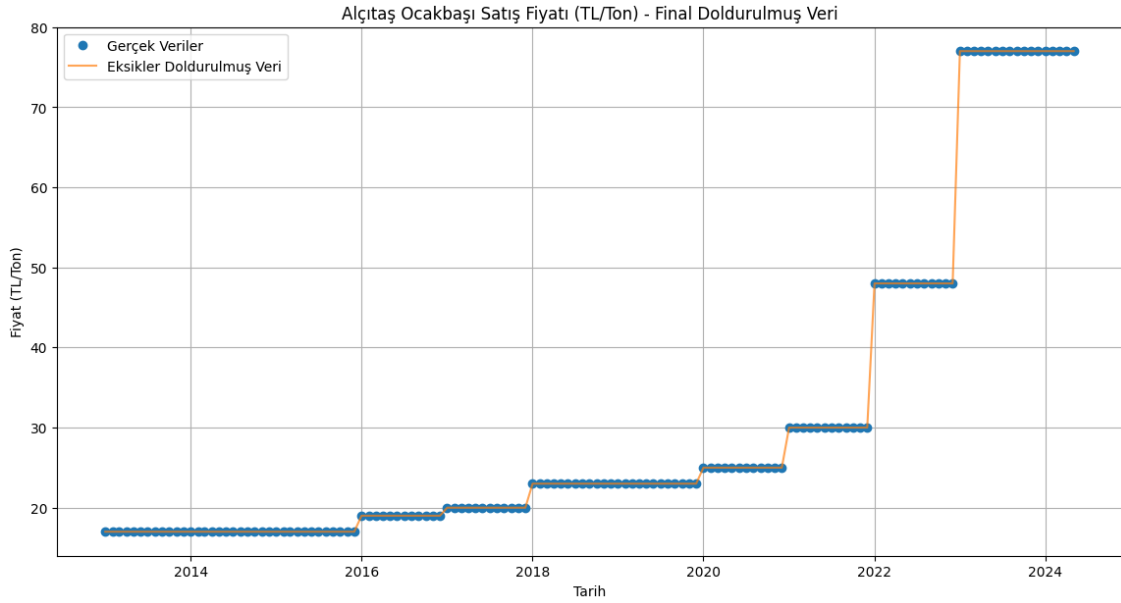
Kalker Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.14'te sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı = 108, eksik veri sayısı = 29 şeklindedir.



Şekil 4.14. Kalker Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Kalker Ocakbaşı Satış Fiyatı (USD/Ton) özniteliği, bazı yıllarda eksik gözlemler içermektedir. TL cinsinden fiyatların yıl içi sabitlik gösterdiği dönemlerde dahi, döviz kuru dalgalanmalarının etkisiyle USD bazında fiyatlar doğal olarak yıllık değişkenlik sergilemektedir. Bu nedenle, eksik yıllara ait gözlemler için lineer interpolasyon yöntemi uygulanmıştır. Bu yöntem sayesinde, eksik yıllara ait fiyatlar, önceki ve sonraki yılların eğilimine uygun biçimde tahmin edilmiş; böylece döviz kuru kaynaklı doğal dalgalanma yapısı korunmuştur. Şekil 4.14'te görüldüğü üzere, lineer interpolasyonla elde edilen değerler, geçmiş yıllardaki eğilimlerle tutarlı bir yapıda olup yapay düzleşmeler içermemektedir. Alternatif olarak değerlendirilen sabit değer atama yöntemi ya da yıl içi sabitlik varsayımı, USD bazlı fiyat serisinin yapısal gerçekliği ile çelişeceği için uygun görülmemiştir. Spline interpolasyon ise sınırlı sayıda veri noktasının olduğu yıllık serilerde aşırı dalgalanmalara neden olabileceğinden tercih edilmemiştir.

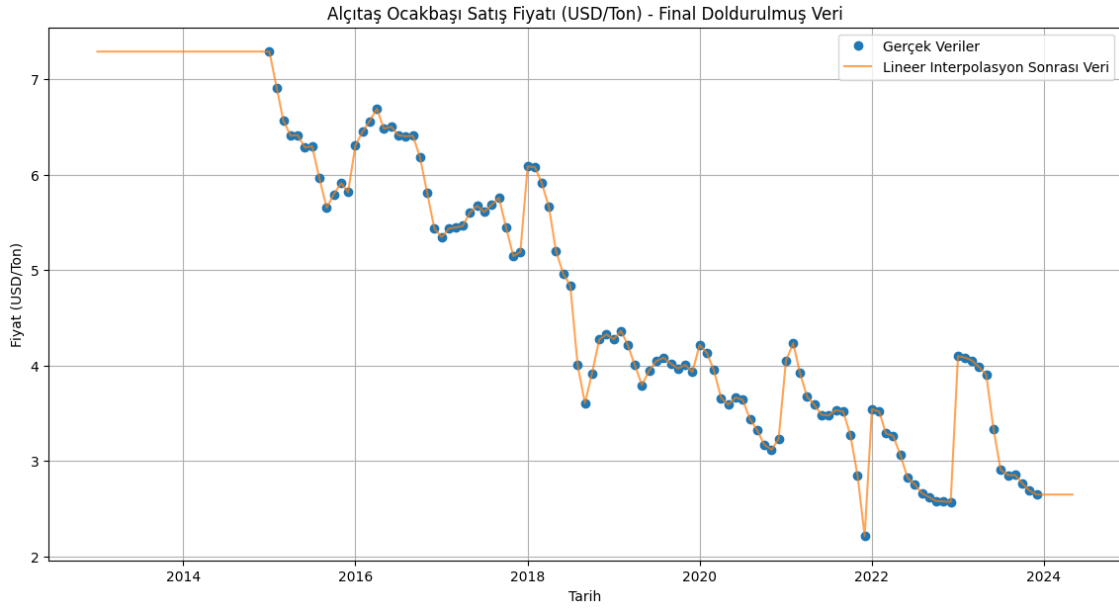
Alçıtaşı Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.15'te sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı = 108, eksik veri sayısı = 29 şeklindedir.



Şekil 4.15. Alçıtaş Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Alçıtaş Ocakbaşı Satış Fiyatı (TL/Ton) özniteliği, bazı yıllarda eksik gözlemler içermekle birlikte, veri setinde mevcut olan dönemlerde bu fiyatların yıl boyunca sabit kaldığı açık biçimde gözlemlenmiştir. Bu yapısal özelliğe uygun olarak eksik yıllar için yıl içi sabit değer tahmini yöntemi uygulanmıştır. Bu yöntemde, her eksik yıl için tek bir fiyat değeri tahmin edilmiş ve yılın tüm aylarına bu sabit değer atanmıştır. Böylece, fiyat serisinin doğal yapısı korunmuş, yıl içi oynaklık yaratılmadan eksik veriler güvenilir biçimde tamamlanmıştır. Şekil 4.15'te görüldüğü üzere, eksik veri doldurma sonrası elde edilen yapı, geçmiş yıllardaki sabit fiyat eğilimleriyle tam bir uyum içerisindedir. Alternatif olarak değerlendirilebilecek interpolasyon yöntemleri, özellikle yıl boyunca sabit kalan fiyat serilerinde yapay geçişler ve anlamsız dalgalanmalar üretebileceği için tercih edilmemiştir.

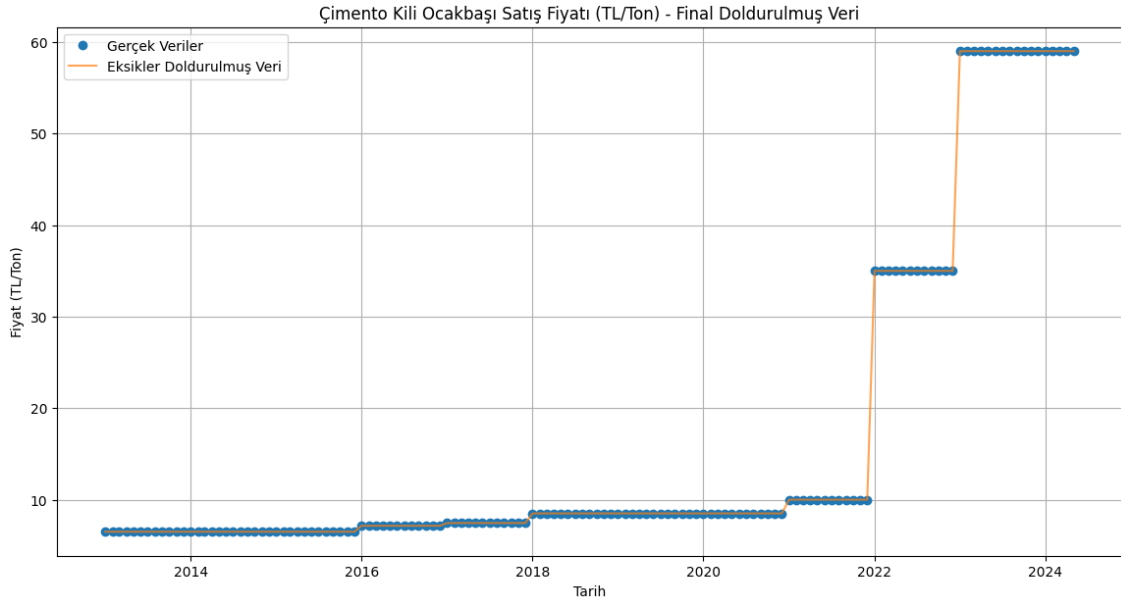
Alçıtaş Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.16'da sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı = 108, eksik veri sayısı = 29 şeklindedir.



Şekil 4.16. Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) özniteliğinde, bazı yıllarda eksik gözlemler tespit edilmiştir. TL bazlı satış fiyatları yıl boyunca sabit bir yapı gösterse de USD cinsinden değerlere döviz kuru etkisi yansıdığından bu seride doğal yıllık dalgalanmalar oluşmaktadır. Bu nedenle, eksik yıllara ait verilerin tahmini amacıyla lineer interpolasyon yöntemi tercih edilmiştir. Bu yaklaşım, USD bazlı fiyatlarda gözlemlenen eğilimsel değişimi yapay oynamalara yol açmadan sürdürmeyi mümkün kılmıştır. Şekil 4.16’da görüleceği üzere lineer interpolasyon sayesinde eksik yıllardaki değerler, çevreleyen yılların eğilimiyle tutarlı bir geçiş oluşturarak serinin yapısal sürekliliğini korumuştur. Yıl içi sabit değer atama yöntemi, döviz etkisini göz ardı edeceği için uygun bulunmamış; spline interpolasyon ise sınırlı veri noktaları üzerinde aşırı duyarlılıkla gereksiz dalgalanmalara neden olabileceği gerekçesiyle tercih edilmemiştir. Eksik veri doldurma işleminin görsel çıktısına yer verilmiştir.

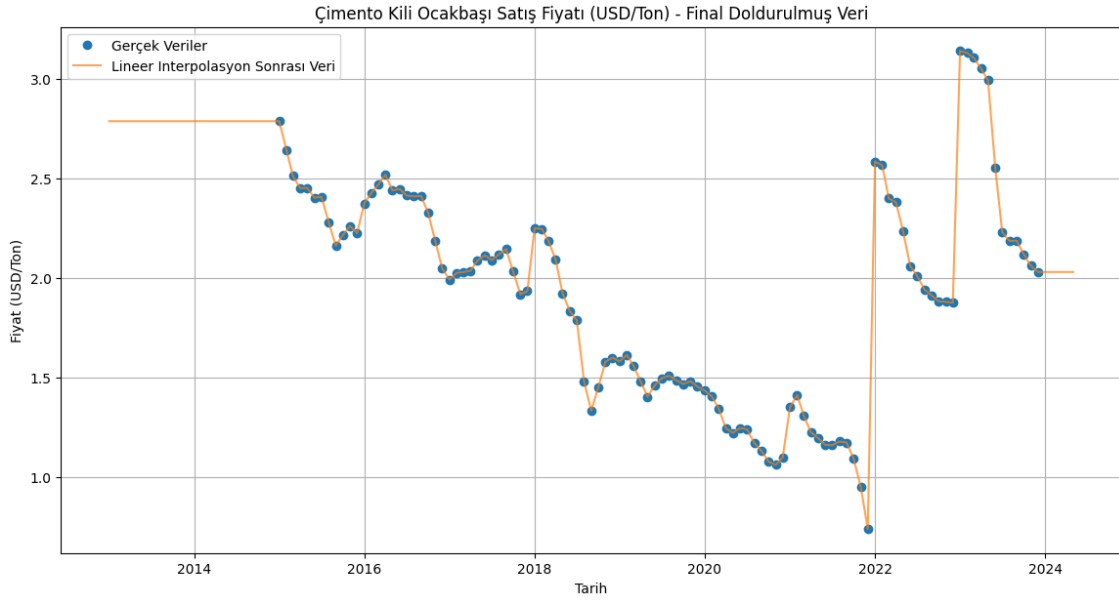
Çimento Kili Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.17’de sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı= 108, eksik veri sayısı = 29 şeklindedir.



Şekil 4.17. Çimento Kili Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Çimento Kili Ocakbaşı Satış Fiyatı (TL/Ton) özniteliği, bazı yıllarda eksik gözlemler içermekle birlikte, mevcut verilerin incelenmesi sonucunda bu fiyatların yıl boyunca sabit kaldığı tespit edilmiştir. Bu bağlamda, eksik yıllarda da fiyatların yıl içinde değişmediği varsayımıyla yıl içi sabit değer tahmini yöntemi uygulanmıştır. Her eksik yıl için tek bir fiyat değeri tahmin edilmiş ve bu değer yılın tüm aylarına aynı şekilde atanmıştır. Bu yöntem, fiyatların yapısal sabitliğini ve sektörel belirleme mekanizmalarını bozmadan veri setine bütünlük kazandırmayı mümkün kılmıştır. Şekil 4.17’de görüldüğü üzere, eksik veri doldurma sonrası elde edilen fiyat eğrisi, mevcut yılların yapısıyla tam bir tutarlılık içindedir. Lineer veya spline interpolasyon gibi alternatif yöntemler ise yıl içi sabitlik özelliğini zedeleyebileceği ve yapay dalgalanmalar üretebileceği gerekçesiyle bu çalışma kapsamında uygun bulunmamıştır.

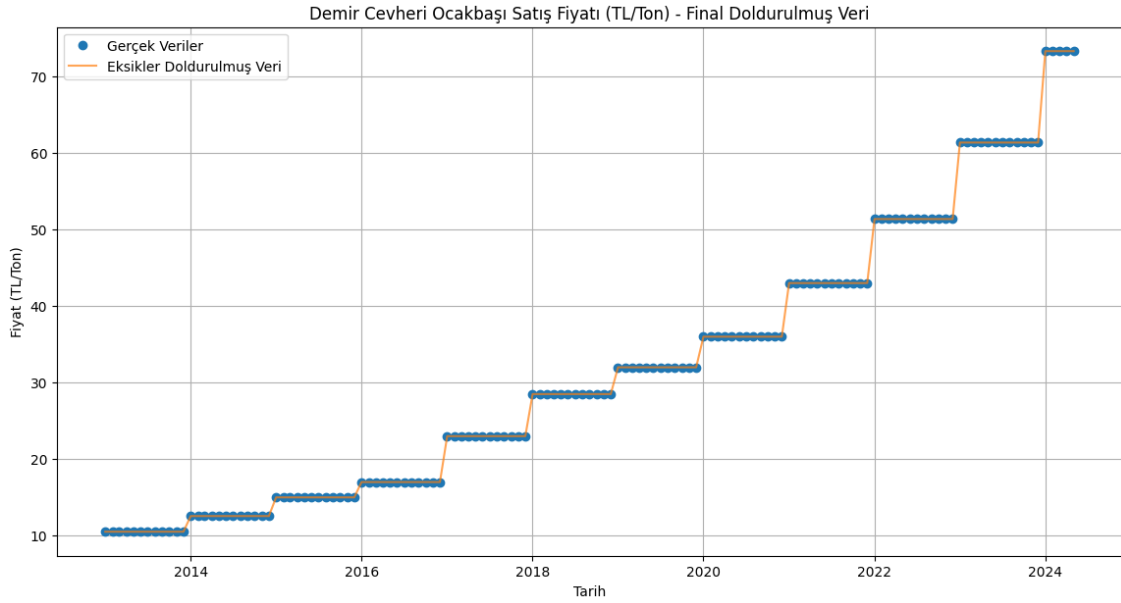
Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.18’de sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı = 108, eksik veri sayısı = 29 şeklindedir.



Şekil 4.18. Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) özniteliğinde, bazı yıllarda eksik veriler bulunmakta olup, bu eksikliklerin giderilmesi için lineer interpolasyon yöntemi uygulanmıştır. Fiyatların TL cinsinden değerlendirildiği dönemlerde yıl içi sabitlik gözlemlenmiş olsa da USD cinsinden fiyatlar döviz kuru etkisine bağlı olarak yıllık bazda doğal dalgalanmalar göstermektedir. Bu nedenle, eksik yıllarda USD bazlı fiyatların doğrusal geçişlerle tahmin edilmesi, hem döviz kuru kaynaklı varyasyonu yansıtmakta hem de serinin yapısal tutarlılığını korumaktadır. Şekil 4.18’de yer alan grafik, interpolasyonla elde edilen değerlerin geçmiş yılların eğilimleriyle örtüştüğünü göstermektedir. Sabit değer atama yöntemi bu tür dinamik serilerde yapay düzleşmelere neden olabileceği için tercih edilmemiş; spline interpolasyon ise sınırlı yıllık veri noktaları üzerinde aşırı oynaklık riski taşıdığı gerekçesiyle uygun bulunmamıştır.

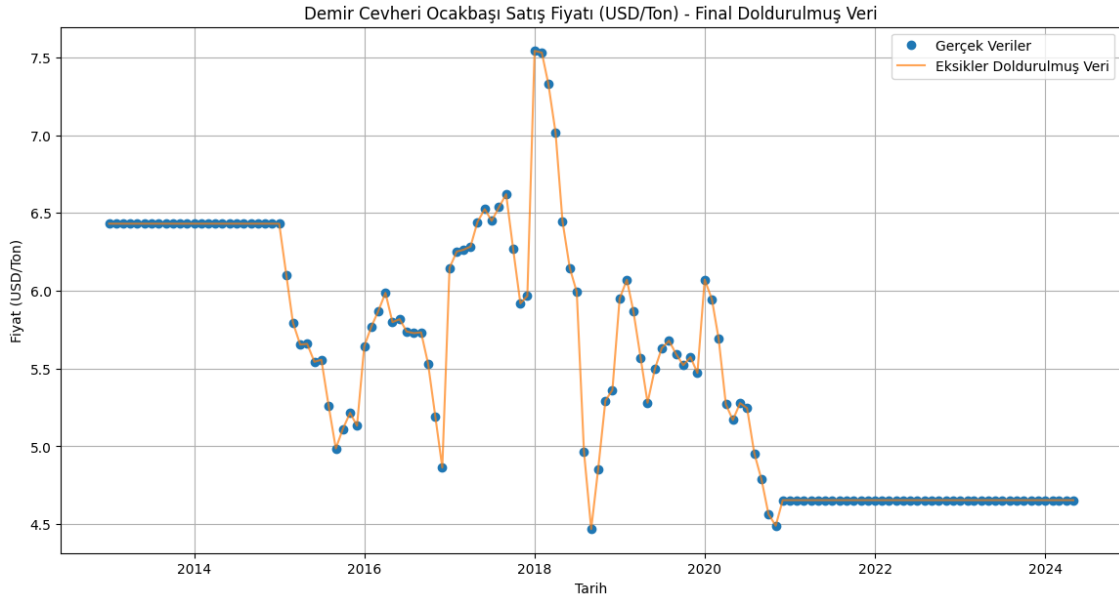
Demir Cevheri Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.19’da sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı = 72, eksik veri sayısı = 65 şeklindedir.



Şekil 4.19. Demir Cevheri Ocakbaşı Satış Fiyatı (TL/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Demir Cevheri Ocakbaşı Satış Fiyatı (TL/Ton) özniteliği, özellikle 2021 – 2024 döneminde eksik veri içermekte olup, bu eksikliklerin giderilmesi için büyüme/küçülme katsayısı ile zincirleme tahmin yöntemi uygulanmıştır. Bu yöntemde, geçmiş yıllarda gözlemlenen artış oranları temel alınarak her yıl için belirli bir büyüme katsayısı hesaplanmış ve bu katsayı kullanılarak eksik yıllara ait tahmini değerler türetilmiştir. Böylece, ekonomik göstergelere özgü kümülatif artış-düşüş eğilimi korunarak, verinin serisel yapısı yapay bozulmalardan arındırılmış şekilde tamamlanmıştır. Bu yaklaşım, özellikle makroekonomik ve sektörel fiyat verilerinde önerilen istatistiksel eksik veri tamamlama tekniklerinden biridir (Little ve Rubin, 2019). Şekil 4.19’da yer alan grafik, bu yöntemle elde edilen tahmini değerlerin geçmiş yıllardaki fiyat trendleriyle yapısal uyum sağladığını göstermektedir. Alternatif olarak değerlendirilen lineer interpolasyon yöntemi ise uzun dönemli eksik verilerde doğrusal sapmalara ve eğilim dışı tahminlere neden olabileceği için uygun bulunmamıştır.

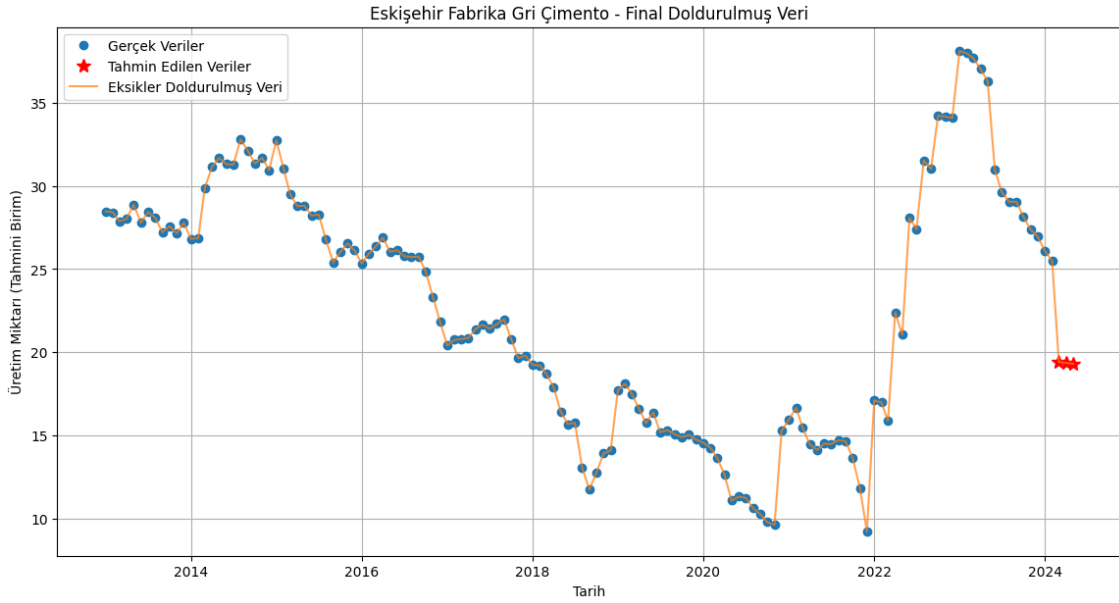
Demir Cevheri Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.20’de sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı= 72, eksik veri sayısı = 65 şeklindedir.



Şekil 4.20. Demir Cevheri Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliği için Yüzeysel Eksik Veri Doldurma Grafiği

Demir Cevheri Ocakbaşı Satış Fiyatı (USD/Ton) özniteliği, 2021 – 2024 döneminde eksik veri içermekte olup, bu eksikliklerin giderilmesinde sabit fiyat tahmini yöntemi uygulanmıştır. Bu tercih, 2015 – 2020 yılları arasındaki gözlemler incelendiğinde USD bazında fiyatların yıllar içinde önemli bir değişim göstermediği, serinin görece olarak durağan bir yapı sergilediği tespitine dayanmaktadır. Bu nedenle, son yıllara ait eksik değerlerin sabit bir fiyat üzerinden tahmin edilmesi, verinin doğallığını koruyan, ekonomik açıdan anlamlı ve yapay oynaklık yaratmayan bir yaklaşım olarak değerlendirilmiştir. Sabit fiyat tahmini yöntemi, özellikle düşük varyanslı ekonomik göstergelerde veri bütünlüğünü sürdürmek için yaygın biçimde önerilen tekniklerden biridir. Lineer ya da spline interpolasyon gibi yöntemler, bu tür durağan serilerde gereksiz varyasyon yaratabileceği ve trend dışı sapmalara neden olabileceği için bu çalışmada uygun bulunmamıştır. Şekil 4.20’de sabit fiyat tahminine dayalı veri üretiminin sonuçları grafiksel olarak sunulmuştur.

Eskişehir Fabrikası Gri Çimento Satın Alım Fiyatı Hedef Değişkeni için Yüzeysel Eksik Veri Doldurma Grafiği Şekil 4.21’de sunulmuştur. Özniteliğe ilişkin toplam satır = 137, mevcut veri sayısı = 134, eksik veri sayısı = 3 şeklindedir.



Şekil 4.21. Eskişehir Fabrikası Gri Çimento Satın Alım Fiyatı Hedef Değişkeni için Yüzeysel Eksik Veri Doldurma Grafiği

Eskişehir Fabrikası Gri Çimento Satın Alım Fiyatı, bu çalışmada tahmin edilmesi amaçlanan hedef değişken olup, Ocak 2013 – Şubat 2024 döneminde eksiksiz olarak veri setinde yer almaktadır ancak Mart 2024 – Mayıs 2024 aralığında gözlemler mevcut değildir. Şekil 4.21’de görüleceği üzere bu kısa vadeli eksikliklerin giderilmesi için geçmiş yıllardaki fiyat trendleri dikkate alınarak trend analizi gerçekleştirilmiş ve buna paralel olarak küçük oranlı büyüme tahmini yöntemi uygulanmıştır. Bu yöntemle, serinin doğal eğilimi ve sektörel mevsimselliği korunarak eksik gözlemler, geçmişe dayalı doğrusal olmayan ancak kontrollü bir artış mantığıyla tahmin edilmiştir. Bu yaklaşım sayesinde, tahmin edilen değerler zaman serisinin mevcut yapısıyla uyumlu olacak şekilde tamamlanmış ve hedef değişkende yapay sıçramalara ya da düzleşmelere neden olunmamıştır. Alternatif olarak değerlendirilen sabit değer atama veya interpolasyon teknikleri, kısa dönemli eksikliklerde serinin doğal trendini yeterince temsil edemeyeceği ve potansiyel doğrusal sapmalar yaratabileceği için tercih edilmemiştir.

Yapılan çalışmalar neticesinde, Ocak 2013 – Mayıs 2024 dönemini kapsayan veri setinde yer alan eksik değerlerin yüzeysel düzeyde tamamlanmasına yönelik kapsamlı ve sistematik bir yaklaşım benimsenmiştir. Öncelikle her bir özneliğin zaman serisi yapısı, sektörel anlamı ve istatistiksel özellikleri dikkate alınarak eksik verilerin doğası ayrıntılı

şekilde analiz edilmiştir. Bu analizler sonucunda, her değişken için en uygun yüzeysel eksik veri doldurma yöntemi belirlenmiş ve seçilen yöntemler, veri setinin içsel dinamiklerine zarar vermeyecek biçimde uygulanmıştır. Uygulanan interpolasyon teknikleri, sabit değer atama stratejileri ya da büyüme tabanlı tahmin yöntemleri, her bir değişkenin karakterine uygun olarak özenle seçilmiş; söz konusu yöntemlerin kullanılma gerekçeleri akademik kaynaklarla desteklenerek ortaya konmuştur. Ayrıca alternatif yöntemlerin neden uygun bulunmadığı da yine literatür temelli argümanlarla açıklanmış, böylece yönteme özgü eleştirel bir değerlendirme zemini oluşturulmuştur.

Gerçekleştirilen bu yüzeysel eksik veri doldurma süreci sonucunda, veri setinde herhangi bir eksik gözlem kalmamış; aylık bazda sürekli ve tutarlı bir veri yapısı elde edilmiştir. Elde edilen eksiksiz veri seti hem korelasyon analizleri hem de öznitelik indirgeme işlemlerine geçiş için sağlam bir istatistiksel temel sunmuştur. Özellikle sonraki aşamalarda gerçekleştirilecek olan öznitelik seçimi ve makine öğrenmesi modelleme adımlarında, eksik veri kaynaklı model başarısı düşüşü riski önemli ölçüde azaltılmıştır. Bu yönüyle, yüzeysel eksik veri doldurma işlemleri yalnızca teknik bir hazırlık aşaması değil, aynı zamanda modelleme sürecinin güvenilirliğini teminat altına alan kritik bir adımdır. İlerleyen bölümlerde detaylandırılacak olan öznitelik seçim süreci, bu veri bütünlüğü sayesinde daha sağlıklı bir zemin üzerinde inşa edilecektir.

4.3. Çalışma 1: Uzman Görüşü Temelli Öznitelik Seçimi

Bu bölümde, yapı kimyasalı üreticisinin farklı departmanlarında görev yapan uzmanların sektörel bilgi ve tecrübelerine dayalı olarak gerçekleştirilmiş öznitelik seçimi süreci ele alınmaktadır. Satın alma, Ar-Ge, satış, üretim ve üst yönetim birimlerinde görev yapan mühendisler, satın alma uzmanları ve genel müdür yardımcılarıyla birebir görüşmeler ve küçük ölçekli toplantılar düzenlenmiş; veri setinde yer alan özniteliklerin anlamlılığı bu uzmanların sektörel değerlendirmeleri doğrultusunda analiz edilmiştir. Görüşmeler sırasında, çalışmanın veri seti Ocak 2013 – Aralık 2024 dönemini kapsadığı ve aylık bazda yalnızca 144 satır içerdiği için, aşırı değişken sayısının model performansına zarar verebileceği vurgulanmış ve toplamda en fazla 10 adet öznitelik belirlenmesi istenmiştir. Bu sınırlama, modelin aşırı öğrenme riskini azaltmak ve Çalışma 3'te uygulanacak olan öznitelik birleştirme stratejisi için de uygun bir esneklik sağlamak amacıyla yapılmıştır.

Yapı kimyasalı sektörü, ekonomik dalgalanmalar, enerji ve hammadde fiyatlarındaki oynamalar, iş gücü maliyetleri ve politik gelişmeler gibi çok çeşitli dış faktörden etkilenebilecek bir yapıya sahiptir. Ayrıca sektörde alınan kararlar zaman zaman sadece istatistiksel eğilimlere değil, piyasa söylentileri gibi rasyonel olmayan faktörlere de dayanabilmektedir. Bu nedenle, hedef değişken olan Eskişehir Fabrikası Gri Çimento Satın Alım Fiyatının tahmininde yalnızca istatistiksel analizlerle değil, sektörel uzmanlıkla desteklenen öznitelik seçimi yapılması önemli bir gereklilik hâline gelmiştir. Bu bağlamda seçilen öznitelikler; üretim hacmi, enerji ve işçilik maliyetleri, politik sinyaller, nüfus dinamikleri ve şirket içi satış endeksleri gibi çok boyutlu faktörleri içermektedir. Aşağıda, uzman görüşleri doğrultusunda belirlenen her bir öznitelik ve seçim gerekçeleri detaylı şekilde sunulacaktır.

Yapı kimyasalı sektörü, yapı ruhsatlarındaki dalgalanmalar, yakıt ve enerji maliyetleri, seçim söylentileri gibi çok sayıda dış faktörden doğrudan etkilenmektedir. Bu nedenle, yapı kimyasallarında kullanılan hammaddelerin satın alım fiyatlarının tahmin edilebilmesi için, sektörel bilgiye dayalı olarak anlamlı ve güçlü tahmin gücüne sahip özniteliklerin belirlenmesi kaçınılmaz hâle gelmiştir. Bu doğrultuda, uzmanlar tarafından seçilen ilk öznitelik “Çimento Üretim Miktarı (ton)” olmuştur. Bu tercihin ardında yatan temel neden, yapı kimyasallarının çimento bazlı ürünlerde kullanılması ve çimento üretimi ile yapı kimyasalı talebi arasında doğrudan bir bağ bulunmasıdır. Üretimdeki dalgalanmalar, doğrudan yapı kimyasalı üreticilerinin satış hacmini ve dolayısıyla satın alım stratejilerini etkilemektedir. Ancak “Çimento Üretim Miktarı” değişkeninde eksik gözlemlerin bulunması, veri bütünlüğü açısından bir risk oluşturmaktadır.

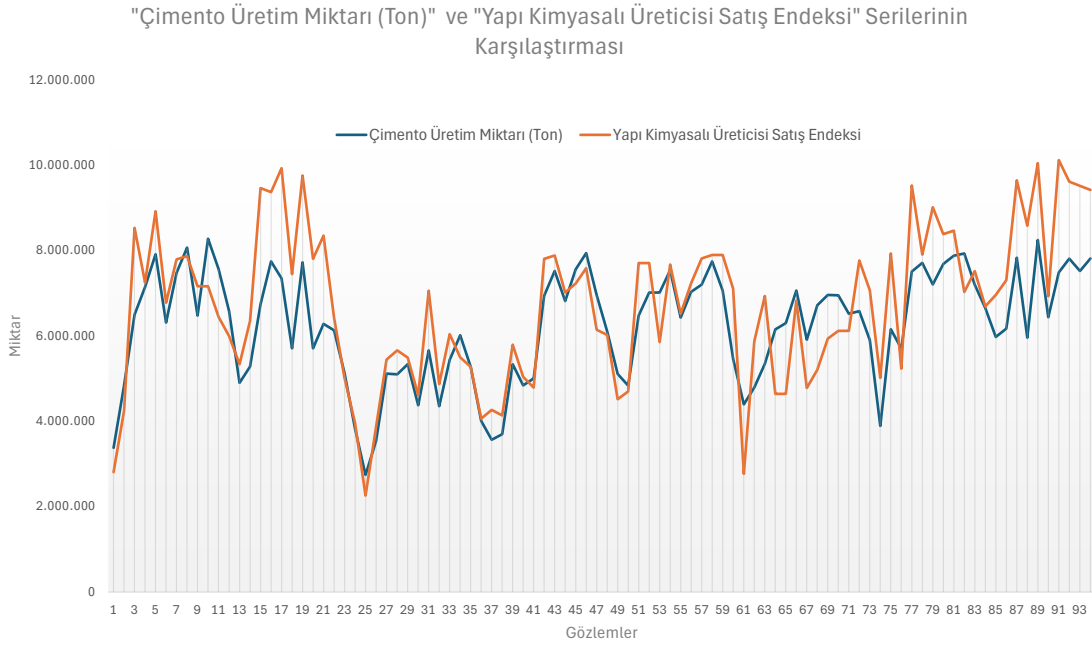
Yapılan ön değerlendirmeler sonrasında, “Çimento Üretim Miktarı (ton)” değişkeninin doğrudan kullanımı yerine, yapı kimyasalı üreticisinin kurumsal kayıtlarından elde edilen “Yapı Kimyasalı Üreticisi Satış Endeksi” özniteliğinin kullanılması tercih edilmiştir. Bu kararın temel gerekçesi, Yapı kimyasalı Üreticisi Satış Endeksi'nin aylık olarak yapı kimyasalı üreticisinin satış hacmini doğrudan temsil etmesi, veri setinde eksik değer barındırmaması ve ölçüm güvenilirliğinin yüksek olmasıdır. Tanımlayıcı istatistikler açısından her iki değişkenin ortalama, medyan ve çeyrekler arası değerleri birbirine oldukça yakın bulunmuştur. Çimento Üretim Miktarı (Ton) ve Yapı Kimyasalı Üreticisi Satış Endeksi için Betimsel İstatistikler Çizelge 4.1’de sunulmuştur.

Çizelge 4.1. Çimento Üretim Miktarı (Ton) ve Yapı Kimyasalı Üreticisi Satış Endeksi için Betimsel İstatistikler

İstatistik	Çimento Üretim Miktarı (Ton)	Yapı Kimyasalı Üreticisi Satış Endeksi
Gözlem Sayısı (n)	94	94
Ortalama	6.252.367	6.726.720
Standart Sapma	1.302.121	1.787.476
Minimum	2.745.103	2.258.797
1. Çeyrek	5.337.445	5.364.408
Medyan	6.457.092	6.947.053
3. Çeyrek	7.313.842	7.859.435
Maksimum	8.277.128	10.120.320

Çizelge 4.1’de görüleceği üzere Çimento Üretim Miktarı (Ton) ortalaması 6.252.367 ton, medyanı 6.457.092 ton, standart sapması 1.302.121 ton iken; Yapı Kimyasalı Üreticisi Satış Endeksi’nin ortalaması 6.726.720, medyanı 6.947.053, standart sapması ise 1.787.476 olarak hesaplanmıştır. Ayrıca her iki değişkenin çeyrekler arası dağılımı da paralellik göstermekte, bu da veri yapılarının benzerliğini desteklemektedir.

İki değişken arasındaki ilişkinin gücünü belirlemek üzere uygulanan Pearson korelasyon analizi sonucunda, $r = 0,814$ düzeyinde oldukça güçlü ve pozitif yönlü bir ilişki elde edilmiştir ($p < 0,001$). Evans (1996) tarafından önerilen sınıflandırmalara göre bu seviye “çok güçlü ilişki” kategorisine girmektedir. Mukaka’nın (2012) rehberliğinde gerçekleştirilen anlamlılık testi sonucunda, elde edilen ilişkinin istatistiksel olarak tesadüfi olmadığı ortaya konmuştur. Çimento Üretim Miktarı (Ton) ve Yapı Kimyasalı Üreticisi Satış Endeksi Karşılaştırması Şekil 4.22’de sunulmuştur.



Şekil 4.22. Çimento Üretim Miktarı (Ton) ve Yapı Kimyasalı Üreticisi Satış Endeksi Karşılaştırması

Şekil 4.22 iki değişkenin gözlem değerlerinin yüksek senkronizasyonla hareket ettiğini görsel olarak kanıtlamaktadır. Bununla birlikte, yapılan uzman görüşmelerinde de Yapı Kimyasalı Üreticisi Satış Endeksi'nin çimento üretimi ile yüksek oranda senkronize olduğu ve sektörel trendleri daha isabetli yansıttığı ifade edilmiştir. Bu nedenle “Çimento Üretim Miktarı (ton)” yerine “Yapı Kimyasalı Üreticisi Satış Endeksi”nin kullanılması hem akademik hem uygulamalı açıdan daha uygun bir temsil biçimi olarak değerlendirilmiştir.

Motorin Fiyatı (Eskişehir/Merkez) (USD/Litre) özniteliği, yapı kimyasalı üretim sektöründe üretim ve lojistik süreçlerinin temel maliyet kalemlerinden biri olan enerji ve nakliye harcamalarının doğrudan bir temsilcisi olması sebebiyle uzmanlar tarafından öznitelik setine dâhil edilmiştir. Özellikle Eskişehir merkezli bir fabrika için bölgesel yakıt fiyatları, üretim girdilerinin taşınması, mamul ürünlerin dağıtımı ve şantiyelere ulaştırılması gibi birçok aşamada maliyet yapısını etkileyen önemli bir değişkendir. Türkiye’de motorin fiyatlarının döviz kuru, uluslararası petrol piyasaları ve vergilendirme politikaları gibi faktörlerden etkilenmesi nedeniyle zaman içinde dalgalanma gösterdiği bilinmektedir. Bu dalgalanmalar, doğrudan şekilde hammadde tedarik stratejilerini ve satın alım fiyatlarını etkileyebilecek niteliktedir.

Özniteliğin USD bazında alınması, döviz kuru etkilerinin doğrudan gözlemlenmesini sağlamakta ve maliyet analizinde uluslararası geçerliliğe sahip bir perspektif sunmaktadır. Aynı zamanda, bu yaklaşım sayesinde farklı yıllar ve dönemler arasında karşılaştırmalı analiz yapılması mümkün hale gelmektedir. Uzman görüşlerine göre, özellikle son yıllarda enerji maliyetlerinin küresel ölçekte artış göstermesi, motorin fiyatlarının tahminleme modellerinde daha belirleyici bir rol oynamasına yol açmıştır. Bu nedenle söz konusu özniteliğin, yapı kimyasalı üreticisinin hammadde satın alma kararlarında etkili olduğu düşünüülerek, modelleme sürecine dâhil edilmesi uygun bulunmuştur.

Petrokok Birim Fiyatı (USD) özniteliği, yapı kimyasalları üretiminde kullanılan çimento türevi ürünlerin üretim maliyetleri üzerinde doğrudan etkisi bulunan enerji girdilerinden biri olan petrokok maliyetlerini temsil etmektedir. Petrokok, yüksek kalorifik değere sahip bir fosil yakıt türü olup özellikle klinker üretiminde ısı işlem ihtiyacını karşılamak amacıyla yaygın şekilde kullanılmaktadır. Çimento üretim süreçlerinin enerji yoğun karakteri göz önüne alındığında, petrokok fiyatlarındaki dalgalanmalar doğrudan çimento üretim maliyetlerine ve dolayısıyla yapı kimyasalı üreticisinin satın alım fiyatlarına yansımaktadır.

Uzmanlar, bu özniteliğin hem doğrudan enerji maliyetlerini temsil etmesi hem de döviz bazlı olması nedeniyle, dış ekonomik etkilerin (örneğin döviz kuru değişimleri, uluslararası emtia piyasaları) fiyat tahminlerine yansımaları daha sağlıklı bir şekilde yakalayabileceğini ifade etmişlerdir. Özellikle Türkiye'nin büyük oranda petrokok ithalatçısı bir ülke olması, yurt dışı fiyatlarının iç piyasadaki maliyet yapısını belirlemesinde kritik rol oynamaktadır. Bu çerçevede, Petrokok Birim Fiyatı (USD) özniteliği, maliyet tahmin modellerinde enerji ve hammadde girdilerinin birlikte modellenenbilmesi açısından oldukça stratejik bir değişken olarak değerlendirilmiş ve öznitelik setine dâhil edilmiştir.

Elektrik Fiyatı (USD/MWh) özniteliği, yapı kimyasalı üretiminde direkt enerji maliyetlerinin önemli bir bileşenidir. Özellikle üretim tesislerinde karıştırma, öğütme, paketleme ve otomasyon gibi proseslerin tamamında elektrik enerjisi yoğun biçimde kullanılmakta olup, bu tüketim toplam üretim maliyetleri içinde hatırı sayılır bir paya sahiptir. Çimento sektörüne dayalı yapı kimyasalı üretiminde, elektrik fiyatlarındaki artışlar

yalnızca üretim hattı maliyetlerini değil, aynı zamanda tedarik zincirindeki taşeron hizmetlerin (örneğin nakliye, depolama) enerjiye dayalı fiyatlamasını da etkilemektedir.

Uzman görüşlerine göre, Türkiye’de elektrik fiyatları hem ulusal enerji politikalarından hem de döviz kuru dalgalanmalarından etkilenerek zaman içinde dalgalı bir seyir izlemektedir. Bu da özellikle döviz bazlı fiyatlandırmanın uygulandığı ürünlerde maliyet kontrolünü zorlaştırmakta ve tahmin modellerinde elektrik fiyatlarının değişkenliğini dikkate almayı gerekli kılmaktadır. Bu bağlamda, Elektrik Fiyatı (USD/MWh) özniteliği, üretim ve lojistik süreçlerinin enerji girdilerine duyarlılığını modellemeye imkân sağlayacak kritik bir değişken olarak uzmanlar tarafından öznitelik setine dâhil edilmiştir.

Üretici Fiyat Endeksi (ÜFE), bir ekonomide üreticilerin mal ve hizmetleri piyasaya sunarken karşılaştığı maliyet baskılarını yansıtan temel makroekonomik göstergelerden biridir. Yapı kimyasalı üretimi gibi hammadde-yoğun ve enerji-yoğun sektörlerde, üretici seviyesindeki maliyet değişimleri doğrudan ürün fiyatlarına yansımaktadır. Bu nedenle, ÜFE’nin zaman içindeki seyri hem kısa vadeli fiyat dalgalanmalarını hem de uzun vadeli maliyet baskılarını modelleyebilmek adına kritik önem taşır.

Uzmanlar, çimento ve yapı kimyasalları gibi entegre sektörlerde ÜFE’nin bir tür "genel maliyet göstergesi" olarak davranabildiğini ifade etmişlerdir. Özellikle hammaddelerin büyük kısmının yurtiçinden temin edilmesi ve ulusal üretim zincirine entegre olması sebebiyle ÜFE, sektörel fiyat hareketlerinin güçlü bir temsilcisi olarak görülmektedir. Ayrıca, devletin vergi politikaları, asgari ücret artışları ve enerji maliyetlerindeki değişim gibi çeşitli makro düzeydeki kararlar da ÜFE’yi etkileyerek yapı kimyasalı sektöründe satın alma kararlarını yönlendirebilmektedir.

Bu gerekçeler doğrultusunda, Üretici Fiyat Endeksi (ÜFE) özniteliği, uzmanlar tarafından hedef değişken üzerindeki yaygın ve sürekli etkisi nedeniyle öznitelik setine dâhil edilmiştir. Bu değişkenin zaman içinde gösterdiği eğilimler, satın alma maliyetlerinin genelleştirilmiş bir göstergesi olarak tahminleme modellerine anlamlı katkı sunma potansiyeli taşımaktadır.

Herhangi bir seçim var mı? (genel veya yerel) özniteliği, yapı kimyasalı üreticisinin sektör dinamikleri ile doğrudan bağlantılı olan politik ve makroekonomik belirsizlik faktörlerini temsil etmesi amacıyla uzmanlar tarafından öznitelik setine dâhil edilmiştir. İnşaat sektörü genel olarak, özellikle de kamu yatırımlarının ve konut teşviklerinin sıklaştığı seçim yıllarında, siyasi takvimlerden etkilenmeye oldukça açık bir yapıya sahiptir. Uzmanlar, seçim dönemlerinin doğrudan ya da dolaylı olarak piyasa koşullarında dalgalanmalara yol açtığını, bu dalgalanmaların ise yapı kimyasalları gibi tamamlayıcı girdilerin tedarikinde fiyat oynaklıklarına sebep olabileceğini belirtmişlerdir. Bu nedenle, belirli bir ayda genel veya yerel bir seçim yapıp yapılmadığı bilgisi, ekonomik beklentilerin, yatırım kararlarının ve kamu harcamalarının yönünü tahmin etmek açısından anlamlı bir girdi olarak değerlendirilmiştir. Özellikle seçimlerin yapı sektörü üzerindeki etkileri, sadece kamusal talep ile sınırlı kalmayıp aynı zamanda özel sektör yatırımcılarının da davranışlarını şekillendirebilmekte; dolayısıyla çimento gibi ana hammadde kalemlerinde kısa vadeli fiyat artışları veya talep sıçramaları gözlenebilmektedir.

“Herhangi bir seçim var mı?” özniteliği, sektörel davranışları anlamada önemli bir gösterge olarak tanımlanmış ve uzman görüşü temelli öznitelik setine alınmıştır ancak uzman görüşü değerlendirmeleri neticesinde, sektördeki fiyat dalgalanmalarının esas kaynağının yalnızca seçimlerin gerçekleşmesi değil, seçim söylentilerinin varlığı olduğu kanaatine varılmıştır. Firmadaki uzmanlar, halk arasında yayılan seçim iddialarının; özellikle yatırımcı davranışlarını, kamu alımlarını ve inşaat faaliyetlerini doğrudan etkilediğini belirtmişlerdir. Bu durum, çimento gibi yapı malzemelerinin talebinde ani sıçramalara veya beklenmedik daralmalara neden olabilmektedir. Dolayısıyla, “Herhangi bir seçim var mı? (genel veya yerel)” özniteliği, doğrudan seçim takvimine dayandığı için bu dalgalanmaları yansıtmakta yetersiz kalabileceğinden, “Herhangi bir seçim söylentisi var mı?” şeklinde yeniden tanımlanmıştır.

Yeni oluşturulan öznitelik, dijital izler aracılığıyla halkın seçim beklentisini veya siyasal gündemle ilgili duyarlılığını ölçmek amacıyla Google Trends verileri kullanılarak sayısallaştırılmıştır. Bu kapsamda, Google Web Arama veri tabanında, “Türkiye” bölgesi ve “Ocak 2013 – Aralık 2024” tarih aralığı seçilerek, belirlenen uygun bir arama terimi üzerinden veri çekilmiştir. Elde edilen değerler, Google Trends’in sağladığı normalize edilmiş arama endeksleri aracılığıyla 0 ile 100 arasında bir aralıkta, aylık düzeyde sayısal

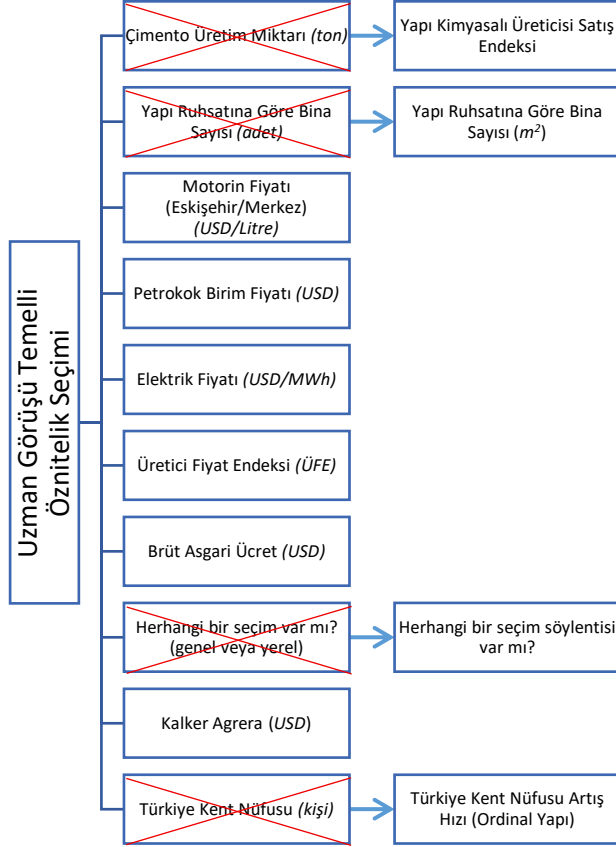
biçime dönüştürülmüş ve analizde kullanılmaya hazır hâle getirilmiştir. Bu sayede, seçimin kendisinden ziyade seçim söylentisinin meydana getirdiği belirsizlik ortamının veri setine daha doğru biçimde yansıtılması sağlanmış ve bu öznitelik, uzman görüşü doğrultusunda çalışma kapsamına dâhil edilmiştir.

Kalker Agrera (USD) özniteliği, başlangıçta oluşturulan öznitelik havuzunda yer almamaktaydı fakat uzmanlar tarafından yürütülen değerlendirme sürecinde, bu girdinin de Eskişehir Fabrikası'nın gri çimento satın alma fiyatlarını etkileyebilecek önemli bir maliyet kalemi olduğu belirtilmiştir. Özellikle çimento üretiminde kullanılan yardımcı hammaddeler arasında yer alan kalker agrerası, üretim reçetelerinde belirli oranlarda yer almakta ve temin maliyetleri doğrudan birim çimento maliyetine yansımaktadır.

Türkiye Kent Nüfusu (kişi) özniteliği, yapı kimyasalı üreticisinin satın alım stratejileri üzerinde dolaylı etkiler yaratabilecek demografik dinamiklerden biri olarak değerlendirilmiştir. Türkiye'de artan kentleşme oranları, konut ve altyapı talebini doğrudan etkileyerek yapı kimyasallarına olan ihtiyacı artırmaktadır. Bu bağlamda, kent nüfusunda meydana gelen artışlar, çimento talebinin önemli bir belirleyicisi olarak öne çıkmaktadır. Bu nedenle söz konusu öznitelik, uzman görüşleri doğrultusunda Çalışma 1 veri setine dâhil edilmiştir ancak bu özniteliğin doğrudan ham hâliyle (yani kişi sayısı olarak) kullanımı, aylar arası aynı değeri tekrar etmesi ve çok büyük sayılar içermesi nedeniyle modelleme açısından sınırlı fayda sağlamaktadır. Ayrıca sürekli değişkenlerin bazı makine öğrenmesi algoritmalarında aşırı öğrenmeye neden olabileceği bilinmektedir (James vd., 2013). Bu nedenle, ham veriler yerine nüfus artış hızı hesaplanmış ve bu oran, ordinal yapıda modele dâhil edilmiştir.

Söz konusu dönüşümün yapılabilmesi adına öncelikle Türkiye kent nüfusunun yıllık artış oranları hesaplanmıştır. Bu oranlar üzerinden histogram analizi gerçekleştirilmiş ve gözlemsel kümelenme yapısı incelenmiştir. Dağılımın %75'inin %1,75'in altında olduğu görülerek bu değer "düşük artış" eşiği olarak belirlenmiştir. %1,75 ile %3,74 arasındaki gözlemler "orta artış", %3,74 üzeri değerler ise "yüksek artış" olarak sınıflandırılmıştır. Bu ordinal dönüşüm hem modelin yorumlanabilirliğini artırmakta hem de sınırlı sayıda gözleme sahip veri setlerinde aşırı öğrenme riskini azaltmaktadır (Kuhn ve Johnson, 2013).

Böylece, uzman görüşüne dayalı olarak gerçekleştirilen öznitelik seçim süreci tamamlanmıştır. Seçilen özniteliklerin görselleştirilmiş hâli Şekil 4.23'te sunulmuş olup, bir sonraki bölümde korelasyon temelli yöntemlerle gerçekleştirilen öznitelik seçim sürecine odaklanılmıştır.



Şekil 4.23. Uzman Görüşü Temelli Öznitelik Seçim Sonuçları

4.4. Çalışma 2: Korelasyon Analizi Temelli Öznitelik Seçimi

Bu bölümde, öznitelik seçim süreci uzman görüşlerinden bağımsız olarak tamamen nicel yöntemler çerçevesinde ele alınmıştır. Korelasyon analizi temelli seçim yaklaşımıyla, model performansını artırmak ve çoklu doğrusal bağlantı gibi olası istatistiksel sorunları önlemek amacıyla korelasyon analizi hesaplamalarına dayalı bir yöntem izlenmiştir. Bu doğrultuda, önce hedef değişkenle yüksek korelasyona sahip öznitelikler belirlenmiş, ardından birbirleriyle yüksek ilişki içinde olan öznitelikler arasından seçim yapılmıştır. Korelasyon matrisleri hem değişkenler arası ilişkilerin yapısını anlamada hem de veri setinin modellemeye uygunluğunu test etmede temel araçlar olarak kullanılmıştır. Elde edilen

bulgular neticesinde, istatistiksel geçerliliği yüksek ve birbirini tamamlayıcı nitelikteki özniteliklerden oluşan alternatif bir öznitelik seti yapılandırılmıştır.

Korelasyon analizi temelli öznitelik seçimi sürecinde ilk adım olarak, veri setindeki öznitelikler ile hedef değişken arasındaki doğrusal ilişkiler Pearson korelasyon katsayısı aracılığıyla analiz edilmiştir. Çalışmada yer alan veri seti 30 öznitelik ve 1 hedef değişkenden oluşmakta olup, analizler Ocak 2013 - Mayıs 2024 dönemine ait toplam 137 gözlem üzerinden gerçekleştirilmiştir.

Korelasyon analizi bulgularının tamamına ilişkin detaylı ısı haritası Ek Açıklama-A, korelasyon matrisinin sayısal hâli ise Ek Açıklama-B'de sunulmuştur. Yürütülen korelasyon analizi sonucunda bazı özniteliklerin hedef değişkenle yüksek düzeyde doğrusal ilişki gösterdiği tespit edilmiştir. Hedef değişken ile en yüksek pozitif korelasyona sahip öznitelikler Çizelge 4.2’de sunulmuştur.

Çizelge 4.2. Hedef Değişken ile En Yüksek Pozitif Korelasyona Sahip Öznitelikler

Öznitelik	Korelasyon Değeri
Çimento Kili Ocakbaşı Satış Fiyatı (<i>USD/Ton</i>)	0,856
Kalker Ocakbaşı Satış Fiyatı (<i>USD/Ton</i>)	0,813
Motorin Fiyatı (Eskişehir/Merkez) (<i>USD/Litre</i>)	0,624
Elektrik Fiyatı (<i>USD/MWh</i>)	0,549
Alçıtaşı Ocakbaşı Satış Fiyatı (<i>USD/Ton</i>)	0,424

Çizelge 4.2 incelendiğinde, “Çimento Kili Ocakbaşı Satış Fiyatı (*USD/Ton*)” ($r = 0,856$), “Kalker Ocakbaşı Satış Fiyatı (*USD/Ton*)” ($r = 0,813$) ve “Motorin Fiyatı (*USD/Litre*)” ($r = 0,624$) gibi öznitelikler hedef değişken ile oldukça güçlü pozitif korelasyon içindedir. Öte yandan hedef değişken ile en yüksek negatif korelasyona sahip öznitelikler Çizelge 4.3’de sunulmuştur.

Çizelge 4.3. Hedef Değişken ile En Yüksek Negatif Korelasyona Sahip Öznitelikler

Öznitelik	Korelasyon Değeri
Çimento Sanayi İstihdam Sayısı (kişi)	-0,132
Çimento İhracatı (ton)	-0,226
Türkiye Nüfusu (kişi)	-0,357
Türkiye Kent Nüfusu (kişi)	-0,359
Klinker İhracatı (ton)	-0,753

Çizelge 4.3 incelendiğinde “Klinker İhracatı (ton)” ($r = -0,753$) değişkeninin hedef değişken ile güçlü negatif bir korelasyona sahip olduğu görülmüştür. Bunun yanı sıra, “Türkiye Nüfusu (kişi)” ($r = -0,357$) ve “Türkiye Kent Nüfusu (kişi)” ($r = -0,359$) gibi bazı değişkenlerin de negatif yönde ilişkiler barındırdığı dikkat çekmiştir. Bu durum, sektörel bağlamda nüfus artışı ve üretim maliyetleri arasındaki dolaylı ilişkileri irdelemek açısından anlamlı bir bulgudur.

Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) özniteliği hedef değişken ile en yüksek pozitif korelasyona ($r = 0,856$) sahip olduğu için seçim sürecinde öncelikli olarak değerlendirilmiştir. Bu bağlamda, aynı hammadde TL bazlı versiyonu olan “Çimento Kili Ocakbaşı Satış Fiyatı (TL/Ton)” özniteliği dikkate alınmamıştır. Bunun temel sebebi, Türkiye’de döviz kurunun yüksek oynaklığı nedeniyle TL bazlı fiyatların dönemsel olarak yanıltıcı olabilmesidir. USD cinsinden veriler, özellikle ithal girdi bağımlılığı olan sektörlerde daha istikrarlı bir analiz zemini sunar. Ayrıca, fiyat değişimlerinin uluslararası piyasalardaki karşılığını yansıtabilme kapasitesi USD cinsinden olan değişkeni analitik olarak daha anlamlı kılmaktadır.

Motorin Fiyatı (Eskişehir/Merkez) (USD/Litre) özniteliği üretim ve nakliye süreçlerinin doğrudan maliyet unsurlarından biri olması nedeniyle hedef değişkenle anlamlı bir ilişki içindedir ($r = 0,624$). Bu değişkenin TL versiyonu olan “Motorin Fiyatı (TL/Litre)” benzer bir eğilim gösterse de USD cinsinden olan verinin tercih edilme nedeni, genel ekonomik analizlerde uluslararası karşılaştırılabilirliğin sağlanması ve kur etkisinden arındırılmış net bir maliyet göstergesi sunmasıdır. Ayrıca, enerji fiyatlarının küresel düzeyde

belirlendiği dikkate alındığında, dolar bazlı fiyatlar modelleme açısından daha tutarlı sonuçlar verebilmektedir. Bu nedenle TL cinsinden alternatif reddedilmiştir.

Elektrik Fiyatı (USD/MWh) özniteliği, üretim maliyetlerinin temel bileşenlerinden biridir ve hedef değişkenle pozitif korelasyon göstermektedir ($r = 0,549$). Bu öznitelik yerine kullanılabilecek “Elektrik Fiyatı (TL/MWh)” değişkeni, döviz kuru değişimlerinden etkilenerek fiyat dalgalanmalarının daha büyük sapmalar göstermesine neden olabilir. Elektrik maliyetinin USD bazında değerlendirilmesi, özellikle dış ticaretle ilişkili sektörlerde daha gerçekçi bir maliyet yapısı sunar. Ayrıca enerji maliyetlerinin küresel enerji fiyatları ile senkronize olarak belirlendiği dikkate alınarak USD bazlı verinin ekonomik tutarlılığı ve tahminleme başarısı açısından daha uygun olduğu düşünülmüştür.

Hedef değişken ile pozitif korelasyon içerisinde olan Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) özniteliği ($r = 0,424$), özellikle ikame hammadde girdilerinden biri olması ve yapı kimyasallarının üretiminde kullanılabilirliği nedeniyle tercih edilmiştir. Alternatif olarak değerlendirilebilecek TL bazlı versiyonu, yine kur oynaklıklarından etkilenmesi nedeniyle tercih edilmemiştir. USD cinsinden fiyatlar, maliyet analizlerinde daha istikrarlı sonuçlar sunarak öngörü gücünü artırmaktadır. Ayrıca, alçıtaşı gibi bazı hammaddelerin yurtiçi arz-talep dengesinden çok döviz bazlı sözleşmelere bağlı olarak fiyatlandığı durumlarda USD veriler daha temsil edici olmaktadır.

Klinker İhracatı (Ton) özniteliği, hedef değişkenle yüksek negatif korelasyon göstermektedir ($r = -0,753$) ve piyasa dengesini ters yönde etkileyebilecek bir faktör olarak önem arz etmektedir. Alternatif olarak değerlendirilebilecek “Klinker Üretim Miktarı (Ton)” gibi öznitelikler ise ya eksik veriler nedeniyle elenmiş ya da hedef değişken üzerindeki doğrudan etkileri istatistiksel olarak daha zayıf kalmıştır. Ayrıca ihracat verisi, yurt içi arzı etkileyerek dolaylı şekilde çimento fiyatlarını belirlediği için tercih edilmiştir. Böylelikle sektördeki dışa açıklığın fiyat üzerindeki etkisi de modele entegre edilebilmiştir.

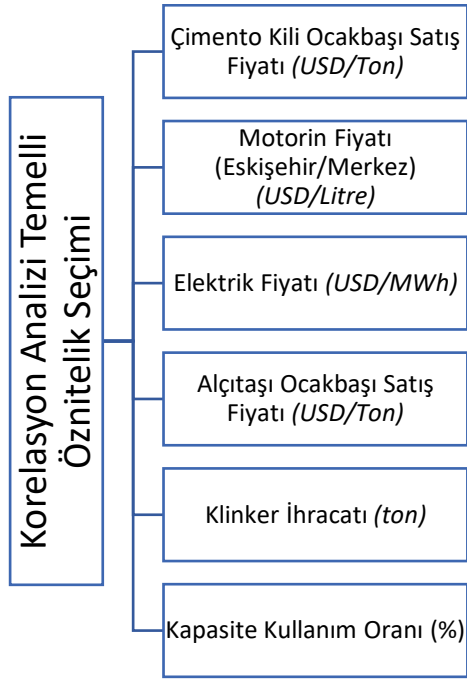
Normalde çalışma başında belirlenen alternatif öznitelikler arasında olmamasına rağmen Kapasite Kullanım Oranı (%) özniteliği, sektörel bilgiye dayalı olarak anlamlı bir öngörücü olabileceği düşüncesiyle modele dâhil edilmiştir. Alternatif olarak değerlendirilen genel ekonomik göstergeler ya da üretim düzeyi ile ilgili bazı diğer öznitelikler, hedef

değişken üzerindeki etkisini dolaylı olarak yansıttığı için elenmiştir. Özellikle çimento üretiminin dâhil olduğu “Diğer metalik olmayan mineral ürünlerin imalatı” alt kategorisine ait aylık kapasite kullanım oranı verisi, sektöre özgü olması nedeniyle öne çıkmıştır. Bu değişkenin seçiminde, üretim düzeyinin sektörel yoğunluğa bağlı olarak maliyet yapısını etkilemesi temel alınmıştır.

Yapı kimyasalı üreticisi ile yapılan görüşmelerde, Eskişehir Fabrikası’nın gri çimento satın alma fiyat endeksinin dövizden arındırılmış biçimde oluşturulduğu bilgisi edinilmiştir. Bu durum, öznitelik seçiminde özellikle fiyat verilerinde TL bazlı değişkenler yerine USD cinsinden değişkenlerin tercih edilmesini gerekçelendirmiştir. Böylelikle hedef değişkenin yapısı ile öznitelikler arasındaki ekonomik bütünlük korunmuş, kur oynaklıklarının model tahmin başarısı üzerindeki etkileri enküçüklenmiştir.

Bu bölümde, korelasyon analizi temelli öznitelik seçimi süreci sistematik bir şekilde yürütülmüş; hedef değişken ile yüksek korelasyona sahip öznitelikler analiz edilerek seçim yapılmıştır. Seçim sürecinde yalnızca korelasyon gücü değil, aynı zamanda sektörel anlamlılık, ekonomik tutarlılık ve modelleme başarısı gibi çok boyutlu kriterler dikkate alınmıştır. Alternatif öznitelikler, gerek istatistiksel açıdan zayıf ilişkiler göstermeleri gerekse ekonomik temsiliyette yetersiz kalmaları nedeniyle tercih edilmemiştir. Bu kararlar alınırken yalnızca korelasyon değerleri değil, değişkenlerin birbirleriyle olan anlamlı ilişkileri ve veri yapısının yapay oynaklıklardan arındırılmış olmasına yönelik hassasiyetler de göz önünde bulundurulmuştur.

Seçilen özniteliklerin görsel temsili Şekil 4.24’te sunulmakta olup, izleyen bölümlerde seçilen özniteliklerle gerçekleştirilen eksik veri doldurma ve modelleme çalışmalarına odaklanılacaktır.



Şekil 4.24.Korelasyon Analizi Temelli Öznitelik Seçim Sonuçları

4.5. Gelişmiş Eksik Veri Doldurma ve Nihai Veri Setlerinin Oluşturulması

Bu bölümde, daha önce tanımlanmış olan Çalışma 1 ve Çalışma 2 kapsamında kullanılacak nihai veri setlerinin oluşturulmasına yönelik gelişmiş eksik veri doldurma süreçleri ele alınmaktadır. Önceki bölümlerde, genel veri yapısı incelenmiş ve yüzeysel doldurma yöntemleriyle ön değerlendirmeler yapılmıştı fakat modelleme aşamasına geçmeden önce, tahminleme başarısını olumsuz yönde etkileyebilecek eksik gözlemlerin daha gelişmiş ve veri yapısına uygun yöntemlerle tamamlanması gerekmektedir. Bu nedenle, her iki çalışmaya özgü öznitelik setleri için uygun bulunan eksik veri tamamlama stratejileri dikkatle planlanmış ve uygulanmıştır.

Bu bağlamda, Çalışma 1 için uzman görüşüyle belirlenen öznitelikler; Çalışma 2 için ise korelasyon analizi temelli korelasyon analizleriyle seçilmiş öznitelikler esas alınarak ayrı veri setleri oluşturulmuştur. Her iki çalışma özelinde eksik verilerin türü, konumu ve yapısı analiz edilmiş, seçilen yöntemlerin dayandığı istatistiksel gerekçeler açıklanmış ve veri setinin homojenliğini bozmayacak şekilde eksik veriler doldurulmuştur. Takip eden alt başlıklarda bu işlemlerin her bir öznitelik bazında ayrıntılı biçimde sunulması amaçlanmaktadır.

4.5.1. Çalışma 1'de kullanılacak veri seti için eksik verilerin ileri yöntemler ile doldurulması

Çalışma 1 kapsamında seçilen öznitelikler incelendiğinde, yalnızca Petrokok Birim Fiyatı (USD) özniteliğinde tek bir gözlemde eksik veri bulunduğu tespit edilmiştir. Diğer tüm özniteliklerin tam ve eksiksiz olması, bu başlık altında kapsamlı bir doldurma prosedürüne ihtiyaç duyulmadığını göstermektedir. Ancak yine de veri setinin bütünlüğünü sağlamak amacıyla söz konusu eksik gözlem için zaman serisi analizlerinde sıklıkla önerilen (Hyndman ve Athanasopoulos, 2018) lineer interpolasyon yöntemi uygulanmıştır. Böylece Çalışma 1 veri seti, herhangi bir eksik gözlem içermeyen, modellemeye hazır ve bütünlüklü bir hâle getirilmiştir.

4.5.2. Çalışma 2'de kullanılacak veri seti için eksik verilerin ileri yöntemler ile doldurulması

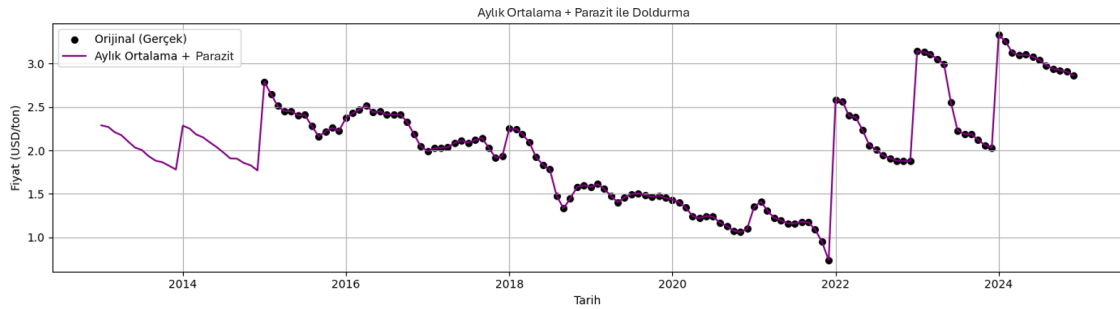
Çalışma 2 kapsamında yürütülen modelleme sürecinde, seçilen özniteliklerin bazıları eksik gözlemler içermektedir. Bu durum, makine öğrenmesi modellerinin veri bütünlüğüne duyarlılığı göz önünde bulundurularak dikkatle ele alınmıştır. İlgili özniteliklerin zaman serisi niteliği taşıması nedeniyle, eksik verilerin ileri düzey tahmin yöntemleriyle tamamlanması uygun görülmüştür. Bu doğrultuda, “Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton)”, “Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton)” ve “Klinker İhracatı (ton)” öznitelikleri için ayrı ayrı değerlendirme yapılmış ve her bir değişkenin yapısına en uygun eksik veri doldurma yöntemi seçilmiştir. Bu yöntemler belirlenirken hem istatistiksel uygunluk hem de sektörel tutarlılık esas alınmış, alternatif yöntemler de dikkate alınarak gerekçeli tercih yapılmıştır.

Şekil 4.25 ,Şekil 4.26 ve Şekil 4.27'de sırasıyla aylık ortalama, aylık ortalama + parazit ve regresyon yöntemleri ile Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) özniteliğinin eksik veri doldurma işlemlerine ilişkin grafikler sunulmuştur.



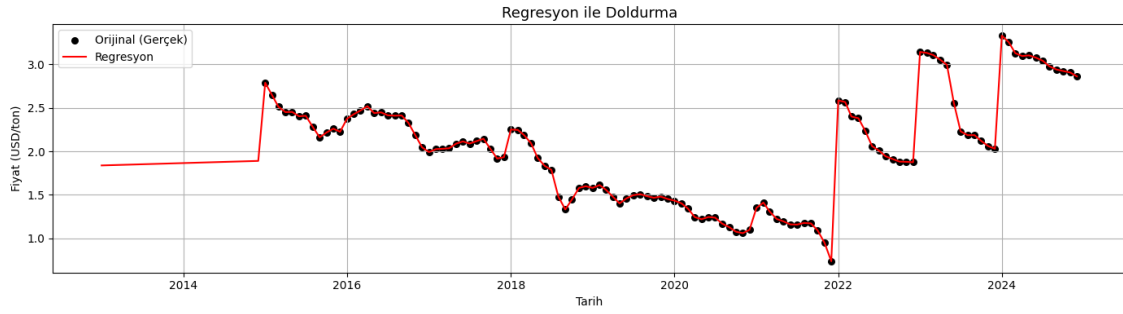
Şekil 4.25. Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) Öznitelığının Aylık Ortalama Yöntemi ile Eksik Verilerin Doldurulması Grafiği

Şekil 4.25'te gösterildiği üzere, eksik verilerin doldurulması için ilk olarak uygulanan aylık ortalama yöntemi, her bir ay için geçmiş yıllarda gözlemlenen değerlerin aritmetik ortalamasını referans alarak eksik gözlemleri tamamlamayı amaçlamaktadır. Bu yaklaşım, mevsimsel desenlerin korunmasına katkı sağlasa da zaman serisi içerisindeki doğal varyasyonu dikkate almamakta ve seri boyunca durağanlık etkisini artırmaktadır. Sonuç olarak, serideki dalgalanmalar bastırılmakta ve aşırı düzleştirilmiş bir yapı ortaya çıkmaktadır.



Şekil 4.26. Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) Öznitelığının Aylık Ortalama + Parazit Yöntemi ile Eksik Verilerin Doldurulması Grafiği

Eksik verilerin doldurulmasında bir diğer yöntem olarak değerlendirilen aylık ortalama + parazit yaklaşımı, ilgili ay için hesaplanan geçmiş dönem ortalamalarına istatistiksel olarak oluşturulmuş rastgele bir hata bileşeni (parazit) eklenmesi esasına dayanmaktadır (Şekil 4.26). Bu yöntem, aylık ortalama yöntemine kıyasla serinin ortalama çizgisine bağlı kalmakla birlikte, zaman serisinin doğal dalgalanma yapısını da koruyarak gerçekçi bir veri örüntüsü sunmaktadır. Böylelikle, serinin yapay olarak düzleştirilmesinin önüne geçilmiş ve modelleme sürecinin gerçek veri dinamiklerine uygunluğu artırılmıştır.

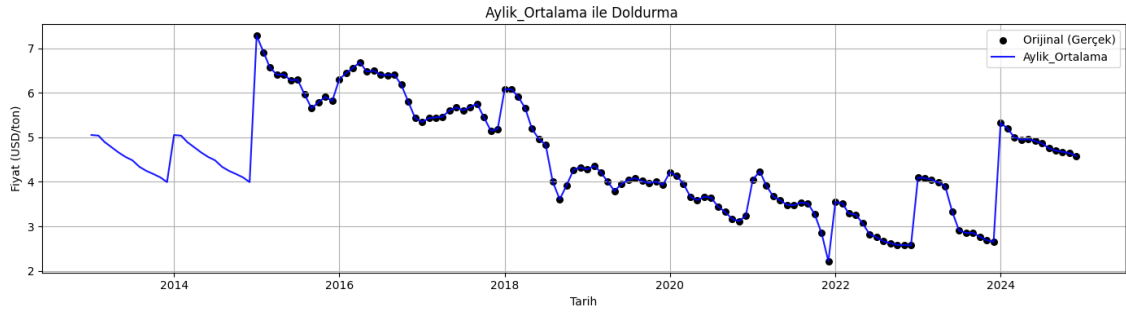


Şekil 4.27. Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Regresyon Yöntemi ile Eksik Verilerin Doldurulması Grafiği

Şekil 4.27’de sunulan tarihe dayalı regresyon yöntemi, tarih bilgisini bağımsız değişken olarak kullanarak eksik veriler için tahminler üretmeyi hedeflemiştir. Bu yöntem kısa vadede kabul edilebilir düzeyde projeksiyonlar sunsa da özellikle 2013–2014 yıllarında hiçbir gözlem verisinin bulunmadığı dönemlerde düşük varyansa sahip, durağan ve düz bir yapı üretmiştir. Bu durum, ilgili yöntemin eğitildiği dönem dışındaki öngörülerde başarısız olduğunu ve serinin yapısını yeterli düzeyde yansıtamadığını göstermektedir.

Gerçekleştirilen teknik ve grafiksel değerlendirmeler neticesinde, eksik gözlemlerin tamamlanması için en uygun yöntemin aylık ortalama + parazit yaklaşımı olduğuna karar verilmiştir. Bu yöntem, yalnızca görsel anlamda serinin doğallığını yansıtmakla kalmamış, aynı zamanda zaman serisi analizleri literatüründe önerilen modern stratejilerle de örtüşmüştür. Böylece, serinin modelleme sürecine uygunluğunun artırılması ve öngörülse doğruluğunun desteklenmesi sağlanmıştır.

Şekil 4.28, Şekil 4.29 ve Şekil 4.30’da sırasıyla aylık ortalama, aylık ortalama + parazit ve regresyon yöntemleri ile Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) özniteliğinin eksik veri doldurma işlemlerine ilişkin grafikler sunulmuştur.

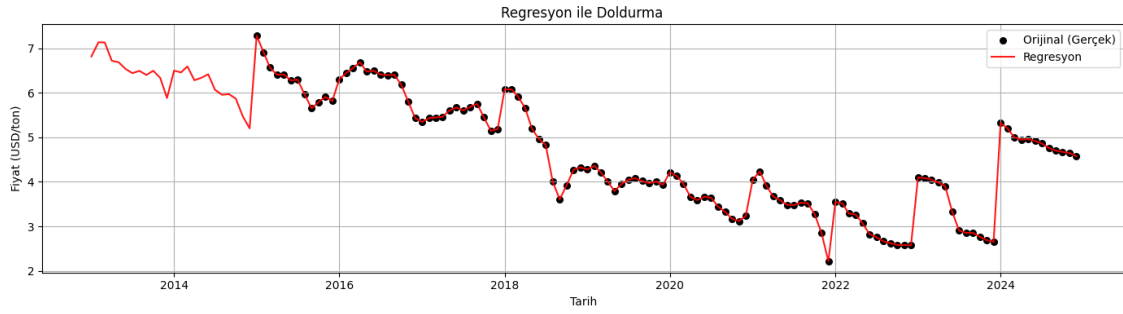


Şekil 4.28. Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Aylık Ortalama Yöntemi ile Eksik Verilerin Doldurulması Grafiği



Şekil 4.29. Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Aylık Ortalama + Parazit Yöntemi ile Eksik Verilerin Doldurulması Grafiği

Şekil 4.28 ve Şekil 4.29'da sunulan grafiklerde, Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) özniteliğine ilişkin eksik verilerin aylık ortalama yöntemi ve bu yöntemin parazit eklenmiş versiyonuyla doldurulması durumları görselleştirilmiştir. Her iki yöntemin de özellikle Ocak 2013–Aralık 2014 döneminde oldukça sabit ve kalıplaşmış değerler ürettiği gözlemlenmiştir. Bu durum, fiyatların zamansal varyasyonlarını yansıtamayan bir yapı ortaya koymakta ve özellikle piyasa koşullarına bağlı olarak beklenen dinamik dalgalanmaların modellenmesinde yetersiz kalmaktadır.



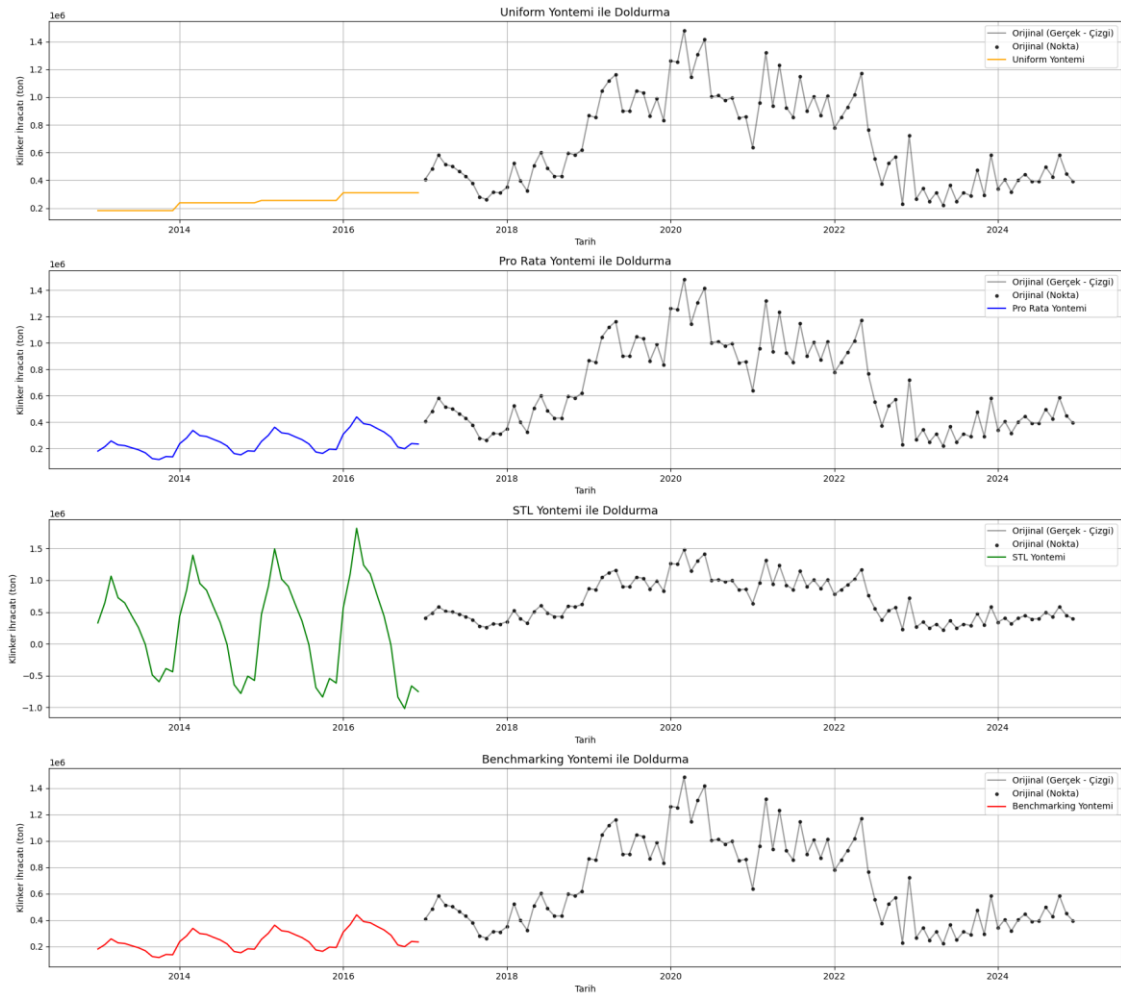
Şekil 4.30. Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) Özniteliğinin Regresyon Yöntemi ile Eksik Verilerin Doldurulması Grafiği

Şekil 4.30’da regresyon tabanlı eksik veri doldurma yönteminin uygulama çıktısı incelendiğinde veri setinde yer alan diğer özniteliklerin açıklayıcı gücünden yararlanılarak eksik dönemlerdeki Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) değerleri tahmin edilmiştir. Elde edilen değerlerin, özellikle 2015 sonrası gözlemlerle benzer eğilimler gösterdiği görülmüş, böylece modelin hem yapısal hem de istatistiksel açıdan anlamlı bir tahminleme performansı sergilediği kanaatine varılmıştır.

Tüm bu teknik ve görsel değerlendirmelerin sonucunda, Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton) özniteliği için regresyon tabanlı eksik veri doldurma yöntemi, modelleme doğruluğu ve veri bütünlüğü açısından en uygun alternatif olarak belirlenmiştir. Bu tercih, sadece görsel benzerliklere değil; aynı zamanda çok değişkenli zaman serisi tahminlemesinde önerilen çağdaş yaklaşımlara da dayanmaktadır. Böylelikle modelleme sürecinde kullanılacak veri setinin yapay etkilerden arındırılarak doğal yapıya uygun olması amaçlanmıştır.

Klinker İhracatı (ton) özniteliği için Ocak 2013–Aralık 2016 dönemine ait yalnızca yıl sonu (Aralık ayı) verilerinin kümülatif olarak sunulmuş olması, eksik alt frekanslı zaman serisi problemi doğurmuştur. Bu tür veri yapıları, literatürde “yıllık düzeyde sağlanan üst frekanslı veriler ile eksik alt frekanslı zaman serileri” olarak tanımlanmakta ve bu yapılar için özel eksik veri tamamlama tekniklerinin uygulanması önerilmektedir (Chow ve Lin, 1971; Denton, 1971). Söz konusu veri eksikliği, klasik interpolasyon yöntemlerinin ötesine geçilerek, aylık tahminlemelerde yapısal ve istatistiksel uyumu sağlayabilecek özel tekniklerle ele alınmıştır.

Şekil 4.31’de sunulan grafik, Uniform, Pro-Rata, STL ve Benchmarking yöntemlerinin uygulanması sonucu elde edilen doldurulmuş serileri karşılaştırmalı olarak göstermektedir. Dört yöntem karşısında yapılan analizlerde özellikle Pro-Rata yönteminin performansı dikkat çekici bulunmuştur. Bu yöntemde yıllık toplam değerler korunarak, geçmiş yılların dağılımına uygun biçimde aylık bazda dağıtım yapılmış ve böylece hem dönem içi süreklilik hem de 2017 sonrası dönem ile kurulan istatistiksel geçiş güçlü biçimde sağlanmıştır.



Şekil 4.31.Klinker İhracatı (ton) Özniteliğinin Çeşitli Yöntemler ile Eksik Verilerin Doldurulması Grafiği

Şekil 4.31’de sunulan grafikler incelendiğinde, STL yöntemi teorik olarak trend ve mevsimsel bileşenleri ayrıştırarak gerçekçi bir yapı sunmayı hedeflese de, uygulamada 2013–2016 döneminde gözlenen sıçramalı yapı yöntemin bu seriye tam uyum sağlayamadığını göstermektedir. Bu durum, kullanılan parametre ayarlarının (örneğin sezon

periyodu) serinin doğasına uygun olmamasından veya eğitim dönemindeki veri yetersizliğinden kaynaklanmış olabilir. STL yöntemi mevsimsel bileşenleri yapay şekilde büyütmüş ve ekonomik gerçeklikten uzak bir varyasyon üretmiştir.

Uniform yöntemi ise tüm eksik dönem boyunca sabit bir değerle doldurma yaptığı için varyasyon içermemekte; bu da modelleme açısından gerçek veriye dayalı öğrenme yetisini sınırlandırmaktadır. Benchmarking yöntemi ise geçmiş benzer dönemlere dayanarak daha yumuşak geçişler sağlamış olsa da, Pro-Rata kadar istikrarlı bir yapı sunamamıştır.

Tüm bu değerlendirmeler sonucunda, “Klinker İhracatı (ton)” özniteliği için eksik verilerin Pro-Rata yöntemi ile doldurulması, serinin doğasına ve modelleme sürecinde gereksinim duyulan süreklilik yapısına en uygun yaklaşım olarak belirlenmiştir.

4.6. Çalışma 1 İçin Modelleme Süreci

Bu bölümde, Çalışma 1 kapsamında uzman görüşü temelli öznitelik seçimiyle oluşturulan veri seti üzerinde gerçekleştirilen makine öğrenmesi uygulamalarına yer verilmektedir. Modelleme süreci, yapı kimyasalı üreticisinin Eskişehir fabrikasına ait gri çimento satın alma fiyat endeksinin tahmini amacıyla yapılandırılmıştır. Süreç boyunca sekiz farklı makine öğrenmesi modeli değerlendirilmiş; her model, önce varsayılan hiperparametreleriyle, ardından GridSearchCV yöntemiyle eniyilenmiş edilmiş parametrelerle çalıştırılmıştır. Bu sayede, hiperparametre ayarlarının model performansı üzerindeki etkileri karşılaştırmalı biçimde analiz edilebilmiştir.

Modelleme sürecinde temel amaç, veri setindeki ilişkileri en doğru şekilde öğrenebilen, geçerli sonuçlar sunabilen ve aşırı öğrenmeden kaçınan modellerin belirlenmesidir. Bu doğrultuda; doğrusal regresyon temelli yöntemlerden topluluk öğrenmesi modellerine, destek vektör makinelerinden örnek tabanlı algoritmalara kadar geniş bir metodolojik yelpazede analizler gerçekleştirilmiştir. Modellerin her biri, belirli varsayımlar ve algoritmik yapılarına uygun olarak yeniden yapılandırılmış ve eğitim-test aşamaları tamamlanmıştır. Kullanılan modeller, varsayılan ve eniyilenmiş hiperparametreler sonucu elde edilen performans ölçütleri ilgili çizelgelerde sunulacaktır.

Çalışma 1 için RF modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.4'te sunulmuştur.

Çizelge 4.4. Çalışma 1 için RF Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
RF	Varsayılan	1,814	6,341	2,518	0,873
RF	Eniyilenmiş	1,818	5,952	2,440	0,881

Çizelge 4.4'ten görüleceği üzere RF modeli hem varsayılan hem de GridSearchCV aracılığıyla eniyilenmiş hiperparametrelerle çalıştırılmış ve model performansı farklı metrikler üzerinden detaylı biçimde değerlendirilmiştir. Eniyilenmiş modelde, MSE 6,34'ten 5,95'e, RMSE ise 2,52'den 2,44'e düşmüştür. Bu gelişme modelin uç değerlerdeki tahmin doğruluğunu artırdığını ve genel hata sapmasını azalttığını göstermektedir. R² değeri 0,873'ten 0,881'e yükselmiş ve modelin hedef değişkendeki varyansın yaklaşık %88'ini açıklayabildiğini ortaya koymuştur. MAE ise eniyilenmiş modelde 1,814'ten 1,818'e çok sınırlı bir artış göstermiştir. Bu durum, modelin genel hata düzeyinde istikrarını koruduğunu ancak RMSE'nin düşmesi ile büyük sapmaların daha iyi yönetildiğini düşündürmektedir. Tüm bu bulgular, hiperparametre eniyilemesinin RF modelinin genelleme yeteneğini geliştirdiğini ve modelleme sürecine katkı sağladığını göstermektedir.

Çalışma 1 için GB modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.5'te sunulmuştur.

Çizelge 4.5. Çalışma 1 için GB Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
GB	Varsayılan	1,992	8,060	2,839	0,838
GB	Eniyilenmiş	2,016	8,268	2,875	0,834

Çizelge 4.5'ten görülebileceği üzere GB algoritması hem varsayılan hem de optimize edilmiş hiperparametrelerle çalıştırılmış; ancak hiperparametre eniyilemesi beklenen performans artışını sağlayamamıştır. Optimize edilen modelde MAE 1,992'den 2,016'ya, MSE 8,06'dan 8,27'ye yükselmiş; bu durum genel hata düzeyinde ve özellikle büyük sapmalarda hafif bir artış olduğunu göstermiştir. RMSE'nin 2,84'ten 2,88'e çıkması,

tahminlerin uç değerlerdeki başarısında da kısmi bir düşüşe işaret etmektedir. R^2 ise 0,838'den 0,834'e gerileyerek, modelin hedef değişkendeki varyansı açıklama gücünde zayıflama yaşandığını ortaya koymuştur. Bu sonuçlar, hiperparametre eniyilemesinin her zaman model performansını artırmadığını ve bazı durumlarda modelin genelleme kabiliyetini olumsuz etkileyebileceğini göstermektedir. Bu bağlamda, elde edilen sonuç, hiperparametre eniyilemesinin dikkatli planlanması gerektiğini ve metodolojik farkındalık açısından önemli bir bulgu olduğunu göstermektedir (Tetko vd., 2024).

Çalışma 1 için XGBoost modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.6'da sunulmuştur.

Çizelge 4.6. Çalışma 1 için XGBoost Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R^2
XGBoost	Varsayılan	1,884	6,581	2,565	0,868
XGBoost	Eniyilenmiş	1,844	7,014	2,648	0,859

Çizelge 4.6'da sunulduğu üzere XGBoost algoritması varsayılan hiperparametrelerle çalıştırıldığında model, $MAE = 1,88$, $MSE = 6,58$ ve $RMSE = 2,57$ değerleri ile düşük hata düzeylerinde bir performans sergilemiştir. $R^2 = 0,868$ ile model, hedef değişkendeki varyansın yaklaşık %86,8'ini öngörebilmiştir. Bu doğrultuda, XGBoost modeli, özellikle RF'ye kıyasla biraz daha düşük varyans açıklamasına rağmen, GB ile benzer hata düzeylerinde güçlü bir alternatif olarak öne çıkmıştır.

Modelin hiperparametre eniyilemesi sonrasında MAE 1,84'e düşerek genel hata seviyesi açısından iyileşme sağlanmış olsa da MSE 'nin 7,01'e ve $RMSE$ 'nin 2,65'e yükselmesi, eniyilenmiş modelin uç değerlerde daha yüksek sapmalar gösterdiğini ortaya koymuştur. Ayrıca R^2 değerinin 0,859'a gerilemesi, varyans açıklama gücünde belirgin bir düşüş yaşandığını göstermektedir. Bu durum, modelin fazla karmaşık hale gelerek aşırı öğrenme riskine girmiş olabileceğine işaret etmektedir. Dolayısıyla, her ne kadar MAE açısından bir miktar iyileşme sağlanmış olsa da genel model başarısı açısından varsayılan yapı daha dengeli bir performans sunmuştur.

Çalışma 1 için LightGBM modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.7’de sunulmuştur.

Çizelge 4.7. Çalışma 1 için LightGBM Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
LightGBM	Varsayılan	2,217	8,972	2,995	0,820
LightGBM	Eniyilenmiş	2,168	8,506	2,917	0,829

Çizelge 4.7’de görüleceği üzere LightGBM modeli varsayılan hiperparametrelerle çalıştırıldığında $MAE = 2,22$, $RMSE = 3,00$ ve $R^2 = 0,82$ değerleri ile ortalama ve uç hata düzeylerinde kabul edilebilir bir performans sergilemiştir. R^2 değeri modelin hedef değişkendeki toplam varyansın yaklaşık %82’sini açıklayabildiğini göstermektedir. Bununla birlikte, aynı veri seti üzerinde çalışan XGBoost ve RF gibi modellerle kıyaslandığında LightGBM’in ortalama hatası biraz daha yüksek kalmıştır. Bu durum, LightGBM’in yaprak-bazında ağaç büyütme stratejisinin daha agresif tahmin davranışları üretme eğilimiyle açıklanabilir.

Modelin hiperparametrelerinin GridSearchCV yöntemiyle optimize edilmesiyle birlikte tüm değerlendirme metriklerinde iyileşme elde edilmiştir. Optimize modelde $MAE = 2,17$, $RMSE = 2,92$ ve $R^2 = 0,83$ olarak hesaplanmıştır. Bu bulgular, hiperparametre ayarlamaları sayesinde hem ortalama hatalarda hem de uç sapmalarda azalma sağlandığını ve modelin varyans açıklama gücünün arttığını ortaya koymaktadır. Dolayısıyla LightGBM, bu çalışmada hem varsayılan hem de optimize haliyle başarılı sonuçlar üretmiş; optimize versiyon ise tüm metriklerde pozitif yönde ilerleme sağlayarak tercih edilebilir bir yapı sunmuştur.

Çalışma 1 için Ridge Regresyon modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.8’de sunulmuştur.

Çizelge 4.8. Çalışma 1 için Ridge Regresyon Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
Ridge Regresyon	Varsayılan	2,348	10,268	3,204	0,794
Ridge Regresyon	Eniyilenmiş	2,261	9,490	3,081	0,810

Çizelge 4.8 incelendiğinde Ridge Regression, doğrusal regresyon modellerinde multicollinearity gibi yapısal sorunlara karşı geliştirilen ve L2 regularizasyonu içeren bir yöntem olarak, veri setinde yer alan öznelilikler arasında yüksek korelasyon olasılığı bulunduğu önemli avantajlar sağlamaktadır. Çalışmada kullanılan varsayılan hiperparametrelerle elde edilen performans metrikleri sırasıyla $MAE = 2,35$, $RMSE = 3,20$ ve $R^2 = 0,794$ olarak hesaplanmıştır. Bu değerler, modelin tahmin başarısının orta seviyede olduğunu ve hedef değişkendeki varyansın yaklaşık %79,4'ünü açıklayabildiğini göstermektedir. Modelin yüksek doğruluk sağlamasa da istikrarlı ve açıklanabilir bir yapı sunduğu, bu yönüyle kompleks modellerin görece dezavantajlarını dengeleyen bir temel model olarak değerlendirilebileceği söylenebilir.

GridSearchCV yöntemiyle gerçekleştirilen hiperparametre eniyilemesi sonucunda modelin performansında anlamlı bir iyileşme gözlemlenmiştir. Optimize edilen model $MAE = 2,26$, $RMSE = 3,08$ ve $R^2 = 0,810$ değerlerine ulaşmıştır. Özellikle $RMSE$ ve MSE değerlerindeki düşüş, modelin uç sapmaları daha etkili biçimde yönetmeye başladığını göstermektedir. Ayrıca varyans açıklama gücünde elde edilen yaklaşık %1,6'lık artış, Ridge Regression'ın ceza parametresi olan alfa değerinin doğru belirlenmesi hâlinde modelin genelleme başarısının anlamlı biçimde artırılabilirliğini ortaya koymaktadır.

Çalışma 1 için Lasso Regresyon modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.9'da sunulmuştur.

Çizelge 4.9. Çalışma 1 için Lasso Regresyon Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R^2
Lasso Regresyon	Varsayılan	4,021	29,168	5,401	0,415
Lasso Regresyon	Eniyilenmiş	2,260	9,490	3,081	0,810

Çizelge 4.9 incelendiğinde görülmektedir ki Lasso Regression modelinin çalıştırılmasıyla, varsayılan hiperparametrelerle elde edilen performans değerleri sırasıyla $MAE = 4,02$, $RMSE = 5,40$ ve $R^2 = 0,415$ olarak belirlenmiştir. Bu metrikler, varsayılan modelin hem ortalama hem uç hatalarda oldukça yetersiz tahminler ürettiğini ve hedef değişkendeki varyansın yalnızca %41,5'ini açıklayabildiğini göstermektedir. Bu durum,

modelin aşırı ceza uygulayarak anlamlı birçok öz niteliği sıfıra indirmesi ve sonucunda ciddi bir yetersiz öğrenme problemi yaşamasıyla açıklanabilir.

Hiperparametre eniyilemesi süreci sonucunda modelin tahmin gücü kayda değer biçimde artmıştır. Optimize model $MAE = 2,26$, $RMSE = 3,08$ ve $R^2 = 0,810$ performans değerlerine ulaşarak, varsayılan modelle karşılaştırıldığında hatalarda yaklaşık %44 azalma ve varyans açıklama oranında neredeyse iki katlık bir iyileşme göstermiştir. Bu sonuçlar, Lasso Regression'ın veri setine özgü parametre ayarlamalarına oldukça duyarlı olduğunu ve uygun koşullarda güçlü bir regresyon aracı hâline gelebildiğini ortaya koymaktadır. Sonuç olarak, bu çalışma kapsamında Lasso Regression modeli, yalnızca yorumlanabilirlik avantajı nedeniyle değil, aynı zamanda doğru ayarlandığında yüksek doğruluk sağlayabilen etkili bir yöntem olarak değerlendirilmiştir.

Çalışma 1 için SVR modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.10'da sunulmuştur.

Çizelge 4.10. Çalışma 1 için SVR Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R^2
SVR	Varsayılan	2,559	12,082	3,476	0,758
SVR	Eniyilenmiş	1,512	4,731	2,175	0,905

Çizelge 4.10 incelendiğinde varsayılan hiperparametrelerle eğitilen SVR modelinin performans metrikleri $MAE = 2,56$, $RMSE = 3,48$ ve $R^2 = 0,758$ olarak elde edilmiştir. Bu sonuçlar, modelin veri setinde yer alan varyansın yaklaşık %75,8'ini açıklayabildiğini ve ortalama hata düzeyinin kabul edilebilir sınırlar içinde olduğunu göstermektedir. Ancak bu düzeydeki performans, çalışmanın yüksek açıklayıcılık hedefi göz önünde bulundurulduğunda yeterli bulunmamış ve hiperparametre eniyilemesine gidilmesi gerekliliği ortaya konmuştur.

SVR modelinin optimize edilmesi sonucunda performans metriklerinde dikkate değer iyileşmeler sağlanmıştır. MAE değeri 1,51'e, RMSE 2,18'e düşerken, R^2 skoru 0,905'e yükselmiştir. Bu değişimler, modelin hem ortalama hem de uç tahmin hatalarında ciddi oranda iyileşme sağladığını göstermekte; aynı zamanda hedef değişkenindeki varyansın

%90'ından fazlasını açıklayabildiğini ortaya koymaktadır. GridSearchCV ile yapılan eniyileme süreci, SVR'nin potansiyel performansını tam anlamıyla ortaya çıkarmıştır. Elde edilen R^2 skoru, çalışmada uygulanan tüm yöntemler arasında en yükseklerden biri olmuş, MAE ve RMSE değerleri ise hata toleransını önemli ölçüde düşürmüştür. Bu bağlamda SVR, gerek doğrusal olmayan ilişkileri modelleme kabiliyeti, gerekse yüksek genelleme performansı ile gri çimento fiyat tahmini problemine etkili bir çözüm sunmuştur.

Çalışma 1 için MLPR modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.11'de sunulmuştur.

Çizelge 4.11. Çalışma 1 için MLPR Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R^2
MLPR	Varsayılan	1,505	4,683	2,164	0,906
MLPR	Eniyilenmiş	1,505	4,690	2,166	0,906

Çizelge 4.11 incelendiğinde çalışma kapsamında uygulanan MLPR modeli, varsayılan hiperparametrelerle dahi son derece başarılı bir tahmin performansı sergilemiştir. Modelin MAE değeri yaklaşık 1,51, RMSE değeri 2,16 ve R^2 skoru 0,906 olarak hesaplanmıştır. Bu metrikler, modelin çimento fiyat tahmininde oldukça düşük hata oranlarıyla çalıştığını ve hedef değişkendeki varyansın %90'ından fazlasını açıklayabildiğini ortaya koymaktadır. MLPR'nin bu yüksek başarısı, doğrusal olmayan ilişkilerin güçlü biçimde varlık gösterdiği veri yapılarında sinir ağlarının etkinliğini bir kez daha kanıtlamaktadır.

Modelin hiperparametre eniyilemesi sonrasında ise MAE, RMSE ve R^2 değerlerinde dikkate değer bir değişim gözlenmemiştir. Optimize edilen modelin MAE değeri 1,504, RMSE değeri 2,165 ve R^2 skoru 0,9059 olarak hesaplanmıştır. Bu sonuçlar, varsayılan modelin zaten veri setine yüksek düzeyde uyum sağladığını ve hiperparametrelerin iyileştirilmesinin anlamlı bir katkı sunmadığını göstermektedir.

Elde edilen yüksek R^2 değeri ve düşük hata metrikleri, MLPR modelini çalışmadaki en başarılı modellerden biri hâline getirmiştir. Bu başarıya karşın modelin aşırı öğrenme göstermemesi ve test verisi üzerindeki başarısını da sürdürebilmesi, genelleme yeteneğinin

yüksek olduğunu göstermektedir. Sonuç olarak, MLPR bu çalışmada yalnızca istatistiksel doğruluk bakımından değil, aynı zamanda modelleme güvenilirliği açısından da öne çıkan yöntemlerden biri olmuştur.

Çalışma 1 kapsamında, satın alma, üretim, satış ve üst yönetim birimlerinde görev yapan uzmanlarla gerçekleştirilen görüşmeler sonucunda belirlenen sektörel bilgiye dayalı özniteliklerle oluşturulmuş veri seti kullanılarak modelleme süreci yürütülmüştür. Bu kapsamda sekiz farklı makine öğrenimi algoritması hem varsayılan hem de optimize edilmiş hiperparametrelerle değerlendirilmiş; modellerin performansları MAE, MSE, RMSE ve R^2 gibi temel regresyon metrikleriyle analiz edilmiştir. Bu çalışma, doğrudan sektör bilgisine dayalı özniteliklerin tahmin gücünü test etmeyi ve karar destek sistemleri için anlamlı içgörüler üretmeyi amaçlamaktadır.

Sonuçlar incelendiğinde, eniyilenmiş MLPR ve SVR modelleri en yüksek açıklayıcılık düzeyine ($R^2 = 0,905$) ulaşarak öne çıkmıştır. Ridge Regresyon ve Lasso Regresyon gibi doğrusal modeller ise özellikle eniyileme sonrası dengeli ancak sınırlı başarı sergilemiştir ($R^2 = 0,810$). Ağaç tabanlı modeller genel olarak güçlü performans göstermiş olsa da, hiperparametre eniyilemesinin performansa etkisi model türüne göre değişiklik göstermiştir. Bu bulgular, sektörel bilgi temelli veri setlerinde doğrusal olmayan ilişkilerin belirgin olduğunu ve esnek modellerin daha başarılı sonuçlar verdiğini göstermektedir. Ayrıca, hiperparametre ayarlarının model başarısına olan katkısının model türüne ve veri yapısına göre önemli ölçüde farklılaştığı da bu çalışma kapsamında gözlemlenmiştir.

4.7. Çalışma 2 İçin Modelleme Süreci

Bu bölümde, korelasyon analizi temelli öznitelik seçimi ile oluşturulan nihai veri seti kullanılarak gerçekleştirilen makine öğrenmesi modelleme süreci açıklanmaktadır. Tıpkı Çalışma 1’de olduğu gibi, bu çalışma kapsamında da RF, GB, XGBoost, LightGBM, Ridge Regresyon, Lasso Regresyon, SVR ve MLPR olmak üzere sekiz farklı makine öğrenmesi algoritması uygulanmıştır. Her bir model, önce varsayılan hiperparametrelerle, ardından GridSearchCV yöntemiyle optimize edilmiş parametrelerle eğitilerek performansları karşılaştırmalı şekilde değerlendirilmiştir. Modellerin hem genel doğrulukları hem de uç değer duyarlılıkları MAE, MSE, RMSE ve R^2 gibi yaygın regresyon metrikleri aracılığıyla

ölçülmüştür. Elde edilen sonuçlara ilişkin sayısal özetler ve yorumlamalar, “Bulgular ve Tartışma” bölümünde detaylı olarak ele alınacaktır. Bu bölümde ise yalnızca her bir modelin kendi içerisindeki değerlendirmelerine yer verilecektir.

Çalışma 2 için RF modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.12’de sunulmuştur.

Çizelge 4.12. Çalışma 2 için RF Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
RF	Varsayılan	2,283	10,215	3,196	0,795
RF	Eniyilenmiş	2,211	9,633	3,104	0,807

Çizelge 4.12 incelendiğinde eniyilenmiş modelin MAE değeri 2,21 iken, varsayılan modelde bu değer 2,28 olarak hesaplanmıştır. Bu düşüş, eniyilenmiş modelin tahminlerinde ortalama sapmayı azaltarak daha tutarlı sonuçlar verdiğini göstermektedir. Eniyilenmiş modelde MSE değeri 10,21’den 9,63’e, RMSE ise 3,20’den 3,10’a düşmüştür. Bu azalmalar, hiperparametre eniyilemesinin özellikle büyük hata değerlerini minimize etmede başarılı olduğunu göstermektedir. Eniyilenmiş RF modelinin R² değeri 0,8069 ile hedef değişkenin yaklaşık %81’lik varyansını açıklayabilmektedir. Varsayılan modelde bu oran %79,5 seviyesindedir. Bu artış, modelin genel olarak daha güçlü bir öğrenme performansı sunduğunu göstermektedir.

Her dört performans metriğinde de eniyilenmiş model, varsayılan modele kıyasla daha iyi sonuçlar vermiştir. Ancak bu iyileşmelerin görece marjinal düzeyde olması, RF algoritmasının zaten varsayılan hiperparametrelerle oldukça başarılı çalıştığını ortaya koymaktadır. Hiperparametre eniyilemesi, özellikle sabit yapıya sahip veri setlerinde küçük fakat istikrarlı iyileştirmeler sağlayarak, karar destek sistemlerinde daha güvenilir öngörüler yapılmasına katkıda bulunabilir.

Bu çalışmada, gri çimento fiyatının tahmini amacıyla hem varsayılan hiperparametrelerle hem de GridSearchCV ile optimize edilmiş hiperparametrelerle uygulanmış ve elde edilen sonuçlar Çalışma 2 için GB modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.13’te sunulmuştur.

Çizelge 4.13. Çalışma 2 için GB Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
GB	Varsayılan	2,631	12,923	3,595	0,741
GB	Eniyilenmiş	2,167	10,097	3,178	0,798

Çizelge 4.13 incelendiğinde eniyilenmiş modelin MAE değeri 2,17 iken, varsayılan modelde bu değer 2,63'tür. Yaklaşık 0,46 birimlik azalma, tahminlerin gerçek değerlere daha yakınlaştığını ve modelin doğruluk açısından daha istikrarlı hale geldiğini göstermektedir. Bu düzeydeki bir MAE, sektörel uygulamalar açısından öngörü başarısının oldukça yüksek olduğunu ortaya koymaktadır. Eniyilenmiş modelde MSE 12,92'den 10,10'a, RMSE ise 3,59'dan 3,18'e düşmüştür. Özellikle RMSE'deki azalma, büyük hataların baskılanabildiğini ve modelin uç tahminlerde daha başarılı hale geldiğini göstermektedir. Bu durum, GB'nin iteratif doğasının ve model ayarlarının tahmin kalitesine olumlu katkı sunduğunu göstermektedir. Varsayılan GB modeli %74,1 oranında varyansı açıklayabilirken, optimize model bu oranı %79,8'e yükseltmiştir. Yaklaşık %6'lık artış, modelin hedef değişkendeki varyasyonu daha etkin şekilde modelleyebildiğini ortaya koymaktadır. R² değerinin bu düzeye çıkması, özellikle karmaşık ekonomik veri setlerinde yüksek model başarısı olarak değerlendirilmektedir.

GB, varsayılan parametrelerle dahi güçlü performans gösteren bir algoritmadır. Ancak hiperparametre eniyilemesi ile MAE, RMSE ve R² metriklerinde kayda değer iyileşmeler elde edilmiştir. Bu durum, modelin hem parametre duyarlılığını hem de veri setine uyum kabiliyetini teyit etmektedir. Özellikle motorin ve elektrik fiyatları, klinker ihracatı gibi karmaşık ekonomik değişkenlerin etkisini yakalama konusunda başarılı olmuştur.

Çalışma 2 kapsamında XGBoost algoritması, gri çimento fiyatının tahmini için hem varsayılan hem de hiperparametreleri eniyilenmiş biçimiyle uygulanmıştır. Elde edilen performans metrikleri Çizelge 4.14'te sunulmuş olup, modeller karşılaştırmalı olarak değerlendirilmiştir.

Çizelge 4.14. Çalışma 2 için XGBoost Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
XGBoost	Varsayılan	2,354	10,268	3,204	0,794
XGBoost	Eniyilenmiş	2,568	12,115	3,481	0,757

Çizelge 4.14 incelendiğinde Varsayılan model, $MAE = 2,35$ ve $RMSE = 3,20$ değerleriyle, tahmin hatalarının makul seviyelerde seyrettiğini göstermektedir. Ayrıca, $R^2 = 0,794$ ile hedef değişkendeki varyansın yaklaşık %79,4'ünü açıklayabilmiştir. Bu sonuçlar, varsayılan XGBoost modelinin yüksek genelleme başarısına sahip olduğunu ortaya koymaktadır. Buna karşılık, eniyilenmiş modelde $MAE = 2,57$, $RMSE = 3,48$ ve $R^2 = 0,757$ seviyelerinde kalmıştır. Hiperparametre ayarları sonrasında modelin hataları artmış ve varyans açıklama gücü düşmüştür. Bu durum, eniyilenmiş modelin daha karmaşık hâle gelerek genelleme kabiliyetini yitirdiğini ve muhtemel bir aşırı öğrenme durumu yaşadığını düşündürmektedir.

Bu çalışmada, LightGBM algoritması, gri çimento fiyatlarının tahmininde hem varsayılan hem de hiperparametreleri eniyilenmiş biçimiyle değerlendirilmiştir. Elde edilen performans metrikleri Çizelge 4.15'te sunulmuştur.

Çizelge 4.15. Çalışma 2 için LightGBM Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
LightGBM	Varsayılan	2,801	14,446	3,801	0,710
LightGBM	Eniyilenmiş	2,736	14,402	3,795	0,711

Çizelge 4.15 incelendiğinde LightGBM'in varsayılan konfigürasyonu, $MAE = 2,80$ ve $RMSE = 3,80$ seviyelerinde bir hata ile tahmin gerçekleştirmiş; bu yapı hedef değişkendeki varyansın yaklaşık %71'ini açıklamıştır. Bu sonuçlar, modelin güçlü bir ilk tahmin yetisine sahip olduğunu göstermektedir. Eniyilenmiş modelde ise $MAE 2,7360$, $RMSE 3,7949$ ve $R^2 0,7113$ olarak elde edilmiştir. Bu değerler, varsayılan yapı ile neredeyse aynı seviyededir. Eniyilenmiş modelde yalnızca marjinal iyileşmeler gözlenmiş, modelin genel tahmin gücünde anlamlı bir sıçrama yaşanmamıştır. Hiperparametre eniyilenmesi sonucunda anlamlı bir gelişme elde edilmemiş olması, bu sürecin gereksiz olduğunu değil;

modelin veriye oldukça uygun olduğunu ve varsayılan ayarların bu veri yapısı için yeterli olduğunu göstermektedir.

Bu çalışmada Ridge modeli, gri çimento fiyatlarının tahminlenmesinde hem varsayılan hem de eniyilenmiş alfa parametresi ile değerlendirilmiştir. Çalışma 2 için Ridge Regresyon modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.16’da sunulmuştur.

Çizelge 4.16. Çalışma 2 için Ridge Regresyon Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
Ridge Regresyon	Varsayılan	3,109	14,563	3,816	0,708
Ridge Regresyon	Eniyilenmiş	3,109	14,563	3,816	0,708

Çizelge 4.16 incelendiğinde elde edilen sonuçlara göre, Ridge Regression modelinin varsayılan ve eniyilenmiş versiyonları tamamen aynı performans metriklerine sahiptir (MAE = 3,11, RMSE = 3,82, R² = 0,708). Bu durum, eniyilenen alfa hiperparametresinin zaten varsayılan değer olan 1,0 olarak belirlenmiş olduğunu ve modelin bu değerde en uygun genel başarıyı sağladığını göstermektedir. Bu sonuçlar, Ridge Regression’ın belirli koşullarda kararlı ve deterministik bir yapı sunduğuna işaret etmektedir. R² değerinin yaklaşık %70 seviyesinde seyretmesi, modelin hedef değişkenin varyansını tatmin edici düzeyde açıkladığını göstermektedir. Ancak modelin düşük varyans ve düşük parametrik esneklikle çalışıyor olması, daha ileri performans iyileştirmelerine açık olmadığını da ortaya koymaktadır. Bu nedenle Ridge Regression, istikrarlı ve açıklanabilir bir tahmin modeli olarak öne çıkarken, karmaşık yapıları öğrenme kapasitesi açısından sınırlı kalmaktadır.

Çalışma 2 için Lasso Regresyon modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.17’te sunulmuştur.

Çizelge 4.17. Çalışma 2 için Lasso Regresyon Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
Lasso Regresyon	Varsayılan	3,044	13,565	3,683	0,728
Lasso Regresyon	Eniyilenmiş	3,132	14,989	3,872	0,700

Çizelge 4.17 incelendiğinde Çalışma 2 kapsamında kullanılan varsayılan hiperparametrelerle çalıştırılan Lasso Regresyon modeli, $MAE = 3,04$, $RMSE = 3,68$ ve $R^2 = 0,728$ değerleriyle makul düzeyde bir performans sergilemiştir. Model, hedef değişken olan gri çimento fiyatlarındaki varyansın yaklaşık %73'ünü açıklayabilmiş ve sektörel uygulamalarda kabul edilebilir sınırlar içerisinde kalmıştır. Hiperparametre eniyilenmesi sonucunda elde edilen model, performans açısından beklenenin aksine gerileme göstermiştir. Eniyilenmiş alfa değeri sonucunda; MAE 3,13'e çıkmış, $RMSE$ 3,87 düzeyine yükselmiş ve R^2 ise %69,95'e düşmüştür. Bu sonuçlar, eniyileme sürecinde belirlenen hiperparametre değerinin fazla yüksek olabileceğini ve bu nedenle modelin bazı anlamlı öznitelikleri de sıfırladığını göstermektedir. Bu durum literatürde “over-regularization” olarak tanımlanır ve Lasso'nun doğasında bulunan aşırı sadeleştirici eğilimle açıklanabilir (James et al., 2021)

Çalışma 2 için SVR modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.18'te sunulmuştur.

Çizelge 4.18. Çalışma 2 için SVR Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R^2
SVR	Varsayılan	2,868	13,177	3,630	0,736
SVR	Eniyilenmiş	1,755	5,198	2,280	0,896

Çizelge 4.18 incelendiğinde eniyilenmiş SVR modeli, Çalışma 2 kapsamında $MAE = 1,75$, $RMSE = 2,28$ ve $R^2 = 0,896$ değerlerine ulaşarak tüm modeller içinde en yüksek doğruluk düzeylerinden birini sağlamıştır. Model, hedef değişkenin varyansının %90'a yakını açıklayabilmiş ve tahmin hatalarını oldukça düşük seviyelerde tutmuştur. Bu başarı, SVR'ın kernel tabanlı öğrenme yapısının, doğrusal olmayan örüntüleri etkin biçimde modelleyebildiğini göstermektedir. Çimento fiyatlarını etkileyen motorin fiyatı, üretici fiyat endeksi, ihracat düzeyleri gibi birbirleriyle karmaşık ilişkiler içeren değişkenlerin model tarafından doğru şekilde işlenmesi bu durumu desteklemektedir.

Varsayılan SVR modeline kıyasla eniyilenmiş modelin; MAE değeri yaklaşık %39, $RMSE$ değeri %37, MSE değeri %60'tan fazla azaltılmış, R^2 değeri ise yaklaşık %16 oranında artırılmıştır. Bu dramatik iyileşmeler, SVR algoritmasının hiperparametrelere olan

duyarlılığını açıkça ortaya koymaktadır. GridSearchCV ile yürütülen eniyileme süreci, modelin hem duyarlılığını artırmış hem de genelleme kapasitesini enbüyüklemiştir.

Çalışma 2 için MLPR modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.19’da sunulmuştur.

Çizelge 4.19. Çalışma 2 için MLPR Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
MLPR	Varsayılan	3,618	19,471	4,413	0,610
MLPR	Eniyilenmiş	2,175	8,182	2,860	0,836

Çizelge 4.19 incelendiğinde eniyilenmiş model; $R^2 = 0,836$ ile hedef değişkendeki varyansın yaklaşık %83,6’sını açıklayabilmiş, $RMSE = 2,86$ ve $MAE = 2,17$ ile tahmin doğruluğu açısından en güçlü modeller arasında yer almıştır. Bu sonuçlar, MLPR’nin karmaşık ve doğrusal olmayan örüntüleri başarılı bir şekilde yakalayabildiğini göstermektedir. Model, SVR gibi güçlü parametrik olmayan algoritmalarla rekabet edebilecek düzeyde düşük hata üretmiştir. Ayrıca eniyilenmiş model; MAE’yi %39,9, RMSE’yi %35.2, R^2 ’yi ise yaklaşık %22,6 oranında iyileştirmiştir. Bu dramatik iyileşme, yapay sinir ağlarının hiperparametrelere olan hassasiyetini açıkça ortaya koymaktadır. Gizli katman sayısı, nöron sayısı, aktivasyon fonksiyonu ve öğrenme oranı gibi parametrelerin doğru seçimi sayesinde modelin genelleme kabiliyeti önemli ölçüde artmıştır.

Çalışma 2 kapsamında, yalnızca akademik literatürde önerilen ve istatistiksel araçlarla belirlenmiş özniteliklerin yer aldığı daha kompakt ve bilimsel temelli bir veri seti ile modelleme süreci yürütülmüştür. Bu veri seti üzerinde sekiz farklı makine öğrenimi algoritması hem varsayılan hem de eniyilenmiş hiperparametrelerle test edilerek dört temel regresyon metriği (MAE, MSE, RMSE, R^2) üzerinden değerlendirilmiştir. Bu çalışmada amaç, doğrusal ve doğrusal olmayan modellerin hem kendi içlerindeki gelişimini hem de bu daha soyut ve sistematik öznitelik kümesine uyumunu gözlemlemektir.

Analiz sonuçlarına göre, en yüksek doğruluk düzeyi SVR modelinde ($R^2 = 0,8958$) elde edilmiştir. SVR’yi RF, XGBoost ve MLPR gibi modeller yakından takip etmiştir. Ridge Regresyon ve Lasso Regresyon gibi doğrusal modeller ise görece düşük açıklama gücü

sunmuştur. Bu durum, veri setinde doğrusal olmayan ilişkilerin baskın olduğunu ve parametrik modellerin sınırlı başarı sağladığını göstermektedir. Ayrıca, MLP ve SVR gibi modellere uygulanan hiperparametre eniyilemesi başarıyı önemli ölçüde artırırken; LightGBM ve Lasso Regresyon gibi modellerde varsayılan parametrelerin dahi daha etkili sonuçlar verdiği gözlemlenmiştir. Sonuç olarak, Çalışma 2, daha kompakt fakat güçlü anlam taşıyan korelasyon analizi temelli özniteliklerin regresyon başarısına etkisini test etme açısından önemli bulgular sunmuştur.

4.8. Çalışma 3: Öznitelik Birleştirme ve Modelleme

Bu çalışma kapsamında, daha önce bağımsız şekilde değerlendirilen Çalışma 1 (uzman görüşü temelli öznitelikler) ve Çalışma 2 (korelasyon analizi temelli öznitelikler) veri setleri birleştirilerek daha zengin bir öznitelik yapısı elde edilmiştir. Amaç hem sektörel bilgi hem de korelasyon analizi temelli çıkarımları aynı modelleme çatısı altında bir araya getirerek tahminleme performansını artırmak ve öngörü gücü yüksek bir model yapısı oluşturmaktır. Birleştirilen veri seti üzerinde aynı sekiz makine öğrenmesi algoritması yeniden eğitilmiş, her iki hiperparametre seti için (varsayılan ve hiperparametre eniyilemesi sonrası elde edilen) performans metrikleri incelenmiştir. Çalışmanın temel amacı, farklı seçim stratejilerinden gelen özniteliklerin sinerjik etkisinin tahminleme doğruluğuna katkısını ölçmektir.

Çalışma 3 için RF modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.20’de sunulmuştur.

Çizelge 4.20. Çalışma 3 için RF Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
RF	Varsayılan	1,840	5,736	2,395	0,885
RF	Eniyilenmiş	1,860	6,060	2,462	0,879

Çizelge 4.20 incelendiğinde; modelin varsayılan hâliyle elde ettiği R² değeri 0,885 olup, hedef değişkenin yaklaşık %88,5’ini açıklayabildiği görülmektedir. Bu oran, sektörel öngörüleme uygulamalarında oldukça yüksek kabul edilir. MAE (1,84) ve RMSE (2,40) değerleri de düşük seviyelerde kalmış, modelin genellikle hedefe yakın tahminler ürettiği

anlaşılmıştır. Eniyilenen modelde ise R^2 değeri 0,879'a gerilemiş, RMSE (2,46) ve MAE (1,86) değerlerinde ise hafif artış gözlemlenmiştir. Bu durum, bazı durumlarda hiperparametre eniyilemesinin modelin genel doğruluğunu sınırlı ölçüde artırabileceğini ya da düşürebileceğini ortaya koymaktadır. Elde edilen sonuçlar, RF algoritmasının hibrit öznitelik seti üzerinde güçlü performans gösterdiğini, ancak hiperparametre eniyilemesinin bu örnekte varsayılan yapıdan daha iyi sonuç vermediğini ortaya koymuştur. Bu durum, modelin yapay olarak karmaşılaştırılmasının her zaman fayda sağlamayabileceğine işaret etmektedir.

Gri çimento fiyatının tahmini amacıyla hem varsayılan hiperparametrelerle hem de GridSearchCV ile eniyilenmiş hiperparametrelerle GB modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.21'de sunulmuştur.

Çizelge 4.21. Çalışma 3 için GB Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R^2
GB	Varsayılan	1,877	7,113	2,667	0,857
GB	Eniyilenmiş	1,849	7,038	2,653	0,859

Çizelge 4.21 incelendiğinde; modelin varsayılan haliyle elde edilen R^2 değeri 0,857 iken, eniyilenen modelde bu değer 0,859'a yükselmiştir. Her iki modelde de hata metrikleri düşük düzeydedir: RMSE yaklaşık 2,65, MAE ise 1,85–1,88 aralığındadır. Eniyilemenin etkisi görece sınırlı kalmış, ancak modelin genelleme kapasitesi hafifçe artmıştır. Bu durum, GB algoritmasının varsayılan parametrelerle dahi oldukça dengeli ve güvenilir sonuçlar üretebildiğini göstermektedir.

Çalışma 3 kapsamında XGBoost algoritması, gri çimento fiyatının tahmini için hem varsayılan hem de hiperparametreleri optimize edilmiş biçimiyle uygulanmıştır. Elde edilen performans metrikleri Çizelge 4.22'de sunulmuş olup, modeller karşılaştırmalı olarak değerlendirilmiştir.

Çizelge 4.22. Çalışma 3 için XGBoost Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R^2
XGBoost	Varsayılan	1,828	6,085	2,467	0,878
XGBoost	Eniyilenmiş	1,996	7,191	2,682	0,856

Çizelge 4.22 incelendiğinde; modelin varsayılan parametrelerle elde ettiği R^2 değeri 0,878 ile oldukça yüksek bir açıklama gücü sergilemiştir. RMSE değeri 2,47 ve MAE değeri 1,83 gibi düşük hata metrikleriyle desteklenen bu sonuçlar, XGBoost'un verideki örüntüleri doğru biçimde öğrenebildiğini göstermektedir ancak hiperparametre eniyilemesi sonrasında performansın beklenenin aksine azaldığı gözlemlenmiştir. Eniyilenmiş modelin R^2 değeri 0,856'ya düşmüş, RMSE ve MAE değerlerinde de artış yaşanmıştır. Bu durum, parametre ayarlarının modelin genelleme gücünü sınırlamış olabileceğini düşündürmektedir. XGBoost, özellikle çok sayıda gözlem içeren, yüksek boyutlu ve parazitli veri setlerinde olağanüstü başarı gösteren bir algoritmadır. Ancak bu çalışmadaki hibrit öznitelik yapısı ve dengeli gözlem sayısı, XGBoost'un karakteristik avantajlarını sınırlamış olabilir.

LightGBM algoritması, gri çimento fiyatlarının tahmininde hem varsayılan hem de hiperparametreleri eniyilenmiş biçimiyle değerlendirilmiştir. Elde edilen performans metrikleri Çizelge 4.23'te sunulmuştur.

Çizelge 4.23. Çalışma 3 için LightGBM Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R^2
LightGBM	Varsayılan	2,186	7,332	2,708	0,853
LightGBM	Eniyilenmiş	2,176	7,505	2,740	0,850

Çizelge 4.23 incelendiğinde; LightGBM varsayılan parametrelerle %85,3 (R^2) açıklama gücüne ulaşırken, eniyilenmiş modelin R^2 değeri %85,0 seviyesinde kalmıştır. Her iki durumda da MAE ve RMSE değerleri oldukça benzer düzeylerde seyretmiş, eniyilemenin model performansına yalnızca marjinal düzeyde etki ettiği görülmüştür. Eniyilenmiş model ile varsayılan model arasındaki farkın düşük olması, varsayılan ayarların zaten bu veri seti için oldukça yeterli olduğunu göstermektedir.

Çalışma 3 için Ridge Regresyom modelinin çalıştırılmasıyla elde edilen performans metrikleri Çizelge 4.24'te sunulmuştur.

Çizelge 4.24. Çalışma 3 için Ridge Regresyon Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
Ridge Regresyon	Varsayılan	2,588	10,656	3,264	0,786
Ridge Regresyon	Eniyilenmiş	2,396	9,848	3,138	0,803

Çizelge 4.24 incelendiğinde varsayılan hiperparametreler ile çalıştırılan Ridge modeli, %78,6 düzeyinde açıklama gücü (R^2) üretmiştir; bu değer, model için tatmin edici seviyededir. Hiperparametre eniyilemesi sonrasında R^2 skoru %80,3'e yükselmiş, MAE ve RMSE değerleri ise düşmüştür. Bu iyileşme, modelin doğruluk düzeyinde ölçülebilir bir gelişim sağladığını göstermektedir.

Çalışma 3 için Lasso Regresyon modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.25'de sunulmuştur.

Çizelge 4.25. Çalışma 3 için Lasso Regresyon Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
Lasso Regresyon	Varsayılan	2,962	12,916	3,594	0,741
Lasso Regresyon	Eniyilenmiş	2,395	9,848	3,138	0,803

Çizelge 4.25 incelendiğinde; varsayılan hiperparametrelerle çalışan Lasso modeli %74,1 düzeyinde açıklama gücü üretmiş ve RMSE değeri 3,59 olarak hesaplanmıştır fakat hiperparametre eniyilemesi sonrasında modelin R^2 değeri %80,3'e, RMSE değeri ise 3,14'e gerilemiştir. Bu gelişme, modelin esnekliğinin artırıldığını ve gereksiz cezalandırmaların önüne geçildiğini göstermektedir. Elde edilen sonuçlar, Ridge modeli ile oldukça benzer bir doğruluk düzeyi sunmuş, ancak Lasso'nun sıfır katsayılı öznitelikler üretmesiyle daha sade bir yapı ortaya koyduğu anlaşılmıştır. Çalışma 3 veri seti çok yüksek boyutlu olmamakla birlikte, Lasso bu yapı içerisinde anlamlı bir öznitelik daraltması uygulayarak modele açıklanabilirlik kazandırmıştır. Bu davranış, özellikle aşırı öğrenme riskine karşı koruyucu bir nitelik taşır. Doğrusal olmayan yapıların öne çıktığı bu çalışmada topluluk öğrenmesi yöntemleri daha yüksek başarı sağlasa da Lasso Regression açıklanabilirlik bakımından önemli bir alternatif olarak öne çıkmıştır.

Çalışma 3 için SVR modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.26'da sunulmuştur.

Çizelge 4.26. Çalışma 3 için SVR Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
R	Varsayılan	2,589	11,470	3,387	0,770
SVR	Eniyilenmiş	1,787	6,860	2,619	0,862

Çizelge 4.26 incelendiğinde SVR modelinin varsayılan parametrelerle çalıştırılması hâlinde R² skoru %77,0 düzeyinde kalmış ve tahmin hataları görece yüksek gözlenmiştir. Ancak hiperparametre eniyilemesi sonrası modelin açıklama gücü %86,2'ye yükselmiş; RMSE ve MAE değerleri anlamlı şekilde düşmüştür. Bu gelişme, hiperparametre eniyilenmesinde hassas ayarların tercih edilmesinin model performansı üzerindeki olumlu etkisini göstermektedir.

Çalışma 3 için MLPR modelinin çalıştırılması ile elde edilen performans metriklerinin sonuçları Çizelge 4.27'de sunulmuştur.

Çizelge 4.27. Çalışma 3 için MLPR Modeli Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
MLPR	Varsayılan	2,899	13,817	3,717	0,723
MLPR	Eniyilenmiş	1,405	4,566	2,137	0,908

Çizelge 4.27 incelendiğinde; MLPR modelinin varsayılan hiperparametreler ile çalıştırılması sonucu elde edilen R² değeri %72,3 seviyesindeyken, eniyilenen modelde bu oran %90,8'e kadar yükselmiştir. Aynı şekilde RMSE 3,72'den 2.14'e, MAE ise 2,90'dan 1,41'e düşerek ciddi oranda hata azalması sağlanmıştır. Bu gelişme, modelin doğru hiperparametrelerle yapılandırıldığında öğrenme kapasitesinin dramatik biçimde arttığını ortaya koymaktadır. Hibrit öznitelik kümesinde yer alan sektörel, ekonomik, politik ve demografik değişkenlerin etkileşimli ve çok boyutlu doğası, MLPR gibi derin öğrenmeye dayalı modeller için oldukça elverişli bir ortam sunmuştur. Sinir ağının yüksek esneklik kabiliyeti sayesinde, bu karmaşık ilişkiler başarıyla öğrenilmiş ve model tahmin gücü açısından SVR, aşamalı öğrenme ve topluluk öğrenmesi algoritmalarını geride bırakmıştır.

Çalışma 3 kapsamında, önceki iki çalışmada kullanılan sektörel ve korelasyon analizi temelli öznitelikler birleştirilerek "hibrit" bir veri seti oluşturulmuştur. Bu yeni özellik

kümesi, ekonomik, demografik, enerji, politik ve sektörel değişkenlerin etkileşimli biçimde modellenmesini amaçlamaktadır. Hibrit veri seti üzerinde toplam sekiz farklı makine öğrenmesi algoritması hem varsayılan hem de eniyilenmiş edilmiş hiperparametrelerle uygulanarak, algoritmaların performansları dört temel ölçüt (MAE, MSE, RMSE, R^2) üzerinden değerlendirilmiştir.

Gerçekleştirilen deneyler sonucunda en yüksek başarıyı hiperparametreleri eniyilenmiş MLPR modeli sağlamış; bu model %90,8 oranında açıklayıcılıkla öne çıkmıştır. RF, GB ve SVR gibi modeller de yüksek doğruluk düzeylerine ulaşmıştır. Parametrik modeller olan Ridge ve Lasso ise görece olarak daha düşük performans göstermiştir. Hiperparametre eniyilemesinin özellikle MLP ve SVR gibi modellerde belirgin iyileşmeler sağladığı, topluluk öğrenmesi modellerinde ise varsayılan parametrelerin dahi yeterli düzeyde başarı sağladığı görülmüştür. Bu bulgular, öznel birleştirimin model başarısı üzerindeki etkisini gözler önüne sermekte ve makine öğrenimi algoritmalarının veri seti yapısına göre farklı başarı seviyeleri gösterebildiğini ortaya koymaktadır.

4.9. Çalışma 4: Kapasite Kullanım Oranı Özneliğinin Modelden Çıkarılması

Bu bölümde, önceki aşamalarda herhangi sektörel bir bilgiye dayanılmadan, sezgiler üzerine veri setine eklenmiş olan “Kapasite Kullanım Oranı” özneliği veri setinden çıkarılarak, makine öğrenimi algoritmalarının bu değişken olmaksızın nasıl performans sergilediği analiz edilmiştir. Amaç, kapasite değişkeninin model doğruluğuna katkısını dolaylı biçimde test etmek ve bu öznelik olmadan modellerin genelleme gücünü ölçmektir. Bu kapsamda toplam sekiz farklı makine öğrenmesi algoritması hem varsayılan hem de hiperparametreleri eniyilenmiş biçimde değerlendirilmiş; modeller MAE, RMSE ve R^2 metrikleri açısından kıyaslanmıştır. Çalışmanın sonunda, özellikle topluluk öğrenmesi tabanlı ve sinir ağı modellerinin yüksek başarı düzeyini koruduğu, doğrusal modellerin ise sıralamada geride kaldığı gözlemlenmiştir.

Çalışma 4 için makine öğrenmesi modellerinin çalıştırılması sonucu elde edilen performans metrikleri Çizelge 4.28’de sunulmuştur.

Çizelge 4.28. Çalışma 4 için Makine Öğrenmesi Modellerinin Çalıştırılması Sonucu Elde Edilen Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
RF	Varsayılan	1,780	5,605	2,367	0,888
RF	Eniyilenmiş	1,871	6,108	2,471	0,878
GB	Varsayılan	1,909	6,979	2,642	0,860
GB	Eniyilenmiş	1,864	6,876	2,622	0,862
XGBoost	Varsayılan	1,788	5,864	2,422	0,882
XGBoost	Eniyilenmiş	2,116	7,787	2,791	0,844
LightGBM	Varsayılan	2,023	6,669	2,582	0,866
LightGBM	Eniyilenmiş	1,941	6,776	2,603	0,864
Ridge Regresyon	Varsayılan	2,576	10,480	3,237	0,790
Ridge Regresyon	Eniyilenmiş	2,395	9,843	3,137	0,803
Lasso Regresyon	Varsayılan	2,938	12,943	3,598	0,741
Lasso Regresyon	Eniyilenmiş	2,394	9,842	3,137	0,803
SVR	Varsayılan	2,609	11,290	3,360	0,774
SVR	Eniyilenmiş	1,839	6,731	2,594	0,865
MLPR	Varsayılan	2,011	7,879	2,807	0,842
MLPR	Eniyilenmiş	1,668	5,932	2,435	0,881

Çizelge 4.28 incelendiğinde ilk dikkat çeken sonuç, RF (Varsayılan) modelinin, kapasite kullanım oranı olmaksızın dahi en yüksek R² skoruna ulaşarak (0,8876) tüm algoritmaları geride bırakmasıdır. Bu bulgu, RF algoritmasının değişkenler arası etkileşimleri doğal olarak modelleyebilme ve bilgi fazlalığını avantaja çevirebilme gücünü destekler niteliktedir. Dahası, modelin varsayılan hiperparametrelerle eniyilenmiş versiyonundan daha iyi performans göstermesi, RF algoritmasının genellikle varsayılan ayarlarla dahi oldukça stabil ve genelleyebilir sonuçlar sunduğuna işaret etmektedir.

XGBoost (Varsayılan) modeli ise 0,8824'lük R² skoru ile ikinci sırada yer almış ve özellikle kapasite gibi güçlü bir açıklayıcı değişkenin yokluğunda dahi yüksek tahmin gücünü koruyarak dikkat çekmiştir. Bu performans, XGBoost'un düzenleme (regularization) yetenekleri sayesinde aşırı öğrenme riskini minimize ederek sağlam tahminler sunabilmesini desteklemektedir. XGBoost'un eniyilenmiş versiyonunun varsayılanın gerisinde kalması, hiperparametre arama alanının yetersizliği ya da modelin aşırı öğrenmeye yönelmiş olmasıyla açıklanabilir.

İkinci sıradaki ağaç tabanlı modellerin ardından, MLPR (Eniyilemiş) modelinin sergilediği yüksek başarı ($R^2 = 0,8811$; $MAE = 1,668$) özellikle dikkat çekicidir. Sinir ağları doğrusal olmayan ilişkileri öğrenmede oldukça etkilidir ve bu modelin başarısı, veri setinde kompleks etkileşimlerin varlığına işaret etmektedir. MAE değerinin en düşük seviyede olması ise, modelin tahminlerinin hedefe daha yakın olduğunu ve sapmaların büyük ölçüde enküçüklendiğini göstermektedir. Bu durum, kapasite kullanım oranı çıkarıldığında dahi sinir ağı yapılarının genelleme gücünü koruyabildiğini ortaya koymaktadır.

Çalışmada ayrıca, hiperparametre eniyilemesinin tüm modeller için fayda sağlamadığı gözlemlenmiştir. Özellikle RF ve XGBoost gibi modellerde, varsayılan parametreler optimize edilmiş versiyonlara kıyasla daha yüksek R^2 skorları sağlamıştır.

Son olarak, LightGBM, Ridge Regresyon ve Lasso Regresyon gibi modeller nispeten daha düşük başarı seviyelerinde kalmıştır. LightGBM modeli varsayılan haliyle %86,6 R^2 skoruna ulaşarak tatmin edici sonuçlar sunsa da XGBoost'un gerisinde kalmıştır. Öte yandan, doğrusal modeller olan Ridge Regresyon ve Lasso Regresyon, sıralamanın sonlarında yer alarak veri setindeki ilişkilerin doğrusal sınırları aştığını bir kez daha ortaya koymuştur. Bu sonuçlar, daha karmaşık ve esnek model yapılarına duyulan ihtiyacı doğrulamaktadır.

4.10. Çalışma 5: Korelasyon Tabanlı Nihai Öznitelik İndirgeme

Bu bölümde, Çalışma 3'te kullanılan geniş kapsamlı hibrit öznitelik kümesinden yola çıkılarak korelasyon analizi sonrasında çoklu bağlantı riski taşıyan öznitelikler elenmiş, sadeleştirilmiş ve daha kompakt bir veri seti oluşturulmuştur, oluşturulan veri setini içeren çalışma, Çalışma 5 olarak isimlendirilmiştir. Amaç, daha az sayıda ancak güçlü etki taşıyan özniteliklerle modellerin ne düzeyde başarılı olabileceğini test etmektir. Bu kapsamda toplam sekiz regresyon algoritması hem varsayılan hem de eniyilenmiş parametrelerle test edilerek MAE, RMSE ve R^2 metrikleri üzerinden değerlendirilmiştir. Elde edilen sonuçlar, bazı modellerin daha az veriyle bile oldukça yüksek başarı sağlayabildiğini, bazılarının ise öznitelik indirgeme sürecinden olumsuz etkilendiğini ortaya koymuştur.

Çalışma 5 için makine öğrenmesi modellerinin çalıştırılması sonucu elde edilen performans metrikleri Çizelge 4.29'da sunulmuştur.

Çizelge 4.29. Çalışma 5 için Makine Öğrenmesi Modellerinin Çalıştırılması Sonucu Elde Edilen Performans Metrikleri

Model	Tür	MAE	MSE	RMSE	R ²
RF	Varsayılan	2,149	7,764	2,786	0,844
RF	Eniyilenmiş	2,150	7,864	2,804	0,842
GB	Varsayılan	2,414	9,434	3,071	0,811
GB	Eniyilenmiş	2,408	9,527	3,087	0,809
XGBoost	Varsayılan	2,019	7,874	2,806	0,842
XGBoost	Eniyilenmiş	2,252	8,637	2,939	0,827
LightGBM	Varsayılan	2,696	12,237	3,498	0,755
LightGBM	Eniyilenmiş	2,749	12,880	3,589	0,742
Ridge Regresyon	Varsayılan	2,573	9,918	3,149	0,801
Ridge Regresyon	Eniyilenmiş	2,573	9,918	3,149	0,801
Lasso Regresyon	Varsayılan	2,771	11,840	3,441	0,763
Lasso Regresyon	Eniyilenmiş	2,581	10,061	3,172	0,798
SVR	Varsayılan	2,755	13,031	3,610	0,739
SVR	Eniyilenmiş	1,578	4,990	2,234	0,900
MLPR	Varsayılan	1,779	5,741	2,396	0,885
MLPR	Eniyilenmiş	2,004	7,149	2,674	0,857

Çizelge 4.29 incelendiğinde SVR - Eniyilenmiş modeli, $R^2 = 0,900$ skoru ile bu çalışmanın en başarılı algoritması olarak öne çıkmıştır. Bu sonuç, SVR'nin az sayıda ancak etkili öz nitelik içeren veri setlerinde yüksek genel doğruluk ve düşük hata düzeyleriyle öne çıkabildiğini göstermektedir.

İkinci olarak, MLPR modeli, hem varsayılan ($R^2 = 0,885$) hem de eniyilenmiş ($R^2 = 0,857$) sürümleriyle yüksek doğruluk sunmuştur. Ancak ilginç bir biçimde, eniyilenmiş sürüm varsayılan sürümden daha düşük performans sergilemiştir. Bu durum, sinir ağlarının çok sayıda hiperparametreye duyarlı yapısından kaynaklanabileceği gibi, GridSearchCV içinde kullanılan parametre aralığının sınırlı olması sebebiyle modelin potansiyelinin tam olarak yansıtılamamış olmasıyla da açıklanabilir. Ayrıca, eniyilenmiş sürümde aşırı öğrenme olasılığı da göz ardı edilmemelidir. Buna rağmen, MLPR'nin düşük MAE değeri (2,004) modele olan güveni korumaktadır.

RF modeli ise önceki çalışmalarda sergilediği lider konumdan bu kez dördüncü ve beşinci sıralara gerilemiştir ($R^2 = 0,844$). Özellikle varsayılan ve eniyilenmiş sürümler arasında çok az fark oluşması, bu modelin hiperparametre duyarlılığının düşük olduğunu ve

veri setinde öznitelik sayısının azalmasının çoklu ağaç yapısında çeşitliliği sınırlamış olabileceğini göstermektedir. RF algoritmasının en büyük gücü, yüksek öznitelik çeşitliliğini kullanarak farklı alt uzaylarda tahminler üretebilmesidir. Ancak Çalışma 5'te bu çeşitlilik azaldığı için modelin performansı doğal olarak bir miktar düşmüştür.

Sonuç olarak, bu çalışmada öznitelik indirgeme yöntemleriyle oluşturulan sadeleştirilmiş veri seti bazı modellerin (SVR, MLP) başarısını korurken, bazı modellerin (özellikle RF) sınırlı performans gösterdiği bir ortam yaratmıştır. Bu durum, makine öğrenmesi algoritmalarının yalnızca veri miktarına değil, aynı zamanda öznitelik kalitesine ve çeşitliliğine ne derece duyarlı olduklarını açıkça ortaya koymaktadır. Dahası, eniyilenmiş modellerin varsayılan sürümlere kıyasla her zaman daha iyi sonuç vermediği; kullanılan hiperparametre aralıklarının, model tipi ve veri yapısına göre dikkatle seçilmesi gerektiği bir kez daha doğrulanmıştır.

5. BULGULAR VE TARTIŞMA

Bu bölümde, farklı veri setleriyle gerçekleştirilen modelleme çalışmalarının sonuçları detaylı biçimde ele alınmıştır. Bölüm 5.1’de her bir makine öğrenmesi modelinin kendi içindeki varsayılan ve eniyilenmiş performansları karşılaştırılmış, Bölüm 5.2’de ise her bir çalışmanın tahminleme başarısı farklı modeller ekseninde değerlendirilmiştir (5.2). Son bölümde, tüm çalışmalar bir araya getirilerek, bütünlük bir performans analizi yapılmış ve en başarılı model yapıları ortaya koymuştur. Elde edilen bu bulgular, öznelik seçiminin etkisi, model yapılandırılmalarının başarısı ve hiperparametre eniyilemesinin katkısı gibi konulara dair kapsamlı bir bakış sunmaktadır. Bu bölümde sunulan analizler, tez çalışmasının temel karar noktalarını oluşturmaktadır.

5.1. Model Bazlı Karşılaştırmalı Performans Analizi

Bu alt bölümde, çalışmada kullanılan sekiz farklı makine öğrenmesi modelinin her biri, beş ayrı çalışma kapsamında elde edilen regresyon başarıları üzerinden bütüncül şekilde değerlendirilmektedir. Her bir modelin varsayılan ve eniyilenmiş sürümleri için dört temel performans metriği (MAE, MSE, RMSE, R^2) ayrı ayrı ele alınarak modelin genel eğilimi, öznelik setlerine verdiği tepkiler ve hiperparametre eniyilemesinin etkisi analiz edilmektedir. Bu yapı sayesinde, modellerin hangi tür veri setlerinde daha başarılı olduğu, ne kadar istikrarlı performans gösterdiği ve varsayılan ayarların ne ölçüde yeterli olduğu akademik bakış açısıyla tartışılmaktadır. Elde edilen çıkarımlar, model seçimi yapacak araştırmacılar veya uygulayıcılar için yol gösterici nitelikte olacaktır.

RF modelinin çalıştırılmasıyla Çalışma 1, Çalışma 2, Çalışma 3, Çalışma 4 ve Çalışma 5 için elde edilen performans metrikleri Çizelge 5.1’de verilmiştir.

Sonuç olarak, RF modeli; yüksek tahmin doğruluğu, düşük varyans ve geniş veri setlerine karşı dayanıklılığı sayesinde beş farklı çalışmanın neredeyse tamamında güçlü performans sergilemiştir. Eniyilenmiş hâlleri, bazı çalışmalar için kısmi avantajlar sağlasa da, varsayılan sürüm dahi çoğu durumda yeterli olmuştur. Bu durum literatürde de sıkça vurgulanan bir özellik olup, algoritmanın aşırı ayar gerektirmeyen yapısını desteklemektedir (Breiman, 2001).

GB modelinin çalıştırılmasıyla Çalışma 1, Çalışma 2, Çalışma 3, Çalışma 4 ve Çalışma 5 için elde edilen performans metrikleri Çizelge 5.2’de verilmiştir.

Çizelge 5.2. GB Modeline İlişkin Performans Metrikleri

GB															
Çalışma 1				Çalışma 2				Çalışma 3				Çalışma 4			
Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş
MAE	1,992	MAE	2,016	MAE	2,631	MAE	2,167	MAE	1,877	MAE	1,849	MAE	1,909	MAE	1,864
MSE	8,060	MSE	8,268	MSE	12,923	MSE	10,097	MSE	7,113	MSE	7,038	MSE	6,979	MSE	6,876
RMSE	2,839	RMSE	2,875	RMSE	3,595	RMSE	3,178	RMSE	2,667	RMSE	2,653	RMSE	2,642	RMSE	2,622
R ²	0,838	R ²	0,834	R ²	0,741	R ²	0,798	R ²	0,857	R ²	0,859	R ²	0,860	R ²	0,862

Çizelge 5.2 incelendiğinde GB modeli, farklı veri setleri üzerinde genel olarak tutarlı ancak sınırlı bir gelişim potansiyeli gösteren bir yapı sergilemiştir. Çalışma 1’de hem varsayılan hem de eniyilenmiş sürümde benzer R² değerleri (sırasıyla 0,838 ve 0,834) elde edilmiş olup, hiperparametre eniyilemesi modelin performansını anlamlı şekilde geliştirememiştir. MAE ve RMSE değerleri de birbirine çok yakın kalarak bu sonucu desteklemiştir. Bu durum, varsayılan ayarların veri setiyle yeterince uyumlu olduğunu göstermektedir.

Çalışma 2’de ise eniyilenmiş modelin varsayılan modele göre kayda değer bir performans artışı sağladığı görülmektedir. R² skoru 0,741’den 0,798’e yükselmiş, MAE 2,631’den 2,167’ye düşmüştür. Bu, daha kompleks ve geniş öznitelik yapılarında GB’in hiperparametre ayarlarına daha duyarlı olduğunu göstermektedir. Modelin aşamalı öğrenme doğası, heterojen öznitelik yapılarıyla daha etkili çalışmasına olanak tanımıştır.

Çalışma 3 ve 4’te ise her iki versiyon da yüksek ve dengeli R² değerleri sunmuştur (sırasıyla 0,857–0,859 ve 0,860–0,862). Bu durum, veri setindeki özniteliklerin GB algoritmasının yapısıyla oldukça uyumlu olduğunu ve modelin güçlü genelleme sunduğunu

Çalışma 2’de ise XGBoost’un varsayılan sürümü $R^2 = 0,794$ ile ılımlı bir açıklayıcılığa sahipken, eniyilenmiş versiyonda bu oran %75,7’ye gerilemiştir. MSE ve RMSE değerlerinde de iyileşme gözlenmemiştir. Bu durum, hiperparametre eniyilemesinin modelin genel yapısıyla yeterince uyumlu yapılmadığında model başarısını düşürebileceğini desteklemektedir. Özellikle az sayıda ve daha seçici özniteliklerin kullanıldığı veri setlerinde, aşamalı öğrenme yapılarının aşırı öğrenmeye yatkın hale gelebileceği unutulmamalıdır.

Çalışma 3, varsayılan modelin $R^2 = 0.878$ ile yine oldukça iyi performans gösterdiği, MAE = 1.828 ve RMSE = 2.467 ile dengeli sonuçlar sunduğu bir senaryo olarak dikkat çekmektedir. Eniyilenmiş sürümde ise R^2 , 0.856’ya gerilemiş, diğer metriklerde de benzer şekilde küçük kayıplar yaşanmıştır. Bu durum, varsayılan modelin bu veri yapısıyla doğal bir uyum yakaladığını göstermektedir.

Çalışma 4’te varsayılan XGBoost modeli $R^2 = 0,882$ ile en yüksek varyans açıklama gücüne sahip modellerden biri olurken, eniyilenmiş sürümde bu değer %84,4’e düşmüştür. MAE ve RMSE’de de eniyilenmiş modelde belirgin kötüleşmeler görülmektedir. Bu gözlem, XGBoost’un varsayılan ayarlarla oldukça istikrarlı çalıştığını ve çoğu durumda eniyilemeye ihtiyaç duymadığını bir kez daha doğrulamaktadır.

Son olarak Çalışma 5’te de benzer bir desen izlenmiştir. Varsayılan model $R^2 = 0,842$ ile başarılı olurken, optimize model $R^2 = 0,827$ ’ye gerilemiştir. Tüm metrikler genel olarak eniyilenmiş modelde daha zayıf kalmıştır. Bu, daha daraltılmış öznitelik setlerinde varsayılan ayarların daha etkili olabileceğini göstermektedir.

Özetle, XGBoost modeli çalışmalarda genel olarak güçlü bir tahmin başarısı sergilemiştir. Ancak varsayılan ayarlar ile eniyilenmiş sürümler arasında çoğu zaman varsayılan lehine bir üstünlük gözlemlenmiş, bu da hiperparametre ayarlarının dikkatli ve geniş aralıklarda yapılması gerektiğini ortaya koymuştur.

LightGBM modelinin çalıştırılmasıyla Çalışma 1, Çalışma 2, Çalışma 3, Çalışma 4 ve Çalışma 5 için elde edilen performans metrikleri Çizelge 5.4’te verilmiştir.

Çizelge 5.4. LightGBM Modeline İlişkin Performans Metrikleri

LightGBM															
Çalışma 1				Çalışma 2				Çalışma 3				Çalışma 4			
Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş
MAE	2,217	MAE	2,168	MAE	2,801	MAE	2,736	MAE	2,186	MAE	2,176	MAE	2,023	MAE	1,941
MSE	8,972	MSE	8,506	MSE	14,446	MSE	14,402	MSE	7,332	MSE	7,505	MSE	6,669	MSE	6,776
RMSE	2,995	RMSE	2,917	RMSE	3,801	RMSE	3,795	RMSE	2,708	RMSE	2,740	RMSE	2,582	RMSE	2,603
R ²	0,820	R ²	0,829	R ²	0,710	R ²	0,711	R ²	0,853	R ²	0,850	R ²	0,866	R ²	0,864

LightGBM modeli, genel olarak hızlı ve kaynak dostu bir aşamalı öğrenme algoritması olmasına rağmen bu çalışmalarda zaman zaman istikrarsız performanslar sergilemiştir. Çizelge 5.4 incelendiğinde Çalışma 1’de, varsayılan model $R^2 = 0,820$ ile makul bir açıklayıcılık sağlamış, eniyilenmiş sürümle birlikte bu oran 0,829’a çıkmıştır. MAE ve RMSE değerlerinde de eniyilenmiş modelin çok az bir iyileşme sunduğu görülmektedir. Bu durum, eniyileme işleminin sınırlı ama olumlu bir katkı sağladığını göstermektedir.

Çalışma 2’de, LightGBM tüm modeller içinde en düşük R^2 değerlerinden birini elde etmiş; hem varsayılan (0,710) hem de eniyilenmiş (0,711) sürümler düşük açıklayıcılık oranlarına sahip olmuştur. Bu bulgu, LightGBM’in bu veri setiyle etkili bir öğrenme gerçekleştiremediğini göstermektedir. MAE ve RMSE değerleri de görece olarak yüksek kalmıştır. Bu da özellikle lineer olmayan karmaşık örüntülerin düşük temsil gücüyle açıklanabilir.

Çalışma 3’te LightGBM’in performansı nispeten daha dengeli gözlemlenmiştir. Varsayılan model $R^2 = 0,853$, eniyilenmiş model ise $R^2 = 0,850$ ile benzer düzeyde performans sergilemiştir. MAE ve RMSE değerleri de oldukça yakın çıkmıştır. Bu tablo, eniyilemenin bu çalışma özelinde anlamlı bir fark meydana getirmediğini ve varsayılan ayarların yeterince güçlü olduğunu göstermektedir.

Çalışma 4’te ise LightGBM görece olarak en güçlü performanslarından birini sergilemiştir. Varsayılan model $R^2 = 0,866$ ile yüksek açıklayıcılığa ulaşırken, eniyilenmiş sürümde bu oran 0,864’e düşmüştür. Burada da varsayılan model bir adım öne çıkmıştır. MAE ve RMSE metrikleri açısından ise eniyilenmiş model daha iyi sonuçlar vermiştir. Bu durum, R^2 dışındaki hata metriklerinin değerlendirilmesi gerektiğini ve sadece tek bir metriğe dayalı karar verilmemesi gerektiğini göstermektedir.

Çalışma 5'te LightGBM'in performansı yeniden düşüş göstermiştir. R^2 değeri varsayılan modelde 0,755 iken, eniyilenmiş sürümde 0,742'ye gerilemiştir. MAE ve RMSE değerleri de diğer modellere kıyasla yüksektir. Bu durum, LightGBM'in daha sadeleştirilmiş ve seçilmiş öznitelikler içeren veri setlerinde yeterli genelleme gücü sağlayamadığını göstermektedir.

Sonuç olarak, LightGBM bazı çalışmalarda yüksek doğruluk sunsa da özellikle daha az ve seçici öznitelik içeren veri setlerinde istikrarını koruyamamıştır. Aşamalı öğrenme yapısının avantajını tam anlamıyla yansıtamadığı bu çalışmalarda, parametre ayarlarının etkisinin sınırlı kaldığı ve varsayılan ayarların genellikle daha dengeli sonuçlar sunduğu görülmektedir.

Ridge Regresyon modelinin çalıştırılmasıyla Çalışma 1, Çalışma 2, Çalışma 3, Çalışma 4 ve Çalışma 5 için elde edilen performans metrikleri Çizelge 5.5'te verilmiştir.

Çizelge 5.5. Ridge Regresyon Modeline İlişkin Performans Metrikleri

Ridge Regresyon																			
Çalışma 1				Çalışma 2				Çalışma 3				Çalışma 4				Çalışma 5			
Varsayılan		Eniyeilenmiş		Varsayılan		Eniyeilenmiş		Varsayılan		Eniyeilenmiş		Varsayılan		Eniyeilenmiş		Varsayılan		Eniyeilenmiş	
MAE	2,348	MAE	2,261	MAE	3,109	MAE	3,109	MAE	2,588	MAE	2,396	MAE	2,576	MAE	2,395	MAE	2,573	MAE	2,573
MSE	10,268	MSE	9,490	MSE	14,563	MSE	14,563	MSE	10,656	MSE	9,848	MSE	10,480	MSE	9,843	MSE	9,918	MSE	9,918
RMSE	3,204	RMSE	3,081	RMSE	3,816	RMSE	3,816	RMSE	3,264	RMSE	3,138	RMSE	3,237	RMSE	3,137	RMSE	3,149	RMSE	3,149
R ²	0,794	R ²	0,810	R ²	0,708	R ²	0,708	R ²	0,786	R ²	0,803	R ²	0,790	R ²	0,803	R ²	0,801	R ²	0,801

Ridge Regresyon, doğrusal modellerin bir türevi olarak çoklu doğrusal bağlantı durumlarında model kararlılığını artırmayı hedefleyen bir yaklaşımdır. Bu çalışmalarda, Ridge Regresyonunun genellikle istikrarlı ancak sınırlı bir performans gösterdiği gözlemlenmiştir.

Çizelge 5.5 incelendiğinde Çalışma 1'de eniyilenen model, R^2 değerinde 0,794'ten 0,810'a yükselerek anlamlı bir gelişim göstermiştir. MSE ve RMSE değerlerinde de gözle görülür iyileşmeler söz konusudur. Bu durum, Ridge Regresyonunun yüksek varyanslı modellerle karşılaştırıldığında düşük riskli ve daha öngörülebilir bir alternatif sunduğunu göstermektedir.

Çalışma 2’de ise varsayılan ve eniyilenmiş modellerin performansları neredeyse aynıdır. R^2 değeri her iki sürümde de 0,708 olarak kalmıştır. MAE ve RMSE değerlerinde de hiçbir farklılık gözlenmemiştir. Bu durum, Ridge Regresyon modelinin kullanılan özniteliklerle fazla esneklik sağlayamadığını ve hiperparametre eniyilemesinin etkisiz kaldığını göstermektedir.

Çalışma 3’te ise eniyilemenin daha belirgin bir katkısı olmuştur. R^2 değeri 0,786’dan 0,803’e yükselmiş, hata metriklerinde de küçük ama tutarlı düşüşler yaşanmıştır. Bu tablo, Ridge Regresyon modelinin özellikle karmaşık yapıdaki veri kümelerinde sınırlı da olsa fayda sağlayabildiğini göstermektedir.

Çalışma 4’te benzer bir eğilim devam etmiştir. Eniyilenmiş modelin R^2 değeri 0,803 ile varsayılan modelin (0,790) üzerinde yer almıştır. MSE ve RMSE’de de iyileşmeler mevcuttur. Ancak yine de Ridge Regresyon, bu çalışmada yer alan aşamalı öğrenme tabanlı modeller veya sinir ağı modelleri kadar yüksek başarı sağlayamamıştır.

Çalışma 5’te, Ridge Regresyon modeli varsayılan ve eniyilenmiş sürümlerle aynı R^2 skorunu (0,801) vermiştir. Bu benzerlik, eniyilenmiş parametrelerin model davranışında anlamlı bir fark meydana getirmediğini göstermektedir. MAE ve RMSE değerleri de birebir aynı kalmıştır. Bu durum, Ridge Regresyon’un özellikle azaltılmış öznitelikli veri setlerinde esnekliğinin sınırlı kaldığını ortaya koymaktadır.

Genel olarak Ridge Regresyon, her çalışmada belirli bir doğruluk düzeyini garanti ederken; esnekliği sınırlı, doğrusal etkileşimlere duyarlı ve karmaşık örüntüleri tam olarak öğrenemeyen bir model olarak karşımıza çıkmıştır. Bununla birlikte, bazı çalışmalar için eniyilenmiş versiyonunun düşük varyanslı ve istikrarlı performans sunduğu da unutulmamalıdır.

Lasso Regresyon modelinin çalıştırılmasıyla Çalışma 1, Çalışma 2, Çalışma 3, Çalışma 4 ve Çalışma 5 için elde edilen performans metrikleri Çizelge 5.6’da verilmiştir.

Çizelge 5.6. Lasso Regresyon Modeline İlişkin Performans Metrikleri

Lasso Regresyon															
Çalışma 1				Çalışma 2				Çalışma 3				Çalışma 4			
Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş	Varsayılan	Eniyilenmiş
MAE	4,021	MAE	2,260	MAE	3,044	MAE	3,132	MAE	2,962	MAE	2,395	MAE	2,938	MAE	2,394
MSE	29,168	MSE	9,490	MSE	13,565	MSE	14,989	MSE	12,916	MSE	9,848	MSE	12,943	MSE	9,842
RMSE	5,401	RMSE	3,081	RMSE	3,683	RMSE	3,872	RMSE	3,594	RMSE	3,138	RMSE	3,598	RMSE	3,137
R ²	0,415	R ²	0,810	R ²	0,728	R ²	0,700	R ²	0,741	R ²	0,803	R ²	0,741	R ²	0,803

Lasso Regresyonu, özellikle değişken seçimi ve model sadeleştirme açısından öne çıkan bir regresyon tekniğidir. Ancak bu yöntemin başarı düzeyi, veri setindeki ilişki yapısına ve özniteliklerin özelliklerine göre ciddi dalgalanmalar gösterebilir. Bu çalışmalarda da benzer bir eğilim gözlemlenmiştir.

Çizelge 5.6 incelendiğinde Çalışma 1'de modelin varsayılan hali oldukça zayıf bir performans sergilemiştir ($R^2 = 0,415$). Ancak eniyilenen sürüm, R^2 skorunu 0,810'a kadar çıkararak dikkat çekici bir gelişim göstermiştir. MAE değeri 4,021'den 2,260'a, RMSE ise 5,401'den 3,081'e düşmüştür. Bu dramatik iyileşme, hiperparametre eniyilemesinin Lasso Regresyon üzerinde ne denli etkili olabileceğini göstermektedir.

Çalışma 2'de ise eniyilenmiş sürüm, varsayılan versiyona göre daha yüksek bir MSE ve daha düşük bir R^2 üretmiştir ($0,700 < 0,728$). Bu durum, Lasso Regresyonun bazen aşırı ceza uygulayarak anlamlı öznitelikleri sıfırlama eğiliminde olduğunu ve bunun model başarısını olumsuz etkileyebileceğini göstermektedir.

Çalışma 3'te ise tam tersi bir eğilim gözlenmiştir. Varsayılan modelin R^2 değeri 0,741 iken, eniyilenmiş sürüm bu skoru 0,803'e çıkarmıştır. MAE ve RMSE değerlerinde de belirgin düşüşler yaşanmıştır. Bu tablo, Lasso Regresyonun uygun hiperparametrelerle doğrusal yapıları iyi temsil edebildiğini ve özellikle daha az parazitli veri setlerinde başarılı olabileceğini göstermektedir.

Çalışma 4'te benzer sonuçlar devam etmiş; eniyilenmiş model, varsayılan versiyona göre daha düşük hata oranları ve daha yüksek R^2 skoru (0,803) sağlamıştır. Bu tekrar eden başarı, Lasso'nun belirli veri setlerinde doğrusal ilişkileri başarıyla yakalayabildiğini ve

düzenleme (regularization) gücünü doğru kullandığında istikrarlı sonuçlar sunduğunu göstermektedir.

Çalışma 5'te ise eniyilenmiş modelin performansı varsayılan sürümünden bir miktar daha yüksek olmuştur ($R^2 = 0,798 > 0,763$). Bununla birlikte, genel hata metrikleri diğer modellere kıyasla daha yüksektir ve bu durum Lasso Regresyonun yüksek boyutlu doğrusal olmayan ilişkilere sahip veri setlerinde sınırlı performans gösterdiğini düşündürmektedir.

Sonuç olarak Lasso Regresyon, özellikle eniyilendiğinde belirli çalışmalarda güçlü performans sergileyebilmiştir. Ancak bu başarı, veri setinin yapısına ve doğru parametre seçimine büyük oranda bağlıdır. Doğrusal sınırları aşamayan yapısı ve agresif değişken eleme mekanizması nedeniyle, karmaşık veri ilişkilerini modellemede zayıf kalabilmektedir.

SVR modelinin çalıştırılmasıyla Çalışma 1, Çalışma 2, Çalışma 3, Çalışma 4 ve Çalışma 5 için elde edilen performans metrikleri Çizelge 5.7'de verilmiştir.

Çizelge 5.7. SVR Modeline İlişkin Performans Metrikleri

SVR															
Çalışma 1				Çalışma 2				Çalışma 3				Çalışma 4			
Varsayılan		Eniyilenmiş		Varsayılan		Eniyilenmiş		Varsayılan		Eniyilenmiş		Varsayılan		Eniyilenmiş	
MAE	2,559	MAE	1,512	MAE	2,868	MAE	1,755	MAE	2,589	MAE	1,787	MAE	2,609	MAE	1,839
MSE	12,082	MSE	4,731	MSE	13,177	MSE	5,198	MSE	11,470	MSE	6,860	MSE	11,290	MSE	6,731
RMSE	3,476	RMSE	2,175	RMSE	3,630	RMSE	2,280	RMSE	3,387	RMSE	2,619	RMSE	3,360	RMSE	2,594
R^2	0,758	R^2	0,905	R^2	0,736	R^2	0,896	R^2	0,770	R^2	0,862	R^2	0,774	R^2	0,865

SVR, bu çalışma kapsamında değerlendirilen tüm modeller arasında eniyilendiğinde en güçlü performansı sergileyen yöntemlerden biri olarak öne çıkmaktadır. SVR, özellikle küçük-orta boyutlu, karmaşık ve doğrusal olmayan yapılar içeren veri setlerinde yüksek doğrulukla genelleme yapabilmesi sayesinde dikkat çekici sonuçlar üretmiştir.

Çizelge 5.7 incelendiğinde Çalışma 1'de SVR'nin varsayılan hali ortalama düzeyde bir performans sergilerken ($R^2 = 0,758$), eniyilenmiş versiyonu R^2 değerini %90,5 seviyesine çıkarmıştır. MAE ve RMSE değerlerinde de dramatik düşüşler gözlemlenmiştir. Bu durum hiperparametre ayarlamalarının SVR üzerindeki güçlü etkisini açıkça göstermektedir. Özellikle C, gamma ve epsilon parametrelerinin doğru seçimi, modelin esnekliğini ve toleransını doğru şekilde dengelemiştir.

Çalışma 2’de de benzer bir iyileşme gözlemlenmiştir. Varsayılan R^2 değeri 0,736 iken, eniyilemeyle bu oran 0,896’ya yükselmiştir. MAE değeri 2,868’den 1,755’e düşerken, RMSE 3,630’dan 2,280’e gerilemiştir. Bu tablo, SVR’nin eniyilendiğinde doğrusal olmayan yapıların hâkim olduğu veri setlerinde etkili bir şekilde öğrenme gerçekleştirdiğini göstermektedir.

Çalışma 3’te SVR’nin varsayılan hali de görece güçlü bir başlangıç yapmış ($R^2 = 0,770$); ancak eniyilenmiş sürüm yine önemli bir sıçrama ile $R^2 = 0,862$ seviyesine ulaşmıştır. Bu artış, SVR’nin özellikle hibrit veri yapılarında (örneğin Çalışma 3 gibi birleştirilmiş öznitelik kümeleri içeren setlerde) performansını eniyilenmiş haliyle daha iyi gösterebildiğini düşündürmektedir.

Çalışma 4’te de modelin eniyilenmiş versiyonu yüksek başarı sağlamış ve $R^2 = 0,865$ ile oldukça güvenilir bir tahmin gücü sergilemiştir. Burada da MAE ve RMSE değerlerinde kayda değer iyileşmeler gözlemlenmiştir. SVR’nin bu başarısı, kapasite kullanım oranı gibi özniteliklerin çıkarılmasının ardından veri setindeki öznitelik ilişkilerini yeniden ve doğru şekilde kurabildiğini göstermektedir.

En dikkat çekici sonuçlardan biri ise Çalışma 5’te ortaya çıkmıştır. Varsayılan sürümde düşük sayılabilecek R^2 değeri (0,739), eniyilenmiş modelde 0,900 seviyesine ulaşmıştır. MAE 2,755’ten 1,578’e, RMSE ise 3,610’dan 2,234’e düşmüştür. Bu sonuç, SVR’nin daha az sayıda öznitelik içeren ve sadeleştirilmiş veri setlerinde dahi son derece güçlü öğrenme yeteneğine sahip olduğunu ortaya koymuştur.

Genel olarak değerlendirildiğinde, SVR eniyilendiğinde tüm çalışmalar arasında en yüksek R^2 skorlarına ulaşan modellerden biri olmuştur. Doğrusal olmayan yapıları başarılı şekilde modelleyebilmesi, aşırı öğrenme riskini kontrollü biçimde yönetebilmesi ve kernel tabanlı esnekliği ile SVR, özellikle yapılandırılmış ve düzenlenmiş veri setlerinde üstün performans göstermiştir.

MLPR modelinin çalıştırılmasıyla Çalışma 1, Çalışma 2, Çalışma 3, Çalışma 4 ve Çalışma 5 için elde edilen performans metrikleri Çizelge 5.8’de verilmiştir.

Çalışma 4'te de eniyilenmiş MLPR modeli oldukça başarılı sonuçlar üretmiş ve $R^2 = 0,881$ 'e ulaşmıştır. RMSE 2,435, MAE 1,668 gibi değerlerle, model bu veri setinde de güvenilir tahminler yapabildiğini göstermiştir. Kapasite kullanım oranı gibi bir değişkenin çıkarılmış olduğu bu çalışma, MLPR'nin yüksek esneklikle farklı veri yapılarına uyum sağlayabildiğini göstermesi açısından önemlidir.

Çalışma 5'te ise varsayılan ve eniyilenmiş modeller arasında dikkat çekici bir denge durumu gözlemlenmiştir. Varsayılan model $R^2 = 0,885$ ile oldukça iyi bir sonuç vermiştir; eniyileme ise bu değeri biraz geriletmiş (0,857), RMSE değerini ise artırmıştır (2,396 → 2,674). Bu durum, bazı durumlarda hiperparametre eniyilemesinin aşırı öğrenme veya dengesiz yapı sorunlarına yol açabileceğini ortaya koymaktadır.

Genel olarak değerlendirildiğinde, MLPR eniyilendiğinde özellikle karmaşık ve birleşik öznitelik yapılarında (özellikle Çalışma 3 gibi) olağanüstü bir başarı göstermiştir. Ancak bazı çalışmalarda varsayılan haliyle de yüksek performans sağladığı görülmüştür. Bu model, güçlü ancak dikkatli ayar gerektiren bir yapay öğrenme algoritması olarak öne çıkmaktadır.

Farklı makine öğrenmesi modelleri üzerinden yürütülen karşılaştırmalı analizler, yapı kimyasal üreticisinin çimento satın alma fiyatlarını öngörmeye hangi algoritmaların daha yüksek performans sunduğunu ortaya koymuştur. Elde edilen sonuçlar, özellikle topluluk öğrenmesi tabanlı yöntemler (RF, GB, XGBoost, LightGBM) ve Kernel tabanlı modellerin (SVR) eniyilendiklerinde oldukça yüksek doğruluk seviyelerine ulaştığını göstermiştir. Öte yandan, MLPR gibi yapay sinir ağı temelli modeller de doğru hiperparametrelerle yapılandırıldığında son derece güçlü tahmin kapasitesi sergilemiştir.

Model bazlı analiz süreci; her bir modelin veri yapısına, karmaşıklığına ve öz nitelik türüne duyarlılığını görme imkânı sunmuş, aynı zamanda hiperparametre eniyilemesinin genel başarı üzerinde ne denli kritik bir rol oynadığını da gözler önüne sermiştir. Ancak tekil model performanslarından daha önemlisi, ilerleyen bölümde detaylandırılacak olan çalışma bazlı analizlerle, hangi veri kümesinin bütüncül olarak daha doğru tahminler sunduğu ortaya konulacaktır. Bu sayede yalnızca model seçimi değil, veri seti tercihinin de modelleme başarısındaki etkisi bütünsel biçimde değerlendirilecektir.

5.2. Çalışma Bazlı Karşılaştırmalı Performans Analizi

Bu bölümde, farklı öznitelik setlerine dayalı olarak oluşturulan beş ayrı çalışma kapsamında elde edilen tahminleme performansları karşılaştırmalı olarak analiz edilmektedir. Her bir çalışmada kullanılan veri setleri, öznitelik mühendisliği ve seçim yaklaşımlarına göre farklılık göstermekte olup; bu farklılıkların makine öğrenmesi modellerinin başarı düzeylerine olan etkileri değerlendirilmektedir. Amaç, hangi veri setinin –dolayısıyla hangi öznitelik seçim stratejisinin– çimento fiyatlarının tahmininde daha yüksek doğruluk ve genelleme kapasitesi sunduğunu belirlemektir. Analizler, temel performans metrikleri olan MAE, MSE, RMSE ve R^2 değerleri dikkate alınarak hem varsayılan hem de eniyilenmiş model versiyonları üzerinden gerçekleştirilmiştir. Elde edilen bulgular, yalnızca model başarısı açısından değil, aynı zamanda veri ön işleme ve öznitelik seçimi süreçlerinin etkisini tartışmak açısından da önemli çıkarımlar sunmaktadır.

Çalışma 1 kapsamında çalıştırılan makine öğrenmesi modelleri sonucunda elde edilen performans metrikleri Çizelge 5.9’da verilmiştir.

Çizelge 5.9. Çalışma 1 Sonucunda Elde Edilen Performans Metrikleri

	Model	MAE	MSE	RMSE	R^2	Bar (R^2)
Çalışma 1	RF (Varsayılan)	1,814	6,341	2,518	0,873	
	RF (Eniyilenmiş)	1,818	5,952	2,440	0,881	
	GB (Varsayılan)	1,992	8,060	2,839	0,838	
	GB (Eniyilenmiş)	2,016	8,268	2,875	0,834	
	XGBoost (Varsayılan)	1,884	6,581	2,565	0,868	
	XGBoost (Eniyilenmiş)	1,844	7,014	2,648	0,859	
	LightGBM (Varsayılan)	2,217	8,972	2,995	0,820	
	LightGBM (Eniyilenmiş)	2,168	8,506	2,917	0,829	
	Rigde Regresyon (Varsayılan)	2,348	10,268	3,204	0,794	
	Rigde Regresyon (Eniyilenmiş)	2,261	9,490	3,081	0,810	
	Lasso Regresyon (Varsayılan)	4,021	29,168	5,401	0,415	
	Lasso Regresyon (Eniyilenmiş)	2,260	9,490	3,081	0,810	
	SVR (Varsayılan)	2,559	12,082	3,476	0,758	
	SVR (Eniyilenmiş)	1,512	4,731	2,175	0,905	
	MLPR (Varsayılan)	1,505	4,683	2,164	0,906	
	MLPR (Eniyilenmiş)	1,505	4,690	2,166	0,906	

Çizelge 5.9'dan görüleceği üzere Çalışma 1 kapsamında kullanılan uzman görüşü temelli öznitelik seti ile gerçekleştirilen tahminleme performansları incelendiğinde, genel olarak modellerin başarılı sonuçlar ürettiği gözlemlenmektedir. Bu çalışma için en dikkat çekici sonuçlar hem varsayılan hem de eniyilenmiş versiyonlarıyla MLPR ve SVR modelleri tarafından elde edilmiştir. Özellikle MLPR, eniyilenmemiş hâliyle 1,505 MAE, 4,683 MSE, 2,164 RMSE ve 0,906 R^2 skorlarıyla oldukça güçlü bir performans sergilemiştir. Aynı şekilde SVR modeli de hiperparametre eniyilemesi sonrasında MAE ve RMSE değerlerinde önemli düşüşler sağlayarak (MAE= 1,512, RMSE= 2,175) en yüksek R^2 skoruna (0,905) ulaşmıştır.

Bununla birlikte, RF ve XGBoost modelleri de eniyilenmiş versiyonlarında yüksek doğruluk sağlamış ve R^2 değerlerinde 0,881 (RF) ve 0,859 (XGBoost) gibi başarılı skorlar yakalamıştır. Bu modeller özellikle dengeli hata metrikleri ile dikkat çekmiştir. Diğer yandan, Lasso Regresyon modelinin varsayılan hâli son derece düşük performans gösterirken ($R^2 = 0,415$), eniyilenmiş sürümünde hatalarını önemli ölçüde azaltmış ve Ridge Regresyon ile benzer bir seviyeye ulaşmıştır.

Genel tablo değerlendirildiğinde, Çalışma 1 için en uygun modellerin eniyilenmiş MLPR, SVR ve RF olduğu söylenebilir. Bu durum, uzman görüşüyle seçilen özniteliklerin makine öğrenimi modelleri açısından anlamlı ve ayrıştırıcı bilgi taşıdığını göstermekte; aynı zamanda iyi bir öznitelik setiyle derin öğrenme tabanlı ve kernel-tabanlı yöntemlerin güçlü performanslar sergileyebileceğini ortaya koymaktadır.

Çalışma 2 kapsamında çalıştırılan makine öğrenmesi modelleri sonucunda elde edilen performans metrikleri Çizelge 5.10'da verilmiştir.

Çizelge 5.10. Çalışma 2 Sonucunda Elde Edilen Performans Metrikleri

	Model	MAE	MSE	RMSE	R ²	Bar (R ²)
Çalışma 2	RF (Varsayılan)	2,283	10,215	3,196	0,795	
	RF (Eniyilenmiş)	2,211	9,633	3,104	0,807	
	GB (Varsayılan)	2,631	12,923	3,595	0,741	
	GB (Eniyilenmiş)	2,167	10,097	3,178	0,798	
	XGBoost (Varsayılan)	2,354	10,268	3,204	0,794	
	XGBoost (Eniyilenmiş)	2,568	12,115	3,481	0,757	
	LightGBM (Varsayılan)	2,801	14,446	3,801	0,710	
	LightGBM (Eniyilenmiş)	2,736	14,402	3,795	0,711	
	Rigde Regresyon (Varsayılan)	3,109	14,563	3,816	0,708	
	Rigde Regresyon (Eniyilenmiş)	3,109	14,563	3,816	0,708	
	Lasso Regresyon (Varsayılan)	3,044	13,565	3,683	0,728	
	Lasso Regresyon (Eniyilenmiş)	3,132	14,989	3,872	0,700	
	SVR (Varsayılan)	2,868	13,177	3,630	0,736	
	SVR (Eniyilenmiş)	1,755	5,198	2,280	0,896	
	MLPR (Varsayılan)	3,618	19,471	4,413	0,610	
	MLPR (Eniyilenmiş)	2,175	8,182	2,860	0,836	

Çizelge 5.10'den görülebileceği üzere korelasyon analizine dayalı olarak belirlenen özniteliklerin kullanıldığı Çalışma 2 kapsamında elde edilen performans sonuçları, bu öznitelik setinin tahminleme açısından genel olarak daha düşük doğruluk sağladığını göstermektedir. Bu durum özellikle MAE, MSE ve RMSE değerlerinde gözlemlenen artışlarla ve birçok modelde düşen R² skorlarıyla açıkça ortaya konmaktadır. Ancak bu genel düşüşe rağmen, eniyilenmiş SVR modeli bu çalışmada öne çıkmaktadır. 1,755 MAE, 5,198 MSE, 2,280 RMSE ve 0,896 R² skorlarıyla yalnızca bu çalışmada değil, aynı zamanda tüm modeller arasında da dikkat çeken bir başarıya ulaşmıştır. Aynı şekilde MLPR modelinin eniyilenmiş versiyonu da 2,175 MAE, 8,182 MSE ve 0,836 R² ile görece güçlü bir performans sergilemiştir. Bu iki model, karmaşık ilişkileri öğrenme yetenekleri sayesinde daha zayıf öznitelik setlerine rağmen başarılı tahminler gerçekleştirebilmiştir.

RF modeli ise bu çalışmada hem varsayılan hem de eniyilenmiş hâlleriyle tutarlı sonuçlar üretmiş; eniyilenmiş sürümde 0,807 R² ile kabul edilebilir bir doğruluk yakalamıştır. Buna karşılık, GB, XGBoost ve LightGBM gibi ağaç tabanlı modellerin çoğu, eniyilenseler dahi hatalarını anlamlı şekilde azaltamamış, bazı durumlarda (örneğin XGBoost Eniyilenmiş) performans kaybı bile yaşamıştır. Bu durum, bu modellerin

kullanılan öznitelik setine karşı daha hassas olduğunu ya da seçilen özelliklerin bu yöntemler için sınırlı bilgi taşıdığını düşündürmektedir.

Öte yandan, Lasso ve Ridge Regresyon gibi doğrusal modellerin performansları hem düşük R^2 hem de yüksek hata metrikleri ile zayıf kalmıştır. Bu da özniteliklerin doğrusal ilişkiler yerine daha karmaşık ve etkileşimli yapılar içerdiğini ve bu nedenle doğrusal modellerin yetersiz kaldığını işaret etmektedir.

Genel değerlendirme yapıldığında, Çalışma 2'de kullanılan korelasyon analizi temelli öznitelik setinin, özellikle SVR ve eniyilenmiş MLPR gibi esnek modellerle birlikte kullanıldığında başarılı sonuçlar verdiği görülmektedir. Ancak bu öznitelik seti, ağaç tabanlı ya da doğrusal modellere kıyasla daha düşük tahmin gücü ortaya koymuş, bu da öznitelik mühendisliği sürecinde uzman görüşünün ne denli kritik olabileceğini bir kez daha vurgulamıştır.

Çalışma 3 kapsamında çalıştırılan makine öğrenmesi modelleri sonucunda elde edilen performans metrikleri Çizelge 5.11'de verilmiştir.

Çizelge 5.11. Çalışma 3 Sonucunda Elde Edilen Performans Metrikleri

	Model	MAE	MSE	RMSE	R^2	Bar (R^2)
Çalışma 3	RF (Varsayılan)	1,840	5,736	2,395	0,885	
	RF (Eniyilenmiş)	1,860	6,060	2,462	0,879	
	GB (Varsayılan)	1,877	7,113	2,667	0,857	
	GB (Eniyilenmiş)	1,849	7,038	2,653	0,859	
	XGBoost (Varsayılan)	1,828	6,085	2,467	0,878	
	XGBoost (Eniyilenmiş)	1,996	7,191	2,682	0,856	
	LightGBM (Varsayılan)	2,186	7,332	2,708	0,853	
	LightGBM (Eniyilenmiş)	2,176	7,505	2,740	0,850	
	Rigde Regresyon (Varsayılan)	2,588	10,656	3,264	0,786	
	Rigde Regresyon (Eniyilenmiş)	2,396	9,848	3,138	0,803	
	Lasso Regresyon (Varsayılan)	2,962	12,916	3,594	0,741	
	Lasso Regresyon (Eniyilenmiş)	2,395	9,848	3,138	0,803	
	SVR (Varsayılan)	2,589	11,470	3,387	0,770	
	SVR (Eniyilenmiş)	1,787	6,860	2,619	0,862	
	MLPR (Varsayılan)	2,899	13,817	3,717	0,723	
	MLPR (Eniyilenmiş)	1,405	4,566	2,137	0,908	

Çalışma 1 ve 2 kapsamında kullanılan özniteliklerin birleştirilmesiyle oluşturulan Çalışma 3 veri seti, tahminleme performansı açısından dikkat çekici iyileşmeler sunmuş ve pek çok modelde istikrarlı başarılar sağlamıştır. Özellikle hem doğrusal olmayan hem de esnek yapıda olan algoritmalar, bu zengin öznitelik setinden yüksek fayda sağlamıştır.

Çizelge 5.11 incelendiğinde eniyilenmiş MLPR, bu çalışmanın en başarılı modeli olarak öne çıkmıştır. 1,405 MAE, 4,566 MSE, 2,137 RMSE ve 0,908 R^2 değeriyle, hem hata metrikleri açısından en düşük değerlere ulaşmış hem de en yüksek doğruluk skorunu elde etmiştir. Bu sonuçlar, çok katmanlı yapay sinir ağlarının karmaşık ilişkileri öğrenme kapasitesini ve bu öznitelik setiyle olan güçlü uyumunu göstermektedir.

Eniyilenmiş SVR modeli de oldukça başarılı bir performans sergilemiş; 1,787 MAE, 6,860 MSE ve 0,862 R^2 değeriyle en güçlü ikinci model konumuna yerleşmiştir. Bu iki modelin başarısı, karmaşık ve etkileşimli yapılar içeren öznitelik setleriyle çalışan algoritmaların avantajını göstermektedir.

RF, GB ve XGBoost gibi ağaç tabanlı modeller de görece güçlü performanslarını sürdürmüştür. Bu modellerin hem varsayılan hem de eniyilenmiş hâlleriyle MAE ve RMSE değerleri düşük, R^2 skorları yüksek düzeyde seyretmiştir. Özellikle varsayılan hiperparametrelerin kullanıldığı RF modeli, 1,840 MAE ve 0,885 R^2 değeriyle eniyilenmiş pek çok modelden daha iyi sonuçlar sunmuştur. Bu durum, modelin doğrudan öznitelik setinden öğrenme gücünün yüksek olduğunu göstermektedir.

Buna karşılık, Ridge ve Lasso Regresyon gibi doğrusal modellerin performansı yine sınırlı kalmıştır. Özellikle varsayılan sürümlerde MAE ve RMSE değerlerinin görece yüksek, R^2 değerlerinin düşük olduğu gözlemlenmiştir. Ancak eniyilenmiş versiyonlarda (özellikle Lasso), bazı gelişmeler gözlemlenmiş ve R^2 değerleri 0,803'e kadar yükselmiştir.

Sonuç olarak, Çalışma 3 veri seti, farklı veri temelli özniteliklerin birleşiminden oluştuğu için modellerin genel performansını artırmıştır. Bu çalışma, öznitelik mühendisliğinde bilgi birikimini birleştirmenin tahmin gücünü yükselttiğini göstermektedir. Özellikle MLPR ve SVR gibi esnek modellerin öne çıkması, karmaşık yapıları öğrenebilen algoritmaların zengin öznitelik setlerinden daha fazla yararlanabildiğini doğrulamaktadır.

Çalışma 4 kapsamında çalıştırılan makine öğrenmesi modelleri sonucunda elde edilen performans metrikleri Çizelge 5.12’de verilmiştir.

Çizelge 5.12. Çalışma 4 Sonucunda Elde Edilen Performans Metrikleri

	Model	MAE	MSE	RMSE	R ²	Bar (R ²)
Çalışma 4	RF (Varsayılan)	1,780	5,605	2,367	0,888	
	RF (Eniyilenmiş)	1,871	6,108	2,471	0,878	
	GB (Varsayılan)	1,909	6,979	2,642	0,860	
	GB (Eniyilenmiş)	1,864	6,876	2,622	0,862	
	XGBoost (Varsayılan)	1,788	5,864	2,422	0,882	
	XGBoost (Eniyilenmiş)	2,116	7,787	2,791	0,844	
	LightGBM (Varsayılan)	2,023	6,669	2,582	0,866	
	LightGBM (Eniyilenmiş)	1,941	6,776	2,603	0,864	
	Rigde Regresyon (Varsayılan)	2,576	10,480	3,237	0,790	
	Rigde Regresyon (Eniyilenmiş)	2,395	9,843	3,137	0,803	
	Lasso Regresyon (Varsayılan)	2,938	12,943	3,598	0,741	
	Lasso Regresyon (Eniyilenmiş)	2,394	9,842	3,137	0,803	
	SVR (Varsayılan)	2,609	11,290	3,360	0,774	
	SVR (Eniyilenmiş)	1,839	6,731	2,594	0,865	
	MLPR (Varsayılan)	2,011	7,879	2,807	0,842	
	MLPR (Eniyilenmiş)	1,668	5,932	2,435	0,881	

Çizelge 5.12 incelendiğinde Çalışma 4 kapsamında gerçekleştirilen modelleme sonuçları, farklı makine öğrenmesi algoritmalarında hiperparametrelerin eniyilenip eniyilenmemesi durumlarına göre performanslarının dikkat çekici biçimde değiştiğini ortaya koymuştur. Varsayılan parametrelerle en düşük MAE (1,780), MSE (5,605) ve RMSE (2,367) değerlerine ulaşan RF, aynı zamanda oldukça yüksek bir R² skoru (0,888) ile güçlü bir tahmin performansı sergilemiştir. Bununla birlikte, MLPR eniyilendiğinde MAE (1,668) ve RMSE (2,435) değerlerinde en iyi sonuçları üretmiş, R² skoru ise 0,881 ile RF’nin hemen gerisinde kalmıştır. Özellikle eniyilenmiş SVR modeli de MAE (1,839) ve R² (0,865) gibi oldukça başarılı sonuçlarla öne çıkmıştır.

Model bazlı genel bir kıyaslama yapıldığında, eniyilenmiş edilmiş MLPR, RF ve SVR modelleri Çalışma 4 veri setinde en yüksek doğrulukla tahmin üretebilen yöntemler olarak öne çıkmaktadır. Buna karşın, Lasso ve Ridge Regresyon gibi doğrusal temelli yöntemlerin performansları hem hata metrikleri hem de R² açısından görece düşük kalmıştır. Bu durum veri setinin doğrusal olmayan ilişkiler içerebileceğini düşündürmektedir. Ayrıca,

XGBoost modelinin eniyilenmesi performansı artırmak yerine düşürmüş, bu durum hiperparametre aramasının modelin genel yapısına uygunluğu açısından dikkatle ele alınması gerektiğini göstermiştir.

Çalışma 5 kapsamında çalıştırılan makine öğrenmesi modelleri sonucunda elde edilen performans metrikleri Çizelge 5.13'te verilmiştir.

Çizelge 5.13. Çalışma 5 Sonucunda Elde Edilen Performans Metrikleri

	Model	MAE	MSE	RMSE	R ²	Bar (R ²)
Çalışma 5	RF (Varsayılan)	2,149	7,764	2,786	0,844	
	RF (Eniyilenmiş)	2,150	7,864	2,804	0,842	
	GB (Varsayılan)	2,414	9,434	3,071	0,811	
	GB (Eniyilenmiş)	2,408	9,527	3,087	0,809	
	XGBoost (Varsayılan)	2,019	7,874	2,806	0,842	
	XGBoost (Eniyilenmiş)	2,252	8,637	2,939	0,827	
	LightGBM (Varsayılan)	2,696	12,237	3,498	0,755	
	LightGBM (Eniyilenmiş)	2,749	12,880	3,589	0,742	
	Rigde Regresyon (Varsayılan)	2,573	9,918	3,149	0,801	
	Rigde Regresyon (Eniyilenmiş)	2,573	9,918	3,149	0,801	
	Lasso Regresyon (Varsayılan)	2,771	11,840	3,441	0,763	
	Lasso Regresyon (Eniyilenmiş)	2,581	10,061	3,172	0,798	
	SVR (Varsayılan)	2,755	13,031	3,610	0,739	
	SVR (Eniyilenmiş)	1,578	4,990	2,234	0,900	
	MLPR (Varsayılan)	1,779	5,741	2,396	0,885	
	MLPR (Eniyilenmiş)	2,004	7,149	2,674	0,857	

Çizelge 5.13 incelendiğinde, Çalışma 5 sonuçları, özellikle eniyilenmiş SVR ve varsayılan parametrelerle çalışan MLPR modellerinin yüksek performansı dikkat çekmiştir. Eniyilenmiş SVR modeli, MAE (1,578), MSE (4,990) ve RMSE (2,234) değerleriyle en düşük hata metriklerine ulaşırken, aynı zamanda en yüksek R² skorunu (0,900) elde ederek, bu çalışma kapsamındaki en başarılı model olarak öne çıkmıştır. Benzer şekilde, varsayılan ayarlarla çalışan MLPR'de MAE (1,779) ve R² = 0,885 ile oldukça güçlü bir tahmin performansı sergilemiştir.

RF modeli hem varsayılan hem de eniyilenmiş hiperparametrelerle tutarlı ve tatmin edici sonuçlar vermiş; her iki durumda da RMSE değeri yaklaşık 2,8 düzeyinde kalmış ve R² skoru 0,842–0,844 arasında gerçekleşmiştir. Bu, modelin geçerli bir başarı sağladığını

göstermektedir. Ancak LightGBM ve Lasso Regresyon gibi modeller, özellikle eniyilenseler dahi diğer modellere göre daha düşük R^2 skorları ve daha yüksek hata metrikleri sergileyerek zayıf bir performans ortaya koymuştur.

Genel olarak değerlendirildiğinde, eniyilenmiş SVR modeli, doğrusal olmayan yapıları yakalamadaki başarısıyla öne çıkarken; varsayılan MLPR ise yapay sinir ağlarının karmaşık veri yapılarına duyarlılığını yansıtmıştır. Bu sonuçlar, Çalışma 5 veri setinin doğrusal olmayan ilişkiler içeren, esnek modellere daha duyarlı bir yapıya sahip olduğunu göstermektedir.

Bölüm kapsamında yapılan karşılaştırmalı analizler, farklı modelleme tekniklerinin yalnızca kendi yapısal kapasiteleriyle değil, aynı zamanda eğitildikleri veri setlerinin niteliğiyle de doğrudan ilişkili olduğunu açıkça ortaya koymaktadır. Bu bağlamda, Çalışma 1 veri seti, çoğu model için en yüksek R^2 skorlarının ve en düşük hata metriklerinin elde edildiği yapı olmuştur. Özellikle MLPR, SVR ve RF gibi modellerin hem varsayılan hem de eniyilenmiş sürümlerinde oldukça başarılı sonuçlar üretmesi, bu veri setinin öğrenmeye uygun, ayırt edici ve bilgi yoğun bir yapıya sahip olduğunu düşündürmektedir.

Diğer yandan, Çalışma 2 veri seti birçok model açısından görece düşük R^2 skorları ve yüksek hata değerleri üretmiştir. Bu durum, veri setindeki özniteliklerin hedef değişkeni açıklama gücünün daha zayıf olabileceğine işaret etmektedir. Çalışma 3 ve Çalışma 4, genellikle Çalışma 1'e yakın performanslar üretmiş; özellikle eniyilenmiş modellerde dikkat çekici iyileşmeler sağlanmıştır. Çalışma 5 ise bazı modellerde yüksek başarılar sağlasa da modelden modele değişen sonuçlarla daha değişken bir yapı sergilemiştir.

Sonuç olarak, modelleme perspektifinden bakıldığında Çalışma 1 veri seti, en dengeli, güçlü ve genellenebilir tahmin performansları sunmuş; dolayısıyla yapı kimyasalı üreticisinin çimento fiyat tahmini bağlamında en öngörülebilir veri seti olma özelliği göstermiştir. Bu durum, veri seti oluşturulurken sektörel uzman görüşlerinin dikkate alınması gerektiğini göstermektedir. Gelecekteki uygulamalarda da bu tarz bilgi destekli öznitelik seçimi yaklaşımlarının dikkate alınması önerilmektedir.

5.3. Genel (Bütünleşik) Değerlendirme ve Stratejik Sonuçlar

Bu bölümde, 5.1 ve 5.2 başlıklarında ayrı ayrı değerlendirilen modellerin ve çalışmaların sonuçları bir araya getirilerek kapsamlı ve bütünleşik bir analiz gerçekleştirilecektir. Amaç, her bir çalışmada elde edilen sonuçları yatay düzlemde karşılaştırmanın ötesine geçerek, tüm modelleme senaryoları ve hiperparametre yapılandırmaları arasında genel anlamda en yüksek başarıyı sağlayan yaklaşımları belirlemektir. Bu kapsamda hem varsayılan hem de eniyilenmiş hiperparametreler ile çalışan modellerin tüm çalışmalar üzerindeki performansları bütünleşik biçimde sıralanacak; ayrıca çalışma bazlı değişkenliğin, model türlerinin ve hiperparametre eniyilemesinin etkileri bir bütün olarak tartışılacaktır. Böylelikle, yapı kimyasalı üreticisinin gri çimento satın alma fiyatını tahmin etmek amacıyla yürütülen bu çalışmada, hangi modelleme stratejilerinin daha istikrarlı ve güçlü sonuçlar sunduğu ortaya konulacaktır.

Bütünleşik performans analizi bulgularının tamamına ilişkin detaylı karşılaştırma tablosu (R^2 değerine göre büyükten küçüğe sıralanmış) Ek Açıklama-C’de sunulmuştur. Metin içinde analiz edilebilirliği kolaylaştırmak adına en yüksek R^2 değerine sahip ilk 10 çalışma-model ikilisi Çizelge 5.14’te sunulmuştur.

Çizelge 5.14. Bütünleşik Karşılaştırma Tablosu (En Yüksek R^2 Değerine Sahip İlk 10 Çalışma-Model İkili)

Çalışma	Model	MAE	MSE	RMSE	R^2	Bar (R^2)
Çalışma 3	MLPR (Eniyilenmiş)	1,405	4,566	2,137	0,908	
Çalışma 1	MLPR (Varsayılan)	1,505	4,683	2,164	0,906	
Çalışma 1	MLPR (Eniyilenmiş)	1,505	4,690	2,166	0,906	
Çalışma 1	SVR (Eniyilenmiş)	1,512	4,731	2,175	0,905	
Çalışma 5	SVR (Eniyilenmiş)	1,578	4,990	2,234	0,900	
Çalışma 2	SVR (Eniyilenmiş)	1,755	5,198	2,280	0,896	
Çalışma 4	RF (Varsayılan)	1,780	5,605	2,367	0,888	
Çalışma 3	RF (Varsayılan)	1,840	5,736	2,395	0,885	
Çalışma 5	MLPR (Varsayılan)	1,779	5,741	2,396	0,885	
Çalışma 4	XGBoost (Varsayılan)	1,788	5,864	2,422	0,882	

Ek Açıklama-C ve Çizelge 5.14’den görüleceği üzere tüm çalışmalar ve modelleme senaryoları dikkate alındığında, MLPR (Eniyilenmiş) modeli Çalışma 3’te elde ettiği MAE= 1,405, MSE= 4,566, RMSE= 2,137 ve R^2 = 0,908 değerleriyle bütünleşik sıralamada en

başarılı model olarak öne çıkmıştır. Bu sonucu Çalışma 1'deki varsayılan haliyle yine MLPR takip etmekte, bu durum modelin hem varsayılan hem de eniyilenmiş versiyonlarının yüksek başarı potansiyeline sahip olduğunu göstermektedir. Bu başarıyı destekleyen bir diğer güçlü model ise SVR (Eniyilenmiş) olup, özellikle Çalışma 1, 2, 3 ve 5'te $R^2 = 0,90$ ve üzeri sonuçlar vererek öne çıkmıştır.

En iyi performans gösteren modellerin büyük çoğunluğu, hem sinir ağı tabanlı (MLPR) hem de kernel temelli (SVR) modellerden oluşmuştur. Bu durum, gri çimento fiyatlarının tahmin edilmesinde doğrusal olmayan ve karmaşık ilişkileri yakalama kapasitesi yüksek modellerin daha etkili olduğunu göstermektedir. Öte yandan, RF ve XGBoost gibi ağaç tabanlı topluluk öğrenmesi temelli modeller, belirli çalışmalarda (Özellikle Çalışma 4 ve 5) rekabetçi sonuçlar üretse de entegre sıralamada genel olarak MLPR ve SVR modellerinin gerisinde kalmıştır.

Hiperparametre eniyilemesinin etkisi genel olarak pozitif yönde olmuş, ancak bu etki model ve çalışmaya bağlı olarak değişkenlik göstermiştir. Örneğin, SVR ve MLPR modelleri eniyilendiğinde anlamlı iyileşmeler sağlarken; bazı durumlarda (Örneğin RF veya GB gibi modellerde), varsayılan yapılandırmalar dahi oldukça güçlü sonuçlar üretebilmiştir. Bu durum, varsayılan yapılandırmaların modelin doğası gereği zaten iyi ayarlanmış olduğunu düşündürmektedir.

Çalışmalar arasında model performansları açısından belirgin farklılıklar gözlenmiştir. Özellikle Çalışma 2, çoğu model için görece daha düşük R^2 değerleri ile ayırırken; Çalışma 1 ve 3 en yüksek doğruluklara ulaşılan çalışmalardır. Bu farkın temelinde, kullanılan öznitelik setlerinin niteliği ve tahmin gücü yüksek değişkenlerin temsil düzeyi yatmaktadır. Nitekim, Çalışma 1'de sektörel uzmanlıkla seçilen değişkenler, Çalışma 3'te ise iki farklı seçimin birleştirilmiş hali kullanılmış ve her iki durumda da model performanslarında belirgin artışlar gözlemlenmiştir.

Özetle, bütünsel değerlendirme sonucunda, gri çimento fiyat tahmininde en başarılı model MLPR (Eniyilenmiş) olarak belirlenmiştir. Bunu SVR (Eniyilenmiş) ve RF (Varsayılan/Eniyilenmiş) takip etmektedir. Bu modellerin, sektörel uygulamalarda öncelikli

olarak deęerlendirilmesi ve ilerleyen alıřmalar iin temel modelleme yapı tařlarını oluřturması nerilmektedir.

Model bazlı performanslar, alıřma bazlı karřılařtırmalar ve nihayetinde btnleřik bir deęerlendirme ile en bařarılı model yapılarına ulařılmıřtır. Bulgular, veri seti ierięinin model bařarısını doęrudan etkiledięini ve bazı modellerin hiperparametre eniyilemesinden yksek derecede fayda saęladıęını gstermiřtir. Elde edilen sonular gerek sektrel gerek akademik baęlamda uygulanabilirlięi yksek model nerilerine zemin hazırlamaktadır. Bu kapsamlı analiz srecinin ardından, alıřmanın son blmne geilmekte ve bulgular ıřıęında genel bir deęerlendirme ile neriler sunulmaktadır.



6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, yapı kimyasalı üreticisinin Eskişehir Fabrikasına ait önemli hammaddelerinden birisi olan gri çimento satın alma fiyatlarının tahmini amacıyla, çeşitli öznitelik setleri ve makine öğrenmesi modelleri kullanılarak kapsamlı bir modelleme süreci yürütülmüştür. Gerçekleştirilen analizlerde hem sektörel bilgiye dayalı hem de korelasyon analizi kaynaklı öznitelik seçimi yöntemleri değerlendirilmiş, eksik veri problemleri yüzeysel gelişmiş yöntemlerle ele alınmış, farklı senaryolar altında modelleme başarısı karşılaştırılmıştır. Elde edilen bulgular, sektörel karar destek sistemlerinin veri bilimi temelli araçlarla güçlendirilmesi konusunda önemli katkılar sunmaktadır.

Modelleme sürecinde, toplam beş farklı çalışma kapsamında sekiz farklı makine öğrenmesi algoritması varsayılan ve eniyilenmiş hiperparametrelerle değerlendirilmiştir. Karşılaştırmalı bütünleşik analizlerde, özellikle Çalışma 3 kapsamında kullanılan öznitelik setiyle eniyilenmiş MLPR modeli en düşük hata değerleri ve en yüksek tahmin doğruluğuna ulaşarak (MAE= 1,405, RMSE= 2,137, R2= 0,908) en başarılı model-çalışma kombinasyonu olmuştur. Bununla birlikte, Çalışma 1’de varsayılan ve eniyilenmiş MLPR modelleri ile eniyilenmiş SVR modeli de benzer şekilde yüksek performans göstermiştir. Bu bulgular, seçilen özniteliklerin tahminleme gücü üzerinde belirleyici bir rol oynadığını ve modele uygun özelliklerle desteklenen yapay sinir ağları ile destek vektör regresyonunun güçlü öngörü araçları sunduğunu göstermektedir.

Öznitelik seçim yöntemlerinin model başarısı üzerindeki etkisi dikkat çekicidir. Sektör temsilcilerinden alınan uzman görüşleri ile oluşturulan öznitelik seti (Çalışma 1), korelasyon analizine dayalı seçimlere (Çalışma 2) kıyasla daha başarılı sonuçlar vermiştir. Bu durum, saha deneyimi ve operasyonel bilgiyle desteklenen değişkenlerin, ekonomik tahminleme problemlerinde oldukça değerli olduğunu ortaya koymaktadır. Diğer yandan, Çalışma 3’te iki farklı yaklaşımın birleştirilmesiyle elde edilen karma öznitelik setinin üstün performansı, disiplinlerarası iş birliğinin veri bilimi süreçlerine sağlayabileceği katkıyı açıkça gözler önüne sermektedir.

Gerçekleştirilen hiperparametre eniyilemesi ise, modellerin tamamında olmasa da birçoğunda anlamlı iyileşmeler sağlamıştır. Özellikle SVR ve MLPR gibi parametre

duyarlılığı yüksek modellerde, eniyilenmiş konfigürasyonların hata oranlarını belirgin ölçüde düşürdüğü görülmüştür. Bu durum, hiperparametre ayarlamalarının sadece teorik bir gereklilik değil, aynı zamanda pratik başarı için vazgeçilmez bir adım olduğunu vurgulamaktadır.

Çalışmanın sonuçları, yapı kimyasalları üretimi sektöründe faaliyet gösteren firmaların satın alma, bütçeleme ve stratejik planlama faaliyetlerine veri temelli destek sağlama potansiyeline sahiptir. Tahmin modelleri, özellikle yüksek doğrulukla çalışan yapay sinir ağı tabanlı sistemlerle entegre edilerek, fiyat dalgalanmalarının önceden öngörülmesi ve satın alma zamanlamalarının daha rasyonel bir temele oturtulması sağlanabilir. Ayrıca bu yaklaşım, benzer yapıda faaliyet gösteren diğer üretici firmalar için de örnek bir karar destek sistemi geliştirme sürecine ışık tutabilir.

Bununla birlikte, yapılan çalışmanın bazı sınırlılıkları da mevcuttur. Kullanılan veri seti yalnızca Eskişehir fabrikasına ait olup, diğer üretim tesislerini kapsamamaktadır. Ayrıca veriler, sektörel gizlilik nedeniyle belirli düzeyde manipülasyona uğramıştır. Bu nedenle, geliştirilen modellerin tüm yapı kimyasalı sektörüne doğrudan genellenebilirliği sınırlı olabilir.

Gelecekte yapılacak çalışmalar kapsamında; modelleme süreci diğer fabrikalara da genişletilebilir, farklı ham madde grupları (selüloz, kalsit, kraft kağıdı, polimerler) için benzer tahminleme sistemleri geliştirilebilir. Ayrıca, öznitelik mühendisliğinde doğal dil işleme (Natural Language Processing - NLP) ve daha güçlü etkileşim yakalayabilen derin öğrenme mimarileri de süreçlere entegre edilerek performans artırılabilir. Bunun yanında, geliştirilen modellerin sadeleştirilmiş versiyonları ile karar vericilere yönelik kullanıcı dostu bir arayüz oluşturularak, bu sistemin gerçek zamanlı bir tahminleme aracı olarak sektöre entegre edilmesi de önerilmektedir.

KAYNAKLAR DİZİNİ

- Afolabi, A. O., & Abimbola, O. (2022). Application of machine learning in cement price prediction through a web-based system. *International Journal of Electrical and Computer Engineering*, 12(5), 5214.
- Allison, P. D. (2002). Missing data. SAGE Publications, Inc., <https://doi.org/10.4135/9781412985079>
- Asgari, M. H., Madanchi Zaj, M., Daneshvar, A., & Manzoor, D. (2024). Bitcoin price forecasting by applying combination of stacking method and Differential Evolution Algorithm. *International Journal of Finance & Managerial Accounting*, 9(35), 93-118.
- Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of clinical epidemiology*, 68(6), 627-636.
- Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Biopsychosocial Science and Medicine*, 66(3), 411-421.
- Başarır, Ç., & Erçakar, M. E. (2016). An analysis of the relationship between crude oil prices, current account deficit and exchange rates: Turkish experiment. *International Journal of Economics and Finance*, 8(11), 48-59.
- Benhelal, E., Zahedi, G., Shamsaei, E., & Bahadori, A. (2013). Global strategies and potentials to curb CO2 emissions in cement industry. *Journal of cleaner production*, 51, 142-161.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The journal of machine learning research*, 13(1), 281-305.
- Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 24.
- Beygo, C., & Cakmak, G. (1998). Spatial analysis of cement production and consumption in Turkey.
- Bilgili, F., Kuşkaya, S., Gençoğlu, P., Ünlü, F., & Magazzino, C. (2024). Exchange rate pass-through at high and low frequencies in Turkey: a continuous wavelet transform model approach. *Journal of Economic Structures*, 13(1), 21.
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.
- Bloem, M. A. M., Dippelsman, M. R., & Mæhle, M. N. Ø. (Eds.). (2001). *Quarterly national accounts manual: concepts, data sources, and compilation*. International Monetary Fund.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

KAYNAKLAR DİZİNİ

- Cai, X., Li, D., & Li, F. (2024). Enhanced Carbon Price Forecasting Using Extended Sliding Window Decomposition with LSTM and SVR. *Mathematics*, 12(23), 3713.
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific model development*, 7(3), 1247-1250.
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & electrical engineering*, 40(1), 16-28.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Childs, C. M. (2021). *Design and Optimization of Cementitious Systems with Machine Learning*. Carnegie Mellon University.
- Chow, G. C., & Lin, A. L. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The review of Economics and Statistics*, 372-375.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *J. off. Stat*, 6(1), 3-73.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press.
- da Silva, C., Barbosa-Povoa, A. P., & Carvalho, A. (2020). Environmental monetization and risk assessment in supply chain design and planning. *Journal of Cleaner Production*, 270, 121552.
- De Boor, C., (2001). *A practical guide to splines* (Rev. ed.). New York: springer.
- Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *Journal of the american statistical association*, 66(333), 99-102.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46.
- Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, 9.
- Eurostat. (2018). *ESS Guidelines on Temporal Disaggregation, Benchmarking and Reconciliation*. European Statistical System.
- Eurostat. (2013). *Handbook on Seasonal Adjustment*. Luxembourg: Publications Office of the European Union.

KAYNAKLAR DİZİNİ

- Eurostat. (2017). *Handbook on Rapid Estimates*. European Union.
- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2). Cambridge: MIT press.
- Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, 84(11), 2809-2815.
- Guennioui, O., Chiadmi, D., & Amghar, M. (2024). Global stock price forecasting during a period of market stress using LightGBM. *International Journal of Computing and Digital Systems*, 15(1), 19-27.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. doi:10.1007/978-0-387-84858-7
- Haykin, S. (2009). *Neural networks and learning machines*, 3/E. Pearson Education India.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- Holz, C. A., & Mehrotra, A. N. (2014). Wage and price dynamics in a large emerging economy: The case of China.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, No. 1). New York: springer.
- Kantanantha, N., & Pattaraumpornchai, P. (2023, November). Machine Learning-Based Price Forecasting for Polypropylene Granules in Thailand. In *Proceedings of the 2023 6th International Conference on Computational Intelligence and Intelligent Systems* (pp. 14-19).
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Kittinaraporn, W., Tuprakay, S., & Prasittisopin, L. (2022). Effective modeling for construction activities of recycled aggregate concrete using artificial neural network. *Journal of Construction Engineering and Management*, 148(3), 04021206.

KAYNAKLAR DİZİNİ

- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, p. 13). New York: Springer.
- Kumar, S., Kumar, D. R., Wipulanusat, W., & Keawsawasvong, S. (2024). Development of ANN-based metaheuristic models for the study of the durability characteristics of high-volume fly ash self-compacting concrete with silica fume. *Journal of Building Engineering*, 94, 109844.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models*. McGraw-hill.
- Lequiller, F. and D. Blades (2014), *Understanding National Accounts: Second Edition*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264214637-en>.
- Li, L. (2021). Prediction of coal prices based on random forest and lasso regression. *Int Core J Eng*, 7, 67-73.
- Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.
- Liu, H., & Motoda, H. (Eds.). (2007). *Computational methods of feature selection*. CRC press.
- Liu, Q., He, P., Peng, S., Wang, T., & Ma, J. (2024). A Survey of Data-Driven Construction Materials Price Forecasting. *Buildings* (2075-5309), 14(10).
- Lu, R., Yao, K., Yang, H., & Dai, W. (2025). Long sequence Time-Series forecasting of rare earth price based on Variational Mode Decomposition and improved Random Forest. *International Journal of Machine Learning and Cybernetics*, 1-25.
- Ma, M., Tam, V. W., Le, K. N., & Osei-Kyei, R. (2024). A systematic literature review on price forecasting models in construction industry. *International Journal of Construction Management*, 24(11), 1191-1200.
- Makridakis, S., Wheelwright, S., & Hyndman, R. J. (1997). *Forecasting: Methods and Applications*, 3rd Ed. John Wiley & Sons.
- Maré, D. C., & Graham, D. J. (2013). Agglomeration elasticities and firm heterogeneity. *Journal of Urban Economics*, 75, 44-56.
- Miller, N. H., Osborne, M., Sheu G., & Sileo G., (2025). *Technology and market power in the cement industry*. RAND Journal of Economics. <https://www.nathanhmilller.org/cementmarkups.pdf>
- Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal*, 24(3), 69-71.
- Ozcan, G., Kocak, B., Gulbandilar, E., & Kocak, Y. (2024). Efficient machine learning models for estimation of compressive strengths of zeolite and diatomite substituting concrete in sodium chloride solution. *Arabian Journal for Science and Engineering*, 49(10), 14241-14256.

KAYNAKLAR DİZİNİ

- Parhi, S. K., Panda, S., Dwibedy, S., & Panigrahi, S. K. (2024). Metaheuristic optimization of machine learning models for strength prediction of high-performance self-compacting alkali-activated slag concrete. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 7(3), 2901-2928.
- Pavankumar, G., & Velmurugan, J. (2023, December). Enhancing Accuracy of Bitcoin Cost Prediction using Ridge Linear Classification over Lasso Regression. In *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)* (pp. 1-4). IEEE.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Polat, S., Kervancı, I. S., & Özceylan, E. (2024) Optimizing Production Planning in Cement Manufacturing Using SVR. *Computational Intelligence and Machine Learning*, 6-11.
- Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 9(3), e1301.
- Raskutti, G., Wainwright, M. J., & Yu, B. (2014). Early stopping and non-parametric regression: an optimal data-dependent stopping rule. *The Journal of Machine Learning Research*, 15(1), 335-366.
- Rodrigue, J. P. (2020). *The geography of transport systems*. Routledge.
- Roosbeh, M., Rouhi, A., Mohamed, N. A., & Jahadi, F. (2023). Generalized support vector regression and symmetry functional regression approaches to model the high-dimensional data. *Symmetry*, 15(6), 1262.
- Rubin, D. B. (Ed.). (1987). *Multiple Imputation for Nonresponse in Surveys*. *Wiley Series in Probability and Statistics*. doi:10.1002/9780470316696
- Saeed, N., Shafi, I., Pervez, S., Thompson, E. B., Castilla, A. K., Samad, M. A., & Ashraf, I. (2025). Intelligent Decision Making for Commodities Price Prediction: Opportunities, Challenges and Future Avenues. *Computational Economics*, 1-59.
- Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19), 2507-2517.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.
- scikit-learn developers. (2025). *GradientBoostingRegressor*. In *scikit-learn: Machine Learning in Python*. Alındığı Yer: scikit-learn documentation (<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html>)
- Shi, M., & Svensson, J. (2006). Political budget cycles: Do they differ across countries and why?. *Journal of public economics*, 90(8-9), 1367-1389.

KAYNAKLAR DİZİNİ

- Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3, p. 4). New York: springer.
- Scornet, E. (2020). Trees, forests, and impurity-based variable importance. arXiv. *arXiv preprint arXiv:2001.04295*.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14, 199-222.
- Tatliyer, M., & Yigit, F. (2016). Does exchange rate volatility really influence foreign trade? Evidence from Turkey. *International Journal of Economics and Finance*, 8(2), 33.
- Tetko, I. V., van Deursen, R., & Godin, G. (2024). Be aware of overfitting by hyperparameter optimization!. *Journal of Cheminformatics*, 16(1), 1-11.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267-288.
- Tran, V. Q., Mai, H. V. T., Nguyen, T. A., & Ly, H. B. (2022). Assessment of different machine learning techniques in predicting the compressive strength of self-compacting concrete. *Frontiers of Structural and Civil Engineering*, 16(7), 928-945.
- Türkiye İstatistik Kurumu (TÜİK). (2023). *Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları*. Erişim adresi: <https://www.tuik.gov.tr>
- Verma, V. (2024). Exploring Key XGBoost Hyperparameters: A Study on Optimal Search Spaces and Practical Recommendations for Regression and Classification. *International Journal of All Research Education and Scientific Methods (IJARESM)*, ISSN, 2455-6211.
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). South-Western Cengage Learning.
- Zhou, W., Yan, Z., & Zhang, L. (2024). A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction. *Scientific Reports*, 14(1), 5905.



EK AÇIKLAMALAR

Ek Açıklama-A: Korelasyon Analizi Temelli Öznitelik Seçimi için Korelasyon Matrisi Isı Haritası	162
Ek Açıklama- B: Korelasyon Analizi Temelli Öznitelik Seçimi için Korelasyon Matrisi	163
Ek Açıklama- C: R^2 Değerine Göre Büyükten Küçüğe Sıralanmış Çalışma-Model İkili Karşılaştırma Tablosu	164

Ek Açıklama-B: Korelasyon Analizi Temelli Öznitelik Seçimi için Korelasyon Matrisi

	Çimento Sanayi İstihdam Sayısı (kişi)	Çimento İhracatı (ton)	Klinker İhracatı (ton)	Çimento Üretim Miktarı (ton)	Klinker Üretim Miktarı (ton)	Yapı Ruhsatına Göre Bina Sayısı (adet)	Motorin Fiyatı (Eskişehir/Merkez) (TL/Litre)	Motorin Fiyatı (Eskişehir/Merkez) (USD/Litre)	Petrolük Birim Fiyatı (USD)	Petrolük İthalatı (kg)	Elektrik Fiyatı (TL/MWh)	Elektrik Fiyatı (USD/MWh)	Türkiye Nüfusu (kişi)	Türkiye Kent Nüfusu (kişi)	Türkiye Kırsal Nüfus (kişi)	Döviz Kuru (TL/EUR)	Döviz Kuru (TL/USD)	Brent Petrol (USD)	Üretici Fiyat Endeksi (ÜFE)	Kalker Ocakbaşı Satış Fiyatı (TL/Ton)	Kalker Ocakbaşı Satış Fiyatı (USD/Ton)	Alçıtaşı Ocakbaşı Satış Fiyatı (TL/Ton)	Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton)	Çimento Kili Ocakbaşı Satış Fiyatı (TL/Ton)	Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton)	Demir Cevheri Ocakbaşı Satış Fiyatı (TL/Ton)	Demir Cevheri Ocakbaşı Satış Fiyatı (USD/Ton)	Brüt Asgari Ücret (TL)	Brüt Asgari Ücret (USD)	Herhangi bir seçim var mı? (genel veya yerel) (Evet/Hayır)	Eskişehir Fabrika Gri Çimento Satın Alın Fiyatı Endeksi
Çimento Sanayi İstihdam Sayısı (kişi)	1,00	0,09	0,13	0,04	-0,05	0,08	0,44	-0,60	0,37	0,00	0,39	0,03	0,61	0,61	-0,55	0,48	0,49	-0,34	0,46	0,43	-0,35	0,46	-0,55	0,44	-0,26	0,55	-0,38	0,47	-0,12	0,02	-0,13
Çimento İhracatı (ton)	0,09	1,00	0,59	0,23	0,35	-0,05	0,46	-0,44	0,52	-0,04	0,56	0,39	0,74	0,75	-0,74	0,56	0,56	0,15	0,51	0,45	-0,51	0,53	-0,75	0,47	-0,43	0,71	-0,71	0,47	-0,48	-0,05	-0,23
Klinker İhracatı (ton)	0,13	0,59	1,00	-0,22	-0,10	-0,28	-0,03	-0,66	0,09	0,02	-0,01	-0,18	0,60	0,61	-0,65	0,09	0,08	-0,29	0,00	-0,09	-0,80	0,01	-0,61	-0,09	-0,79	0,36	-0,38	0,06	-0,21	-0,05	-0,75
Çimento Üretim Miktarı (ton)	0,04	0,23	-0,22	1,00	0,78	0,36	0,19	0,02	0,01	-0,07	0,16	0,09	0,03	0,03	-0,04	0,19	0,18	0,07	0,19	0,16	0,00	0,16	-0,07	0,16	0,01	0,10	-0,13	0,14	-0,11	0,00	0,13
Klinker Üretim Miktarı (ton)	-0,05	0,35	-0,10	0,78	1,00	0,32	0,13	-0,05	-0,02	-0,04	0,12	0,07	0,09	0,09	-0,13	0,16	0,14	0,08	0,13	0,08	-0,11	0,11	-0,11	0,09	-0,10	0,12	-0,22	0,11	-0,04	-0,03	0,01
Yapı Ruhsatına Göre Bina Sayısı (adet)	0,08	-0,05	-0,28	0,36	0,32	1,00	0,15	0,11	0,03	0,02	0,15	0,13	-0,08	-0,07	0,01	0,14	0,13	0,07	0,15	0,11	0,11	0,10	0,06	0,11	0,14	0,01	-0,01	0,10	-0,05	0,10	0,21
Motorin Fiyatı (Eskişehir/Merkez) (TL/Litre)	0,44	0,46	-0,03	0,19	0,13	0,15	1,00	-0,09	0,69	-0,03	0,85	0,48	0,64	0,63	-0,42	0,97	0,98	0,31	0,98	0,93	-0,06	0,93	-0,65	0,94	0,03	0,85	-0,58	0,94	-0,30	-0,05	0,24
Motorin Fiyatı (Eskişehir/Merkez) (USD/Litre)	-0,60	-0,44	-0,66	0,02	-0,05	0,11	-0,09	1,00	-0,06	-0,03	-0,11	0,26	-0,74	-0,75	0,76	-0,26	-0,25	0,69	-0,18	-0,13	0,78	-0,22	0,71	-0,12	0,76	-0,51	0,59	-0,24	0,13	0,06	0,62
Petrolük Birim Fiyatı (USD)	0,37	0,52	0,09	0,01	-0,02	0,03	0,69	-0,06	1,00	-0,02	0,77	0,65	0,59	0,58	-0,45	0,67	0,68	0,49	0,68	0,66	-0,08	0,68	-0,57	0,70	0,05	0,69	-0,43	0,58	-0,54	-0,04	0,14
Petrolük İthalatı (kg)	0,00	-0,04	0,02	-0,07	-0,04	0,02	-0,03	-0,03	-0,02	1,00	-0,06	-0,08	-0,03	-0,03	0,03	-0,02	-0,02	-0,06	-0,03	-0,06	-0,08	-0,06	-0,02	-0,06	-0,09	-0,03	-0,03	0,00	-0,01	0,00	-0,11
Elektrik Fiyatı (TL/MWh)	0,39	0,56	-0,01	0,16	0,12	0,15	0,85	-0,11	0,77	-0,06	1,00	0,82	0,64	0,64	-0,53	0,82	0,84	0,35	0,86	0,82	-0,06	0,84	-0,63	0,85	0,07	0,79	-0,60	0,72	-0,53	-0,06	0,37
Elektrik Fiyatı (USD/MWh)	0,03	0,39	-0,18	0,09	0,07	0,13	0,48	0,26	0,65	-0,08	0,82	1,00	0,24	0,24	-0,23	0,41	0,44	0,58	0,48	0,46	0,20	0,46	-0,23	0,51	0,32	0,36	-0,29	0,28	-0,56	-0,05	0,55
Türkiye Nüfusu (kişi)	0,61	0,74	0,60	0,03	0,09	-0,08	0,64	-0,74	0,59	-0,03	0,64	0,24	1,00	1,00	-0,87	0,74	0,74	-0,14	0,68	0,62	-0,65	0,72	-0,96	0,64	-0,56	0,93	-0,74	0,70	-0,36	-0,07	-0,36
Türkiye Kent Nüfusu (kişi)	0,61	0,75	0,61	0,03	0,09	-0,07	0,63	-0,75	0,58	-0,03	0,64	0,24	1,00	1,00	-0,89	0,74	0,74	-0,14	0,68	0,62	-0,66	0,71	-0,96	0,63	-0,57	0,92	-0,75	0,68	-0,37	-0,07	-0,36
Türkiye Kırsal Nüfus (kişi)	-0,55	-0,74	-0,65	-0,04	-0,13	0,01	-0,42	0,76	-0,45	0,03	-0,53	-0,23	-0,87	-0,89	1,00	-0,52	-0,53	0,28	-0,47	-0,38	0,70	-0,47	0,84	-0,41	0,58	-0,72	0,72	-0,45	0,35	0,10	0,34
Döviz Kuru (TL/EUR)	0,48	0,56	0,09	0,19	0,16	0,14	0,97	-0,26	0,67	-0,02	0,82	0,41	0,75	0,74	-0,52	1,00	1,00	0,22	0,99	0,95	-0,18	0,97	-0,75	0,95	-0,09	0,93	-0,68	0,97	-0,31	-0,04	0,11
Döviz Kuru (TL/USD)	0,49	0,56	0,08	0,18	0,14	0,13	0,98	-0,25	0,68	-0,02	0,84	0,44	0,75	0,74	-0,53	1,00	1,00	0,22	0,99	0,95	-0,17	0,97	-0,75	0,95	-0,08	0,93	-0,67	0,97	-0,33	-0,04	0,13
Brent Petrol (USD)	-0,34	0,15	-0,29	0,07	0,08	0,07	0,31	0,69	0,49	-0,06	0,35	0,58	-0,13	-0,14	0,28	0,22	0,22	1,00	0,26	0,27	0,42	0,24	0,11	0,29	0,45	0,09	0,07	0,16	-0,26	0,06	0,42
Üretici Fiyat Endeksi (ÜFE)	0,46	0,51	0,00	0,19	0,13	0,15	0,98	-0,18	0,68	-0,03	0,86	0,48	0,70	0,68	-0,47	0,99	0,99	0,26	1,00	0,97	-0,07	0,97	-0,68	0,97	0,01	0,90	-0,63	0,96	-0,30	-0,03	0,23
Kalker Ocakbaşı Satış Fiyatı (TL/Ton)	0,43	0,45	-0,09	0,16	0,08	0,11	0,93	-0,13	0,66	-0,06	0,82	0,46	0,63	0,62	-0,38	0,95	0,95	0,27	0,97	1,00	0,09	0,99	-0,58	1,00	0,17	0,86	-0,57	0,92	-0,25	0,00	0,33
Kalker Ocakbaşı Satış Fiyatı (USD/Ton)	-0,35	-0,51	-0,80	0,00	-0,11	0,11	-0,06	0,78	-0,08	-0,08	-0,06	0,20	-0,65	-0,66	0,70	-0,18	-0,17	0,42	-0,07	0,09	1,00	-0,02	0,74	0,08	0,98	-0,39	0,52	-0,13	0,32	0,11	0,81
Alçıtaşı Ocakbaşı Satış Fiyatı (TL/Ton)	0,46	0,53	0,01	0,16	0,11	0,10	0,93	-0,22	0,68	-0,06	0,84	0,46	0,72	0,71	-0,47	0,97	0,97	0,24	0,97	0,99	-0,02	1,00	-0,67	0,99	0,06	0,91	-0,63	0,93	-0,28	-0,02	0,24
Alçıtaşı Ocakbaşı Satış Fiyatı (USD/Ton)	-0,55	-0,75	-0,61	-0,07	-0,11	0,06	-0,65	0,71	-0,57	-0,02	-0,63	-0,23	-0,96	-0,96	0,84	-0,75	-0,75	0,11	-0,68	-0,58	0,74	-0,67	1,00	-0,59	0,66	-0,89	0,81	-0,68	0,52	0,08	0,42
Çimento Kili Ocakbaşı Satış Fiyatı (TL/Ton)	0,44	0,47	-0,09	0,16	0,09	0,11	0,94	-0,12	0,70	-0,06	0,85	0,51	0,64	0,63	-0,41	0,95	0,95	0,29	0,97	1,00	0,08	0,99	-0,59	1,00	0,18	0,86	-0,58	0,91	-0,29	-0,01	0,35
Çimento Kili Ocakbaşı Satış Fiyatı (USD/Ton)	-0,26	-0,43	-0,79	0,01	-0,10	0,14	0,03	0,76	0,05	-0,09	0,07	0,32	-0,56	-0,57	0,58	-0,09	-0,08	0,45	0,01	0,17	0,98	0,06	0,66	0,18	1,00	-0,31	0,43	-0,07	0,22	0,08	0,86
Demir Cevheri Ocakbaşı Satış Fiyatı (TL/Ton)	0,55	0,71	0,36	0,10	0,12	0,01	0,85	-0,51	0,69	-0,03	0,79	0,36	0,93	0,92	-0,72	0,93	0,93	0,09	0,90	0,86	-0,39	0,91	-0,89	0,86	-0,31	1,00	-0,74	0,90	-0,32	-0,06	-0,11
Demir Cevheri Ocakbaşı Satış Fiyatı (USD/Ton)	-0,38	-0,71	-0,38	-0,13	-0,22	-0,01	-0,58	0,59	-0,43	-0,03	-0,60	-0,29	-0,74	-0,75	0,72	-0,68	-0,67	0,07	-0,63	-0,57	0,52	-0,63	0,81	-0,58	0,43	-0,74	1,00	-0,61	0,47	0,07	0,17
Brüt Asgari Ücret (TL)	0,47	0,47	0,06	0,14	0,11	0,10	0,94	-0,24	0,58	0,00	0,72	0,28	0,70	0,68	-0,45	0,97	0,97	0,16	0,96	0,92	-0,13	0,93	-0,68	0,91	-0,07	0,90	-0,61	1,00	-0,12	-0,05	0,09
Brüt Asgari Ücret (USD)	-0,12	-0,48	-0,21	-0,11	-0,04	-0,05	-0,30	0,13	-0,54	-0,01	-0,53	-0,56	-0,36	-0,37	0,35	-0,31	-0,33	-0,26	-0,30	-0,25	0,32	-0,28	0,52	-0,29	0,22	-0,32	0,47	-0,12	1,00	0,00	0,04
Herhangi bir seçim var mı? (genel veya yerel) (Evet/Hayır)	0,02	-0,05	-0,05	0,00	-0,03	0,10	-0,05	0,06	-0,04	0,00	-0,06	-0,05	-0,07	-0,07	0,10	-0,04	-0,04	0,06	-0,03	0,00	0,11	-0,02	0,08	-0,01	0,08	-0,06	0,07	-0,05	0,00	1,00	0,08
Eskişehir Fabrika Gri Çimento Satın Alın Fiyatı Endeksi	-0,13	-0,23	-0,75	0,13	0,01	0,21	0,24	0,62	0,14	-0,11	0,37	0,55	-0,36	-0,36	0,34	0,11	0,13	0,42	0,23	0,33	0,81	0,24	0,42	0,35	0,86	-0,11	0,17	0,09	0,04	0,08	1,00

Ek Açıklama-C: R^2 Değerine Göre Büyükten Küçüğe Sıralanmış Çalışma-Model İkilisi Karşılaştırma Tablosu

Çalışma	Model	MAE	MSE	RMSE	R^2	Bar (R^2)
Çalışma 3	MLPR (Eniyilenmiş)	1,405	4,566	2,137	0,908	
Çalışma 1	MLPR (Varsayılan)	1,505	4,683	2,164	0,906	
Çalışma 1	MLPR (Eniyilenmiş)	1,505	4,690	2,166	0,906	
Çalışma 1	SVR (Eniyilenmiş)	1,512	4,731	2,175	0,905	
Çalışma 5	SVR (Eniyilenmiş)	1,578	4,990	2,234	0,900	
Çalışma 2	SVR (Eniyilenmiş)	1,755	5,198	2,280	0,896	
Çalışma 4	RF (Varsayılan)	1,780	5,605	2,367	0,888	
Çalışma 3	RF (Varsayılan)	1,840	5,736	2,395	0,885	
Çalışma 5	MLPR (Varsayılan)	1,779	5,741	2,396	0,885	
Çalışma 4	XGBoost (Varsayılan)	1,788	5,864	2,422	0,882	
Çalışma 4	MLPR (Eniyilenmiş)	1,668	5,932	2,435	0,881	
Çalışma 1	RF (Eniyilenmiş)	1,818	5,952	2,440	0,881	
Çalışma 3	RF (Eniyilenmiş)	1,860	6,060	2,462	0,879	
Çalışma 3	XGBoost (Varsayılan)	1,828	6,085	2,467	0,878	
Çalışma 4	RF (Eniyilenmiş)	1,871	6,108	2,471	0,878	
Çalışma 1	RF (Varsayılan)	1,814	6,341	2,518	0,873	
Çalışma 1	XGBoost (Varsayılan)	1,884	6,581	2,565	0,868	
Çalışma 4	LightGBM (Varsayılan)	2,023	6,669	2,582	0,866	
Çalışma 4	SVR (Eniyilenmiş)	1,839	6,731	2,594	0,865	
Çalışma 4	LightGBM (Eniyilenmiş)	1,941	6,776	2,603	0,864	
Çalışma 3	SVR (Eniyilenmiş)	1,787	6,860	2,619	0,862	
Çalışma 4	GB (Eniyilenmiş)	1,864	6,876	2,622	0,862	
Çalışma 4	GB (Varsayılan)	1,909	6,979	2,642	0,860	
Çalışma 1	XGBoost (Eniyilenmiş)	1,844	7,014	2,648	0,859	
Çalışma 3	GB (Eniyilenmiş)	1,849	7,038	2,653	0,859	
Çalışma 3	GB (Varsayılan)	1,877	7,113	2,667	0,857	
Çalışma 5	MLPR (Eniyilenmiş)	2,004	7,149	2,674	0,857	
Çalışma 3	XGBoost (Eniyilenmiş)	1,996	7,191	2,682	0,856	
Çalışma 3	LightGBM (Varsayılan)	2,186	7,332	2,708	0,853	
Çalışma 3	LightGBM (Eniyilenmiş)	2,176	7,505	2,740	0,850	
Çalışma 5	RF (Varsayılan)	2,149	7,764	2,786	0,844	
Çalışma 4	XGBoost (Eniyilenmiş)	2,116	7,787	2,791	0,844	
Çalışma 5	RF (Eniyilenmiş)	2,150	7,864	2,804	0,842	
Çalışma 5	XGBoost (Varsayılan)	2,019	7,874	2,806	0,842	
Çalışma 4	MLPR (Varsayılan)	2,011	7,879	2,807	0,842	
Çalışma 1	GB (Varsayılan)	1,992	8,060	2,839	0,838	
Çalışma 2	MLPR (Eniyilenmiş)	2,175	8,182	2,860	0,836	
Çalışma 1	GB (Eniyilenmiş)	2,016	8,268	2,875	0,834	
Çalışma 1	LightGBM (Eniyilenmiş)	2,168	8,506	2,917	0,829	

Ek Açıklama-C: R^2 Değerine Göre Büyükten Küçüğe Sıralanmış Çalışma-Model İkilişi Karşılaştırma Tablosu (devamı)

Çalışma	Model	MAE	MSE	RMSE	R^2	Bar (R^2)
Çalışma 5	XGBoost (Eniyilenmiş)	2,252	8,637	2,939	0,827	
Çalışma 1	LightGBM (Varsayılan)	2,217	8,972	2,995	0,820	
Çalışma 5	GB (Varsayılan)	2,414	9,434	3,071	0,811	
Çalışma 1	Rigde Regresyon (Eniyilenmiş)	2,261	9,490	3,081	0,810	
Çalışma 1	Lasso Regresyon (Eniyilenmiş)	2,260	9,490	3,081	0,810	
Çalışma 5	GB (Eniyilenmiş)	2,408	9,527	3,087	0,809	
Çalışma 2	RF (Eniyilenmiş)	2,211	9,633	3,104	0,807	
Çalışma 4	Lasso Regresyon (Eniyilenmiş)	2,394	9,842	3,137	0,803	
Çalışma 4	Rigde Regresyon (Eniyilenmiş)	2,395	9,843	3,137	0,803	
Çalışma 3	Lasso Regresyon (Eniyilenmiş)	2,395	9,848	3,138	0,803	
Çalışma 3	Rigde Regresyon (Eniyilenmiş)	2,396	9,848	3,138	0,803	
Çalışma 5	Rigde Regresyon (Varsayılan)	2,573	9,918	3,149	0,801	
Çalışma 5	Rigde Regresyon (Eniyilenmiş)	2,573	9,918	3,149	0,801	
Çalışma 5	Lasso Regresyon (Eniyilenmiş)	2,581	10,061	3,172	0,798	
Çalışma 2	GB (Eniyilenmiş)	2,167	10,097	3,178	0,798	
Çalışma 2	RF (Varsayılan)	2,283	10,215	3,196	0,795	
Çalışma 1	Rigde Regresyon (Varsayılan)	2,348	10,268	3,204	0,794	
Çalışma 2	XGBoost (Varsayılan)	2,354	10,268	3,204	0,794	
Çalışma 4	Rigde Regresyon (Varsayılan)	2,576	10,480	3,237	0,790	
Çalışma 3	Rigde Regresyon (Varsayılan)	2,588	10,656	3,264	0,786	
Çalışma 4	SVR (Varsayılan)	2,609	11,290	3,360	0,774	
Çalışma 3	SVR (Varsayılan)	2,589	11,470	3,387	0,770	
Çalışma 5	Lasso Regresyon (Varsayılan)	2,771	11,840	3,441	0,763	
Çalışma 1	SVR (Varsayılan)	2,559	12,082	3,476	0,758	
Çalışma 2	XGBoost (Eniyilenmiş)	2,568	12,115	3,481	0,757	
Çalışma 5	LightGBM (Varsayılan)	2,696	12,237	3,498	0,755	
Çalışma 5	LightGBM (Eniyilenmiş)	2,749	12,880	3,589	0,742	
Çalışma 3	Lasso Regresyon (Varsayılan)	2,962	12,916	3,594	0,741	
Çalışma 2	GB (Varsayılan)	2,631	12,923	3,595	0,741	
Çalışma 4	Lasso Regresyon (Varsayılan)	2,938	12,943	3,598	0,741	
Çalışma 5	SVR (Varsayılan)	2,755	13,031	3,610	0,739	
Çalışma 2	SVR (Varsayılan)	2,868	13,177	3,630	0,736	
Çalışma 2	Lasso Regresyon (Varsayılan)	3,044	13,565	3,683	0,728	
Çalışma 3	MLPR (Varsayılan)	2,899	13,817	3,717	0,723	
Çalışma 2	LightGBM (Eniyilenmiş)	2,736	14,402	3,795	0,711	
Çalışma 2	LightGBM (Varsayılan)	2,801	14,446	3,801	0,710	
Çalışma 2	Rigde Regresyon (Varsayılan)	3,109	14,563	3,816	0,708	
Çalışma 2	Rigde Regresyon (Eniyilenmiş)	3,109	14,563	3,816	0,708	
Çalışma 2	Lasso Regresyon (Eniyilenmiş)	3,132	14,989	3,872	0,700	

Ek Açıklama-C: R^2 Değerine Göre Büyükten Küçüğe Sıralanmış Çalışma-Model İkilisi Karşılaştırma Tablosu (devamı)

Çalışma	Model	MAE	MSE	RMSE	R^2	Bar (R^2)
Çalışma 2	MLPR (Varsayılan)	3,618	19,471	4,413	0,610	
Çalışma 1	Lasso Regresyon (Varsayılan)	4,021	29,168	5,401	0,415	