

T.C.
ERZİNCAN BİNALİ YILDIRIM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

MİKROSİMÜLASYON PROGRAMI KULLANILARAK
ÜRETİLEN KENT İÇİ TRAFİK KORİDORUNA AİT VERİLER
İLE MAKİNE ÖĞRENMESİ TAHMİN MODELLERİNİN
OLUŞTURULMASI

Fatih Ahmet DENİZ

Danışman: Dr. Öğr. Üyesi Muhammed Ali ÇOLAK

İNŞAAT MÜHENDİSLİĞİ
ANA BİLİM DALI

ERZİNCAN
2023

Her Hakkı Saklıdır.

ÖZET

Yüksek Lisans

MİKROSİMÜLASYON PROGRAMI KULLANILARAK ÜRETİLEN KENT İÇİ TRAFİK KORİDORUNA AİT VERİLER İLE MAKİNE ÖĞRENMESİ TAHMİN MODELLERİNİN OLUŞTURULMASI

Fatih Ahmet DENİZ

Erzincan Binali Yıldırım Üniversitesi
Fen Bilimleri Enstitüsü
İnşaat Mühendisliği Ana Bilim Dalı

Danışman: Dr. Öğr. Üyesi Muhammed Ali ÇOLAK

Günümüzde özel araç sahipliğinin hızla artması nedeniyle trafik sorunları özellikle kent merkezlerinde ciddi boyutlara ulaşmıştır. Bu sorunlardan en önemlileri trafikteki gecikme süreleri ve buna bağlı olarak seyahat süresindeki artışlardır. Bu gecikmeler sonucunda yakıt tüketimi ve sera gazı salınımı artışı, ekonomik ve çevresel anlamda zararlara yol açmaktadır.

Bu çalışmada, simülasyon verileri ile sentetik veriler üretilerek makine öğrenmesi modelleri oluşturulması hedeflenmiştir. Bu amaçla, Erzincan ili merkezindeki 7 kavşaktan oluşan bir trafik koridoru seçilerek, kavşak tipleri, sinyalizasyon süreleri, taşıt hacimleri ve şerit sayılarındaki değişikliklerin kombinasyonu ile 27.648 senaryo oluşturulmuş, AIMSUN trafik simülasyon programı kullanılarak modellenmiş ve simüle edilmiştir. Simülasyon sonucu tüm senaryolar için elde edilen toplam seyahat süresi, gecikme süresi ve IEM CO₂ parametrelerine ait veriler ile makine öğrenmesi modelleri kurulmuştur. Modellerde oluşturulan Lineer Regresyon, Destek Vektör Makinesi, Karar Ağacı, Rastgele Orman, AdaBoost, Gradyan Artırma ve XGBoost algoritmalarının R², MAE, MSE ve RMSE hata metriklerine göre performansları karşılaştırılmıştır. Ayrıca 27.648 senaryo içinden seçilen üç farklı senaryo ile algoritmalar kendi aralarında ve simülasyon sonuçlarıyla karşılaştırılmıştır.

Hata metriklerine göre başarılı modeller olarak gecikme süresi için Lineer Regresyon ve Destek Vektör Makinesi, IEM CO₂ için XGBoost ve toplam seyahat süresi için Destek Vektör Makinesi ve XGBoost seçilmiştir. Bu modellerde incelenen katsayı değerlerine göre kavşak geometrisi ile otomobil trafik hacmi değişkenlerinin, üç parametre için de hayati öneme sahip olduğu saptanmıştır. Bu çalışmada simülasyon verileri kullanımı ile sentetik veriler üretilerek başarılı makine öğrenmesi modellerinin kurulabileceği görülmüştür.

2022, 176 Sayfa

Anahtar Kelimeler: AIMSUN, makine öğrenmesi, tahmin modelleri, trafik simülasyonu, veri üretimi, yapay zekâ algoritmaları

ABSTRACT

Master Thesis

DEVELOPING MACHINE LEARNING FORECAST MODELS USING THE DATA OF THE URBAN TRAFFIC CORRIDOR GENERATED VIA A MICROSIMULATION PROGRAM

Fatih Ahmet DENİZ

Erzincan Binali Yıldırım University
Graduate School of Natural and Applied Sciences
Department of Civil Engineering

Supervisor: Assist. Prof. Muhammed Ali ÇOLAK

With the rapid growth in private vehicle ownership in recent years, traffic problems have reached a serious level, especially in urban areas. The most significant of these problems are traffic delays and the consequent increase in travel time. These delays result in an increase in fuel consumption and greenhouse gas emissions, causing economic and environmental harm.

The purpose of this study was to develop machine learning models based on synthetic data generated from simulation data. A traffic corridor comprising seven intersections in Erzincan's city center was selected for this study, where 27,648 scenarios were created using a combination of intersection types, signaling times, vehicle volumes, and lane changes, and the AIMSUN traffic simulation program was used to model and simulate the scenarios. A machine learning model was developed based on data obtained from simulations regarding total travel time, delay time and IEM CO₂ parameters for each scenario. R², MAE, MSE, and RMSE error metrics were used to compare the performance of Linear Regression, Support Vector Machine, Decision Tree, Random Forest, AdaBoost, Gradient Boosting, and XGBoost algorithms included in the models. Furthermore, three different scenarios selected from 27,648 scenarios were compared with one another and with the results of the simulations.

According to error metrics, SVM and Linear Regression were selected as successful models for Delay Time, XGBoost for IEM CO₂, and SVM and XGBoost for Total Travel Time. Based on the coefficient values analyzed in these models, it was determined that the intersection geometry and automobile traffic volume variables were of paramount importance for all three parameters. The results of this study suggested that successful machine learning models were possible to be developed through the use of synthetic data along with simulation data.

2022, 176 pages

Keywords: AIMSUN, machine learning, forecast models, traffic simulation, data production, artificial intelligence algorithms

TEŞEKKÜR

Tez çalışmam boyunca bana yol gösteren ve desteğini esirgemeyen saygıdeğer danışman hocam Dr. Öğr. Üyesi Muhammed Ali ÇOLAK'a,

Araştırmamın her aşamasında yolumu aydınlatan, zamanını ve bilgisini benden esirgemeyen hocam Dr. Fatih HATTATOĞLU'na,

Araştırma sürecinde tecrübeleri ve önerileriyle araştırmaya katkı sağlayan Dr. Öğr. Üyesi Fatih İrfan BAŞ'a ve Arş. Gör. Ahmet Oğuz DEMİRİZ'e,

Beni akademik anlamda teşvik eden ve bu alanda gelişmeme destek olan ablam Öğr. Gör. Dr. Hacer DENİZ'e,

Tüm yaşamım boyunca olduğu gibi araştırma sürecinde de desteklerini benden esirgemeyen annem, babam ve diğer ablalarım sonsuz teşekkür ederim.

Fatih Ahmet DENİZ

Haziran, 2023.

İÇİNDEKİLER

	Sayfa
ÖZET.....	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER	iv
ŞEKİLLER LİSTESİ.....	vii
TABLolar LİSTESİ.....	xiii
SİMGELER ve KISALTMALAR	xv
1. GİRİŞ.....	1
2. KAYNAK ÖZETLERİ	2
3. KURAMSAL TEMELLER.....	6
3.1. Kavşaklar ve Sinyalizasyon	6
3.1.1. Kavşaklar.....	6
3.1.1.1. Eş düzeyli kavşaklar.....	6
3.1.1.2. Farklı düzeyli kavşaklar	10
3.1.2. Sinyalizasyon	17
3.1.2.1. Sinyalizasyon ile ilgili temel kavramlar.....	17
3.1.3. Servis seviyesi	18
3.2. Simülasyon.....	19
3.2.1. AIMSUN	21
3.2.1.1. İncelenen parametreler	22
3.2.1.2. GEH analizi	24
3.3. Yapay Zekâ	25
3.3.1. Python.....	25
3.3.1.1. NumPy	26
3.3.1.2. Pandas	26
3.3.1.3. Matplotlib.....	27
3.3.1.4. Scikit-Learn.....	27
3.3.1.5. Seaborn.....	27
3.3.2. Sentetik veri.....	28
3.3.3. Makine öğrenmesi	29

3.3.3.1. Lineer Regresyon	30
3.3.3.2. Destek Vektör Makinesi.....	31
3.3.3.3. Karar Ağacı	33
3.3.3.4. Rastgele Orman.....	34
3.3.3.5. AdaBoost (Adaptive Boosting).....	35
3.3.3.6. Gradyan Artırma	36
3.3.3.7. XGBoost (xtreme gradient boosting) (ekstrem gradyan artırma)	36
3.3.3.8. Aşırı uyum (overfitting)	37
3.3.3.9. Veri sayısallaştırma yöntemleri.....	38
3.3.3.10. Pipeline Yapısı	39
3.3.3.11. GridSearchCV	40
3.3.3.12. Hata metrikleri.....	41
4. MATERYAL ve YÖNTEM.....	44
4.1. Materyal	44
4.1.1. Çalışmanın yapıldığı koridor hakkında bilgiler	44
4.1.2. Çalışmanın kullanılan koridora ait trafik verileri.....	48
4.1.2.1. Taşıt sayımları.....	48
4.1.2.2. Sinyalizasyon verileri.....	51
4.1.3. Simülasyon Modellemesi	54
4.1.4. Mevcut Durum Simülasyon Modellemesi	54
4.1.4. GEH analizi	58
4.1.5. Oluşturulan senaryolar	59
4.1.5.1. Kavşaklarda geometrik değişiklikler.....	62
4.1.5.2. Kavşaklarda mevcut sinyalizasyon sürelerinde yapılan değişiklikler...64	
4.1.5.3. Senaryoların istatistiksel dağılımı	66
4.2. Yöntem.....	69
4.2.1. Kütüphanelerin ve Fonksiyonların İçeri Aktarılması.....	71
4.2.2. Gecikme Süresi	71
4.2.2.1. Açıklayıcı istatistiklerin hesaplanması ve analizi	71
4.2.2.2. Verilerin hazırlanması.....	74
4.2.2.3. Makine öğrenmesi algoritmalarının modellenmesi.....	75
4.2.3. IEM CO ₂	76
4.2.3.1. Açıklayıcı istatistiklerin hesaplanması ve analizi	76

4.2.3.2. Verilerin hazırlanması.....	78
4.2.3.3. Makine öğrenmesi algoritmalarının modellenmesi.....	79
4.2.4. Toplam Seyahat Süresi.....	79
4.2.4.1. Açıklayıcı İstatistiklerin Hesaplanması ve Analizi.....	79
4.2.4.2. Verilerin hazırlanması.....	81
4.2.4.3. Makine öğrenmesi algoritmalarının modellenmesi.....	82
5. BULGULAR	83
5.1. Gecikme Süresi Parametresi İçin Bulgular	83
5.1.1. GridSearchCV Öncesi ve Sonrası Performans Değerlendirme.....	83
5.1.2. Test ve eğitim skorları arasındaki yüzdesel farklar.....	85
5.1.3. Final model sonrası değerlendirme	90
5.1.4. Modellerin karşılaştırılması.....	103
5.1.5. Modellere ait tahmin değerlerinin karşılaştırılması	105
5.1.6. Başarılı Modellerin Etki Katsayılarının Değerlendirilmesi	106
5.2. IEM CO ₂ parametresi için bulgular	112
5.2.1. GridSearchCV öncesi ve sonrası performans değerlendirme	112
5.2.2. Test ve Eğitim Skorları Arasındaki Yüzdesel Farklar	114
5.2.3. Final Model Sonrası Değerlendirme	120
5.2.4. Modellerin Karşılaştırılması.....	134
5.2.5. Modellere Ait Tahmin Değerlerinin Karşılaştırılması	136
5.2.6. Başarılı Modellerin Etki Katsayılarının Değerlendirilmesi	136
5.3. Toplam Seyahat Süresi Parametresi İçin Bulgular	138
5.3.1. GridSearchCV öncesi ve sonrası performans değerlendirme	138
5.3.2. Test ve eğitim skorları arası yüzdesel farklar.....	139
5.3.3. Final Model Performans Değerlendirme.....	146
5.3.4. Modellerin Karşılaştırması.....	160
5.3.5. Tahmin Değerleri Karşılaştırılması.....	162
5.3.6. Başarılı modellerin etki katsayılarının değerlendirilmesi	163
6. SONUÇ ve ÖNERİLER.....	169
KAYNAKLAR	172

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 3.1. Üç kollu eş düzeyli kavşak tipleri	7
Şekil 3.2. Dört kollu kavşaklar.....	8
Şekil 3.3. Çok kollu kavşak	9
Şekil 3.4. Sırası ile yarım dönel, mini dönel ve modern dönel kavşak	10
Şekil 3.5. Trompet kavşak.....	11
Şekil 3.6. Tipik tek noktalı kavşak	11
Şekil 3.7. Baklava (diamond) kavşak örneği	12
Şekil 3.8 Yarım yonca kavşak örneği	13
Şekil 3.9 Yonca kavşak örneği.....	13
Şekil 3.10 Dairesel dönel kavşak	14
Şekil 3.11. Eliptik dönel kavşak	15
Şekil 3.12. Tam direksiyonel kavşak	16
Şekil 3.13. Yarım direksiyonel kavşaklar	16
Şekil 3.14. Servis seviyelerinin sınırları	19
Şekil 3.15 Sentetik veri üretim sürecine ait diyagram	29
Şekil 3.16. Doğrudan postalamaya cevap	34
Şekil 3.17. Bir AdaBoost akış şeması örneği.....	36
Şekil 4.1. Çalışmanın yapıldığı koridor	44
Şekil 4.2. Dört yol Kavşağı.....	45
Şekil 4.3. Merplaza Kavşağı	45
Şekil 4.4. Eski Terminal Kavşağı	46
Şekil 4.5. Terzibaba Kavşağı	46
Şekil 4.6. Eski Polis Okulu Kavşağı	47
Şekil 4.7. TEDAŞ Kavşağı	47
Şekil 4.8. Yeni Terminal Kavşağı.....	48
Şekil 4.9. Dört yol Kavşağı'na ait sinyalizasyon verileri.....	51
Şekil 4.10. Merplaza Kavşağı mevcut durum faz süreleri	52
Şekil 4.11. Eski Terminal Kavşağı faz süreleri.....	52
Şekil 4.12. Terzibaba Kavşağı faz süreleri.....	53
Şekil 4.13. Eski Polis Okulu Kavşağı faz süreleri	53

Şekil 4.14. TEDAŞ Kavşağı faz süreleri.....	53
Şekil 4.15. Yeni Terminal Kavşağı faz süreleri	54
Şekil 4.16. Mevcut durum modellemesi	55
Şekil 4.17. Mevcut durum modellemesi	55
Şekil 4.18. Mevcut durum modellemesi	56
Şekil 4.19. Yeni Terminal Kavşağı mevcut durum modellemesi	56
Şekil 4.20. Çalışma yapılacak koridorun mevcut durum modeli	57
Şekil 4.21. Kavşakların modern dönel kavşak olarak modellenmesi	63
Şekil 4.22. Terzibaba Kavşağı altgeçit modellemesi	63
Şekil 4.23. Dört yol Kavşağı sinyalizasyon düzenlemesi sonucu faz süreleri	65
Şekil 4.24. Merplaza Kavşağı sinyalizasyon düzenlemesi sonucu faz süreleri	66
Şekil 4.25. Eski Polis Okulu Kavşağı sinyalizasyon düzenlemesi sonucu faz süreleri ..	66
Şekil 4.26. Kavşaklardaki geometrik ve sinyalizasyon değişikliklerine göre senaryoların dağılımı	67
Şekil 4.27. Veri setindeki taşıt hacimlerinin dağılımlarını belirten dairesel grafikler	68
Şekil 4.28. Veri setindeki senaryoların doğu-batı ve batı-doğu yönündeki şerit sayısına göre dağılımı belirten dairesel grafikler	68
Şekil 4.29. Makine öğrenmesi model akış diyagramı	70
Şekil 4.30. Gecikme süresi parametresine ait sütun dağılım grafiği.....	73
Şekil 4.31. Gecikme süresi parametresine ait kutu grafiği.....	74
Şekil 4.32. Gecikme süresi parametresine ait unique (eşsiz) değerlerin sayısı.....	75
Şekil 4.33. IEM CO ₂ parametresine ait sütun dağılım grafiği	77
Şekil 4.34. IEM CO ₂ parametresine ait kutu grafiği	78
Şekil 4.35. IEM CO ₂ parametresine ait unique (eşsiz) değerlerin sayısı	79
Şekil 4.36. Toplam seyahat süresi parametresine ait sütun dağılım grafiği.....	80
Şekil 4.37. Toplam seyahat süresi parametresine ait kutu grafiği.....	81
Şekil 4.38. Toplam seyahat süresi parametresine ait “unique” (eşsiz) verilerin sayısı...	82
Şekil 5.1. Lineer Regresyon için gecikme süresi eğitim ve test verileri arasında yüzdelik fark	85
Şekil 5.2. Destek Vektör Regresörü için gecikme süresi eğitim ve test verileri arasında yüzdelik fark.....	86

Şekil 5.3. Karar Ağacı için gecikme süresi eğitim ve test verileri arasında yüzdelik fark	86
Şekil 5.4. Rastgele Orman için gecikme süresi eğitim ve test verileri arasında yüzdelik fark	87
Şekil 5.5. AdaBoost için gecikme süresi eğitim ve test verileri arasında yüzdelik fark ..	88
Şekil 5.6. Gradyan Artırma için gecikme süresi eğitim ve test verileri arasında yüzdelik fark	88
Şekil 5.7. XGBoost için gecikme süresi eğitim ve test verileri arasında yüzdelik fark ..	89
Şekil 5.8. Gecikme süresi için Lineer Regresyon modeli tahmin hatası grafiği	90
Şekil 5.9. Gecikme süresi için Lineer Regresyon modelinde artık değerler dağılımı.....	91
Şekil 5.10. Gecikme süresi için Destek Vektör Regresörü modeli tahmin hatası grafiği	92
Şekil 5.11. Gecikme süresi için Destek Vektör Regresörü modelinde artık değerler dağılımı	93
Şekil 5.12. Gecikme süresi için Karar Ağacı modeli tahmin hatası grafiği	94
Şekil 5.13. Gecikme süresi için Karar Ağacı modelinde artık değerler dağılımı	95
Şekil 5.14. Gecikme süresi için Rastgele Orman modeli tahmin hatası grafiği.....	96
Şekil 5.15. Gecikme süresi için Rastgele Orman modelinde artık değerler dağılımı	97
Şekil 5.16. Gecikme süresi için AdaBoost modeli tahmin hatası grafiği.....	98
Şekil 5.17. Gecikme süresi için AdaBoost modelinde artık değerler dağılımı	99
Şekil 5.18. Gecikme süresi için Gradyan Artırma modeli tahmin hatası grafiği	100
Şekil 5.19. Gecikme süresi için Gradyan Artırma modelinde artık değerler dağılımı..	101
Şekil 5.20. Gecikme süresi için XGBoost modeli tahmin hatası grafiği	102
Şekil 5.21 Gecikme süresi için XGBoost modelinde artık değerler dağılımı	103
Şekil 5.22. Gecikme süresi için makine öğrenme modelleri RMSE hata oranlarının karşılaştırılması	104
Şekil 5.23. Gecikme süresi için makine öğrenme modelleri R^2 skorlarının karşılaştırılması	105
Şekil 5.24. Gecikme süresi Lineer Regresyon modelinde değişkenlerin katsayı değerleri	107
Şekil 5.25 Gecikme süresi Lineer Regresyon modelinde değişkenlerin katsayı değerleri sütun grafiği	108

Şekil 5.26. Gecikme süresi Destek Vektör Regresörü modelinde değişkenlerin katsayı değerleri.....	109
Şekil 5.27. Gecikme süresi Destek Vektör Regresörü modelinde değişkenlerin katsayı değerleri sütun grafiği	110
Şekil 5.28. Lineer Regresyon için IEM CO ₂ eğitim ve test verileri arasında yüzdelik fark	115
Şekil 5.29. Destek Vektör Regresörü için IEM CO ₂ eğitim ve test verileri arasında yüzdelik fark.....	116
Şekil 5.30. Karar Ağacı için IEM CO ₂ eğitim ve test verileri arasında yüzdelik fark ..	116
Şekil 5.31. Rastgele Orman için IEM CO ₂ eğitim ve test verileri arasında yüzdelik fark	117
Şekil 5.32. AdaBoost için IEM CO ₂ eğitim ve test verileri arasında yüzdelik fark.....	118
Şekil 5.33. Gradyan Artırma için IEM CO ₂ eğitim ve test verileri arasında yüzdelik fark	118
Şekil 5.34. XGBoost için IEM CO ₂ eğitim ve test verileri arasında yüzdelik fark.....	119
Şekil 5.35. IEM CO ₂ için Lineer Regresyon modeli tahmin hatası grafiği.....	120
Şekil 5.36. IEM CO ₂ için Lineer Regresyon modelinde artık değerler dağılımı	121
Şekil 5.37. IEM CO ₂ için Destek Vektör Regresörü modeli tahmin hatası grafiği	122
Şekil 5.38. IEM CO ₂ için Destek Vektör Regresörü modelinde artık değerler dağılımı	123
Şekil 5.39. IEM CO ₂ için Karar Ağacı modeli tahmin hatası grafiği	124
Şekil 5.40. IEM CO ₂ için Karar Ağacı modelinde artık değerler dağılımı	125
Şekil 5.41. IEM CO ₂ için Rastgele Orman modeli tahmin hatası grafiği	126
Şekil 5.42 IEM CO ₂ için Rastgele Orman modelinde artık değerler dağılımı.....	127
Şekil 5.43. IEM CO ₂ için AdaBoost modeli tahmin hatası grafiği	128
Şekil 5.44. IEM CO ₂ için AdaBoost modelinde artık değerler dağılımı.....	129
Şekil 5.45. IEM CO ₂ için Gradyan Artırma modeli tahmin hatası grafiği.....	130
Şekil 5.46. IEM CO ₂ için Gradyan Artırma modelinde artık değerler dağılımı	131
Şekil 5.47. IEM CO ₂ için XGBoost modeli tahmin hatası grafiği	132
Şekil 5.48. IEM CO ₂ için XGBoost modelinde artık değerler dağılımı.....	133
Şekil 5.49. IEM CO ₂ için makine öğrenme modelleri RMSE hata oranlarının karşılaştırılması	135

Şekil 5.50. IEM CO ₂ için makine öğrenme modelleri R ² skorlarının karşılaştırılması	135
Şekil 5.51. XGBoost modeli için bağımsız değişkenlerin nitelik önemi değerleri.....	136
Şekil 5.52. XGBoost modeli için bağımsız değişkenlerin nitelik önemi değerleri sütun grafiği.....	137
Şekil 5.53. Lineer Regresyon için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark.....	140
Şekil 5.54. Destek Vektör Regresörü için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark	140
Şekil 5.55. Karar Ağacı için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark	141
Şekil 5.56. Rastgele Orman için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark.....	142
Şekil 5.57. AdaBoost için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark	143
Şekil 5.58. Gradyan Artırma için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark.....	144
Şekil 5.59. XGBoost için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark	145
Şekil 5.60. Toplam seyahat süresi için Lineer Regresyon modeli tahmin hatası grafiği	146
Şekil 5.61. Toplam seyahat süresi için Lineer Regresyon modelinde artık değerler dağılımı	147
Şekil 5.62. Toplam seyahat süresi için Destek Vektör Regresörü modeli tahmin hatası grafiği.....	148
Şekil 5.63. Toplam seyahat süresi için Destek Vektör Regresörü modelinde artık değerler dağılımı	149
Şekil 5.64. Toplam seyahat süresi için Karar Ağacı modeli tahmin hatası grafiği.....	150
Şekil 5.65 Toplam seyahat süresi için Karar Ağacı modelinde artık değerler dağılımı	151
Şekil 5.66. Toplam seyahat süresi için Rastgele Orman modeli tahmin hatası grafiği.	152
Şekil 5.67. Toplam seyahat süresi için Rastgele Orman modelinde artık değerler dağılımı	153
Şekil 5.68. Toplam seyahat süresi için AdaBoost modeli tahmin hatası grafiği.....	154

Şekil 5.69. Toplam seyahat süresi için AdaBoost modelinde artık değerler dağılımı ..	155
Şekil 5.70. Toplam seyahat süresi için Gradyan Artırma modeli tahmin hatası grafiği	156
Şekil 5.71 Toplam seyahat süresi için Gradyan Artırma modelinde artık değerler dağılımı	157
Şekil 5.72. Toplam seyahat süresi için XGBoost modeli tahmin hatası grafiği	158
Şekil 5.73. Toplam seyahat süresi için XGBoost modelinde artık değerler dağılımı ...	159
Şekil 5.74. Toplam seyahat süresi için makine öğrenme modelleri RMSE hata oranlarının karşılaştırılması	161
Şekil 5.75. Toplam seyahat süresi için makine öğrenme modelleri R^2 skorlarının karşılaştırılması	161
Şekil 5.76. Destek Vektör Regresörü modeli için bağımsız değişkenlere ait katsayı değerleri.....	163
Şekil 5.77. Destek Vektör Regresörü modeli için bağımsız değişkenlere ait katsayı değerleri sütun grafiği	164
Şekil 5.78. XGBoost modeli için bağımsız değişkenlerin nitelik önemi değerleri.....	166
Şekil 5.79. XGBoost modeli için bağımsız değişkenlerin nitelik önemi değerleri sütun grafiği	167

TABLÖLER LİSTESİ

	Sayfa
Tablo 3.1. Simülasyon yazılımları genel özellik karşılaştırması	21
Tablo 3.2 Kırmızı, yeşil ve mavi renklerinin One-Hot Encoder yardımıyla sayısallaştırılmış hali	38
Tablo 4.1. Dört yol Kavşağı taşıt sayımları	49
Tablo 4.2. Merplaza Kavşağı taşıt sayımları	49
Tablo 4.3. Eski Terminal Kavşağı taşıt sayımları	49
Tablo 4.4. Terzibaba Kavşağı taşıt sayımları	49
Tablo 4.5. Eski Polis Okulu Kavşağı taşıt sayımları	50
Tablo 4.6. TEDAŞ Kavşağı taşıt sayımları	50
Tablo 4.7 Yeni Terminal Kavşağı taşıt sayımları	50
Tablo 4.8. GEH analizi sonuçları	58
Tablo 4.9. Çalışma için kavşaklarda yapılan değişiklikler	60
Tablo 4.10. Çalışma için taşıt hacminde ve koridor şerit sayılarında yapılan değişiklikler	60
Tablo 4.11. Bazı senaryolara ait kavşak geometrik ve sinyalizasyon bilgileri.	61
Tablo 4.12. Gecikme süresi parametresi için tanımlayıcı istatistikler	72
Tablo 4.13 IEM CO ₂ için tanımlayıcı istatistikler.	76
Tablo 4.14. Toplam seyahat süresi için tanımlayıcı istatistikler	79
Tablo 5.1. Gecikme süresi parametresi için GridSearchCV öncesi performans değerleri	83
Tablo 5.2. Gecikme süresi parametresi için GridSearchCV sonrası performans değerleri	84
Tablo 5.3. Gecikme süresi parametresi için test ve eğitim skorları arası yüzdelik fark .	85
Tablo 5.4. Gecikme süresi için makine öğrenmesi modellerinin hata metrikleri değerleri	104
Tablo 5.5 Gecikme süresi için performansa göre sıralanmış RMSE ve R ² skorları	104
Tablo 5.6. Karşılaştırma yapılacak senaryolara ait bilgiler	106
Tablo 5.7 Gecikme süresi için model tahmin değerlerinin karşılaştırılması	106
Tablo 5.8. IEM CO ₂ parametresi için GridSearchCV öncesi performans değerleri	112
Tablo 5.9. IEM CO ₂ parametresi için GridSearchCV sonrası performans değerleri	113

Tablo 5.10 IEM CO ₂ parametresi için test ve eğitim skorları arası yüzdelik fark	114
Tablo 5.11 IEM CO ₂ için makine öğrenmesi modellerinin hata metrikleri değerleri...	134
Tablo 5.12 IEM CO ₂ için performansa göre sıralanmış RMSE ve R ² skorları.....	134
Tablo 5.13 IEM CO ₂ için model tahmin değerlerinin karşılaştırılması	136
Tablo 5.14. Toplam seyahat süresi parametresi için GridSearchCV öncesi performans değerleri.....	138
Tablo 5.15. Toplam seyahat süresi parametresi için GridSearchCV sonrası performans değerleri.....	139
Tablo 5.16. Toplam seyahat süresi parametresi için test ve eğitim skorları arası yüzdelik farklar	139
Tablo 5.17. Toplam seyahat süresi için makine öğrenmesi modellerinin hata metrikleri değerleri.....	160
Tablo 5.18. Toplam seyahat süresi için performansa göre sıralanmış RMSE ve R ² skorları	160
Tablo 5.19. Karşılaştırma yapılacak senaryolara ait bilgiler.....	162
Tablo 5.20. Toplam seyahat süresi için model tahmin değerlerinin karşılaştırılması...	162

SİMGELER ve KISALTMALAR

Simgeler

b	Regresyon sabiti
ε	Regresyon Hata Terimi
α	Güvenirlilik Katsayısı
x	Bağımsız değişken
y	Bağımlı değişken
λ	Lagrange Çarpanı
Φ	Non-lineer haritalama fonksiyonu
w	Ayarlanabilir katsayı
M	Mean (ortalama)
n	Senaryo sayısı

Kısaltmalar

AIMSUN	Advanced Interactive Microscopic Simulator for Urban and Non-urban Networks
AdaBoost	Adaptive Boosting
DT	Decision Tree
IEA	International Energy Agency
IEM	Instantaneous Emission Model
KGM	Karayolları Genel Müdürlüğü
MAE	Mean Absolute Error
MSE	Mean Squared Error
R^2	Regresyon Değeri
RF	Random Forest
RMSE	Root Mean Squared Error
SVM	Support Vector Machine
SVR	Support Vector Regressor
TÜİK	Türkiye İstatistik Kurumu
WHO	World Health Organisation
XGBoost	Extreme Gradient Boosting

1. GİRİŞ

Ulaşım tarih boyunca insan hayatını etkileyen unsurlardan biri olmuştur. Sanayi Devrimi sonucu motorlu taşıtların icadıyla büyük önem kazanmış ve günümüz dünyasının nüfus artışıyla her geçen gün önemini artırmıştır.

Dünyada 1.446 milyar araç bulunmaktadır (Hedges & Company, 2022). Araçların çokluğu, yılda yaklaşık 1.3 milyon insanın ölümüne sebep olan trafik kazalarını da beraberinde getirmektedir (WHO, 2022). Dünya geneli 36,8 Gt olan CO₂ salınım değerinin önemli bir kısmını motorlu taşıtlar oluşturmaktadır (IEA, 2022). Bu verilere bağlı olarak trafikteki gecikme ve seyahat sürelerinin yanında artışlar ile yakıt tüketiminde ve sera gazı salınımında artışlar meydana gelmektedir. Bu sorunlar ekonomik ve çevresel anlamda zararlara yol açmaktadır. Ayrıca güvensizlik ve rahatsızlık vererek seyahat deneyimini olumsuz etkilemektedir. Bu bilgilerden yola çıkarak yol ağlarının planlanması ve düzenlenmesi ile seyahat süresindeki artışlar ve gecikmeler minimize edilmelidir Böylece yol ağlarının gereken oranda güvenli, konforlu ve sürdürülebilir hale getirilmesi mümkün olabilecektir.

Çalışmanın amacı; gecikme süresi (delay time), IEM CO₂ (instantaneous emission model of CO₂) (CO₂ anlık emisyon modeli) ve toplam seyahat süresi (total travel time) parametrelerinin; seçilen trafik koridoruna ait kavşaklarda ve şeritlerde yapılan geometrik düzenlemeler, kavşaklarda gerçekleştirilen sinyalizasyon düzenlemeleri ve taşıt hacimleri artışlarının değişken olarak kabul edildiği tahmin modelleri oluşturmaktır.

Erzincan'ın en yoğun trafiğine sahip güzergâhlarından biri olan Dört Yol Kavşağı'ndan Yeni Terminal Kavşağı'na kadar uzanan koridorda yedi sinyalizasyon kavşağına ait taşıt ve sinyalizasyon verileri elde edilmiştir. Bu veriler AIMSUN trafik simülasyon programına aktarılmış, mevcut durum ve alternatif senaryolar olmak üzere toplam 27.648 senaryo oluşturulmuştur. Oluşturulan senaryolarla elde edilen veri seti ile çeşitli makine öğrenmesi algoritmaları eğitilmiş ve bir dizi tahmin modeli oluşturulmuştur. Bu tahmin modelleri ile söz konusu koridorun optimizasyonu için gerekli bulgulara erişilmiştir.

2. KAYNAK ÖZETLERİ

Jones (2020), trafik tahmini ve olay tespit sistemlerini iyileştirmek için bir Akıllı Trafik Yönetimi çerçevesi önermektedir. Bu çerçevede trafik ile ilgili "tweet"leri belirlemek ve makine öğrenimi algoritmaları kullanarak eksik dönel kavşakları ve kentsel yolların akış dağılımını simüle etmeye çalışmıştır. Yapılan analizlerde "tweet"lerin %40'ında stres tespit edilmiştir.

Purbarani vd. (2019), Sürekli Koşullu Rasgele Alanlar ve Aşırı Öğrenme Makinesi algoritmaları ile karayolu trafik verilerini karşılaştırmış, çalışma sonucunda Ortalama Mesafeye Sürekli Koşullu Rastgele Alanlar Yönteminin, standart Sürekli Koşullu Rasgele Alanlar yönteminden iki katı kadar performans gösterdiği bulgusuna erişmişlerdir.

Zhu vd. (2022), biri China-V ve diğeri China-VI standartlarını karşılayan iki hafif hizmet benzinli araçtan gerçek dünyadaki karbon emisyonlarını araştırmak için Taşınabilir Emisyon Ölçüm Sistemleri kullanmışlardır. Kademeli Regresyon ve LightGBM makine öğrenmesi metotları kullanılan çalışmada China-VI standardına geçişte karbon emisyonu değerinin azaldığını belirlemişlerdir.

Zahid vd. (2020), kısa vadeli gelecekteki trafik durumu tahminini doğru bir şekilde modellemek için farklı makine öğrenimi sınıflandırıcıları kullanmışlardır. Rastgele tarama yoluyla hiperparametre optimizasyonunun başarılı sonuçlar verdiğini bulmuşlardır. Çalışma sonucunda Karar Ormanlarının diğer yöntemlere göre üstün performans gösterdiğini saptamışlardır.

Manibardo vd. (2021), Akıllı Ulaşım Sistemleri kapsamında bir trafik tahmini vaka çalışması tasarlamışlardır. Çalışma, derin öğrenmenin her durumda en iyi modelleme tekniği olmayabileceğini ortaya koymaktadır. Ayrıca yol trafiği tahmininde yeni zorluklar ve araştırma fırsatları sıralanmıştır. Çalışma, bu alanda gelecekteki araştırmalar için yol göstermektedir.

Hou vd. (2019), kentsel alanlarda düşük karbonlu seyahat bloklarını belirlemek için bir metodoloji önermişlerdir. Araştırmada, işlevsel karma, yoğun yol ağları ve kamu hizmeti tesislerinin, sakinlerin düşük karbonlu seyahatleri için faydalı olduğunu göstermektedir.

Araştırma alanı içindeki E bloğunun, en düşük seyahat karbon emisyonlarına sahip olduğunu bulmuşlardır.

Alatawneh vd. (2023), bir lojistik firmasında 12 sürücünden 284014 veri örneği toplamışlar ve CO₂ emisyonundan ne zaman sorumlu olduklarını belirlemek için bir metodoloji önermişlerdir. Önerilen metodoloji, yöneticiler için araç filosunun enerji verimliliğini artırmak ve CO₂ emisyonunu azaltmak için potansiyel alanı basit ve hızlı bir şekilde belirlemelerini sağlamaktadır.

Garmei ve Kashi (2019) AIMSUN yazılımını kullanarak yayaların ve geçen araçların hacmini değerlendirmişler ve yayaların ortalama hızının 4,25 km/s olduğunu ve 1,35 s olan varsayılan tepki süresinin 0,9 s olarak değiştiğini bulmuşlardır. İki modun karşılaştırma sonuçlarında, kalibre edilmiş durumdaki ortalama hızın %48 ve gecikme süresinin %5 azaldığını gözlemlemişlerdir.

Tu vd. (2019), yol envanterlerini iyileştirmek için mikroskobik emisyon modellemesini bölgesel trafik atama modellemesiyle bütünleştiren CLEVER yöntemini önermişlerdir. CLEVER metodunun, Büyük Toronto Bölgesi'ndeki sabah zirve sera gazı emisyon değerlerini tahmin etmede geleneksel bir ortalama hız yaklaşımından daha iyi olduğu sonucuna varmışlardır.

Elhenawy (2015), hava durumu ve görünürlük değişkenlerine göre trafik sıkışıklığını, seyahat sürelerini ve sürücü güvenliğini tahmin eden modeller önermiştir. Çalışma, bu modellerin; tıkanıklığı belirleme, seyahat sürelerini tahmin etme ve sürücü güvenliğini artırma konularında iyileştirmeler sağladığını bulmuştur.

Mecheva vd. (2022), Plovdiv'deki trafik için en iyi araç takip modeli ve yönlendirme algoritmasını belirlemesi adına bir metodoloji önermektedir. Çalışma SUMO ve Python aracılığıyla sıralı ve paralel uygulaması kullanmışlardır. 7000'den fazla simülasyon gerçekleştirmişler ve varsayılan yönlendirme adımlı Daraltma Hiyerarşileri yönlendirme algoritması ve varsayılan yapılandırma parametreleriyle Krauss araba takip modeli kullanmanın uygun olduğunu bulmuşlardır.

Masiero vd. (2011), bir alanda herhangi iki nokta arasındaki bir araç tarafından harcanacak zamanı tahmin etmek için makine öğrenimi tekniği uygulamışlardır. Tahmini

süreyi doğru bir şekilde elde etmek için bir dizi semantik değişkeni hesaba katmış ve araçların gerçekleştirdiği hareketler hakkındaki tarihsel verileri kullanmışlardır. Ön çalışma niteliğinde olmasına rağmen, bu çalışmadan tatmin edici sonuçlar elde edilmiştir.

Varga vd. (2023), trafik akışlarını farklı ayrıntı düzeylerinde işlemek için makroskobik, mezoskopik ve mikroskobik simülasyonları aralarında doğrusal bir algoritma ile birleştiren yeni bir mezoskopik trafik modeli geliştirmişlerdir. Bu algoritmayı iki senaryoda SUMO ve Carla simülatörleri kullanarak test etmişler ve yalnızca mikroskobik simülasyon kullanmaya kıyasla modelleme doğruluğunu korurken simülasyon performansının %200-500 oranında arttığını tespit etmişlerdir.

Jiber vd. (2020), Fas'ta bir yol bölümünün gelecekteki saatlik trafiğini tahmin etmek için hibrit bir model önermişlerdir. Model, 5 yıllık bir süre boyunca gerçek tarihsel verilere uygulanmış ve diğer iyi bilinen algoritmalarla karşılaştırılmıştır. Önerilen model ile hata ölçümlerine ve Akaike'nin Bilgi Kriteri tekniğine dayanan tahmin doğruluğu açısından diğerlerinden daha iyi performans gösterdiği belirlenmiştir.

Bahuleyan ve Vanajakshi (2017), kentsel arter ağlarında seyahat süresi tahmini için bir metodoloji önermiştir. Metodoloji, gerçek zamanlı bilgilere dayanarak K-En Yakın Komşu ve Rastgele Orman algoritmalarıyla seyahat süresini tahmin etmekte ve herhangi bir başlangıç-varış çifti adına tüm olası rotalarda seyahat edilebilmesi için gereken süreyi bulmaktadır. Model, Hindistan, Chennai'deki toplu taşıma otobüslerinden temin edilen GPS verileri kullanılarak bir çalışma ağı üzerinde doğrulanmıştır.

Aboah ve Arthur (2021), seyahat süresini tahmin etme amacıyla seyahat süresi ve hız değerlerini kullanarak Bütünleşik Otoregresif Hareketli Ortalama, Yinelemeli Sinir Ağı, Otoregresif Model, Uzun Kısa Süreli Bellek ve Geçitli Tekrarlayan Birim algoritmaları ile tahmin modelleri oluşturmuşlardır. Her bir modelin oluşturulması adına kullanılan öğrenme oranı 0,0001-0,01, en iyi öğrenme oranı 0,001 olarak bulmuşlar ve Bütünleşik Otoregresif Hareketli Ortalama Modeli'nin seyahat süresi tahmini için en iyi model mimarisi olduğunu saptamışlardır.

Vanajakshi ve Rilett (2007), San Antonio, Teksas'taki TransGuide Traffic Management merkezinden alınan verileri kullanarak, Destek Vektör Makinelerinin performansını diğer

tahmin yöntemleriyle karşılaştırmıştır. Destek Vektör Makineleri; geçmiş yöntemler, gerçek zamanlı yöntemler ve yapay sinir ağları gibi diğer yöntemlerden daha iyi performans göstermiştir. Destek Vektör Makinelerinin, kısa vadeli tahmin için güvenilir bir seçenek olabileceğine dair bulgulara erişmişlerdir.

Chikaraishi vd. (2020), Temmuz 2018'de Hiroşima'da meydana gelen büyük bir ulaşım ağı kesintisinden elde edilen verileri kullanarak tekrarlanmayan olaylar sonucu trafik ağlarındaki sıkışmayı tahmin etmeye çalışmışlardır. Çalışmada Rastgele Orman, Destek Vektör Makinesi, XGBoost, Sığ İleri Beslemeli Sinir Ağı ve Derin İleri Beslemeli Sinir Ağı modelleri kullanılmıştır. Araştırma sonucunda Rastgele Orman ve XGBoost'un en iyi tahmin doğruluğunu ürettiğini, ancak derin sinir ağı modellerinin daha yorumlanabilir olduğunu ve mevcut trafik akışı teorisi açısından mantıksal olarak açıklanabileceğini saptamışlardır.

Hou ve Edara (2020), Missouri, Saint Louis şehrinde bir ulaşım ağı üzerinde seyahat süresini tahmin etmek için Uzun-Kısa Süreli Bellek ve Evrişimli Sinir Ağı derin öğrenme modelini kullanmışlardır. Çalışmada, derin öğrenmenin hem sıkışık hem de sıkışık olmayan trafik koşulları için doğru tahmin sağlayabildiği ve beklenmeyen olayların veya özel olayların trafik dinamiklerini başarılı bir şekilde yakalayabildiği bulgusuna ulaşmışlardır. Bu bulgular, derin öğrenmenin bir ağ ölçeğinde seyahat sürelerinin gerçek zamanlı tahminine yönelik umut verici bir yaklaşım sunduğunu göstermektedir.

3. KURAMSAL TEMELLER

3.1. Kavşaklar ve Sinyalizasyon

3.1.1. Kavşaklar

Kavşaklar, iki ya da daha fazla kara yolunun kesiştiği, birleştiği ve ayrıldığı alanlarda trafiğin akıcı bir şekilde ilerlemesi ve farklı yollar arasında geçiş yapılması için tasarlanan mühendislik yapılarıdır. Genel olarak üç ya da dört kol şeklinde tasarlanırlar. Dörtten fazla kol barındıran kavşaklar trafik akışını karmaşık hale getirdiği için istenmeyen bir kavşak tipi olarak bilinir (KGM, 2016).

Bağlı oldukları yolların kritik noktaları olarak kapasite ve performansında önemli etkiler bırakan kavşaklar en az iki kara yolunun kesiştiği unsurlar olduğundan kavşaklarda trafik yoğunluğuna bağlı olarak araç etkileşimi artmaktadır (Kırımlı, 2019). Bu etkileşimin sonucunda ya trafik akışı yavaşlamakta ya da trafik kazaları meydana gelmektedir. Trafik akışının yavaşlaması araçların gecikme süresini ve seyahat süresini artırarak ciddi zaman kaybı ve yakıt tüketimini artırarak ulaşım maliyetinin yükselmesine neden olmaktadır. Ayrıca yakıt tüketimindeki bu artış birçok zararlı gazın salınımına yâni çevre kirliliğine sebep olmaktadır.

Kent içi ulaşımındaki gecikmelerin yaklaşık %70'i, trafik kazalarının %40-60'ı kavşaklarda meydana geldiği gerçeği göz önüne alınırsa kavşak tasarımının güvenlik ve konfor açısından dikkatlice yapılması gerekliliği ortaya çıkmaktadır (Erol, 2018; aktaran Demircioğlu 2021). Her ne kadar güvenlik ve konfor açısından tatmin edici kavşak tasarımı istense de ekonomik parametreleri düşünmeden kavşak tasarımı yapmanın gerçekçi olmadığını da göz önünde bulundurmak gerekmektedir.

3.1.1.1. Eş düzeyli kavşaklar

İki ya da daha fazla kara yolunun aynı düzlemde kesişmesi sonucu oluşan eş düzeyli kavşaklar, genellikle trafik hacmi az olan kent içi ve kent dışı kara yollarında kullanılan kavşak tipidir. Her ne kadar ekonomik olsa da kazaların %50'den fazlasının eş düzeyli kavşaklarda meydana gelmesi, farklı düzeyli kavşakları daha güvenli kılmaktadır

(Gedizlioğlu, 2002; aktaran Korkmaz, 2019). Ayrıca trafik akışının belli bir geçiş sıralaması olan eş düzey kavşaklarda kesintiye uğraması kaçınılmazdır (Korkmaz, 2019).

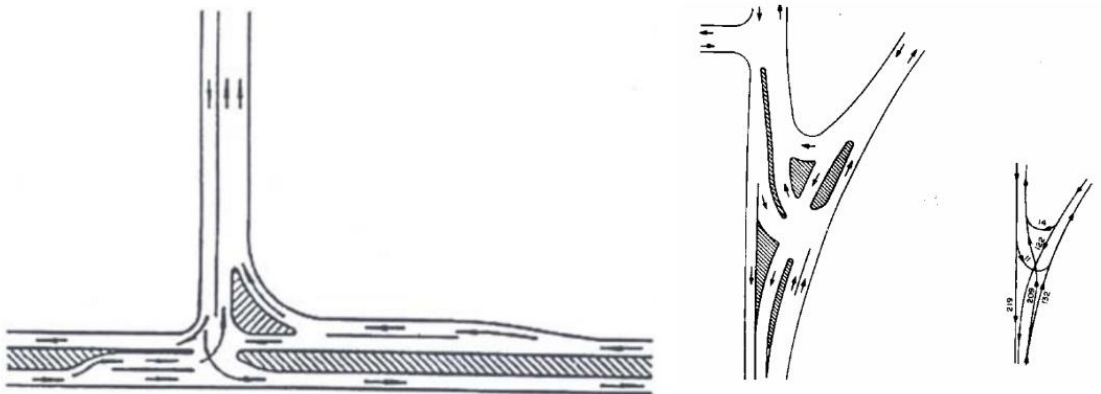
Eş düzeyli kavşaklarda, kavşakla bağlantılı olan kollardaki trafik hacmi, arazi şartları, yaya ve taşıt güvenliği önem gerektiren hususlardır (KGM, 2016).

Eş düzeyli kavşaklar;

- Kol sayısına göre üç kollu, dört kollu ve çok kollu olarak;
- Trafik kontrol sistemine göre sinyalize ve sinyalize olmayan olarak;
- Kapasiteyi artırmak amacıyla, işaretleme veya adalar ile yapılan yönlendirmeye göre kanalize edilmiş ve kanalize edilmemiş kavşaklar olarak sınıflandırılır.

3.1.1.1.1. T ve Y şeklinde üç kollu kavşaklar

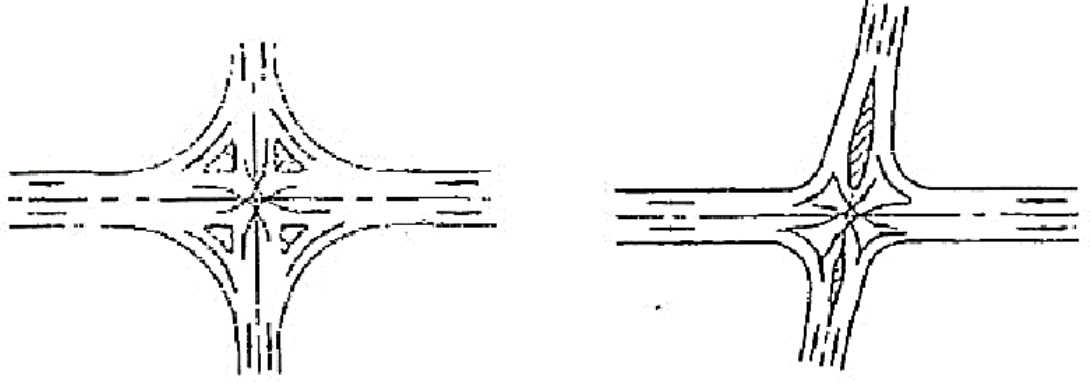
T tipi kavşaklarda tali yolun ana yol ile kesişme açısı 60° ile 120° arasında olmalıdır. Kollardaki trafiğin hacmine göre 2x1 ve 2x2 şeritli yollarda, gerekirse kavşak girişleri kanalize edilerek kullanılabilir (KGM, 2016). Y tipi kavşaklar ise, ana yol – tali yol kesişme açısının 60° ile 120° arasında imar, kamulaştırma gibi nedenlerle düzenlenemediği durumlarda uygulanan ve yol kesişmelerinin dar açılı olduğu kavşaklardır (KGM, 2016). Trafik hacmi artışı söz konusu olduğu durumlarda ışıklı ikaz sisteminin uygulanması gerekir. Kollardaki U-dönüşü için orta ada genişliği, dönüş yapacak taşıtların minimum dönüş yarıçapına göre düzenlenmelidir.



Şekil 3.1. Üç kollu eş düzeyli kavşak tipleri (KGM, 2016)

3.1.1.1.2. Dört kollu kavşaklar

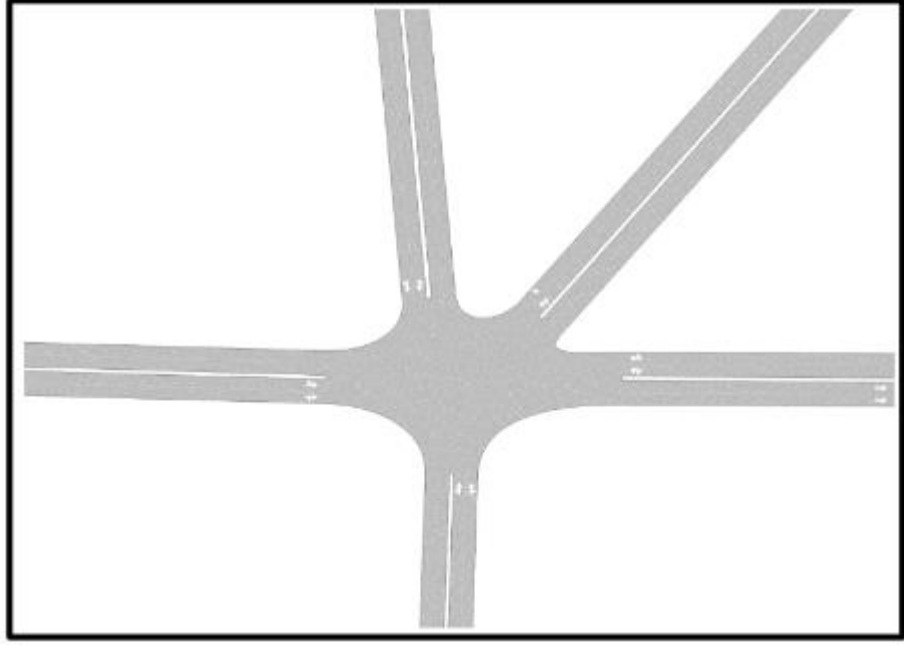
Dört kollu kavşaklar, iki yolun birbirlerini kestiği bölgelerde oluşan kavşaklardır. Üç kollu eş düzey kavşaklarda olduğu gibi, dört kollu eş düzey kavşaklarda da ana yol – tali yol kesişme açısı 60° ile 120° arasında olmalıdır (KGM, 2016).



Şekil 3.2. Dört kollu kavşaklar (KGM, 2016)

3.1.1.1.3. Çok kollu kavşaklar

Çok kollu kavşaklar, en az 5 koldan oluşan ve teknik açıdan istenmeyen bir kavşak tipidir. Ancak bu tür bir kavşak tasarımı sadece zorunlu olduğu durumlarda tercih edilmelidir. Bu durumda bile trafiğin az olduğu ve DUR kontrolünün uygulandığı yerler tercih edilmelidir.



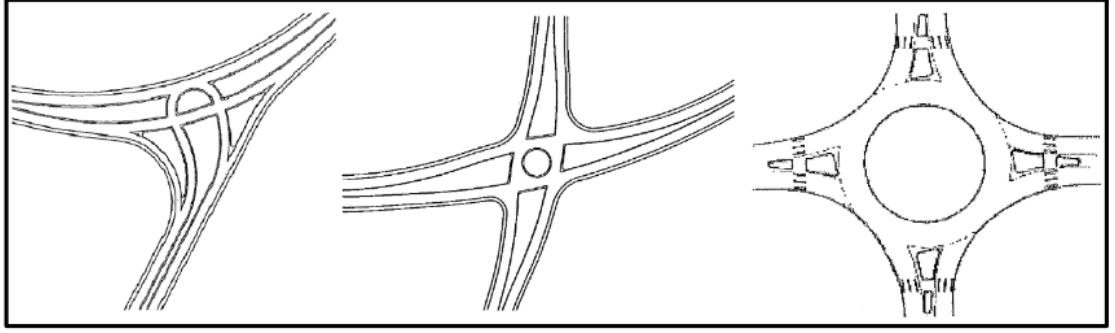
Şekil 3.3. Çok kollu kavşak (Demiriz, 2019)

3.1.1.1.4. Dönel kavşaklar

Eş düzeyli kavşak tiplerinden biri de dönel kavşaktır. Kavşak ortasında bulunan ada çapına ve şekline göre mini dönel, modern dönel ve yarım dönel kavşaklar olmak üzere 3 gruba ayrılırlar.

Yarım dönel kavşaklar; 3 kollu olurken, mini ve modern dönel kavşakların dörtten fazla kollu tasarlanmaları arzu edilmez. Mini dönel kavşaklarda ada çapı en az 8 metre, modern dönel kavşaklarda ise en az 20-25 metre olacak şekilde tasarlanmalıdır.

Modern dönel kavşaklarda araçların kavşağa girmeden önce yavaşlaması için yol bölümü bir kulp ile kavşağa bağlanır. Kavşağa giren kollara “yol ver” trafik tanzim işareti konur ve kavşak içi hareketler öncelikli hareket olarak tanımlanır (KGM 2016).



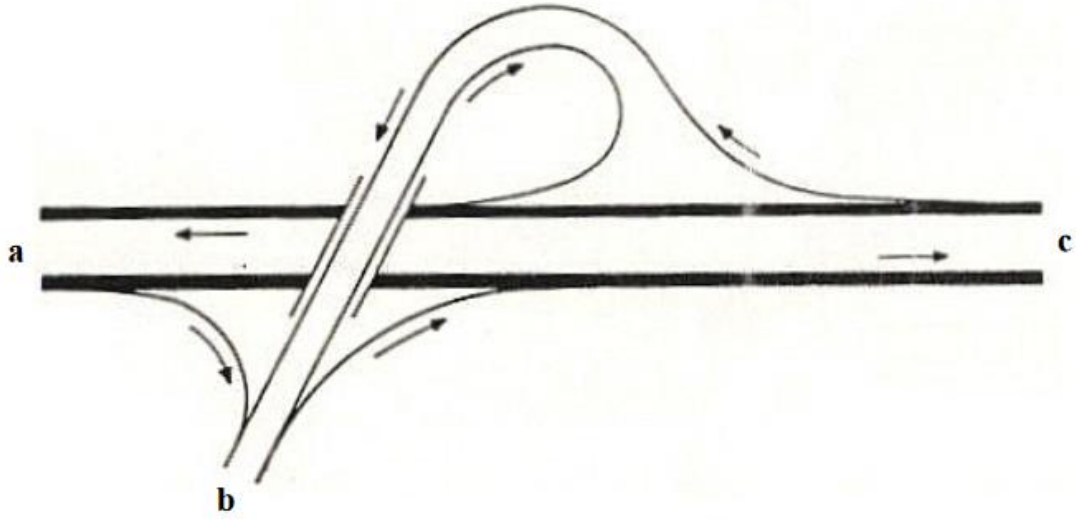
Şekil 3.4. Sırası ile yarım dönel, mini dönel ve modern dönel kavşak (KGM 2016).

3.1.1.2. Farklı düzeyli kavşaklar

Trafik sıkışıklığının fazla olduğu, kazaların yoğun bir şekilde gerçekleştiği otoyol veya ekspres inşaatlarında ve topoğrafyanın kullanımında sınırlamalar olduğu yerlerde trafik akımlarını farklı kotlara dağıtmak için farklı düzeyli kavşaklar kullanılır (Güldamlası, 2007). Katlı kavşaklar trompet, tek nokta şehir içi, baklava (diamond), yarım yonca, yonca, dönel ve direksiyonel – yarı direksiyonel tipli kavşaklar olmak üzere yedi ayrı grupta sınıflandırılır (Tunalıoğlu, 2005).

3.1.1.2.1. Trompet kavşaklar

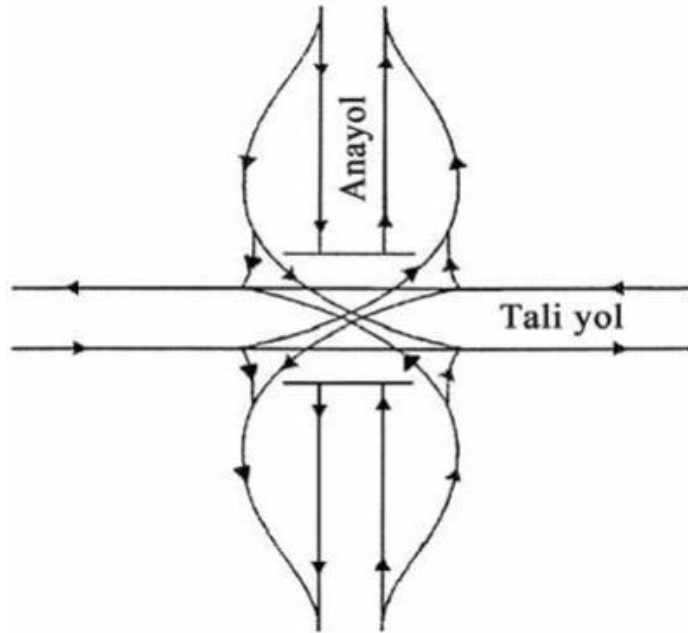
Bu tip köprülülü kavşaklar, birbiriyle kesişen ana yol ve ekspres yolu bağlamak için kullanılır. Trompet kavşak, serbest akım hızının yüksek olduğu ana yolu kesen hızın daha düşük olduğu tali yol için düşünülmelidir. Sola dönüş trafiğinin çok olduğu yöne direksiyonel rampa, az olduğu yöne ise lup (döngü) yerleştirilmelidir (KGM, 2016).



Şekil 3.5. Trompet kavşak (KGM, 2016)

3.1.1.2.2. Tek noktalı kavşaklar

Kamulaştırma kısıtlılığı olan yerlerde ana yola kesintisiz akım sağlandığı ve rampa ve tali yolların dört kollu eş düzeyli kavşaklar şeklinde düzenlendiği kavşak şeklidir. Çakışmalar diğer kavşaklardan farklı olarak kavşak içinde meydana gelmektedir (Demircioğlu, 2021).

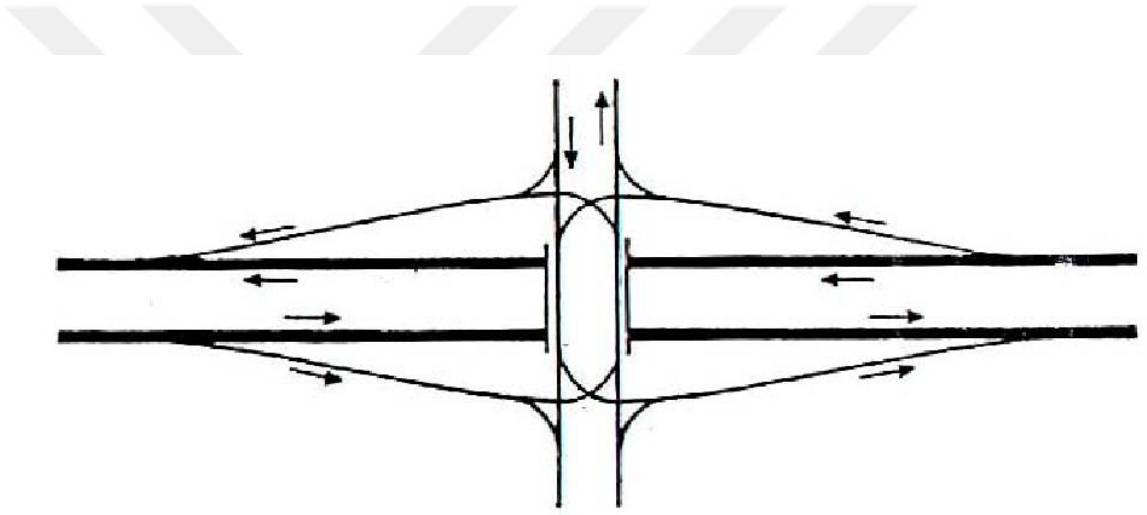


Şekil 3.6. Tipik tek noktalı kavşak (KGM, 2016)

Tek Nokta Kavşak diamond kavşağa benzemektedir. Ancak diamond kavşakta diyagonal rampalar tali yolda iki adet sinyalizasyon (veya kanalize) eş düzey kavşak gerektirirken, bu tipte tali yolda kavşağın içinde bir adet sinyalizasyonlu eş düzey kavşak yeterli olmaktadır.

3.1.1.2.3. Baklava (diamond) kavşaklar

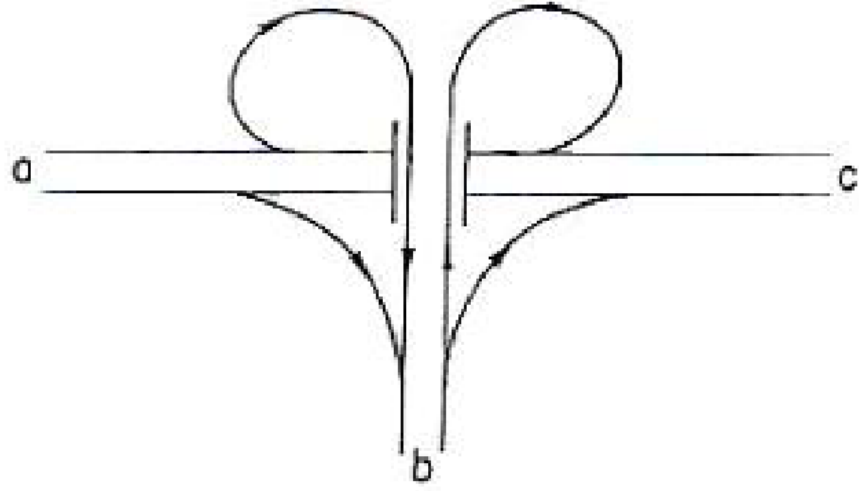
Köprülü kavşak tipi olarak yaygın bir uygulama alanına sahiptir ve her yöne dönüşü mümkün kılmaktadır. Ana yol hacmi yüksek ve tali yol hacmi düşük yollarda verimli çalışan ekonomik bir kavşak modelidir. En büyük dezavantajı ise sola dönüş kapasitesinin düşük olmasıdır (KGM, 2016).



Şekil 3.7. Baklava (diamond) kavşak örneği (KGM, 2016)

3.1.1.2.4. Yarım yonca kavşaklar

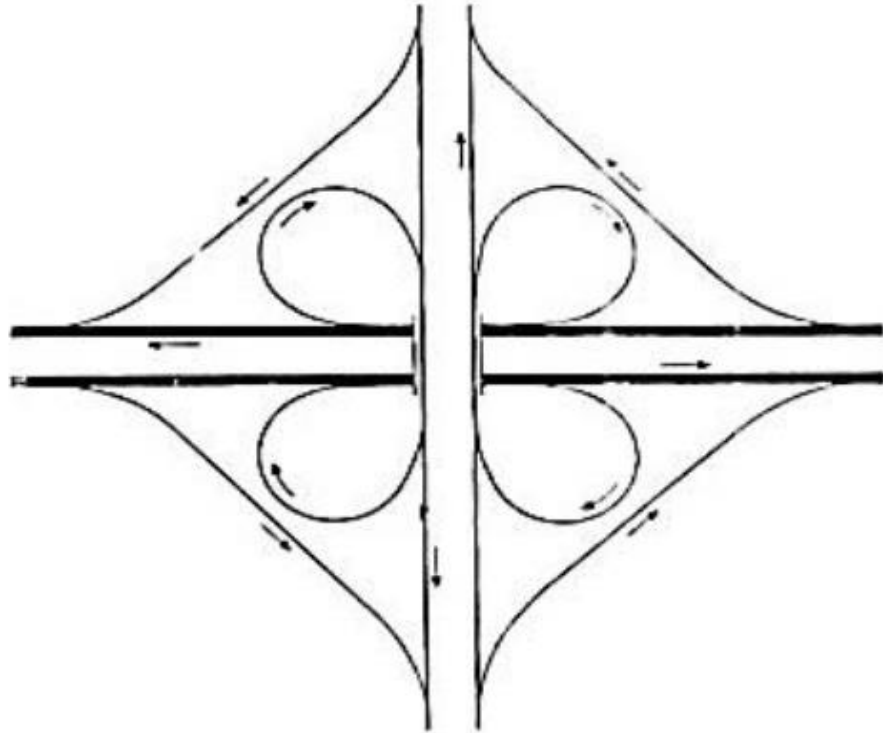
Luplar (döngü rampaları) vasıtasıyla iki yönden sola dönüşlerin sağlandığı kavşaklardır. Yarım yonca kavşaklar, tam yonca kavşakların başlangıç aşaması için zemin hazırlar. Bir sonraki aşamada tali yol diğer yöne uzatılıp eksik olan rampaların eklenmesi suretiyle tam yonca kavşak tipine dönüştürülebilir (KGM, 2016).



Şekil 3.8 Yarım yonca kavşak örneği (KGM, 2016)

3.1.1.2.5. Yonca kavşaklar

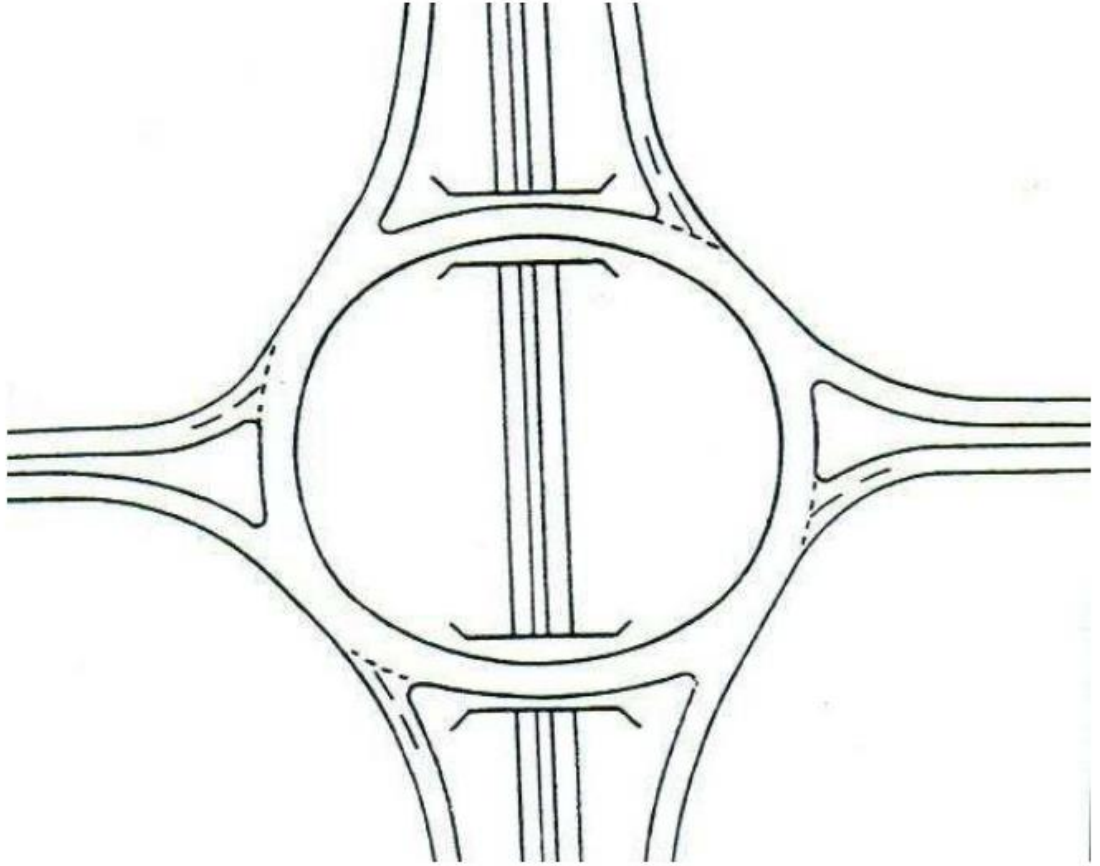
Sola dönüşlerde çakışma ihtimalinin sıfır olduğu kavşaklardır. Çok fazla alan işgal ettiğinden az tercih edilmektedir.



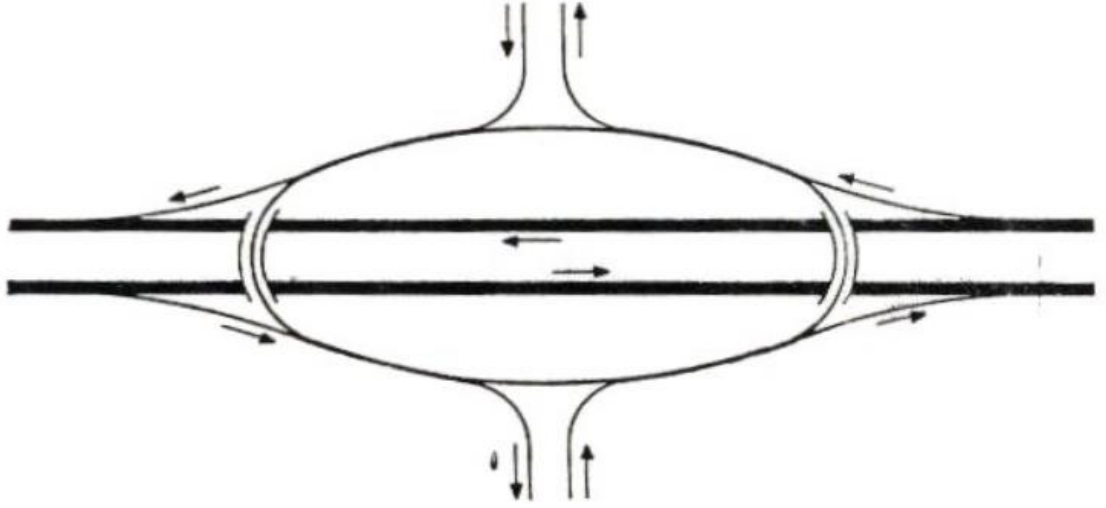
Şekil 3.9 Yonca kavşak örneği (KGM, 2016)

3.1.1.2.6. Farklı düzeyli dönel kavşaklar

Kesişen kol sayısının ve sol/sağ/U dönüş trafik hacminin fazla olduğu yerlerde alternatif bir kavşak tipi olarak tasarlanabilmektedir. Dönel kavşak, kesişen yolları dönel ada etrafına yönlendirerek bu yollardan ayrılan ana güzergâhın transit geçiş yapmasına olanak sağlamaktadır. Bu kavşak tipi, ana güzergâh dışında kalan yolların kapasiteyi azaltan örtülme hareketlerinin sorun teşkil etmediği alanlarda uygulanması önerilmektedir. Uygulanabilirliği kısıtlı olması sebebiyle alternatif olan kavşak tiplerinin uygulanamadığı ve yukarıda belirtilen özelliklerin olduğu yerlerde tercih edilmektedir (KGM, 2016).



Şekil 3.10 Dairesel dönel kavşak (KGM, 2016)



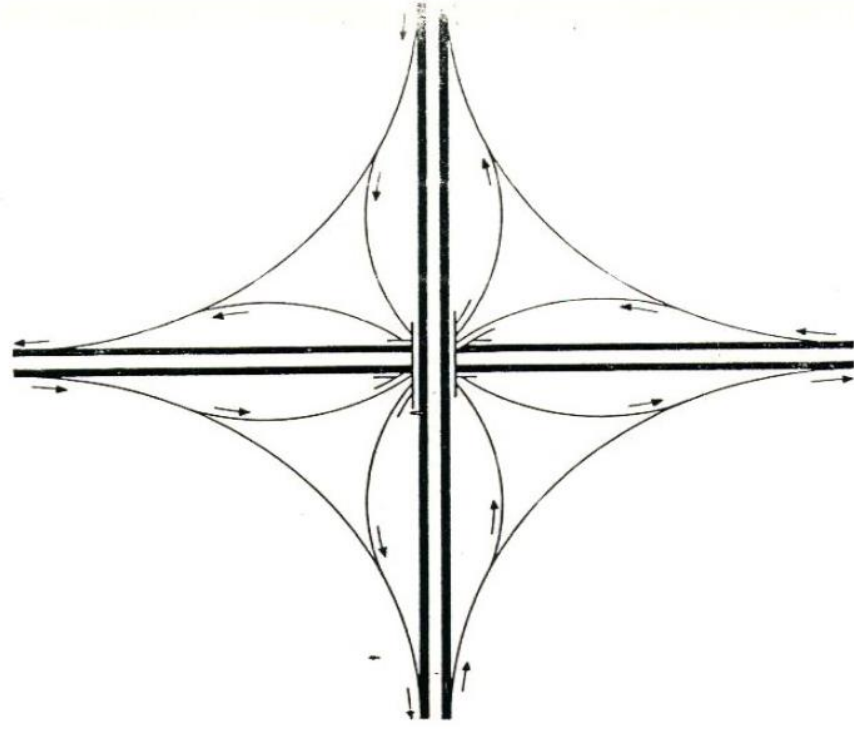
Şekil 3.11. Eliptik dönel kavşak (KGM, 2016)

Şekil 3.10’da dairesel dönel kavşak, Şekil 3.11’de eliptik dönel kavşak verilmiştir. Ana yol trafik hacmi ve tali yol trafik hacmi birbirine yakın olduğu durumlarda dairesel dönel kavşak; ana yol trafik hacminin daha fazla olduğu durumlarda ise eliptik dönel kavşak tercih edilmektedir.

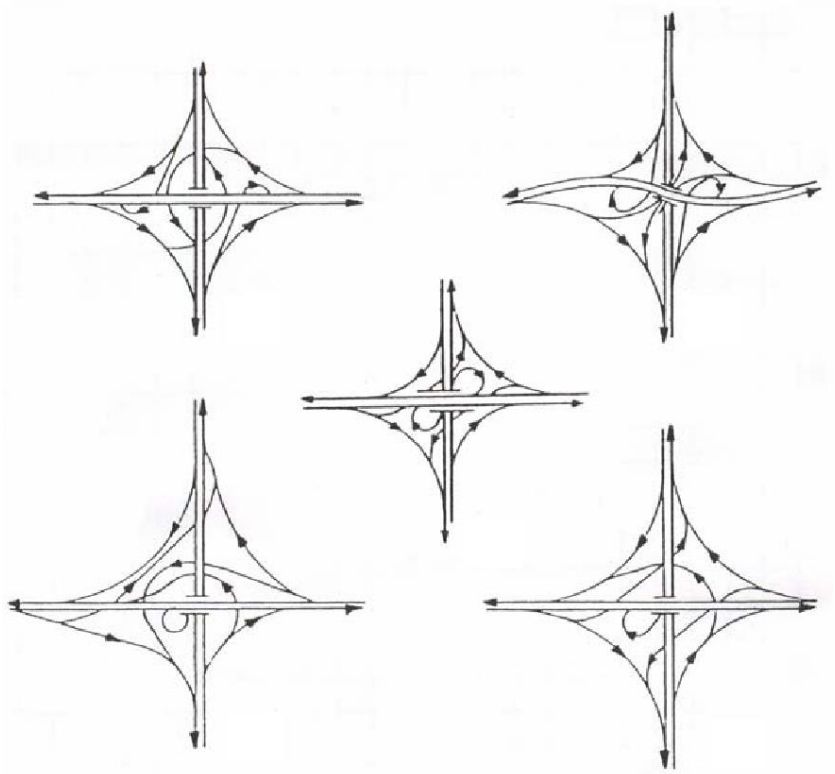
3.1.1.2.7. Direksiyonel ve yarı direksiyonel kavşaklar

Diğer kavşak türleri ile kıyaslandığında işletme hızı, kapasite ve konfor açısından büyük fark oluşturmaktadır. Bu sebeple yapım maliyeti diğer katlı kavşaklardan fazladır. Çok şeritli kombinasyonları, trafik kapasiteleri birbirine yakın olan iki ana yolun kesişiminde kullanılır.

Tam direksiyonel ve yarı direksiyonel kavşakların farkına bakılacak olursa; tam direksiyonel kavşaklarda sola dönüş hareketleri doğrudan bağlantı ile sağlanırken yarı direksiyonel kavşaklarda en az bir doğrudan bağlantı için bir dönüş rampası eklenmektedir (Tunalıoğlu, 2005).



Şekil 3.12. Tam direksiyonel kavşak (KGM, 2016)



Şekil 3.13. Yarım direksiyonel kavşaklar (KGM, 2016)

3.1.2. Sinyalizasyon

Trafik akımlarının yeşil, sarı ve kırmızı ışıklar kullanılarak taşıtların ve/veya yayaların geçiş haklarını paylaştırdığı ve kontrol altına aldığı uyarı sistemine sinyalizasyon; sinyalizasyon uygulanan kavşağa da sinyalize ya da ışıklı kavşak olarak adlandırılmaktadır. Doğru bir şekilde uygulanması, trafik kapasitesini artırmaya ve kazaların önlenmesine büyük bir katkı sağlamaktadır (Demiriz, 2019).

3.1.2.1. Sinyalizasyon ile ilgili temel kavramlar

Faz; kavşak kollarının oluşturduğu trafik akımlarından en az birinin belli süreyle geçiş hakkı alması ve süre bitiminde bu hakkın sonlanmasıdır.

Periyot; kavşaktaki tüm fazların toplamını oluşturur. Tüm fazların süresi sonlanınca başa dönerek kendini tekrar eder.

Yeşil süre; herhangi bir yönde verilen geçiş hakkının süresini ifade eder.

Sarı süre; geçiş hakkının başlatılması ya da sonlandırılması için müsaade edilen süredir. Yeşil süre ile kırmızı sürenin arasında yer alır.

Kırmızı süre; bir fazda geçiş hakkının olmadığı zamanda geçen süredir.

Boşaltma süresi; bir hareket bittikten sonra diğer hareketin başlaması için geçen süredir ve koruma kırmızısı olarak da bilinir.

3.1.2.1.1. Sinyal ayarlamaları

Sinyal ayarlamaları; ön zamanlı sinyal, yarı trafik uyarmalı sinyal ve tam trafik uyarmalı sinyal olmak üzere üç sınıfa ayrılmaktadır.

Ön zamanlı sinyal; yeşil, sarı ve kırmızı sürelerin ve periyot süresinin önceden planlanarak kavşaklarda uygulanmasına denir.

Yarı trafik uyarmalı sinyal; yeşil, sarı ve kırmızı sürelerin ve periyot süresinin tali yollardaki trafik hacmindeki değişikliğe bağlı olarak değiştirilmesidir.

Tam trafik uyarmalı sinyal; yeşil, sarı ve kırmızı sürelerin ve periyot süresinin, bir kavşakta bulunan tüm kollardan gelen trafik hacminin değişmesine bağlı olarak ayarlanması işlemidir (Tunç, 2003; aktaran Demiriz, 2019).

3.1.2.1.2. Işıklı dönel kavşaklar

Işıklı dönel kavşaklar, gecikme süresini azaltmak ve kazaları engellemek gibi amaçlardan dolayı uygulanır. Gereken kıstasları uygulamaktan kaçıldığında gecikme süresini ve trafik kazalarını artırmaktadır. Bundan dolayı sinyalizasyon periyot süresi ve her bir kavşak kolu için faz süresi hesapları iyi yapılmalı; gerekli durumlarda dönüş adaları ya da cepleri tasarlanmalıdır. Işıklı dönel kavşakların tasarımında genellikle ada içi trafik ışığı konulmasından kaçınılmaktadır (Demiriz, 2019).

3.1.3. Servis seviyesi

Yolculuk sırasında sürücülerin ve yolcuların memnuniyetini ölçen servis seviyesi; ölçüm yaparken trafik hızı, trafik hacmi, konfor, trafikteki kesiklikler, manevra serbestlikleri ve seyahat süresi parametrelerini göz önünde bulundurmaktadır (Kök, 2019). A en yüksek ve F en düşük seviye olarak kararlaştırılmış altı adet seviye bulunmaktadır.

A servis seviyesi, taşıtların birbirinin varlığından ve davranışlarından etkilenmediği serbest akım şartlarını ve ufak çaplı aksamalar olsa bile seyahat hızının değişmediği durumu ifade eder.

B servis seviyesi, taşıtların varlığı A servis seviyesine göre biraz öne çıksa da serbest akımın bulunduğu ve manevra kabiliyeti sınırlamasının önemsiz derecede ufak olduğu durumdur.

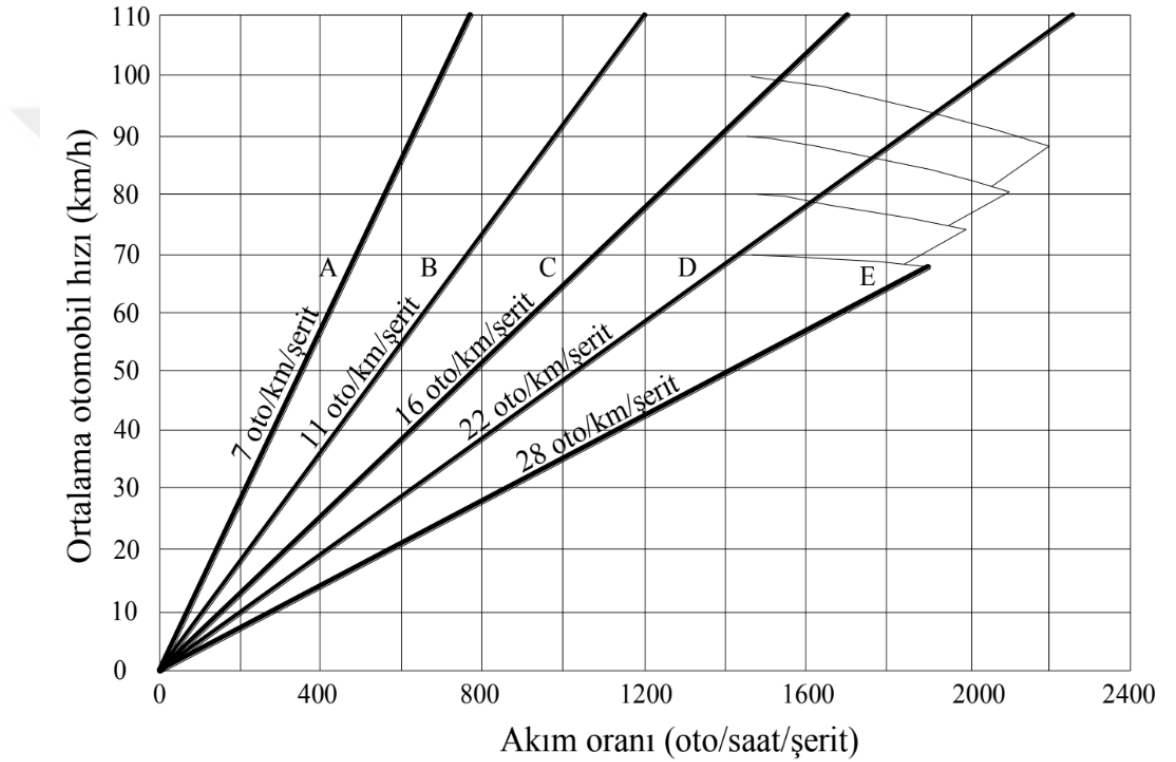
C servis seviyesi, trafik yoğunluğu, işletme koşulları ve manevra serbestliğinin diğer taşıtlardan dolayı önemli ölçüde etkilendiği durumdur. Küçük aksamalar trafik kuyruklarının oluşmasına sebebiyet verebilir.

D servis seviyesi, manevra serbestliğinin büyük ölçüde sınırlandığı trafik sıkışıklıkları olan durumu ifade eder. Sadece çok küçük aksaklıklara toleransı olan bir hizmet düzeyidir.

E servis seviyesi, trafik hacminin kapasiteye ulaşmasına çok yakın olduğu seviyeyi ve kararsız akım durumunu ifade etmektedir. Manevra serbestliği tamamen kısıtlanmış durumdadır. Aksaklıkların giderilmesi zordur ve kuyruklar kolayca oluşabilir.

F servis seviyesi, trafik hacmi ile trafik kapasitesi eşit ya da trafik hacminin daha yüksek olduğu düzeydir. Zorlamalı ve durarak ilerleme söz konusudur.

Şekil 3.14'te servis seviyelerinin sınırları verilmiştir.



Şekil 3.14. Servis seviyelerinin sınırları (Kök, 2019)

3.2. Simülasyon

Bilimsel bir çalışma ortaya koymak için gereken temel unsurlardan biri, çalışma için gereken verilerin toplanmasıdır. Bu verileri toplamak için farklı yöntemler kullanılabilir:

- Mevcut kaynaklardan yararlanma (eski kayıtlar, arşivler, raporlar, yıllıklar...),
- Gözlem yapma,
- Görüşme yapma,

- Ölçek oluşturma – deney yapma,
- Simülasyon modelleme.

Bu çalışmada, veri seti elde edilirken veri toplama yöntemi olarak simülasyon kullanılmıştır. Ulaşım düzenlemelerinin etkisini planlamak ve analiz etmek için bilgisayar tabanlı simülasyonlar yaygın olarak kullanılmaktadır (Nguyen vd., 2016).

Trafik simülasyonu, gerçek trafik verilerinin bilgisayar ortamında modellenmesi ve analizi sonucu elde edilen verilerle trafik ağlarında düzenleme ve iyileştirme yapmayı sağlamaktadır. Uygulanacak bir planlamanın ya da düzenlemenin ne derece verimli olacağının tespiti bu yöntemle güvenli, ucuz ve kolay uygulanabilir olacağı için kullanımı yaygındır (Hattatoğlu vd., 2016).

Trafik simülasyonu sayesinde gecikme süresi, seyahat süresi, yakıt tüketimi, CO₂ ve NO_x vb. gazların salınımları gibi parametrelerin gerçek verilere yakın sonuçları elde edilebilmektedir. Bu verilerin edinimi ile yapılacak çalışmalar sayesinde gecikme süresinde ve seyahat süresinde azalmanın oluşturulması zaman tasarrufuna katkı sağlayacaktır. Yakıt tüketiminin azalması ile yenilenemeyen enerjinin tükenmesinin önüne geçilebilmektedir. CO₂, NO_x gibi gazların salınımlarının azalmasıyla da çevre kirliliğini önlenilebilecektir.

Trafik simülasyon modelleri mikroskobik, mezoskopik ve makroskobik olarak üç gruba ayrılırlar.

- Mikroskobik simülasyon modellemede araç hareketleri doğrudan izlenir ve her birinin bağlantı performansları ayrı ayrı ifade edilir.
- Mezoskopik modellemede mikroskobik modellemelerde olduğu gibi araç hareketleri doğrudan izlenir ancak bağlantı performansları bir bütün olarak ifade edilir.
- Makroskobik modelli seviyede ise araç hareketleri dolaylı olarak izlenir ve mezoskopik modellemelerdeki gibi bağlantı performansları bir bütün olarak ifade edilir. Mikroskobik model küçük, mezoskopik orta ve makroskobik büyük boyutlu ağlarda kullanılmaktadır (Celikoglu ve Dell’Orco 2007).

Günümüzde trafik simülasyon programlarına AIMSUN, CORSIM, PARAMICS, PTV VISSIM, SIMTRAFFIC ve SUMO örnek olarak verilebilir. Belirtilen programlara ait genel özellik karşılaştırmaları Tablo 3.1’de verilmiştir.

Tablo 3.1. Simülasyon yazılımları genel özellik karşılaştırması (Akkaya ve Engin, 2022).

	SUMO V:0.10.3	Aimsun V:6.0.4	Paramics V:6.4.1	Sim Traffic V:6	CORSIM V:6.1	PTV Vissim V:9
Açık Kaynak Kodu	✓	X	X	X	X	X
Windows tabanlı	✓	✓	✓	✓	✓	✓
Linux tabanlı	✓	✓	X	X	X	✓
MacOS tabanlı	X	✓	X	X	X	X
Yol ağı oluşturma	✓	✓	✓	✓	✓	✓
Araç modelleme yeteneği	✓	✓	X	X	X	✓
Grafik Özellikleri 2D	✓	✓	✓	✓	✓	✓
Grafik Özellikleri 3D	X	✓	✓	X	X	✓
Simülasyon Raporlama	✓	✓	X	✓	X	✓
Geniş Yol Ağı Modelleme	✓	✓	X	X	X	✓
Maks. CPU kullanımı	%17	%40	%50	%50	%50	%50
Araç tipleri	✓	✓	✓	✓	✓	✓
Sinyalizasyon Modelleme	✓	✓	✓	✓	✓	✓
Toplu Taşıma Özelliği	✓	✓	✓	X	✓	✓

Tablo 3.1’e göre SUMO, AIMSUN ve VISSIM programların daha kullanışlı programlar olduğu görülmektedir. Ancak üç boyutlu görselleştirme özellikleri, Windows OS, MacOS ve Linux tabanlı dağıtımlar ile uyumu ve tablodaki birçok programdan daha az CPU kullanılması özellikleri ile AIMSUN en kullanışlı simülasyon programı sayılabilir.

Makroskobik, mezoskopik ve mikroskobik simülasyon modelleme yeteneği, bir koridoru modelleyip analiz yapma özellikleri ile bu çalışma için gerekli olan özelliklere sahip olması nedeniyle geniş bir kullanım alanına sahip AIMSUN programı tercih edilmiştir.

3.2.1. AIMSUN

AIMSUN Advanced Interactive Microscopic Simulator for Urban and Non-urban Networks (Kentsel ve Kentsel Olmayan Ağlar İçin Gelişmiş Etkileşimli Mikroskobik Simülör) ifadesinin kısaltmasıdır. İlk olarak 1981 yılında Katalonya Üniversitesi

Simülasyon ve Operasyon Araştırma Laboratuvarı üyeleri tarafından geliştirilen bu program, 2018 yılından itibaren Siemens tarafından geliştirilmeye devam etmektedir. AIMSUN ilk olarak mikroskobik simülasyon modelleme programı şeklinde geliştirilmiştir. Ancak bu program günümüzde makroskobik, mezoskopik ve mikroskobik modelleri de içermektedir (Demiriz, 2019).

3.2.1.1. İncelenen parametreler

3.2.1.1.1. Gecikme süresi (Delay Time)

Simülasyon programı tarafından üretilen araçların tamamına ait gecikme süresinin ortalamasının kilometre başına düşen değeridir. Birimi saniye/kilometredir. Gecikme süresini hesaplamak için Denklem 3.1 kullanılmaktadır (AIMSUN, 2023).

$$DT_{sys} = \frac{\sum_{i=1}^{N_{sys}} DT_i}{N_{sys}} \quad (3.1)$$

DT_{sys} : Gecikme süresi (s/km)

N_{sys} : I istatistik aralığı (sn), sürecinde ağdan çıkan araç sayısı

DT_i : Ağdaki i'ninci aracın kilometre başına gecikme süresi (sn/km),

Görüldüğü gibi DT_{sys} değerini bulmak için her araca ait DT_i değerinin bulunması gerekmektedir. Bunun için Denklem 3.2 ve Denklem 3.3 kullanılmaktadır.

$$DT_i = \frac{TDT_i}{D_i} \quad (3.2)$$

DT_i : Ağdaki i'ninci aracın kilometre başına gecikme süresi (s/km)

TDT_i : Ağdaki i'ninci araca kadarki araçların kümülatif gecikme süresi (s)

D_i : i'ninci aracın ağdaki seyahat uzunluğu (km)

$$F_{sys} = \frac{N_{sys}}{I} \quad (3.3)$$

F_{sys} : Ortalama trafik akışı değeri (s)

I: İstatistik aralığı

N_{sys} : *I* istatistik aralığı (*s*), sürecinde ağdan çıkan araç sayısı

Bu parametre, bir ulaşım ağında eğitim ve çalışma gibi faktörlerden dolayı özellikle pik saatlerde ortaya çıkabilen beklenmedik tıkanıklık ve gecikmelerin bu ulaşım ağını kullanan vatandaşlar için sorun olmaktan çıkarılması için bir veri kaynağı olarak kullanılmaktadır. Yani, simülasyon modelindeki yapılan değişiklikler sonucu bu parametreye ait değişiklikleri izleyerek doğru bir trafik modellemesine erişim imkânı doğmaktadır.

3.2.1.1.2. CO₂ anlık emisyon modeli (IEM CO₂) (instantaneous emission model of CO₂)

Panis vd. (2006) modeline göre trafik ağından çıkmış tüm araçların yaydığı toplam CO₂ salınımın gram cinsinden hesaplanmış halidir. Herhangi bir kirletici gaz salınım değerini hesaplamak için Denklem 3.4 kullanılmaktadır (AIMSUN, 2023).

$$PollEm_{sys,j} = \sum_{i=1}^{N_{sys}} (TPE_{i,j}) \quad (3.4)$$

$TPE_{i,j}$: *j*'ninci kirleticinin *i*'ninci araç tarafından yaptığı toplam emisyon değeri (g)

$PollEm_{sys,j}$: *j*'ninci kirleticiye ait emisyon değeri (g)

Trafik ağlarındaki konfora ve sürdürülebilirliğe ait yorum yapabilmeye olanak sağlayan bu parametre trafikteki araçların da bu konforu ve sürdürülebilirliği etkileyen yakıt tüketimi ve emisyon değerleri ile ilgili fikirler yürütmenin önünü açmaktadır. Böylelikle mevcut durumda Trafik Mühendisliği kullanılarak bir iyileştirilme yapılırken, yakıt tüketimi ve emisyon konusunda da devlet politikalarıyla Otomotiv Mühendisliğinde gerçekleştirilmesi gereken düzenlemeler konusunda yol haritası çıkarılabilir.

3.2.1.1.3. Toplam seyahat süresi (Total Travel Time)

Aralığın sonunda veya simülasyonun sonunda ağdan geçen tüm araçların deneyimlediği toplam seyahat süresidir ve birimi saat (h) olarak ifade edilir. Toplam seyahat süresini hesaplamak için Denklem 3.5. kullanılmaktadır (AIMSUN, 2023).

$$TotalTravelTime_{sys} = \sum_{i=1}^{N_{sys}} (TEX_i - TEN_i) \quad (3.5)$$

TEX_i : i'ninci aracın ağdan çıkış zamanı (saniye),

TEN_i : i'ninci aracın ağdan çıkış zamanı (saniye),

$TotalTravelTime_{sys}$: ağdaki araçların toplam seyahat süresi

Araçların bir ulaşım ağında harcadıkları süreyi ifade ettiğinden bu parametrenin analizi; yol ağının kalitesi, konforu, güvenilirliği ve sürdürülebilirliği konusunda fikir sahibi olmaya olanak sağlamaktadır. Çünkü bir aracın ağdan çıkma süresi uzadığında araç için ulaşım ağındaki riskler uzayan süre kadar devam edecektir. Bu süreçte yakıt tüketiminin olmaması mümkün olmadığından yakıt tüketimi ve emisyon konusunda da ulaşım ağında etkileri olacaktır.

3.2.1.2. GEH analizi

1970 senesinde Geoffrey E. Harvers tarafından bulunan GEH formülü; trafik tahmini ve modellemesinde, trafik hacmi kıyaslamada ve trafik mühendisliğinde kullanılmaktadır.

GEH analizi, modellenen simülasyonun analiz sonuçlarını gözlemlenen değerlerle karşılaştırmak için yaygın olarak kullanılan bir yöntemdir (Sekhar vd. , 2013; Paz vd. 2015). GEH analizinde kullanılan formül Denklem 3.6 ile gösterilmiştir.

$$GEH = \sqrt{2 * \frac{(M-C)^2}{(M+C)}} \quad (3.6)$$

M: Model değeri

C: Gözlem değeri

Formül ile yapılan işlem sonucu bulunan GEH değerinin 5'in altında olması gerekli bir kısıttır (Bayrak vd., 2016; de Villa vd., 2014).

Bu çalışmada GEH analizi yapılarak modellenen mevcut durum senaryosunun kıstasa uyduğu tespit edilmiştir. Bu analize ait ayrıntılar Bölüm 4.1.4'te verilmiştir.

3.3. Yapay Zekâ

Yapay zekâ, insan beyninin öğrenme algoritmasının bilgisayar ortamında kodlanarak oluşturulan bir taklittir. İlk olarak 1950'lerde ortaya çıkan bu terimin kökeni, makinelerin düşünebilme yetisine sahip olup olmadığını araştıran ve bir makinenin zekâ seviyesini ölçmek için Turing Testini geliştiren Alan Turing'e dayanmaktadır. 1956'da ise John McCarthy "yapay zekâ" ismini ortaya atarak bu konuda ilk konferansı düzenlemiş ve bu konferansa katılan bilim insanları bu konuda çalışmalar yapılması hakkında karar almışlardır. 1959 yılında ise Arthur Samuel, bilgiyi direkt bilgisayarlara kodlamak ya da formüle etmek yerine bilgisayarın bilgiyi kendi kendine öğrenebilme yetisine sahip olacak şekilde kodlayarak oluşturulan ve yapay zekânın geniş bir alt dalı olan makine öğrenmesi kavramını ortaya atmıştır.

Yapay zekâ ve makine öğrenmesi modellemeleri ya da kodlamalarında genelde MATLAB, Python, R gibi programlama dilleri kullanılmaktadır. Dünyada bu alanda büyük bir farkla en çok kullanılan dil Python'dur. Bu çalışmada makine öğrenmesi modellemeleri için Python programa dili kullanılmıştır.

3.3.1. Python

Python bilgisayar için yazılmış ve ona ne yapacağını söyleyen bir dizi cümle olan kodlardan oluşan programlama dillerinden biridir (Python, 2021). Python, yüksek seviyeli, genel amaçlı bir programlama dilidir. Guido van Rossum (1991) tarafından geliştirilmeye başlanmıştır. Python, basit ve anlaşılır sözdizimiyle bilinen bir dil olarak öne çıkmaktadır.

Python, istatistik ve veri madenciliği, bilimsel ve numerik hesaplamalar, makine öğrenmesi, nesnelerin interneti, robotik uygulamaları, oyun ve uygulama geliştirme,

masaüstü grafik ara yüzleri gibi pek çok alanda yaygın bir kullanıma sahiptir. Çeşitli amaçlarla, dünyaca bilinen pek çok bilimsel ya da ticari kuruluşlarda kullanılmaktadır.

Geliştiricilerin hızlı ve etkili kodlar yazmasını sağlayan, bazı kodların tekrar tekrar yazılması zahmetinden kurtaran Python, hazır fonksiyonlar ve araçlar içeren kütüphanelere sahiptir. Ayrıca bu kütüphanelerin açık kaynaklı olma eğilimi sayesinde geliştiricilerin kodları özelleştirmesi suretiyle kendi projelerinde kullanmasına izin vermektedir. Bu çalışmada Python'a ait Numpy, Pandas, Matplotlib, Scikit-Learn ve Seaborn kütüphaneleri kullanılmıştır.

3.3.1.1. NumPy

NumPy, sayısal hesaplama için kullanılan bir Python kitaplığıdır. Büyük çok boyutlu veri diziler üzerinde çalışmak için çeşitli matematiksel işlevler sağlayan bir kütüphanedir. NumPy;

- Çok boyutlu diziler üzerinde çalışmak için,
- Dizilerde eleman bazlı hesaplamalar için,
- Diziler arasında matematiksel işlemler için,
- Dizi (2012).tabanlı veri kümelerinin okunması ve diske yazdırılması için,
- Farklı matematiksel işlemler için çeşitli fonksiyonlar sağlamaktadır (McKinney, 2012).

3.3.1.2. Pandas

Pandas, veri analizi ve veri işleme araçları sağlayan bir Python kütüphanesidir. NumPy'ın dizi işleme özelliklerini tabloların ve ilişkisel veri tabanlarının esnek veri işleme özellikleri ile birleştirmektedir. Çeşitli biçimlerdeki verileri okumak ve yazmak ve DataFrame adı verilen tablolı bir yapı kullanarak verileri işlemek ve analiz etmek için araçlar sağlamaktadır. Veri bilimi ve analitik projelerinde yaygın olarak kullanılmaktadır (McKinney, 2012).

3.3.1.3. Matplotlib

Grafikler ve diğer 2 boyutlu veri görselleştirmeleri kullanmak için Tkinter, wxPython, Qt veya GTK gibi genel amaçlı grafik arayüz (GUI) araç setlerini sağlayan bir kütüphanedir. Matplotlib, Python'da çok çeşitli statik, animasyonlu ve etkileşimli görselleştirmeler oluşturabilme özelliklerine sahiptir. Matplotlib ile çizgi grafikleri, dağılım grafikleri, çubuk grafikler, histogramlar, 3B grafikler oluşturulabilmektedir (McKinney, 2012).

3.3.1.4. Scikit-Learn

Scikit-Learn makine öğrenimi modellemeleri için araçlar sağlayan Python kitaplığıdır. Açık kaynaklı bir kitaplıktır. Scikit-learn;

- Makine öğrenmesi havuzuna ait tipik makine öğrenmesi veri setlerine erişimi sağlayan veri yönetimi özelliği,
- Verileri makine öğrenimi algoritmalarıyla kullanımına hazırlayan ölçekleme, normalleştirme ve özellik çıkarma gibi ön işleme araçları,
- Hâlihazır veriler için optimum modeli seçmek için GridSearchCV gibi fonksiyonları,
- Makine öğrenimi modellerinin performansının değerlendirilmesine olanak tanıyan bir dizi değerlendirme ölçütü,
- Denetimli ve denetimsiz öğrenme modelleri için özel araçlar sağlamaktadır (Kramer, 2016).

3.3.1.5. Seaborn

Seaborn, Python programlama dili için yazılmış bir veri görselleştirme kütüphanesidir. Matplotlib kütüphanesi üzerine inşa edilmiştir ve istatistiksel grafikler oluşturmak için yüksek seviyeli bir arayüz sağlamaktadır (Bisong, 2019).

Scatter plot, line plot, bar plot, heatmap gibi birçok grafik türü Seaborn kütüphanesi içinde bulunmaktadır. Ayrıca renklendirme, ölçeklendirme ve yazı tipi seçimi gibi değişiklikleri de sağlamaktadır.

Grafik çizdirme yeteneği dışında, Seaborn aynı zamanda doğrusal regresyon, lojistik regresyon ve ANOVA gibi istatistiksel modellerle çalışmak için de işlevler içermektedir. Bu özellikleri sayesinde, veri keşfi yapmak ve verilerden elde edilen görüşleri iletmek için kullanışlı bir kütüphanedir.

3.3.2. Sentetik veri

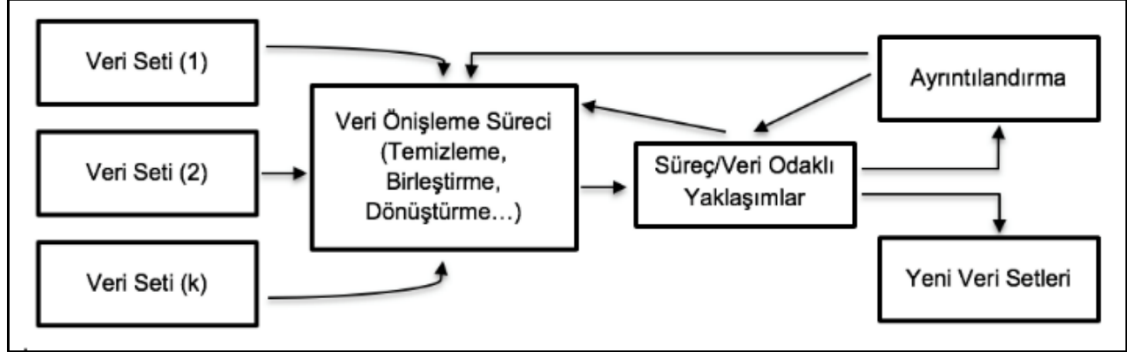
Gerçek verilerden yapay zekâ algoritmaları ile gerçek verilerin yerini almak için yeterliliğe sahip olan sentetik veriler, bize sonsuz miktarda eğitim verisi verebilme yeteneğine sahiptir (Jadeberg vd., 2014).

Sentetik veri üretimi;

- Kişisel ve hassas nitelikteki verilerin ihlal riskinin bulunduğu durumlarda,
- Veri kümelerinin büyüklüğünün yetersizliği durumunda,
- İhtiyaç duyulan verilerin var olmaması ya da erişilebilir olmaması durumunda,
- Makine öğrenmesinde kullanılacak eğitim verilerinin oluşturulmasının maliyetli olduğu durumlarda,
- Veri setlerinin homojen olduğu ancak, heterojen veri setlerine ihtiyaç duyulduğu durumlarda,
- Veri setlerinin geçerli olsa da düzenlenmeye ve birleştirilmeye ihtiyaç duyulduğu durumlarda,
- Geçerli veri setleri olsa da düzenlenmemiş farklı veri setlerinin birleştirilmesi veya bir araya getirilmesine ihtiyaç olduğu durumlarda,
- Gerçek verilerin ya da geliştirilen bir yazılımın test edilip doğrulanmasında kullanılabilmektedir (Gartner, 2020; aktaran Deveci ve Esen, 2022; Nikolenko, 2021).

Sentetik veri üretiminde kabaca iki grup bulunmaktadır; süreç odaklı ve veri odaklı yaklaşımlar. Süreç odaklı yaklaşımlar herhangi bir olayın süreçlerine hesaplamalı ve matematiksel modeller kullanarak yaklaşmaktadır. Bu yaklaşımlara örnek olarak kesikli olay benzetimi, etmen tabanlı modelleme ve Monte Carlo benzetimi verilebilir. Veri odaklı yaklaşımda ise Destek Vektör Regresörü, Bayes Ağları, Derin Öğrenme Metotları

vs. makine öğrenimi yöntemleri ile mevcut verilerin öğrenimi sonucu çıkarımlarda bulunmaktadır.



Şekil 3.15 Sentetik veri üretim sürecine ait diyagram (Deveci ve Esen, 2022).

Şekil 3.15'te görüldüğü gibi sentetik veri üretimine giren veri setleri ilk aşamada bir ön işlemden geçirilerek veri sentezine hazır hale getirilir. Sonrasında ise yukarıda anlatılan yaklaşımlarla veriler işlenerek yeni veri setleri oluşturulur.

3.3.3. Makine öğrenmesi

Makine öğrenmesi, bilgisayar donanımlarının güçlenmesi ve veri hacimlerinin artışı ile günümüzde hemen hemen her alanda kullanılmaktadır. Büyük verilerin işlenmesini kolaylaştıran makine öğrenmesi, eldeki verileri geçmiş edinim olarak kabul edip gelecekteki verilere dair tahmin modellemeleri oluşturulmasını sağlamakta ve bu modeller doğrultusunda problemlere çözüm üretilmesi için kullanılmaktadır.

Makine öğrenmesi temelinde Sınıflandırma, Kümeleme, Regresyon ve Tahmin işlemlerini gerçekleştirir. Makine öğrenmesini kullanabilmek için mevcut verileri doğru ve anlamlı bir hale getirmek gerekmektedir. Bu verileri anlamlı hale getirdikten sonra kurulan algoritma ile makine öğrenmesi tahmin modeli oluşturulmaktadır. Çeşitli analizlerle makine öğrenmesinin doğruluğu ölçülür, eğer istenilen sonuç elde edilmezse tekrar bir algoritma oluşturmak gerekmektedir. Doğruluğu en yüksek tahminlemeye ulaşıncaya kadar modelleme tekrar edilir (Altıntaş, 2020).

Makine Öğrenmesi denetimli öğrenme (supervised learning), denetimsiz öğrenme (unsupervised learning) ve pekiştirmeli öğrenme (reinforcement learning) olmak üzere üç sınıfa ayrılmaktadır.

- Denetimli öğrenme, verilerin sınıflandırılmasında ve yüksek doğruluk oranlı tahmin algoritmaları eğitmek için veri kümelerinin etiketli olarak kullanılması şeklinde tanımlanabilir.
- Denetimsiz öğrenme ise insan müdahalesine gerek duyulmadan saklı kalıpları veya veri gruplamalarını keşfedebildiklerinden, sınıflandırılmamış veya etiketsiz veri kümelerini kümelemek için kullanılır.
- Pekiştirmeli öğrenme de deneme yanılma yoluyla öğrenerek belirtilen problem için en iyi öneri ya da ilkeyi sunan davranışsal makine öğrenmesi modelidir (Özgün, 2021).

Bu çalışmada denetimli makine öğrenmesi algoritmaları olan Lineer Regresyon (Linear Regression), Destek Vektör Makinesi (Support Vector Machine), Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), AdaBoost, Gradyan Artırma (Gradient Boosting), XGBoost algoritmaları kullanılmıştır.

3.3.3.1. Lineer Regresyon

Lineer Regresyon basitliği, yorumlanabilirliği ve sağlamlığı nedeniyle popüler olarak kullanılan bir regresyon tekniğidir (Aytekin, 2021). Bir bağımlı değişken ile en az bir bağımsız değişkenin arasındaki ilişkiyi inceleyen bir istatistiksel analiz yöntemidir. Bu regresyon türü her bağımsız değişkenin bağımlı değişkenle doğrusal bir ilişki olduğu varsayılmaktadır (Skyles, 1993).

Lineer Regresyon, bağımlı değişken (y) ve n tane bağımsız değişken ($x_1, x_2, \dots, x_n$) arasındaki ilişki hata terimi ile (ϵ) birlikte Denklem 3.7'deki gibi ifade edilebilir

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n + \epsilon \quad (3.7)$$

y : bağımlı değişken

b_0 : kesme noktası

$b_1, b_2, \dots, b_n$: bağımsız değişkenlerin katsayıları

$x_1, x_2, \dots, x_n$: bağımsız değişkenlerin değerleri

Bu denklem bu çalışmada Destek Vektör Regresörü algoritması için de kullanılmıştır.

Denklemdaki katsayılar ($b_0, b_1, \dots$) lineer regresyon modelinin eğitimi sırasında hesaplanır ve modele en iyi uyan doğruyu veya düzlemi belirlemek için kullanılır. Modelin amacı, bağımsız değişkenlerin değerlerine dayanarak bağımlı değişkenin tahminini yapmaktır.

Lineer Regresyon modellerinin birkaç temel varsayımı vardır, bunlar 5 noktada özetlenebilir (Aytekin, 2021).

- Bağımsız bir değişkenin değerindeki değişiklik, bağımlı değişkenin değerindeki değişikliklerle doğrusal bir şekilde değişmelidir.
- Hata teriminin asimptotik ortalama değeri 0 olmalı ve normal dağılım göstermelidir.
- Hata teriminin varyansı tüm durumlar için sabit ve aynı olmalıdır.
- Bağımsız değişken otokorelasyon olmamalıdır.
- Bağımsız değişkenler birbirinden doğrusal olarak bağımsız olmalıdır. Bire bir korelasyonları ihmal edilebilir olmalı ve bir bağımsız değişken, değişkenlerin doğrusal bir kombinasyonu olmamalıdır.

3.3.3.2. Destek Vektör Makinesi

Destek Vektör Makinesi temeli istatistiksel öğrenme teorisine dayanan hem sınıflandırma hem de regresyon problemlerinde kullanılan bir denetimli makine öğrenmesi algoritmasıdır (Çoban ve Demir, 2021). Destek Vektör Makinesi; sınırlı numuneler, doğrusal olmayan ve yüksek boyuttaki problemler için avantajlı makine öğrenmesi yöntemidir (Li ve Xu, 2018).

Destek Vektör Makinesinin temel fikri sınıflandırma aralığını mümkün olduğu kadar büyük tutarken iki veri kategorisinin doğru ayrılmasını sağlamak için iki hiper düzlem

bulmaktır. Bu hiper düzlem, sınıflandırma işleminde kullanılırken, yeni bir veri noktasının hangi sınıfa ait olduğunu belirlemek için kullanılır.

Destek Vektör Regresörü ise Destek Vektör Makinesinin temel ilkelerinin regresyon problemine genelleştirilerek kullanılmasıdır. Eğitim veri setine yaklaşan bir fonksiyon bulabilmek için tahmin hatalarını en az seviyede tutmaya çalışır. Yapısı basittir ve küçük numuneleri desteklemek için iyi bir yeteneğe sahiptir. Bu nedenle Destek Vektör Regresörü küçük örnekler, doğrusal olmayan, yüksek boyutlar ve yerel optimallik gibi sorunları çözebilmektedir.

Gelecek vektörü $x_i \in R^n$ ve hedef çıktısı $y_i \in R$ olan bir $\{(x_1, y_1), \dots, (x_l, y_l)\}$ veri setinde girdi ve çıktı verileri arasındaki doğrusal olmayan ilişkinin doğrusal bir fonksiyonla belirtilmesi gerekmektedir. Bu ilişkiyi gösteren fonksiyon Denklem 3.8’de verilmiştir.

$$f(x) = w^T * \Phi(x) + b \quad (3.8)$$

$f(x)$: tahmin değeri,

Φ : non-linear haritalama fonksiyonu

w ($w \in R_n$), b ($b \in R$): ayarlanabilir katsayılar

Destek vektör regresyonu standart formu ise $C > 0$ ve $\epsilon > 0$ olmak üzere Denklem 3.9’da tanımlanmaktadır.

Kısıtlar;

$$\begin{cases} w^T \Phi(x_i) + b - y_i \leq \epsilon + \xi_i \\ y_i - w^T \Phi(x_i) - b \leq \xi_i^* \\ \xi_i, \xi_i^* \geq 0; i = 1, 2, \dots, \ell \end{cases} \quad (3.9)$$

ξ_i^* : ϵ üzerindeki eğitim hataları

ξ_i : ϵ altındaki eğitim hataları

Eşitsizlik kısıtlamaları ile yukarıdaki ikinci dereceden optimizasyon problemi çözüldükten sonra, Denklem 3.8’deki parametre vektörü w ; Denklem 3.10 ile bulunur.

$$w = \sum_{i=1}^{\ell} (\lambda_i^* - \lambda_i) \Phi(x_i) \quad (3.10)$$

λ_i^* , λ_i : Lagrange çarpanları

Denklem 3.8, Denklem 3.10 ile düzenlenecek olursa;

$$f(x) = \sum_{i=1}^{\ell} (\lambda_i^* - \lambda_i) K(x_i, x_j) + b \quad (3.11)$$

Denklem 3.11'deki $K(x_i, x_j)$ fonksiyonu radyal temelli kernel fonksiyonunu ifade etmektedir.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (3.12)$$

3.3.3.3. Karar Ağacı

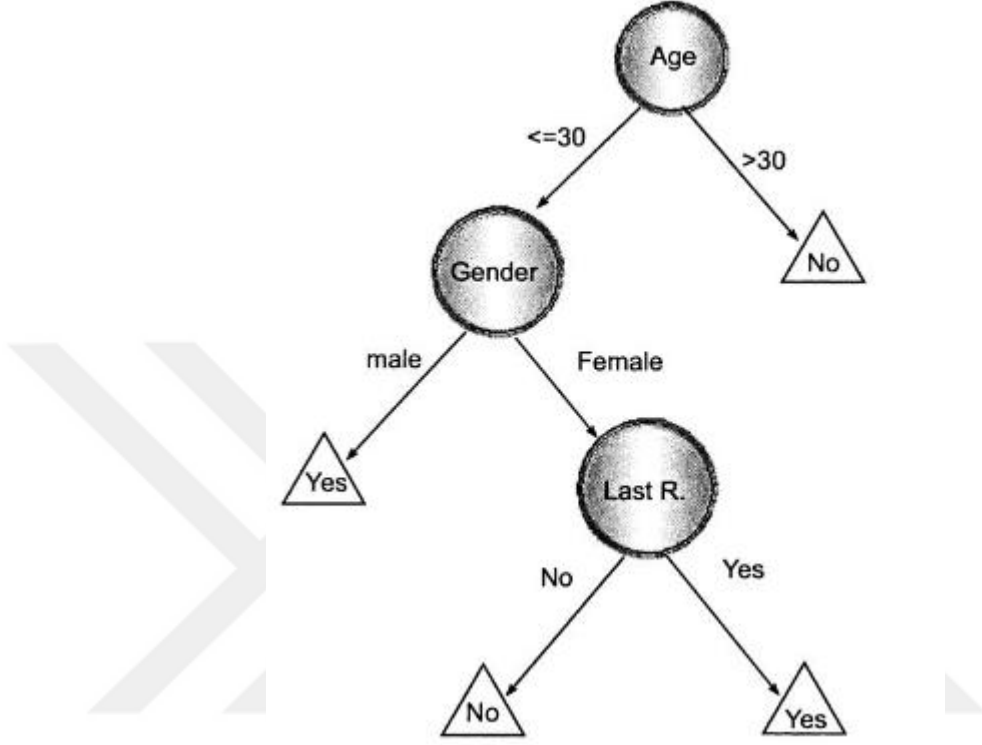
Karar Ağacı, sınıflandırma ya da tahmin problemlerinde kullanılan bir makine öğrenmesi algoritmasıdır. Karar Ağacı, özellikle doğrusal olmayan ilişkilerin olduğu karmaşık verilerde etkili bir şekilde çalışabilir.

Karar ağacının genel yapısı şu şekildedir;

- *Kök Düğümü (Root Node)*: Karar ağacında kendisine doğru akış olmayan tek bir giden kenardan (outgoing edge) oluşur,
- *Dâhili/Test Düğümü (Internal/Test Node)*: Bir tane gelen kenarla birlikte giden kenarlara da sahip düğüm çeşididir,
- *Uç/Yaprak Düğümü (Terminal/Leaf Node)*: Sadece gelen kenarı olan final çıktısını belirten düğümdür (Kuzey, 2012).

Karar Ağacı algoritması, bir seçim yapma sürecindeki karar ağacı yapısına benzer şekilde inşa edilir. Karar Ağacı yapısı, içindeki düğümlerin birbirleriyle bağlandığı bir hiyerarşiye sahiptir. Kök düğümünden başlayarak her bir dalda bir soru ya da test ile verileri bölerek ilerleyen ve her bir dalın sonunda bir karar ya da tahmin yapan bir yapıya sahiptir.

Karar Ağacı, veri kümesindeki en önemli özellikleri belirlemek ve bu özelliklerin sınıflandırma veya regresyon performansını artırmak için değişken seçimi yapabilen bir algoritmadır. Şekil 3.16’da Karar Ağacı modeline bir örnek verilmiştir.



Şekil 3.16. Doğrudan postalamaya cevap (Rokach ve Marmon, 2005; aktaran Kuzey, 2012)

3.3.3.4. Rastgele Orman

Sınıflandırma ya da regresyon ağaç topluluklarından Rastgele Orman, bir Karar Ağacı topluluğu (ensemble) olarak çalışır ve bu nedenle birden fazla Karar Ağacı kullanarak daha iyi sonuçlar verme eğilimindedir (Şirin, 2020).

1. *Verilerin belirlenmesi:* Her alt kümedeki veriler için rastgele özellikler seçilir. Ancak bazı varyasyonlarda veri setindeki en önemli özellikleri seçmek gibi gereklilikler için özellik seçimi daha karmaşık hale getirilebilir.
2. *Karar Ağacı modellerinin oluşturulması:* Her bir alt kümedeki veriler kullanılarak bir Karar Ağacı oluşturulur. Ağaç derinliği belirli bir maksimum değer ile sınırlanır. Karar Ağacı'nın özellikleri ve sınıfları kullanılarak, verileri

sınıflandırmaya çalışılır. Ancak, bazı durumlarda, minimum örnek sayısı veya düğüm başına minimum bilgi kazancı gibi başka sınırlamalar da uygulanabilir.

3. *En iyi bölünmenin seçilmesi:* Karar Ağacı, her düğümde en iyi (en yüksek bilgi kazancı sağlayacak) bölünmeyi belirlemek için kriterler kullanır. Bunlar arasında bilgi kazancı, Gini endeksi ve Gini azalması gibi kriterler bulunur.
4. *Tekrarlanan adımlar:* 2. ve 3. adımlar, istenen sayıda farklı alt kümelerden farklı Karar Ağacı modeli üretilene kadar tekrarlanır.
5. *Sınıf tahmini:* Oluşturulan tüm Karar Ağacı modellerinin sonuçları bir araya getirilerek sınıf tahmini yapılır. Yöntemin gücü, farklı alt kümelerin farklı Karar Ağacı modelleri tarafından sınıflandırılmasıdır. Bu sayede aşırı uyum (overfitting) problemleri minimize edilir ve daha doğru sonuçlar elde edilir.

3.3.3.5. AdaBoost (Adaptive Boosting)

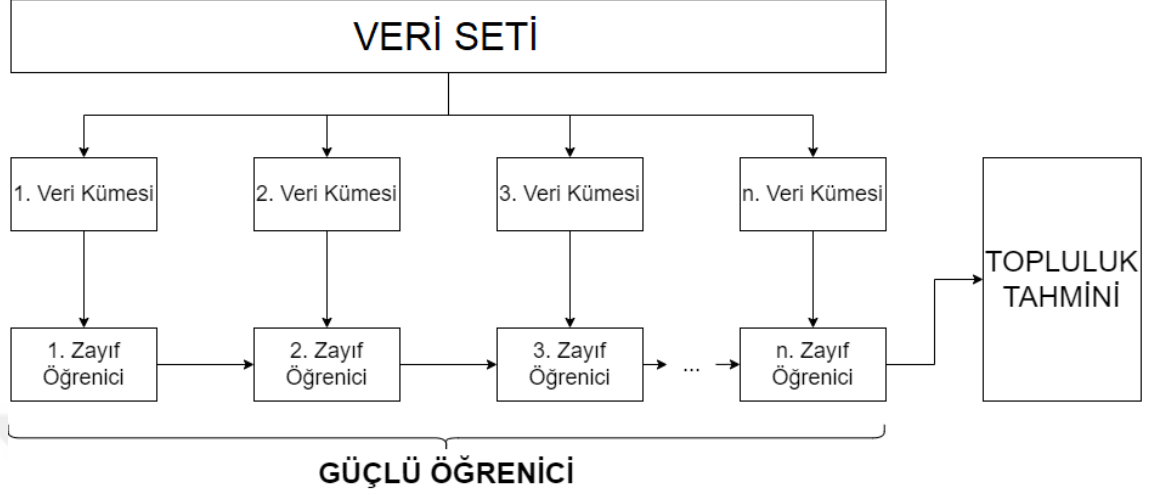
İlk olarak 1990 yılında birbirinden bağımsız olarak Robert E. Schapire ve Yoav Freund tarafından geliştirilen boosting algoritmaları, genel olarak zayıf öğrenici (weak learner) olarak bilinen bir dizi modeli bir araya getirerek güçlü bir öğrenci (strong learner) oluşturmak üzere tasarlanmışlardır (Mayr vd., 2014). Boosting algoritmaları regresyon ve sınıflandırma ağaçlarında kullanıldığında daha fazla performans gösterdiği bilinmektedir (Yürük, 2022).

AdaBoost, yüksek boyutlu veriler üzerinde etkili bir şekilde çalışabilir. Genellikle diğer sınıflandırma algoritmalarına göre daha hızlı sonuçlar üretir. Bununla birlikte, aşırı uyum (overfitting) riski yüksektir ve bu nedenle çoğu zaman aşırı uyumu önlemek için düzenleme yöntemleri kullanılmaktadır.

AdaBoost modellenmesinde ilk iş eğitim kümesini zayıf bir öğrenci ile eğitmektir. Yanlış tahmin yapan verilerin ağırlığı artırılırken, daha doğru tahmin yapan verilerin ağırlığı azaltılır. Zayıf öğrenci çıkışı diğer bir öğrencide giriş olarak eğitilir, farklı zayıf öğrencilerden gelen bilgi bir araya getirilir. Bu sayede, daha fazla dikkat gerektiren veriler üzerinde daha fazla çalışarak daha iyi bir sonuç elde edilmektedir.

AdaBoost, her bir zayıf öğrencinin tahminleri yapıldıktan sonra ağırlıkları yeniden hesaplar ve ardından bu ağırlık tanımlanmış tahminlerin bir araya getirilmesiyle nihai

sınıflandırma yapar. Bu sayede, farklı zayıf öğrenicilerden gelen bilgiyi bir araya getirerek daha doğru bir tahmin yapabilmektedir.



Şekil 3.17. Bir AdaBoost akış şeması örneği

3.3.3.6. Gradyan Artırma

Gradyan Artırma, çeşitli makine öğrenimi uygulamalarında genel olarak verilerden parametrik olmayan regresyon ya da sınıflandırma modeli oluşturmak için kullanılan çeşitli uygulamalarda önemli başarı göstermiş bir makine öğrenimi tekniğidir (Natekin ve Knoll, 2013).

Gradyan Artırma algoritmasının nasıl işlediğine bakılacak olursa;

- Öncelikle basit bir model, veriler üzerinde eğitilir, genellikle başlangıç modeli en basit model olmaktadır.
- Bu modelden çıkan hatalar hesaplanır ve bu hatalar üzerinden yeni bir model eğitilir
- Bu modeller, üçüncü bir model eğitmede kullanılır.
- Bu işlem istenen doğruluk düzeyine ulaşmaya kadar tekrarlanır.

3.3.3.7. XGBoost (xtreme gradient boosting) (ekstrem gradyan artırma)

XGBoost, Gradyan Artırma yönteminin optimize edilecek şekilde aşırı öğrenmeye karşı dirençli bir varyasyonudur. Ayrıca boş verileri yönetebilmesi, hızlı olması, hata fonksiyonunu minimize eden ve yüksek performanslı tahmin modelleri oluşturabilen bir

yöntem olması ve iyi hesaplama yapılabilmesi; XGBoost'u diğer makine öğrenmesi yöntemleri ve algoritmalarından avantajlı bir konuma getirmektedir (Yürük, 2022). Daha az bellek kullanarak daha yüksek doğruluk oranlarına ulaşabilmesi, genellikle büyük boyutlu veri setleri üzerinde çalışan makine öğrenimi projelerinde tercih edilme sebebidir. XGBoost'un diğer avantajları arasında:

- Otomatik eksik veri işlemesi yapar.
- Entegrasyon kolaylığı ile Python, R, Java, Scala ve C++ gibi birçok farklı programlama diliyle kullanılabilir.
- Birçok özelleştirilebilir parametre seçeneği sunar.
- GPU desteği sayesinde hesaplama hızını artırır.

3.3.3.7.1. Nitelik önem katsayısı (feature importance)

Nitelik önem katsayısı belirli bir bağımsız değişkenin toplam nitelik önem katsayısı değeri içindeki payını temsil eder. Nitelik önem katsayısı, özelliklerin kullanım sıklığı, tahmin hataları üzerindeki etkileri ve düğüm bölünmelerindeki önem gibi faktörlere dayanarak hesaplanır. Bu değerler, modelin eğitimi tamamlandıktan sonra elde edilir ve her bir özelliğin önem sırasını belirleyerek modelin çalışmasını daha iyi anlamak için kullanılabilir.

XGBoost'un nitelik önem katsayısı değerleri, bir modelin tahminler üzerinde hangi bağımsız değişkenlerin daha büyük bir etkiye sahip olduğunu anlamak için değerli bir araç sağlar.

3.3.3.8. Aşırı uyum (overfitting)

Aşırı öğrenme veya aşırı uyum, (ezberleme) bir modelin eğitim verileri üzerinde çok iyi eğitildiği ve yeni/görünmeyen verilere genelleme yapmak yerine ezber yapması, makine öğrenimi modellerinin yaygın bir sorunudur. Bu sorun, modelin daha önce görülmemiş yeni veriler üzerinde doğru tahminler yapma yeteneğinin bir ölçüsü olan test verilerinde düşük performansa neden olmaktadır. Aşırı uyum, model çok karmaşık olduğunda veya modelin karmaşıklığını desteklemek için yeterli veri olmadığında ortaya çıkabilmektedir.

Aşırı uyumu önleyerek optimum modeli belirlemek için farklı bölgeler için çapraz doğrulama yöntemi kullanılmakta veya aşırı uyuma dirençli hibrit modeller geliştirilmektedir (Rahmati vd., 2018; Huang ve Perry, 2016). Bu çalışmada ise aşırı uyumu önlemek ve en iyi hiperparametreleri seçmek için tüm makine öğrenmesi modellerinde çapraz doğrulama yöntemi kullanılmıştır (Deng vd., 2015).

3.3.3.9. Veri sayısallaştırma yöntemleri

Makine Öğrenmesi programlanırken iki tip sayısallaştırma yöntemi kullanılır; Label Encoder ve One-Hot Encoder.

Label Encoder genellikle aralarında hiyerarşik ilişki bulunan değişkenleri değişken sayısına göre (0, 1, ...) şeklinde tek bir kategoride sayısallaştırmayı sağlayan bir yöntemdir.

One-Hot Encoder kategorik değişkenleri sayısal değerler olarak temsil etmek için kullanılan bir tekniktir. Bu teknikte, orijinal sütundaki her benzersiz kategori için yeni bir sütun oluşturularak ve ardından karşılık gelen kategoriyi içeren satır için yeni sütuna 1 ve diğer tüm satırlar için 0 atanarak yapılır.

Örneğin; One-Hot Encoder yardımıyla kırmızı [1,0,0], yeşil [0,1,0] ve mavi [0,0,1] renkleri Tablo 3.2’de sayısallaştırılmıştır.

Tablo 3.2 Kırmızı, yeşil ve mavi renklerinin One-Hot Encoder yardımıyla sayısallaştırılmış hali

RENK	KIRMIZI	YEŞİL	MAVİ
KIRMIZI	1	0	0
YEŞİL	0	1	0
MAVİ	0	0	1

Cinsiyet (kadın – erkek) değişkeni gibi sınıflar arası fark olmayan başka bir deyişle nominal değişkenlerde, Label Encoder kullanılması durumunda verilen sayısal temsiller, bir büyüklük küçüklük ifade etmesinden kaynaklı olarak ordinal değişkenlerde yapılan sayısallaştırma tipine benzer şekilde sıralama durumunun oluşmasına sebep olmaktadır. Sayısallaştırılan nominal kategorik değişkenin sınıfları arasında sınıfsal bir farklılık olmamasına karşın, Label Encoder kullanımı sonucu bir sınıfsal fark varmışçasına

betimlemeyi ortaya çıkarmaktadır. Bu nedenle Label Encoder tekniğini nominal değişkenler özelinde kullanmak doğru bir yaklaşım olarak nitelendirilmeyebilir. Dolayısıyla nominal değişkenlerin gözlemlendiği durumlarda doğru bir değer dönüşümü ve temsil şekli uygulamak adına Label Encoder tercih edilmemektedir. Nominal değişkenleri bir ölçüm problemine sebep olmadan oluşturmak adına One-Hot Encoder tekniği uygulamak çok daha doğru bir yaklaşım olmaktadır. Nominal değişkenlerin olduğu veri setlerinde veri kategorileri yeni bir sütuna dönüştürülür ve değişkenlere 1 veya 0 atanır, bu metot One-Hot Encoder metodudur. Bu metodun avantajı, bir değeri kolaylıkla doğru bir şekilde ağırlıklandırmaktadır, ancak veri kümesine ek sütunlar ekleme dezavantajı vardır (Anh vd., 2022).

Çalışmada toplanan verilere ait değişkenler arasında bir ordinallik (Kavşak Adı, Taşıt Tipi, Şerit Sayısı) durumu bulunmamasından dolayı Label Encoder yerine One-Hot Encoder tercih edilmiştir. (Jiang vd., 2020; Gnat, 2021)

3.3.3.10. Pipeline Yapısı

Pipeline yapısı, birden fazla komutun paralel olarak aynı anda yürütülmesini sağlayan bir yapıdır. Sadece makine öğrenmesine ait bir kavram değildir. Ancak Python'daki Pipeline komutu bu kavram temel alınarak oluşturulmuştur.

Pipeline yapısının verimli çalışabilmesi için; farklı veriler üzerinde defalarca tekrarlanan görevler tanımlanır ve görevler paralel yürütülebilen küçük alt görevlere bölünür.

Pipeline komutu;

- Komutun alınması (Fetch)
- Çözümleme (Decode)
- Çalıştırma (Execute)

Olmak üzere üç temel adımda işler.

Fetch aşamasında komutlar, bellekte belirli bir adresten başlar. İşlemci, program sayacının işaret ettiği bellek adresindeki komutu yürütmek için bu adrese erişir ve işlemi gerçekleştirir. Program sayacı her adımda bir artış göstererek sıradaki komutun bellek

adresini tutar ve işlemci tarafından yürütülen komutlar sıralı bir şekilde işlem basamağından alınır.

Decode aşaması, komutun operand değerlerinin çözümlendiği basamaktır. İşlemci, işlem basamağından aldığı komutu decode basamağında analiz eder ve komutun hangi işlemi yapacağını, hangi verilerle çalışacağını belirler. Operand değerleri bellekten veya işlemcinin kayıtlarından okunur ve işleme konur. Decode basamağı, komutların doğru bir şekilde yürütülmesini sağlamak için oldukça önemlidir ve işlemci performansı üzerinde de etkili olabilir.

Execute aşamasında; CPU'nun ilgili işlevsel bölgelerine çözümlenen komutlar iletilir ve iletilen her komut, belirli bir görevi yerine getirmesi için CPU'daki ilgili birimlere yönlendirilir. Komutların gerektirdiği aritmetik/lojik işlemler CPU'nun ALU biriminde gerçekleştirilir.

Birçok makine öğrenmesi çalışmasında olduğu gibi bu çalışmada da Pipeline, verilerin işlenmesinden makinelerin eğitilmesine kadar birçok aşamanın kolayca yönetilmesini sağlamaktadır. Bu çalışma için ayrıntılı bakılacak olursa “Pipeline” komutu model uyumunun sağlanması, kullanılacak metotla makinenin eğitilmesini ve eğitim sonucu hata metrikleri ile performans değerlendirmesini yönetecek şekilde kullanılmıştır.

3.3.3.11. GridSearchCV

GridSearchCV, makine öğrenimi modellerinde kullanılan hiperparametre optimizasyon yöntemlerinden biridir. Hiperparametreler, modelin performansını etkileyen faktörlerdir ve manuel olarak ayarlanması zaman alıcı ve yorucu bir iş olabilir. GridSearchCV, bu süreci otomatikleştirerek, verilen bir ızgarada (grid) tüm hiperparametre kombinasyonlarını deneyerek en iyi performansa sahip hiperparametreleri seçer. GridSearchCV 4 temel aşama ile anlatılırsa;

- Öncelikle, kullanılacak olan modelin hiperparametreleri belirlenir. Bu parametreler, modelin performansını etkileyen farklı faktörlerdir. Örneğin, bir Rastgele Orman modeli için ağaç sayısı, maksimum derinlik, minimum örnek bölme sayısı vb. gibi hiperparametreler kullanılabilir.

- Daha sonra, her hiperparametrenin olası deęerleri belirlenir. Bu olası deęerler, bir ızgara (grid) üzerinde belirlenir. Örneęin, bir Rastgele Orman modelinde ağaç sayısı için [100, 200, 300] ve maksimum derinlik için [10, 20, 30] gibi bir ızgara oluşturulabilir.
- GridSearchCV, ızgaradaki her bir hiperparametre kombinasyonunu kullanarak modeli eğitir ve performans metriklerini hesaplar. Performans metrięi, seçilen problem tipine göre deęişebilir. Örneęin, sınıflandırma problemlerinde doğruluk (accuracy), regresyon problemlerinde ortalama karesel hata (mean squared error) gibi.
- GridSearchCV, tüm denemelerin sonucunda en iyi performans gösteren hiperparametre kombinasyonunu seçer ve bu parametreleri kullanarak bir final modeli oluşturur.

GridSearchCV'nin faydaları şunlardır;

- Manuel olarak hiperparametrelerin ayarlanmasından daha az zaman harcamayı sağlar.
- Daha iyi bir model performansı elde etmeye yardımcı olur.
- Farklı hiperparametrelerin etkilerini keşfetmeyi sağlar.

GridSearchCV'nin bazı dezavantajları da şunlardır;

- Çok sayıda hiperparametre olduğunda, işlem süresi uzayabilir.
- Hiperparametre ızgarasının yanlış belirlenmesi, performansın düşük olmasına neden olabilir.

3.3.3.12. Hata metrikleri

Hata metrikleri, oluşturulan modellemelerin performansını ve hata oranlarını belirlemek için kullanılmaktadır. Bu çalışmada makine öğrenmesi modellemelerinin performansını deęerlendirmek için R^2 , MAE, MSE ve RMSE metrikleri tercih edilmiştir.

R^2 skoru, MAE (Mean Absolute Error) (Ortalama Mutlak Hata), MSE (Mean Squared Error) (Ortalama Karesel Hata) ve RMSE (Root Mean Squared Error) (Ortalama Karesel Hatanın Karekökü) gibi metrikler, makine öğrenimi modellerinin performansını

değerlendirmek için kullanılan yaygın metriklerdir. Bu metrikler, modelin tahminlerinin gerçek değerlere ne kadar yakın olduğunu ölçmek için kullanılmaktadır.

3.3.3.12.1. R^2

Regresyon modellerinde bağımsız değişkenlerin bağımlı değişken üzerindeki varyansının yüzde kaçını açıkladığını ölçmekte ve genel olarak r^2 olarak bilinmektedir. R^2 değeri korelasyon katsayısı r 'nin karesi alınarak bulunmaktadır. Değerler 0 (ilişkinin olmama durumu) ile 1 (tam ilişki) arasında değişmekte ve bu değer 1'e yaklaşması regresyon modelinin veriyi açıklama yeteneğinin yüksek olduğu anlamına gelmektedir (Günel, 2003).

3.3.3.12.2. MAE

Ortalama mutlak hata (MAE) gözlenen değerler ile tahmin edilen değerler arasındaki farklarının mutlak değerlerinin ortalama hatasıdır. Zaman serileri ve regresyon modellemesinde sıkça kullanılan bir hata metriğidir (Özen vd., 2021). MAE'nin değeri 0'a ne kadar yakınsa, modelin tahminleri gerçek değerlere o kadar yakındır. MAE, özellikle aykırı değerlerin olduğu durumlarda kullanışlı bir metriktir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (3.13)$$

y_i : gözlenen değer

y'_i : tahmin edilen değer

n : gözlem sayısı

3.3.3.12.3. MSE

MSE, tahmin değerlerinin gerçek değerlerden farklarının karelerinin ortalamasıdır. MSE'nin değeri 0'a ne kadar yakınsa, modelin tahminleri gerçek değerlere o kadar yakındır. MSE, genellikle hata karelerini büyük hataların daha fazla etkili olmasını sağlayarak modelin performansını değerlendirmek için kullanılmaktadır.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (3.14)$$

3.3.3.12.4. RMSE

Ortalama kare hatasının kökü tanımının kısaltılışı olan RMSE, MSE'nin kareköküdür, yani gözlenen değerler ve tahmin edilen değerler arasındaki farkın standart sapmasıdır. RMSE'nin değeri 0'a ne kadar yakınsa, modelin tahminleri gerçek değerlere o kadar yakındır.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (3.15)$$

RMSE'nin karesi MSE aykırı değerlere karşı duyarlı bir hata metriğidir, ancak RMSE bu duyarlılığa sahip değildir. Öte yandan MSE büyük değerlerin performans yorumlamasında kolayca karşılaştırılmaz ve bu durumda RMSE'nin kolay yorumlama ve hassasiyet özelliklerine başvurmak gerekmektedir.

RMSE değerinin performansını yüzdelik olarak ifade etmek için RMSE değerini veri setinin ortalamasına (mean) bölmek gerekmektedir.

$$RMSE \text{ Hata Oranı} = \frac{RMSE}{M} \quad (3.16)$$

RMSE: RMSE metrik değeri

M: hedef parametrenin ortalaması

4. MATERYAL ve YÖNTEM

4.1. Materyal

Erzincan ili merkezindeki taşıt sayımları kullanılarak AIMSUN programında oluşturulan simülasyon ve bu simülasyon üzerinden elde edilen senaryolarla veri seti üretilmiştir.

Erzincan ili Doğu Anadolu Bölgesi'nin Yukarı Fırat Bölümü'nde konumlanmış bir şehirdir. Erzincan sınırları içerisinde merkez dâhil 9 ilçe, 15 belediye ve 522 köye sahiptir. Erzincan'ın yüz ölçümü 11.974 km² ve 1185 m denizden olan yüksekliğe sahiptir. Erzincan'ın nüfusu 2021 yılında 237.351 iken 2022 239.223'ye yükselmiş ve motorlu taşıtlar sayısı 2021'de 63.158 iken 2022'de 65.156 olmuştur (TÜİK, 2022).

4.1.1. Çalışmanın yapıldığı koridor hakkında bilgiler

Çalışmada kullanılan koridor; batıda Dört Yol Kavşağı ile başlamakta ve Merplaza, Eski Terminal, Terzibaba, Eski Polis Okulu, TEDAŞ ve Yeni Terminal kavşaklarını kapsamaktadır. Koridorun uzunluğu yaklaşık 4.150 metredir. Bu güzergâh şehrin trafik hacmi en yüksek güzergâhlarından biridir. Bu yoğunluk trafikte gecikme süresi ve kuyruk uzunluğu gibi sorunlar oluşturmaktadır. Koridorun uydu görüntüsü Şekil 4.1'de gösterilmiştir.



Şekil 4.1. Çalışmanın yapıldığı koridor (Google Earth, 2023)

Bu çalışmada koridorun batıdaki ilk kavşağı olan Dört yol Kavşağı, Erzincan şehrinin merkezinde yer alan sinyalizasyonlu dönel kavşaktır. Erzincan şehrinin en yoğun kavşaklarından biridir. Kavşağın uydur görüntüsü Şekil 4.2’de gösterilmiştir.



Şekil 4.2. Dört yol Kavşağı (Google Earth, 2023)

Merplaza Kavşağı’na bakılacak olursa; bu kavşakta sola dönüşler yasaktır. Bu sebeple sadece sağa dönüşlerin yapıldığı ya da doğu batı istikametinde trafiğin seyrettiği bir kavşaktır. Kavşağa ait uydur görüntüsü Şekil 4.3’te gösterilmiştir.



Şekil 4.3. Merplaza Kavşağı (Google Earth, 2023)

Eski Terminal Kavşağı ise şehrin eski sanayi bölgesine ulaşımında tercih edilen sinyalizasyonlu bir kavşaktır. Bu kavşağın uydu görüntüsü Şekil 4.4'te gösterilmiştir.



Şekil 4.4. Eski Terminal Kavşağı (Google Earth, 2023)

Sinyalizasyonlu bir kavşak olan Terzibaba Kavşağı'nda sağa dönüşler sinyalizasyonlu olarak değil kontrollü olarak verilmiştir. Bu kavşağın uydu görüntüsü Şekil 4.5'te gösterilmiştir.



Şekil 4.5. Terzibaba Kavşağı (Google Earth, 2023)

Eski Polis Okulu Kavşağı sinyalize kavşaktır. Bu kavşakta ana cadde üzerindeki batı yönünden gelen araçlar sağa dönüşleri kontrollü olarak yapmaktadır. Bu kavşağın uydu görüntüsü Şekil 4.6’da gösterilmiştir.



Şekil 4.6. Eski Polis Okulu Kavşağı (Google Earth, 2023)

TEDAŞ Kavşağı girişleri sinyalize iken ada içi manevraları kontrollü olarak yapılan bir kavşaktır. Bu kavşağın uydu görüntüsü Şekil 4.7’de gösterilmiştir.



Şekil 4.7. TEDAŞ Kavşağı (Google Earth, 2023)

Yeni Terminal Kavşağı çalışma için seçilen koridorun en doğusundaki kavşağıdır. Modern Dönel Kavşak geometrisine sahip bir sinyalize dönel kavşaktır. Bu kavşağın uydu görüntüsü Şekil 4.8’de gösterilmiştir



Şekil 4.8. Yeni Terminal Kavşağı (Google Earth, 2023)

4.1.2. Çalışmanın kullanılan koridora ait trafik verileri

Yapılan bu çalışmada otomobil, otobüs ve kamyon olmak üzere taşıt sayımları ve hareketleriyle birlikte kavşaklara ait sinyalizasyon süreleri Demiriz’in (2019) tez çalışmasından alınmıştır.

4.1.2.1. Taşıt sayımları

Çalışmanın yapıldığı alanlardaki taşıt sayımı otomobil, otobüs ve kamyon olarak gözlemciler tarafından ayrı ayrı yapılmıştır. Gözlemciler, taşıt sayımları ile taşıtların kavşaklara hangi yönden girdiklerini ve kavşaklardan hangi yönden çıktıklarını belirlemişlerdir. Pazartesi, Çarşamba, Cuma ve Pazar günleri sabah 07.45-08.45 saatleri arası, öğlen 12.00-13.00 arası ve akşam 16.45-17.45 arasında sayımları yapmışlardır. Sayımlar sonucunda en yoğun trafik hacminin Pazartesi 07.45-08.45 arasında olduğu tespit edilmiştir. Bu sebeple bu çalışmada yapılan mevcut durum senaryosu için girilen taşıt sayıları Pazartesi 07.45-08.45 arası olarak seçilmiştir. Tablo 4.1-4.8 ile güzergâhta bulunan kavşakların otomobil, otobüs ve kamyon sayımları her kavşak için verilmiştir.

Tablo 4.1. Dört yol Kavşağı taşıt sayımları

Yönler	OTOMOBİL				OTOBÜS				KAMYON			
	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı
Batı	66	462	243	31	14	48	13	0	1	1	0	0
Güney	8	68	229	123	0	13	29	12	0	0	1	0
Doğu	170	28	52	448	16	0	11	49	0	0	0	1
Kuzey	313	186	43	71	33	12	0	16	1	0	0	0

Tablo 4.2. Merplaza Kavşağı taşıt sayımları

Yönler	OTOMOBİL				OTOBÜS				KAMYON			
	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı
Batı	206	648	0	0	3	70	0	0	0	1	0	0
Güney	0	138	360	0	0	3	4	0	0	0	0	0
Doğu	0	0	158	482	0	0	4	70	0	0	0	1
Kuzey	442	0	0	100	3	0	0	5	0	0	0	0

Tablo 4.3. Eski Terminal Kavşağı taşıt sayımları

Yönler	OTOMOBİL				OTOBÜS				KAMYON			
	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı
Batı	77	363	62	15	4	53	9	0	0	1	0	0
Güney	0	83	92	61	0	2	22	12	0	1	1	0
Doğu	113	4	36	511	4	0	7	67	0	0	0	1
Kuzey	249	135	3	55	9	4	0	8	2	0	0	0

Tablo 4.4. Terzibaba Kavşağı taşıt sayımları

Yönler	OTOMOBİL				OTOBÜS				KAMYON			
	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı
Batı	9	389	175	5	3	37	14	0	0	2	0	0
Güney	0	27	43	5	0	0	0	1	0	1	1	0
Doğu	27	2	41	435	2	0	6	56	0	0	1	1
Kuzey	114	24	0	156	0	1	0	19	2	0	0	0

Tablo 4.5. Eski Polis Okulu Kavşağı taşıt sayımları

Yönler	OTOMOBİL				OTOBÜS				KAMYON			
	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı
Batı	4	335	40	12	5	24	7	0	0	3	0	0
Güney	0	14	35	14	0	1	1	5	0	0	1	0
Doğu	9	0	14	378	0	0	3	49	0	0	0	2
Kuzey	41	34	0	113	2	2	0	7	2	0	0	0

Tablo 4.6. TEDAŞ Kavşağı taşıt sayımları

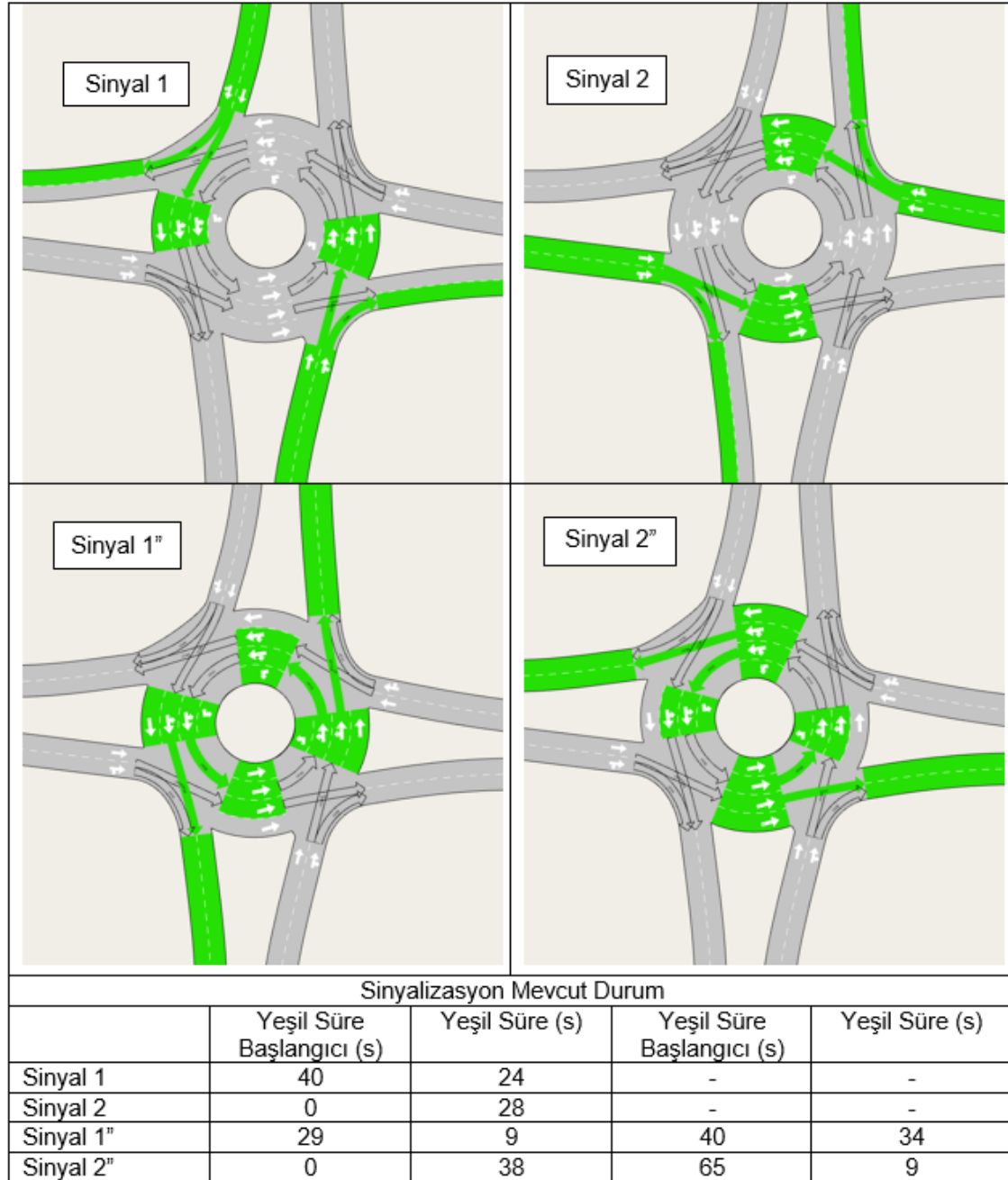
Yönler	OTOMOBİL				OTOBÜS				KAMYON			
	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı
Batı	48	512	48	4	3	12	4	0	0	3	0	0
Güney	4	144	172	56	0	2	3	2	0	4	3	0
Doğu	96	5	70	460	2	0	2	14	0	0	0	2
Kuzey	62	228	3	4	1	2	3	3	2	0	0	0

Tablo 4.7 Yeni Terminal Kavşağı taşıt sayımları

Yönler	OTOMOBİL				OTOBÜS				KAMYON			
	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı	Güney	Doğu	Kuzey	Batı
Batı	15	283	-	12	2	8	-	0	0	8	-	0
Güney	6	34	-	65	0	3	-	6	0	0	-	0
Doğu	57	6	-	344	1	0	-	8	0	0	-	5
Kuzey	-	-	-	-	-	-	-	-	-	-	-	-

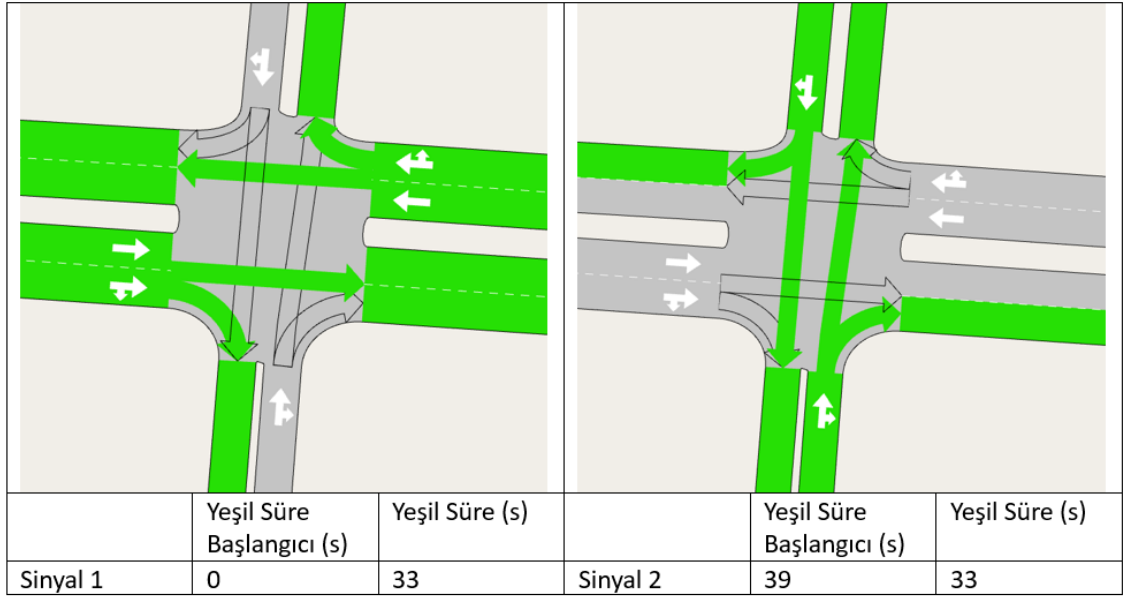
4.1.2.2. Sinyalizasyon verileri

Çalışmada gerçek kavşaklara ait sinyalizasyon verileri kullanılmıştır. Bu veriler kavşakların batıdan doğuya sıralandığı şekilde Şekil 4.9-4.15’de verilmiştir.



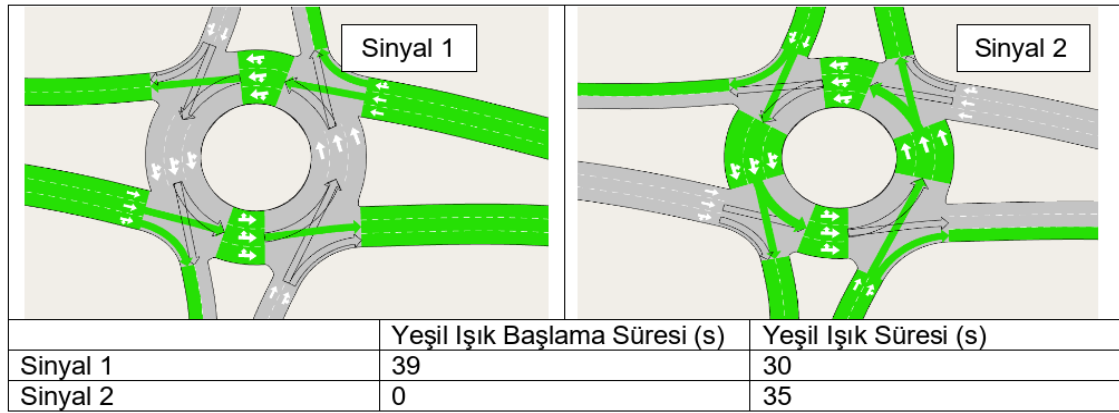
Şekil 4.9. Dört yol Kavşağı'na ait sinyalizasyon verileri

Dört yol kavşağında mevcut durumda periyot 76 saniye, kırmızı ile yeşil ışık arasındaki sarı ışık süresi 1 saniye, koruma kırmızısı 1 saniyedir.



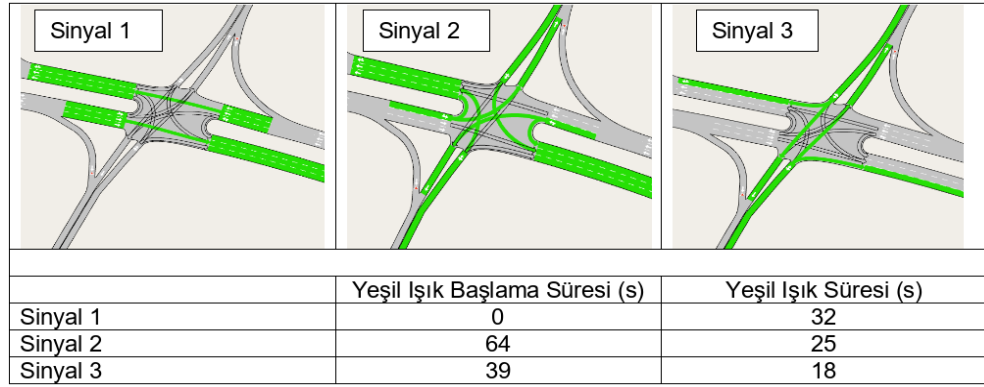
Şekil 4.10. Merplaza Kavşağı mevcut durum faz süreleri

Merplaza Kavşağı'nda mevcut durumda periyot süresi 78 saniye, kırmızı ışık ile yeşil ışık arasındaki sarı ışık süresi 4 saniye, koruma kırmızısı süresi 2 saniyedir.



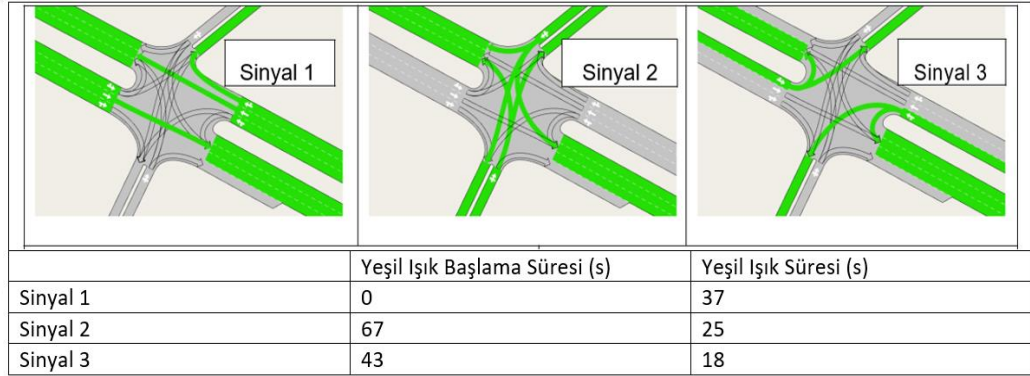
Şekil 4.11. Eski Terminal Kavşağı faz süreleri

Eski Terminal Kavşağı'nda mevcut durumda periyot süresi 73 saniye, kırmızı ışık ile yeşil ışık arasındaki sarı ışık süresi 2 saniye, koruma kırmızısı süresi 1 saniyedir.



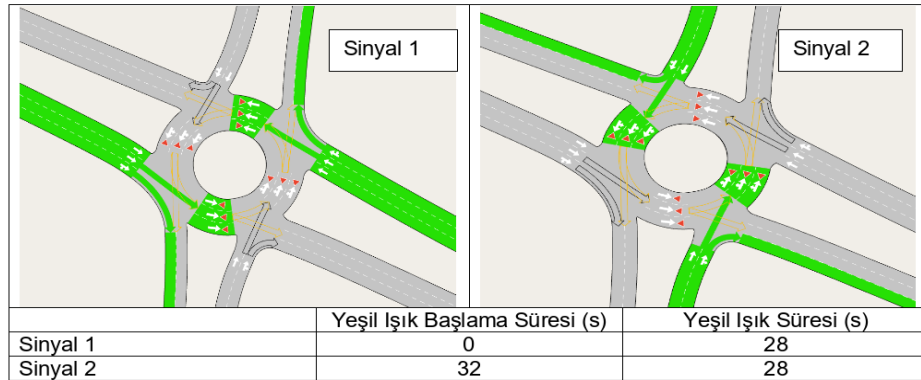
Şekil 4.12. Terzibaba Kavşağı faz süreleri

Terzibaba Kavşağı'nda mevcut durumda periyot süresi 96 saniye, kırmızı ışık ile yeşil ışık arasındaki sarı ışık süresi 4 saniye, koruma kırmızısı süresi 2 saniyedir.



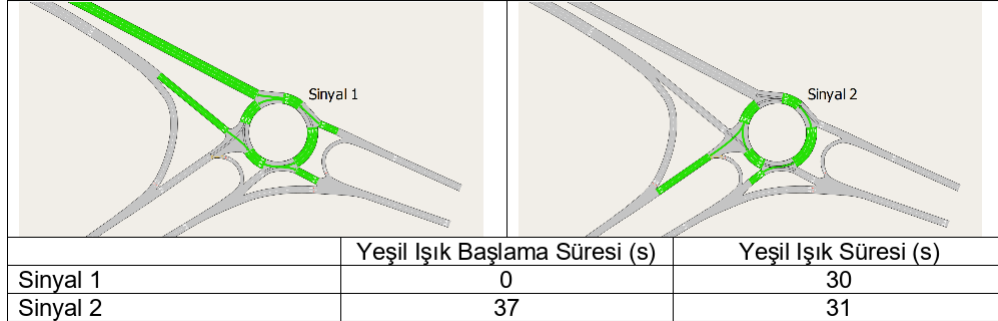
Şekil 4.13. Eski Polis Okulu Kavşağı faz süreleri

Eski Polis Okulu Kavşağı'nda mevcut durumda periyot süresi 98 saniye, kırmızı ışık ile yeşil ışık arasındaki sarı ışık süresi 4 saniye, koruma kırmızısı süresi 2 saniyedir.



Şekil 4.14. TEDAŞ Kavşağı faz süreleri

TEDAŞ Kavşağı'nda mevcut durumda periyot süresi 64 saniye kırmızı ışık ile yeşil ışık arasındaki sarı ışık süresi 2 saniye koruma kırmızısı süresi 1 saniyedir.



Şekil 4.15. Yeni Terminal Kavşağı faz süreleri

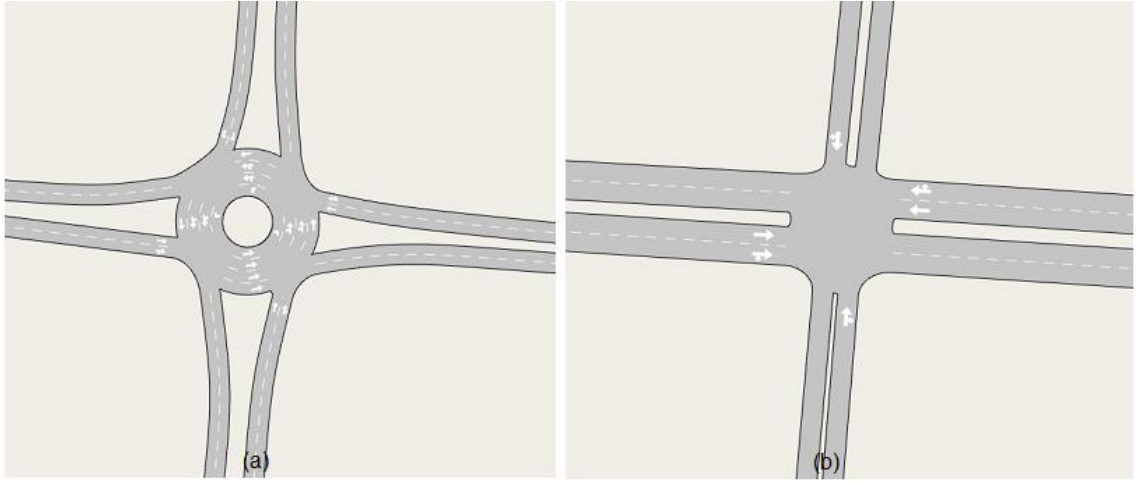
Yeni Terminal Kavşağı'nda mevcut durumda periyot süresi 75 saniye, kırmızı ışık ile yeşil ışık arasındaki sarı ışık süresi 4 saniye, koruma kırmızısı süresi 2 saniyedir.

4.1.3. Simülasyon Modellemesi

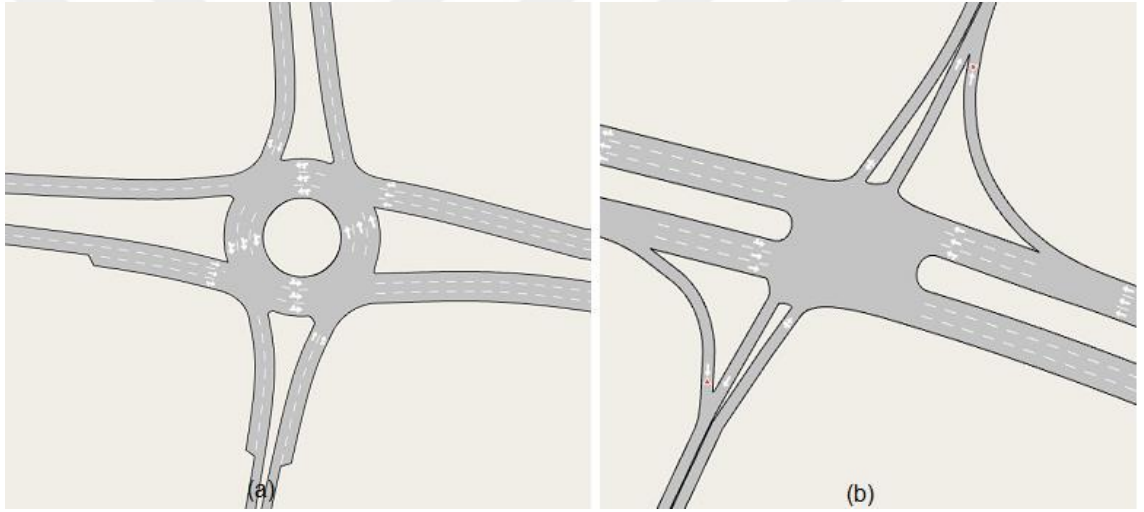
AIMSUN yardımı ile oluşturulan trafik simülasyonu modeli kullanılarak farklı senaryolar oluşturulmuş ve bu senaryoların analizi sonucunda elde edilen verilerle makine öğrenmesi için veri hazırlanmıştır. Bu veri seti, çalışmanın nihai materyalidir ve bu nedenle bu nihai materyale erişmek için kullanılan simülasyon modellemesinin önemi yüksektir.

4.1.4. Mevcut Durum Simülasyon Modellemesi

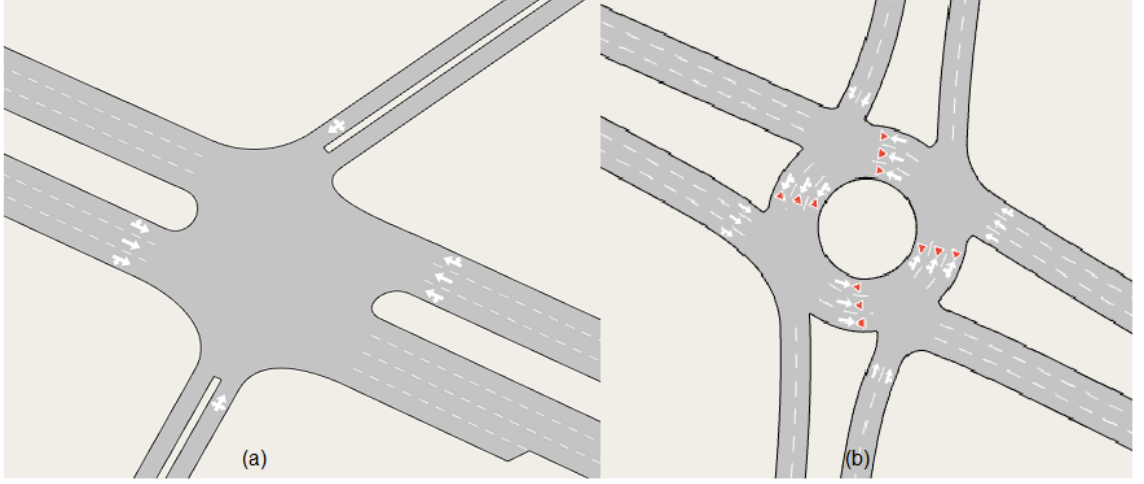
Trafik ağının bilgisayar ortamında birebir oluşturulabilmesi için OpenStreetMap platformundan Erzincan şehrinin ulaşım haritası AIMSUN programına aktarılmış ve koridorun modellemesi bu harita üzerinden oluşturulmuştur. Ayrıca Google Earth platformu ile koridordaki trafik elemanlarının doğru konumlandırıldığı teyit edilmiştir. Şekil 4.16, Şekil 4.17, Şekil 4.18 ve Şekil 4.19'da modellenen kavşaklar; Şekil 4.20'de ise koridorun modellenmiş hali verilmiştir.



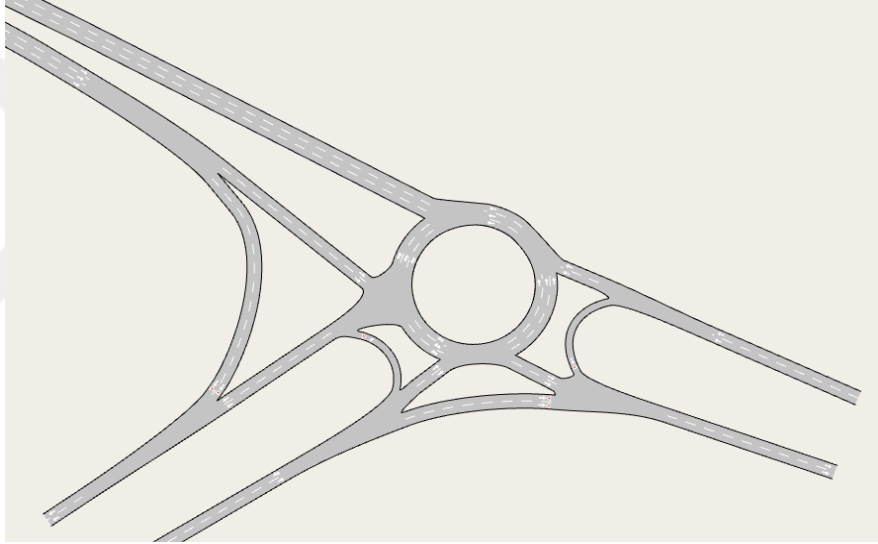
Şekil 4.16. Mevcut durum modellemesi **(a)** Dört Yol Kavşağı **(b)** Merplaza Kavşağı



Şekil 4.17. Mevcut durum modellemesi **(a)** Eski Terminal Kavşağı **(b)** Terzibaba Kavşağı



Şekil 4.18. Mevcut durum modellemesi **(a)** Eski Polis Okulu Kavşağı **(b)** TEDAŞ Kavşağı



Şekil 4.19. Yeni Terminal Kavşağı mevcut durum modellemesi



Şekil 4.20. Çalışma yapılacak koridorun mevcut durum modeli

4.1.4. GEH analizi

Simülasyon modellemesinin doğruluğunu test etmek için GEH analizi her kavşak için ayrı ayrı hesaplanmış ve Tablo 4.8’de gösterilmiştir.

Tablo 4.8. GEH analizi sonuçları

Dört Yol Kavşağı													
GEH	Batı-Doğu			Doğu-Batı			Kuzey-Güney			Güney-Kuzey			Toplam
	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	
Gözlenen Değerler	802	75	2	698	76	1	613	61	1	428	54	1	2812
Simülasyon Değerleri	769	89	2	684	61	0	658	76	2	421	54	1	2817
GEH Değeri	1,18	1,55	0,00	0,53	1,81	1,41	1,79	1,81	0,82	0,34	0,00	0,00	0,09
Kriter GEH<5	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet

Merplaza Kavşağı													
GEH	Batı-Doğu			Doğu-Batı			Kuzey-Güney			Güney-Kuzey			Toplam
	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	
Gözlenen Değerler	854	73	1	640	74	1	542	8	-	498	7	-	2698
Simülasyon Değerleri	808	86	1	623	60	0	577	6	-	483	10	-	2654
GEH Değeri	1,60	1,46	0,00	0,68	1,71	1,41	1,48	0,76	-	0,68	1,03	-	0,85
Kriter GEH<5	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	-	Evet	Evet	-	Evet

Eski Terminal Kavşağı													
GEH	Batı-Doğu			Doğu-Batı			Kuzey-Güney			Güney-Kuzey			Toplam
	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	
Gözlenen Değerler	517	66	1	664	78	1	442	21	1	236	36	2	2065
Simülasyon Değerleri	476	75	1	667	64	1	471	20	1	225	32	1	2034
GEH Değeri	1,84	1,07	0,00	0,12	1,66	0,00	1,36	0,22	0,00	0,72	0,69	0,82	0,68
Kriter GEH<5	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet

Terzibaba Kavşağı													
GEH	Batı-Doğu			Doğu-Batı			Kuzey-Güney			Güney-Kuzey			Toplam
	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	
Gözlenen Değerler	578	54	2	505	58	2	294	20	2	75	1	2	1593
Simülasyon Değerleri	599	67	1	504	59	4	288	16	2	76	0	5	1621

Tablo 4.8. GEH analizi sonuçları (devam)

GEH Değeri	0,87	1,67	0,82	0,04	0,13	1,15	0,35	0,94	0,00	0,12	1,41	1,60	0,70
Kriter GEH<5	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet
Eski Polis Okulu Kavşağı													
GEH	Batı-Doğu			Doğu-Batı			Kuzey-Güney			Güney-Kuzey			Toplam
	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	
Gözlenen Değerler	391	36	3	401	52	2	188	11	2	63	7	2	1158
Simülasyon Değerleri	384	41	3	408	41	4	201	16	0	40	4	0	1142
GEH Değeri	0,36	0,81	0,00	0,35	1,61	1,15	0,93	1,36	2,00	3,20	1,28	2,00	0,47
Kriter GEH<5	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet
TEDAŞ Kavşağı													
GEH	Batı-Doğu			Doğu-Batı			Kuzey-Güney			Güney-Kuzey			Toplam
	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	
Gözlenen Değerler	612	19	3	631	18	2	297	9	2	376	7	7	1983
Simülasyon Değerleri	607	19	3	640	14	4	276	8	2	368	7	9	1957
GEH Değeri	0,20	0,00	0,00	0,36	1,00	1,15	1,24	0,34	0,00	0,41	0,00	0,71	0,59
Kriter GEH<5	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet	Evet
Yeni Terminal Kavşağı													
GEH	Batı-Doğu			Doğu-Batı			Kuzey-Güney			Güney-Kuzey			Toplam
	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	Araba	Otobüs	Kamyon	
Gözlenen Değerler	310	10	8	407	9	5	-	-	-	105	9	-	863
Simülasyon Değerleri	268	11	10	395	9	7	-	-	-	75	8	-	783
GEH Değeri	2,47	0,31	0,67	0,60	0,00	0,82	-	-	-	3,16	0,34	-	2,7886
Kriter GEH<5	Evet	Evet	Evet	Evet	Evet	Evet	-	-	-	Evet	Evet	-	Evet

Çizelge 4.1’de görüldüğü gibi hesaplanan GEH değerleri genel olarak 0 ila 2 arasında seyretmektedir. Bu durum mevcut durum simülasyon modelinin gerçeğe yakın ve kullanılabilir bir simülasyon modeli olduğunu teyit etmektedir.

4.1.5. Oluşturulan senaryolar

Aşağıdaki tablolardaki gibi her kavşak için ayrı ayrı yapılan ve geometrik değişiklikler ve sinyalizasyon süresindeki değişiklikler, doğu-batı yöndeki yollara eklenen şeritler ve

kavşaklardaki taşıt hacimlerini artırıp bu değişikliklerin kombinasyonu ile 27.648 senaryo hazırlanmıştır. Her bir parametre için oluşturulan yeni durumlara ait bilgiler Tablo 4.9-4.10’da verilmiştir.

Tablo 4.9. Çalışma için kavşaklarda yapılan değişiklikler

Dört Yol Kavşağı	Merplaza Kavşağı	Eski Terminal Kavşağı	Terzibaba Kavşağı	Eski Polis Okulu Kavşağı	TEDAŞ Kavşağı	Yeni Terminal Kavşağı
Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum
Modern Dönel Kavşak	-	Modern Dönel Kavşak	Altgeçit	Modern Dönel Kavşak	Modern Dönel Kavşak	Modern Dönel Kavşak
Doğu – Batı Yöndeki Faz süresine 5 saniye Ekleme	Doğu – Batı Yöndeki Faz süresine 5 saniye Ekleme			Doğu – Batı Yöndeki Faz süresine 5 saniye Ekleme		

Tablo 4.10. Çalışma için taşıt hacminde ve koridor şerit sayılarında yapılan değişiklikler

Otomobil Trafik Hacmi	Otobüs Trafik Hacmi	Kamyon Trafik Hacmi	Batı-Doğu Yönünde Şerit Ekleme	Doğu-Batı Yönünde Şerit Ekleme
Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum
%5 Artış	%5 Artış	%50 Artış	1 şerit ekleme	1 şerit ekleme
%10 Artış	%10 Artış	%100 Artış		
		%150 Artış		

Bu değişkenlerin kombinasyonlarıyla oluşturulmuş 3 tane örnek senaryo Tablo 4.11.’de verilmiştir.

Tablo 4.11. Bazı senaryolara ait kavşak geometrik ve sinyalizasyon bilgileri.

Senaryo No	Dört Yol Kavşağı	Merplaza Kavşağı	Eski Terminal Kavşağı	Terzibaba Kavşağı	Eski Polis Okulu Kavşağı	TEDAŞ Kavşağı	Yeni Terminal Kavşağı	Trafik Hacmi Değişimi Araba (%)	Trafik Hacmi Değişimi Otobüs%	Trafik Hacmi Değişimi Kamyon%	Batı-Doğu istikametine (1) şerit eklenmesi	Doğu-Batı istikametine (1) şerit eklenmesi
19987	Mevcut Durum	Mevcut Durum	Dönel Kavşak	Altgeçit	Dönel Kavşak	Mevcut Durum	Dönel Kavşak	Mevcut Trafik	%5 artış	%50 artış	Ekle	Şerit Mevcut Durumu
20068	Doğu-Batı yönünde yeşil süresi 5 s arttır	Mevcut Durum	Mevcut Durum	Mevcut Durum	Dönel Kavşak	Mevcut Durum	Dönel Kavşak	Mevcut Trafik	%5 artış	%50 artış	Ekle	Şerit Mevcut Durumu
27572	Mevcut Durum	Doğu-Batı yönünde yeşil süresi 5 s arttır	Dönel Kavşak	Mevcut Durum	Dönel Kavşak	Mevcut Durum	Mevcut Durum	%5 artış	%10 artış	%100 artış	Ekle	Ekle

Örnek olarak senaryo 19.987'ye bakılacak olursa; Dört Yol, Merplaza, TEDAŞ kavşaklarında herhangi bir değişiklik olmamıştır. Terzibaba kavşağı altgeçit olarak; Eski Terminal, Eski Polis Okulu ve Yeni Terminal kavşakları ise modern dönel kavşak olarak modellenmiştir. Trafik hacimlerinde ise otomobil hacmi aynı tutulmuş, otobüs hacmi %5 ve kamyon hacmi ise %50 artırılmıştır. Batı-Doğu istikametinde bir şerit eklenmiş ve Doğu-Batı istikametinde ise mevcut durum korunmuştur.

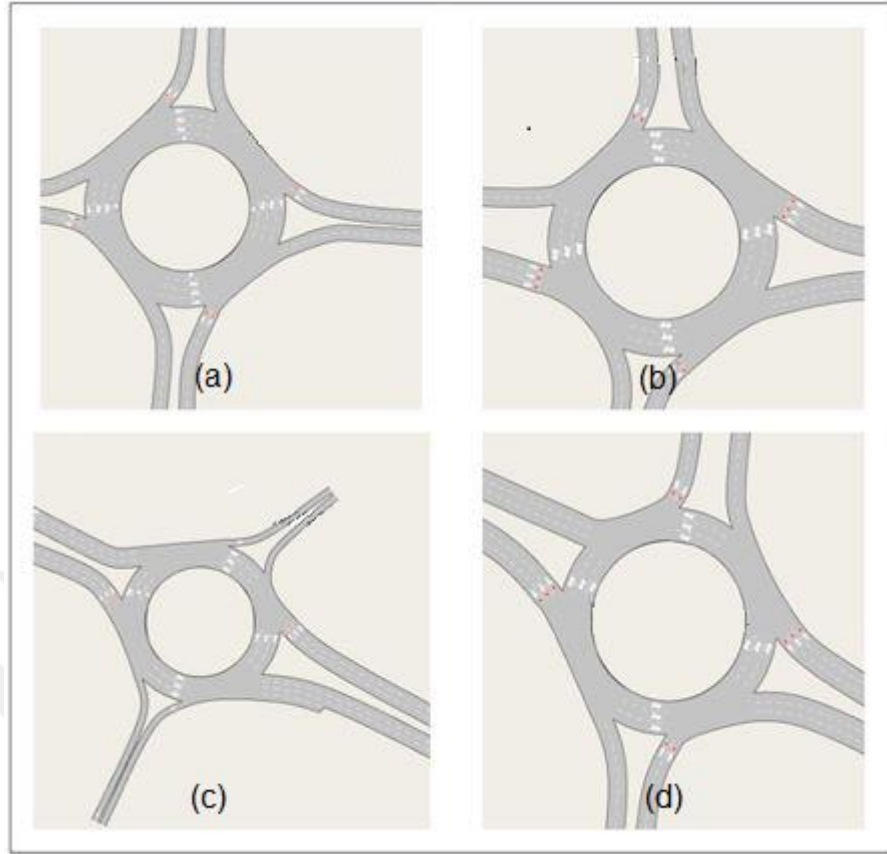
Diğer bir örnek olarak Senaryo 20.068'de ise Merplaza, Eski Terminal, Terzibaba, TEDAŞ kavşaklarında herhangi bir değişiklik olmamış; Eski Polis Okulu ve Yeni Terminal kavşakları modern dönel kavşak olarak modellenmiştir. Dört Yol Kavşağı'nda ise Batı-Doğu yönündeki sinyalizasyon süresi 5 s artırılmıştır. Trafik hacimlerinde ise otomobil hacmi aynı tutulmuş, otobüs hacmi %5 ve kamyon hacmi ise %50 artırılmıştır. Batı-Doğu istikametinde bir şerit eklenmiş, Doğu-Batı istikametinde ise mevcut durum korunmuştur.

Diğer bir örnek olarak Senaryo 27.572'teki değişikliklere bakılacak olursa; Dört yol, Terzibaba, TEDAŞ ve Yeni Terminal kavşaklarında herhangi bir değişiklik olmamış; Eski Terminal ve Eski Polis Okulu kavşakları modern dönel kavşak olarak modellenmiştir. Trafik hacimlerinde otomobilde %5, otobüste %10 ve kamyonunda %100 artış yapılmıştır.

4.1.5.1. Kavşaklarda geometrik değişiklikler

Mevcut durum trafik ağında yoğun trafiğe sahip ve geometrik olarak koşulları sağlayan Dört yol, Eski Terminal, Eski Polis Okulu ve TEDAŞ kavşakları kavşak ada çapları 25 m olacak şekilde modern dönel kavşak olarak tasarlanmıştır. Yeni Terminal Kavşağı ise geometrik olarak zaten modern kavşak örneğini karşıladığından sadece sinyalizasyonu kaldırılarak bu kavşağa kontrollü geçiş tanımlanmıştır.

Terzibaba Kavşağı'nda ise en yüksek trafik hacme sahip Doğu-Batı yönünde 2x2 şeritli olarak bir altgeçit modellenmiştir.



Şekil 4.21. Kavşakların modern döneel kavşak olarak modellenmesi (a) Dört yol Kavşağı (b) Eski Terminal Kavşağı (c) Eski Polis Okulu Kavşağı (d) TEDAŞ Kavşağı

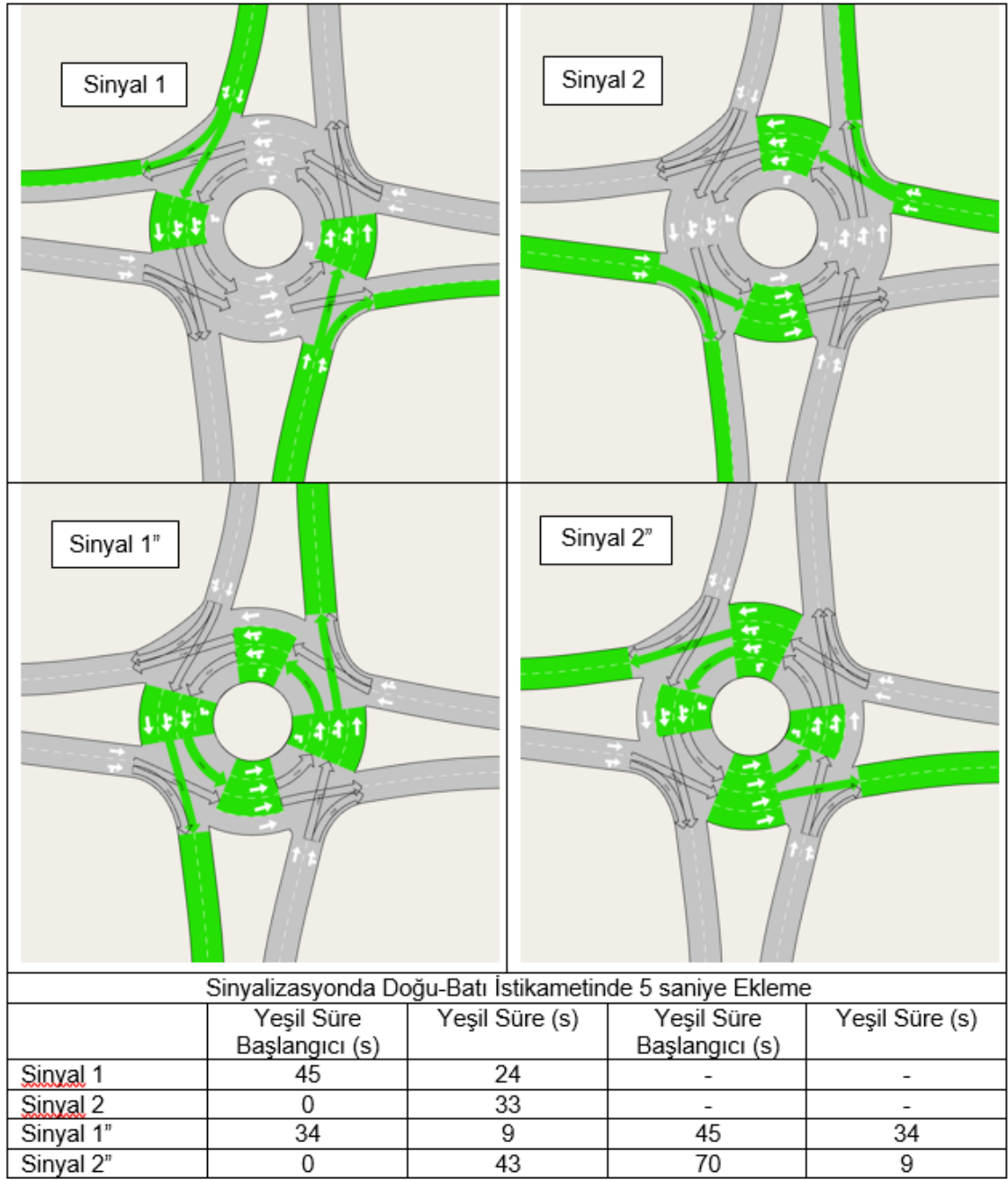


Şekil 4.22. Terzibaba Kavşağı altgeçit modellemesi

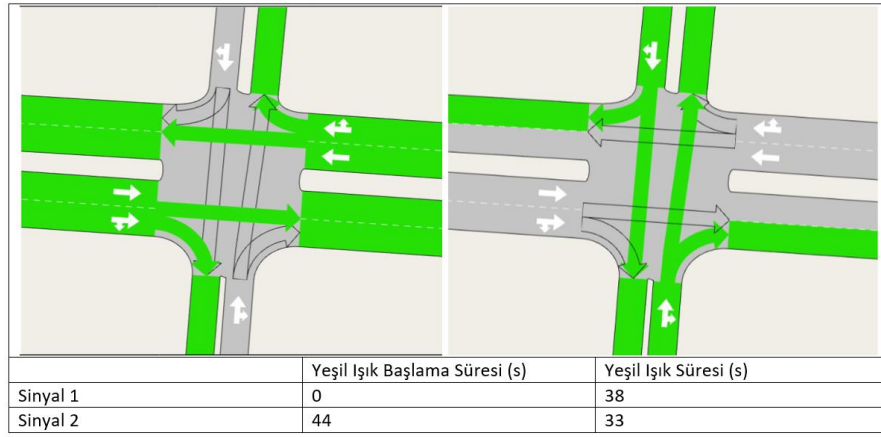
4.1.5.2. Kavşaklarda mevcut sinyalizasyon sürelerinde yapılan değişiklikler

Bu çalışmada Dört Yol Kavşağı, Merplaza Kavşağı ve Eski Polis Okulu Kavşağı için doğu-batı yöndeki faz süreleri 5 saniye artırılmış, böylece periyot süreleri de 5 saniye artmıştır. Dört Yol Kavşağı'nın periyot süresi 76 saniye iken 81 saniye, Merplaza Kavşağı'nın periyot süresi 78 saniye iken 83 saniye ve Eski Polis Okulu Kavşağı'nın periyot süresi 98 saniye iken 103 saniye olarak değiştirilmiştir. Şekil 4.23-4.25'te bu değişiklikler belirtilmiştir.

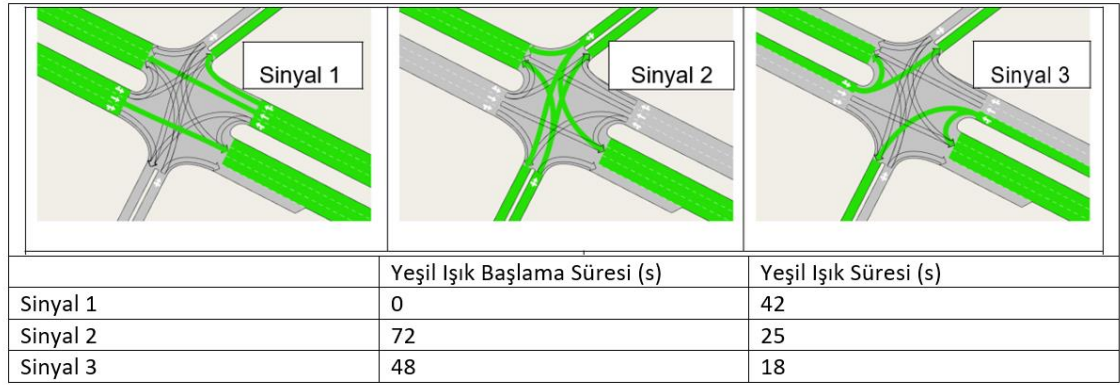




Şekil 4.23. Dört yol Kavşağı sinyalizasyon düzenlemesi sonucu faz süreleri



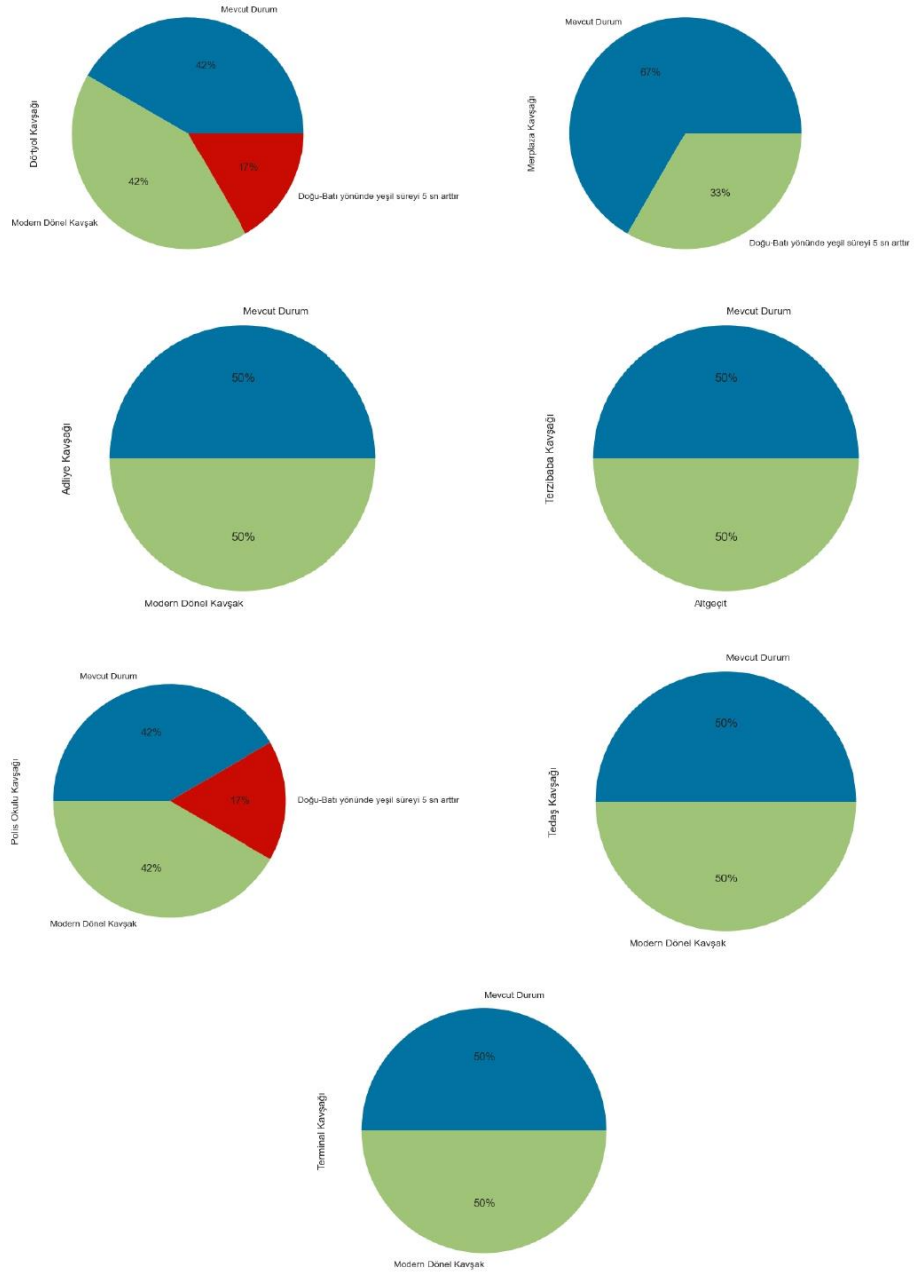
Şekil 4.24. Merplaza Kavşağı sinyalizasyon düzenlemesi sonucu faz süreleri



Şekil 4.25. Eski Polis Okulu Kavşağı sinyalizasyon düzenlemesi sonucu faz süreleri

4.1.5.3. Senaryoların istatistiksel dağılımı

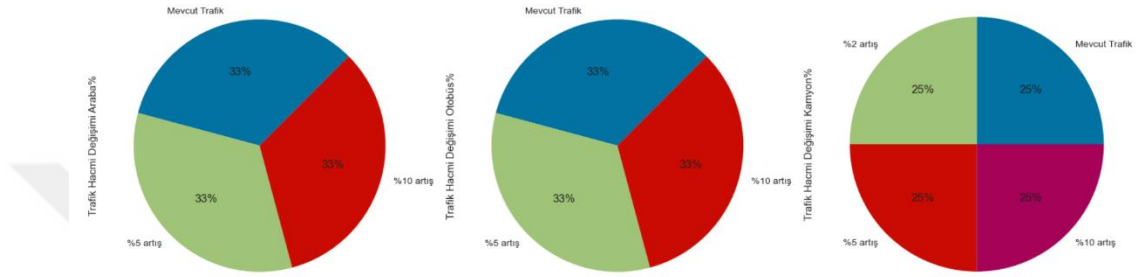
Şekil 4.26’da her kavşak için geometrik değişikliklere ve sinyalizasyondaki değişikliklere ait dairesel grafik verilmiştir.



Şekil 4.26. Kavşaklardaki geometrik ve sinyalizasyon değişikliklerine göre senaryoların dağılımı

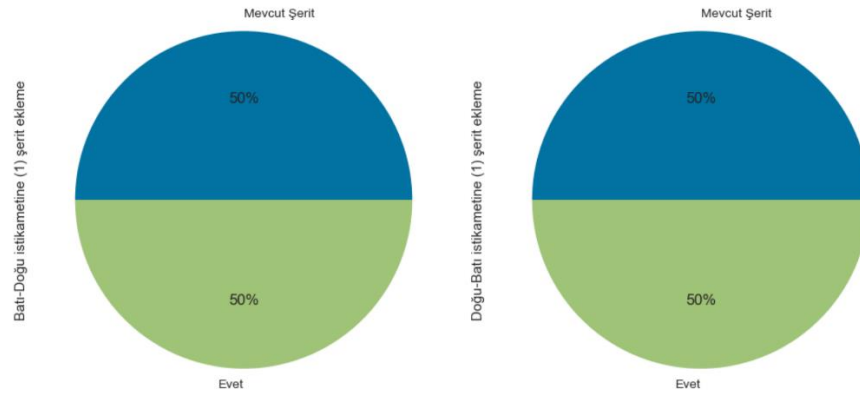
Şekil 4.26'da görüldüğü gibi, sadece modern dönel kavşak veya alt geçit tasarımı eklenen kavşaklara ait veriler, “mevcut durum” ve “modern dönel kavşak / alt geçit” olarak yarı

yarıya dağılmıştır. Bu kavşaklar Adliye Kavşağı, Terzibaba Kavşağı, Tedaş Kavşağı ve Terminal Kavşağı'dır. Dört yol Kavşağı'nda ve Eski Polis Okulu Kavşağı'nda ise veriler %42 mevcut durum, %42 modern dönel kavşak ve %16 doğu-batı istikamette yeşil süreye 5 saniye eklenen senaryolardan oluşmaktadır. Merplaza Kavşağı için dağılım ise %67 mevcut durum, %33 doğu-batı istikamette yeşil süreye 5 saniye eklenen senaryolardan oluşmaktadır.



Şekil 4.27. Veri setindeki taşıt hacimlerinin dağılımlarını belirten dairesel grafikler

Şekil 4.27’de ise otomobil, otobüs ve kamyon hacimleri ile ilgili değişkenlere ait dairesel grafik verilmiştir. Bu şekilde görülmektedir ki veri setinde taşıt hacimlerindeki değişiklikler eşit oranda dağılım göstermektedir.



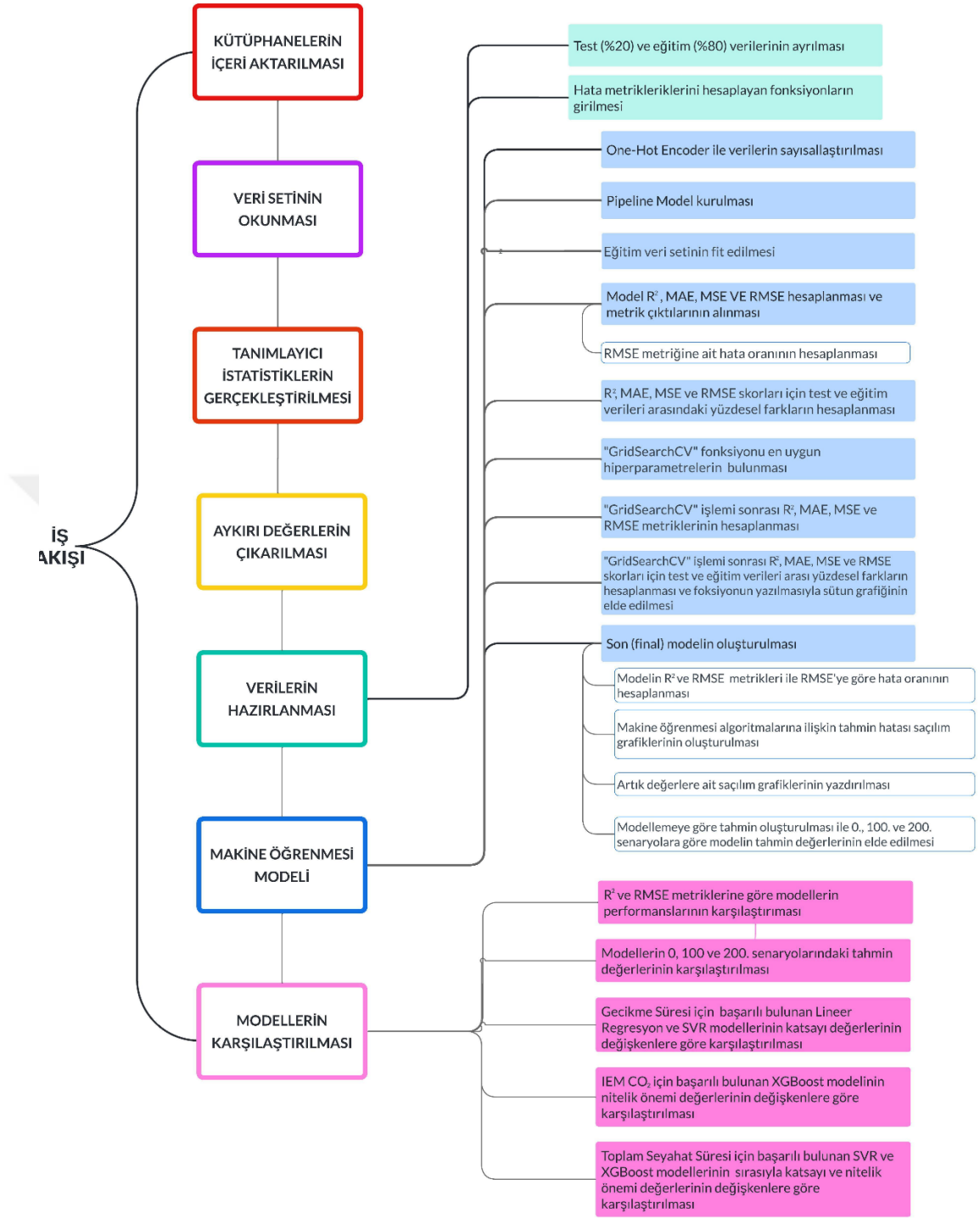
Şekil 4.28. Veri setindeki senaryoların doğu-batı ve batı-doğu yönündeki şerit sayısına göre dağılımı belirten dairesel grafikler

Şekil 4.28’e bakıldığında ise senaryolar iki istikamet için de şerit sayısına göre eşit dağılmışlardır.

4.2. Yöntem

Çalışmada Lineer Regresyon, Destek Vektör Regresörü, Karar Ağacı, Rastgele Orman, AdaBoost, Gradyan Artırma, XGBoost metotlarıyla tahmin modellemeleri oluşturulmuş, bu modellemeleri oluştururken ve grafikler hazırlanırken Numpy, Pandas, Matplotlib, Scikit-Learn ve Seaborn kütüphaneleri kullanılmıştır. Python ortamında gecikme süresi, IEM CO₂ ve toplam seyahat süresi parametreleri için bahsi geçen makine öğrenme algoritmaları kullanılarak modelleme yapıldığı için bu üç parametre üzerinde yapılan işlemler açıklanacak ve değerlendirmeler detaylandırılacaktır. Belirtilen bu işlemler her parametre için ayrı ayrı yapılacaktır.

Çalışmadaki her parametre için de geçerli olan iş akışı Şekil 4.29'da verilmiştir.



Şekil 4.29. Makine öğrenmesi model akış diyagramı

4.2.1. Kütüphanelerin ve Fonksiyonların İçeri Aktarılması

Python ile makine öğrenimi algoritmalarının modellenmesinde, öncelikle model için oluşturulan “.ipynb” uzantılı dosyaya gerekli fonksiyonları sağlayan kütüphaneleri içe aktarmak gereklidir. Bu amaçla belirtilen dosyalara NumPy, Pandas, Matplotlib, Scikit-Learn ve Seaborn kütüphaneleri dâhil edilmiştir.

Ayrıca, verileri eğitim ve test olarak ayırmak için kullanılan; *train_test_split* fonksiyonu, sırasıyla Lineer Regresyon için kullanılan *LinearRegression* fonksiyonu ve model performansını değerlendirmek için kullanılan metrik fonksiyonları olan *mean_absolute_error*, *mean_squared_error* ve *r2_score* gibi fonksiyonlar Scikit-Learn kütüphanesinden içeri aktarılmıştır.

Çapraz doğrulama yapmak için kullanılan *cross_validate* ve *cross_val_score* fonksiyonları, hiperparametre optimizasyonu için kullanılan GridSearchCV fonksiyonu Scikit-Learn kütüphanesinden içeri aktarılan diğer fonksiyonlardır.

ColumnTransformer ve *Pipeline*, veri işleme ve modelleme için bir dizi dönüştürme işlemi yürütmek üzere bir arada kullanılan komutlardır. *OneHotEncoder*, kategorik nitelikleri sayısal verilere dönüştürmek için kullanılır. Belirtilen tüm bu fonksiyonların içe aktarımı yapılmıştır.

Girdi özelliklerinin veri tiplerine dayalı olarak dönüştürücüleri otomatik olarak çıkarmak suretiyle sütun dönüştürücüsü oluşturma işlemini basitleştirmeye yarayan *make_column_transformer* içe aktarılan bir diğer fonksiyondur.

Son olarak, *yellowbrick* kütüphanesinin bazı modülleri, olan *PredictionError*, *RadViz* ve *ResidualsPlot* fonksiyonları dâhil edilmiştir. Bu araçlar modelin tahmin etme gücünü ve hatalarını görsel olarak analiz etmek için kullanılan fonksiyonlardır.

4.2.2. Gecikme Süresi

4.2.2.1. Açıklayıcı istatistiklerin hesaplanması ve analizi

Modellemeler oluşturulurken öncelikle, AIMSUN kullanılarak oluşturulan senaryolarla elde edilen veri seti Python ortamına aktarılmış ve bu verilerin ortalaması, standart

sapması, minimum değeri, maksimum değeri, medyan değeri (%50), birinci kartil (%25) ve üçüncü kartil (%75) değerleri gibi açıklayıcı istatistikler tespit edilmiştir.

Tablo 4.12. Gecikme süresi parametresi için tanımlayıcı istatistikler

	n	M	std	Min.	25%	50%	75%	Maks.
Gecikme Süresi	27648,0	75,802	15,739113	37,22	66,2375	76,395	88,56	169,53

Tablo 4.12'ye göre gecikme süresi parametresine ait 27,648 değerın ortalaması 75,802'dir ve medyan (%50) 76,395 değeri ile yakınlık göstermesi veri setinin bu parametre için normal bir dağılıma sahip olduğu söylenebilir.

Bir veri setinden;

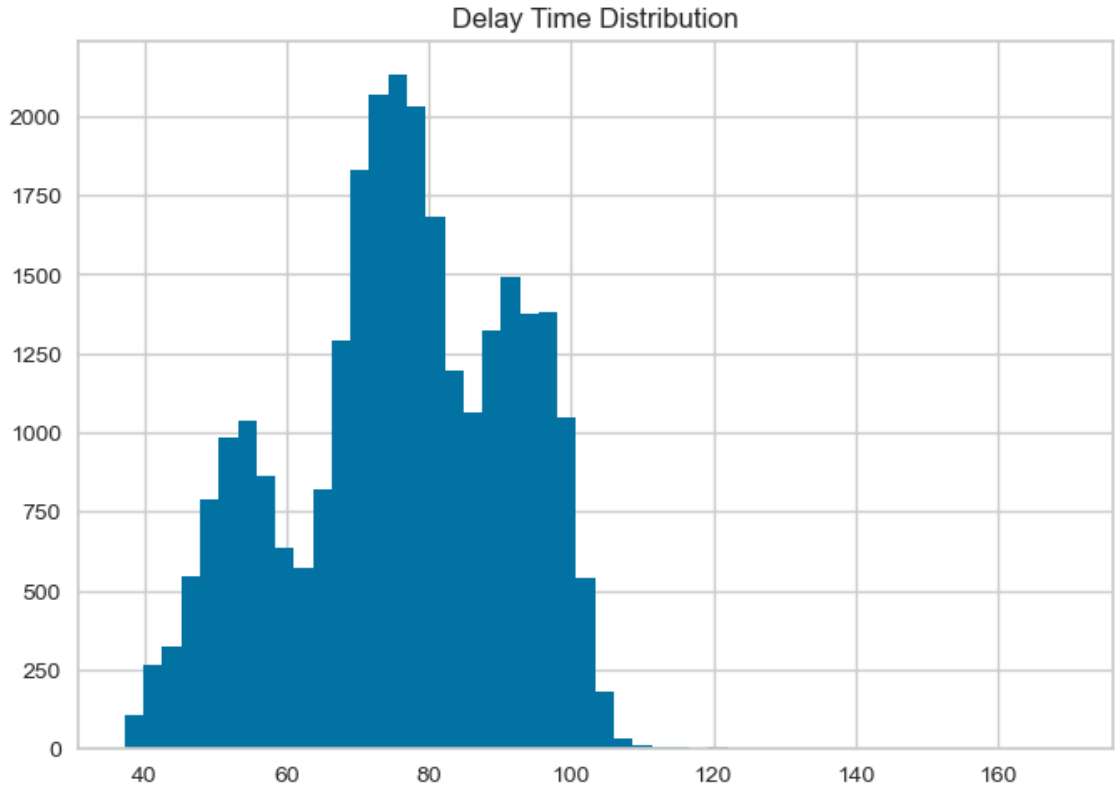
$$M - 3 * std < 3 * std < M + 3 * std \quad (4.1)$$

Koşulunu karşılaması beklenir ve bu koşulu sağlamayan değerler aykırı (outlier) değer olarak tabir edilir (Şener, 2020).

Bu parametre için bu koşul kontrol edildiğinde;

$$M - 3 * std = 28,58423 < 3 * std = 47,217339 < M + 3 * std = 123,018908 \quad (4.2)$$

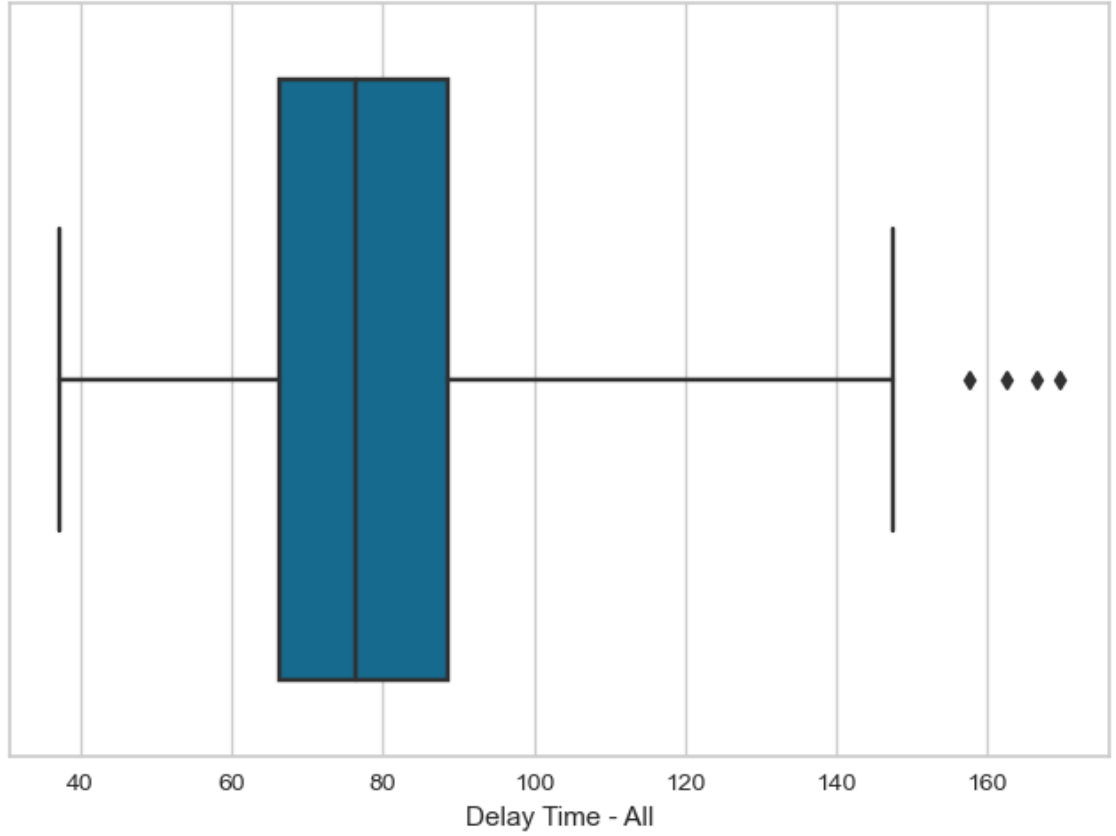
Durumu ile Denklem 4.1'deki koşul sağlandığından dolayı aykırı (outlier) veri analizine ihtiyaç yoktur. Ayrıca (Q1) birinci kartil ve medyan arasındaki fark ile üçüncü kartil (Q3) ve medyan arasındaki fark fazla olmamasından dolayı veriler düzgün dağılıma sahiptir denebilir.



Şekil 4.30. Gecikme süresi parametresine ait sütun dağılım grafiği

Şekil 4.30, gecikme süresi parametresi için sütun grafiğine bakıldığında ise genel olarak hafif sağdan çarpıklık görülmektedir. Grafikte farklı pik noktaları olması sebebiyle bu grafik bimodal dağılım olarak sınıflandırılmalıdır.

Denklem 4.2 ile gecikme süresi parametresinin aykırı değer veri analizine ihtiyaç duyulmadığı önceden saptandığı halde daha sağlıklı modellemeler yapmak için aykırı değerler tespit edilmiştir. Bu parametre için aykırı değerlerin sayısının 4 olduğu bulunmuştur. Bu değerler ve tanımlayıcı istatistiklere ait özellikler ile veri kümesini açıklayan kutu grafiği Şekil 4.31’de gösterilmiştir.



Şekil 4.31. Gecikme süresi parametresine ait kutu grafiği

Tablo 4.12 yorumlanırken belirtilen, 1. Kartil ve 3. Kartil değerlerinin ortanca değere (medyan) yakın olma durumu Şekil 4.31’de de açıkça görülmektedir. Bu grafikte ortanca değer ve 1.Kartil ve 3. Kartil değerlerinin minimum değere hafif meyilli olduğu yani hafif sağdan çarpık olduğu da görülmektedir. Yine de genel hatlarıyla verilerin normal dağılıma sahip olduğu söylenebilir.

4.2.2.2. Verilerin hazırlanması

Verilerin hazırlanmasında ilk aşamada “unique” (eşsiz) değerler tespit edilmiştir.

Sonrasında ise test verileri toplam veri setinin %20’sini, eğitim verileri de %80’ini oluşturacak şekilde ayrılmış; ayrılan bu verilere ait hata metriklerinin hesaplayıp sonucunu yazdıran fonksiyonlar girilmiş ve veri hazırlama aşaması sonlandırılmıştır.

```
Dört Yol Kavşağı has 3 unique value
Merplaza Kavşağı has 2 unique value
Adliye Kavşağı has 2 unique value
Terzibaba Kavşağı has 2 unique value
Polis Okulu Kavşağı has 3 unique value
Tedaş Kavşağı has 2 unique value
Terminal Kavşağı has 2 unique value
Trafik Hacmi Değişimi Araba% has 3 unique value
Trafik Hacmi Değişimi Otobüs% has 3 unique value
Trafik Hacmi Değişimi Kamyon% has 4 unique value
Batı-Doğu istikametine (1) şerit ekleme has 2 unique value
Doğu-Batı istikametine (1) şerit ekleme has 2 unique value
```

Şekil 4.32. Gecikme süresi parametresine ait unique (eşsiz) değerlerin sayısı

4.2.2.3. Makine öğrenmesi algoritmalarının modellenmesi

Bütün makine öğrenmesi modellerinde aşağıdaki işlemler yapılmıştır.

- One-Hot Encoder işlemi yapılarak veriler sayısallaştırıldı.
- Pipeline Model kuruldu.
- Veri setinin %80 ile oluşturulan eğitim (train) veri seti fit edildi.
- Modelin R^2 , MAE, MSE ve RMSE değerlerini hesaplayan komutlar yazıldı, bu metriklerin çıktıları alındı.
 - Metrikleri yorumlamak için hedef parametrenin ortalaması alındı.
 - RMSE metriğinin hata oranının hesabı için Denklem 3.16 kullanıldı.
- Test ve eğitim (Train) verileri arasındaki yüzdesel metrik farkları için bir fonksiyon yazıldı.
- R^2 , MAE, MSE ve RMSE skorları arasındaki yüzdesel farklar hesaplandı.
- GridSearchCV işlemi uygulandı ve en uygun parametreler bulundu.
- GridSearchCV işlemi sonrası R^2 , MAE, MSE ve RMSE metrikleri tekrar hesaplandı.
- GridSearchCV işlemi sonrası R^2 , MAE, MSE ve RMSE skorları arasındaki yüzdesel farklar hesaplandı ve fonksiyon yazılarak sütun grafiği elde edildi.
- Son (Final) Model kuruldu,
 - Bu modelin R^2 ve RMSE metrikleri hesaplandı. RMSE'ye göre hata oranı hesaplandı,
 - Modelin ortalaması (mean) hesaplandı.

- Makine öğrenmesi algoritmalarına ait tahmin hatasını belirten ve aşırı uyum (overfitting) olup olmadığının kontrolünde kullanılan nokta saçılım grafiği çizdirildi.
- Artık (residuals) değerlere ait saçılım grafiği çizildi
- Modellemeye göre tahmin yapıldı. 0., 100. Ve 200. örnek senaryolara göre modelin tahmin ettiği değerler yazdırıldı.

4.2.3. IEM CO₂

4.2.3.1. Açıklayıcı istatistiklerin hesaplanması ve analizi

Bu parametrede de verilerin ortalaması, standart sapması, minimum değeri, maksimum değeri, ortanca değeri (%50), birinci kartil (%25) ve üçüncü kartil (%75) değerleri gibi açıklayıcı istatistikler tespit edilmiştir.

Tablo 4.13 IEM CO₂ için tanımlayıcı istatistikler.

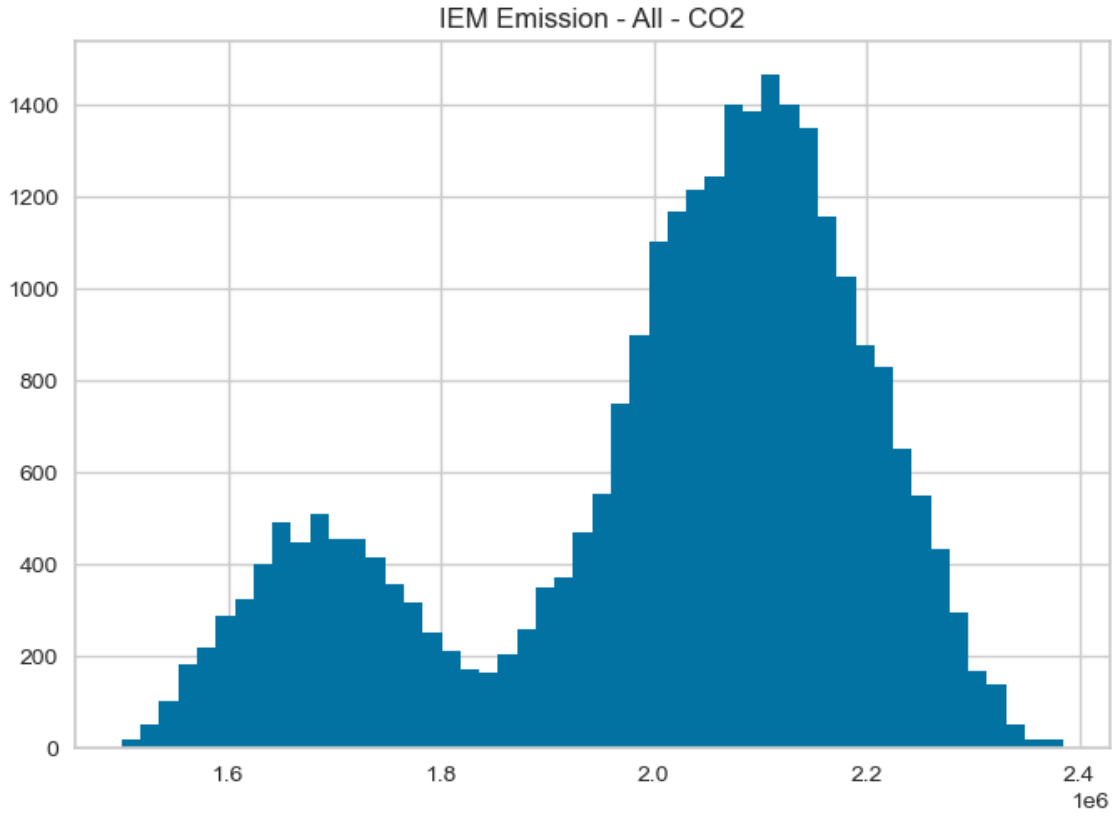
	n	M	std	Min.	25%	50%	75%	Mals.
IEM CO ₂	27648,0	2,01e+06	196792,430892	826,37	1919507,505	2,06e+06	2144987,94	2384365,83

IEM CO₂ parametresine ait 27.648 değerlerin ortalaması 2,01e+06'dır ve medyan (%50) 2,06e+06 değeri ile yakınlık göstermemesi veri setinin bu parametre için normal bir dağılıma sahip olmadığını söyleyebilir.

Bu parametrenin Denklem 4.1.'deki koşulları sağlayıp sağlamadığı kontrol edilirse;

$$mean - 3 * std = 1417303,807 > 3 * std = 590378,1927 \quad (4.3)$$

Olduğundan bu parametre aykırı değerlere sahiptir. Buna rağmen medyan ve ikinci kartil arasındaki fark ile medyan ve üçüncü kartil arasındaki fark çok fazla değildir.

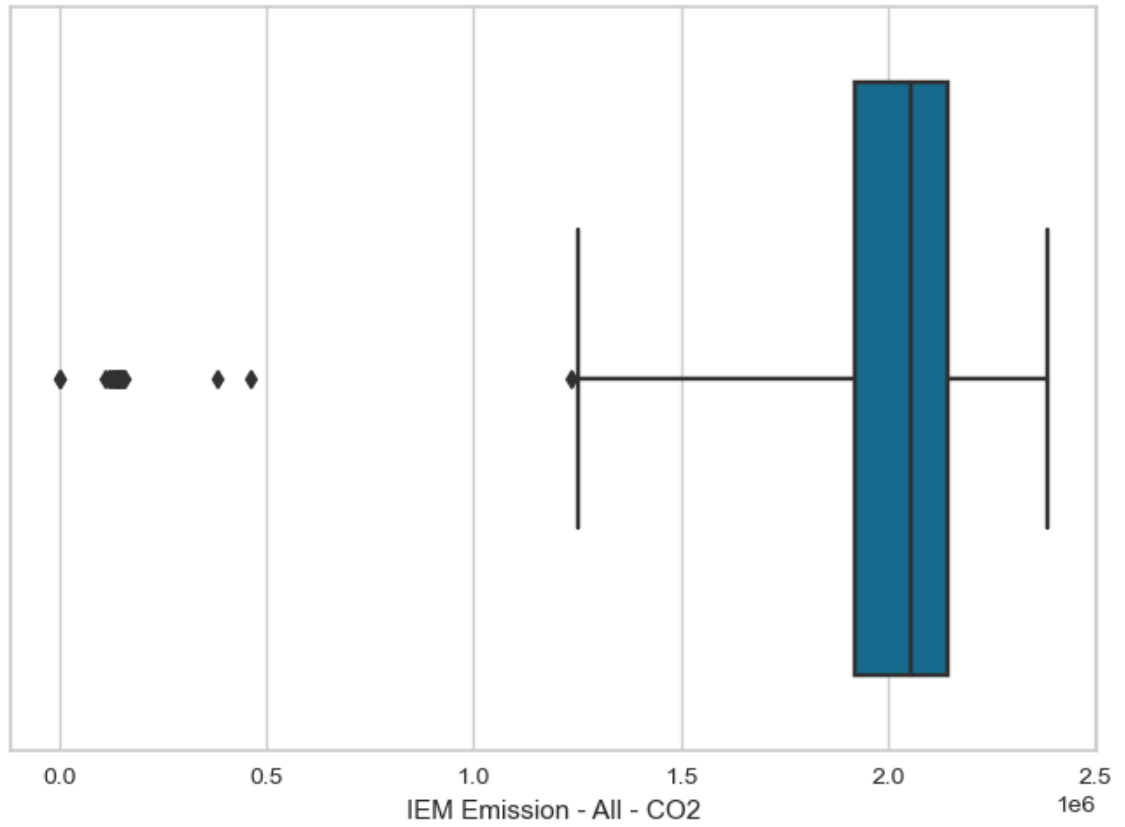


Şekil 4.33. IEM CO₂ parametresine ait sütun dağılım grafiği

Şekil 4.33'te veriler iki ayrı tepe noktası etrafında dağılmıştır. Bu durum bimodal bir dağılım olarak ifade edilebilir.

Ayrıca parametreye ait bu grafiğe bakılacak olursa dağılımda soldan çarpıklık olduğu (Left Skewed Distribution) görülmektedir. Yani açıkça veriler kesinlikle düzgün bir şekilde dağılmamıştır.

Denklem 4.3 ile IEM CO₂ parametresinin aykırı veri analizine ihtiyaç duyulduğu görüldüğü için, daha sağlıklı modellemeler yapmak amacıyla aykırı değerler tespit edilmiş ve bu parametre için aykırı değer sayısının 20 olduğu bulunmuştur. Bu değerleri ve tanımlayıcı istatistiklere ait özellikler ile veri kümesini açıklayan kutu grafiği Şekil 4.34'te gösterilmiştir.



Şekil 4.34. IEM CO₂ parametresine ait kutu grafiği

Tablo 4.13 yorumlanırken belirtilen 1. Kartil ve 3.Kartil değerlerinin ortanca değere (medyan) yakın olma durumu Şekil 4.34'te açıkça görülmektedir. Ayrıca ortanca değerin ve 1.Kartil ve 3. Kartil değerlerinin maksimum değere yakın olduğu, yani daha önce de belirtildiği gibi soldan çarpıklık olduğu da görülmektedir.

4.2.3.2. Verilerin hazırlanması

Verilerin hazırlanmasında ilk aşamada eşsiz (unique) değerler tespit edilmiştir.

Sonrasında ise test verileri toplam veri setinin %20'sini, eğitim verileri de %80'ini oluşturacak şekilde ayrılmış; ayrılan bu verilere ait hata metriklerini hesaplayıp sonucunu yazdıran fonksiyonlar girilmiş ve veri hazırlama aşaması sonlandırılmıştır.

```

Dört yol Kavşağı has 3 unique value
Merplaza Kavşağı has 2 unique value
Adliye Kavşağı has 2 unique value
Terzibaba Kavşağı has 2 unique value
Polis Okulu Kavşağı has 3 unique value
Tedaş Kavşağı has 2 unique value
Terminal Kavşağı has 2 unique value
Trafik Hacmi Değişimi Araba% has 3 unique value
Trafik Hacmi Değişimi Otobüs% has 3 unique value
Trafik Hacmi Değişimi Kamyon% has 4 unique value
Batı-Doğu istikametine (1) şerit ekleme has 2 unique value
Doğu-Batı istikametine (1) şerit ekleme has 2 unique value

```

Şekil 4.35. IEM CO₂ parametresine ait unique (eşsiz) değerlerin sayısı

4.2.3.3. Makine öğrenmesi algoritmalarının modellenmesi

Gecikme süresi parametresinde yapılan işlemler IEM CO₂ Parametresinde de tekrar edilmiştir.

4.2.4. Toplam Seyahat Süresi

4.2.4.1. Açıklayıcı İstatistiklerin Hesaplanması ve Analizi

Tablo 4.14'te Toplam seyahat süresi parametresi için tanımlayıcı istatistiklere ait değerler verilmiştir.

Tablo 4.14. Toplam seyahat süresi için tanımlayıcı istatistikler

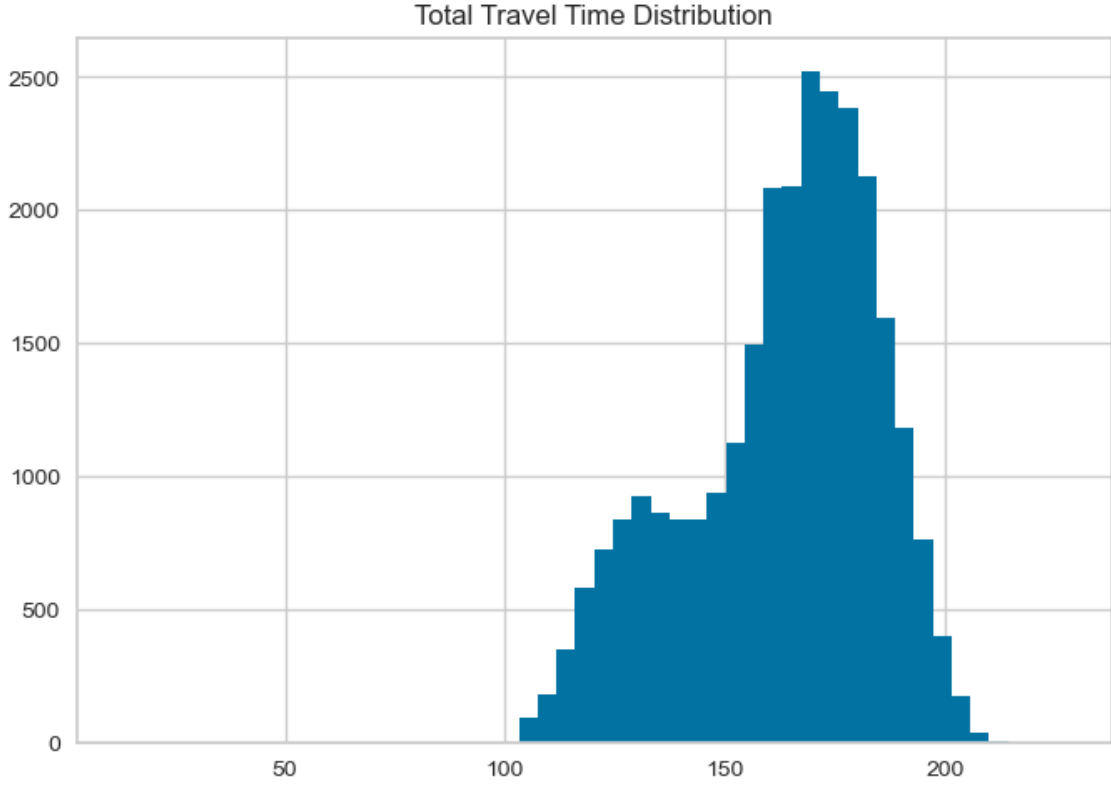
	n	M	std	Min.	25%	50%	75%	Maks.
Toplam Seyahat Süresi	27648,0	162,72	21,884862	13,31	148,84	166,92	178,92	227,12

Toplam seyahat süresi parametresi dağılımına bakılacak olursa; ortalaması (mean) 162,72 ve medyan değeri 166,92'dir. Değerler birbirlerine yakın olduğu için düzgün bir dağılıma sahip bir parametre olarak varsayılabilir.

Bu parametrenin Denklem 4.1.'deki koşulları sağlayıp sağlamadığı kontrol edilirse;

$$M - 3 * std = 97,060515 > 3 * std = 65,654586 \quad (4.4)$$

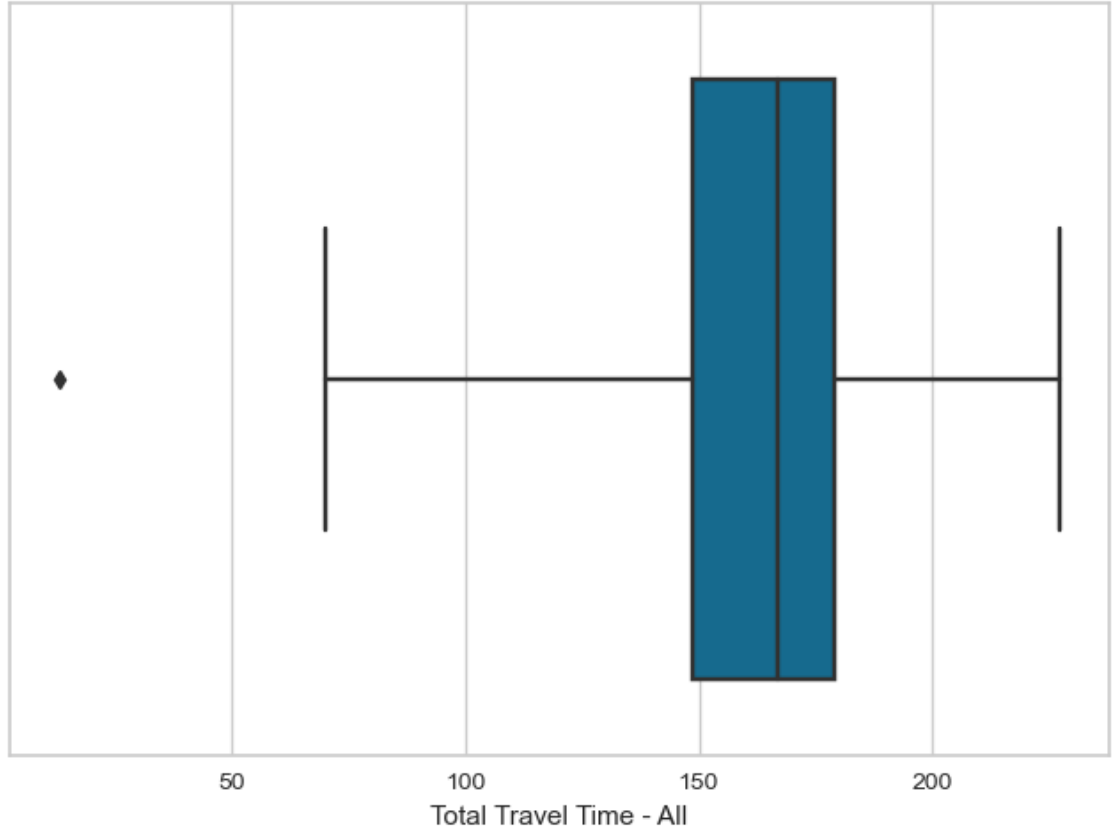
Olduğundan bu parametre aykırı değerlere sahiptir. Buna rağmen medyan ve birinci kartil arasındaki fark ile medyan ve üçüncü kartil arasındaki fark çok fazla değildir.



Şekil 4.36. Toplam seyahat süresi parametresine ait sütun dağılım grafiği

Şekil 4.36’da verilen parametreye ait sütun dağılım grafiği analiz edilecek olursa soldan çarpıklık (Left Skewed Distribution) görülmektedir. Yani açıkça veriler normal bir dağılıma sahip değildir.

Denklem 4.4 ile toplam seyahat süresi parametresinin aykırı değer veri analizine ihtiyaç duyulduğu bulunduğundan daha sağlıklı modellemeler yapmak için aykırı değerler tespit edilmiştir. Bu parametre için aykırı değer sayısının 1 olduğu bulunmuştur. Aykırı değeri ve tanımlayıcı istatistiklere ait özellikler ile veri kümesini açıklayan kutu grafiği Şekil 4.37’de gösterilmiştir.



Şekil 4.37. Toplam seyahat süresi parametresine ait kutu grafiği

Tablo 4.14 yorumlanırken belirtilen, 1. Kartil ve 3.Kartil değerlerinin ortanca değere (medyan) yakın olma durumu Şekil 4.37’de de açıkça görülmektedir. Ayrıca ortanca değerin ve 1. Kartil ve 3. Kartil değerlerinin maksimum değere yakın olduğu, yani bariz soldan çarpıklık olduğu da görülmektedir.

4.2.4.2.Verilerin hazırlanması

Verilerin hazırlanmasında ilk aşamada “unique” (eşsiz) tabir edilen değerler tespit edilmiştir.

Sonrasında ise test verileri toplam veri setinin %20’sini, eğitim verileri de %80’ini oluşturacak şekilde ayrılmış; ayrılan bu verilere ait hata metriklerini hesaplayıp sonucunu yazdıran fonksiyonlar girilmiş ve veri hazırlama aşaması sonlandırılmıştır.

```
Dört Yol Kavşağı has 3 unique value
Merplaza Kavşağı has 2 unique value
Adliye Kavşağı has 2 unique value
Terzibaba Kavşağı has 2 unique value
Polis Okulu Kavşağı has 3 unique value
Tedaş Kavşağı has 2 unique value
Terminal Kavşağı has 2 unique value
Trafik Hacmi Değişimi Araba% has 3 unique value
Trafik Hacmi Değişimi Otobüs% has 3 unique value
Trafik Hacmi Değişimi Kamyon% has 4 unique value
Batı-Doğu istikametine (1) şerit ekleme has 2 unique value
Doğu-Batı istikametine (1) şerit ekleme has 2 unique value
```

Şekil 4.38. Toplam seyahat süresi parametresine ait “unique” (eşsiz) verilerin sayısı

4.2.4.3. Makine öğrenmesi algoritmalarının modellenmesi

Gecikme süresi parametresinde yapılan işlemler toplam seyahat süresi parametresinde de tekrar edilmiştir.

5. BULGULAR

5.1. Gecikme Süresi Parametresi İçin Bulgular

5.1.1. GridSearchCV Öncesi ve Sonrası Performans Değerlendirme

Tablo 5.1. Gecikme süresi parametresi için GridSearchCV öncesi performans değerleri

	R ² Eğitim	R ² Test	MAE Eğitim	MAE Test	MSE Eğitim	MSE Test	RMSE Eğitim	RMSE Test
Lineer Regresyon	0,9423	0,9483	1,8866	1,8724	14,2697	12,8508	3,7775	3,5848
SVR	0,9556	0,9616	0,9180	0,9591	10,9784	9,5433	3,3134	3,0892
Karar Ağacı	1,0	0,9514	0,0	1,2884	0,0	12,0936	0,0	3,4776
Rastgele Orman	0,9954	0,9708	0,3974	1,0915	1,1356	7,2623	1,0656	2,6949
AdaBoost	0,7675	0,7700	5,1920	5,2336	57,5443	57,2086	7,5858	7,5636
Gradyan Artırma	0,9552	0,9605	1,4484	1,4480	11,0862	9,8289	3,3296	3,1351
XGBoost	0,9549	0,9603	1,4495	1,4490	11,1521	9,8838	3,3395	3,1439

Tablo 5.1’de tüm algoritmalarda hem test hem de eğitim verileri için R² skorlarının 1’e yakın olduğu saptanmıştır. Değerler gayet makul bir aralıkta çıkmıştır. AdaBoost R² skorları 0,7 aralığında hesaplanmıştır. Bu modelin diğer algoritmalarından daha az bir değerde çıktığını görülmüştür. Karar Ağacı modelinde ise bu değer 1,0 çıkmıştır. Sonuçlar, modelin genelleme özelliğini kaybettiğini gösterebileceğinden aşırı uyum durumuna işaret ediyor olabilir.

MAE değeri, verilerin değer büyüklüğü ile kıyaslandığında her modelleme yöntemi için tahmin değerleri ile kısmen başarılı bir model oluşturulduğu görülmektedir. Burada Karar Ağacı 0 değeri alarak aşırı uyum olasılığını yükseltmektedir.

MSE ve RMSE metrikleri, MAE gibi tahmin değerlerinin gözlenen değerlere uzaklığını temsil eder. Bu metrikler, gerçek değerlerle karşılaştırıldığında kabul edilebilir bir aralıkta çıkmıştır. Ayrıca Karar Ağacı diğer metriklerde olduğu gibi bu iki metrikte de 0 çıkmıştır. Yani aşırı uyum durumuna göre işlem yapılması gerekmektedir.

Tablo 5.2. Gecikme süresi parametresi için GridSearchCV sonrası performans değerleri

	R² Eğitim	R² Test	MAE Eğitim	MAE Test	MSE Eğitim	MSE Test	RMSE Eğitim	RMSE Test
Lineer Regresyon	0,9424	0,9485	1,8841	1,8686	14,2533	12,8069	3,7753	3,5787
SVR	0,9556	0,9616	0,9180	0,9591	10,9784	9,5433	3,3134	3,0892
Karar Ağacı	0,8977	0,9043	3,5283	3,4790	25,3279	23,7895	5,0327	4,8775
Rastgele Orman	0,8979	0,9046	3,5226	3,4725	25,2630	23,7272	5,0262	4,8711
AdaBoost	0,8488	0,8525	4,6629	4,6932	37,4268	36,6705	6,1177	6,0556
Gradyan Artırma	0,8879	0,8936	3,7587	3,7969	27,7336	26,4519	5,2663	5,1431
XGBoost	0,9189	0,9248	2,9161	2,9222	20,0751	18,7043	4,4805	4,3248

GridSearchCV ile fit edilen test ve eğitim (train) R^2 değerleri Karar Ağacı, Rastgele Orman, Gradyan Artırma ve XGBoost parametreleri için azalmıştır. Lineer Regresyon ve AdaBoost parametrelerinde GridSearchCV kullanımı ile R^2 değerinde artış gözlenmiştir. Destek Vektör Regresörü parametresi için ise bir değişiklik gözlenmemiştir.

MAE, MSE ve RMSE değerlerinde ise GridSearchCV ile Karar Ağacı, Rastgele Orman, Gradyan Artırma ve XGBoost parametreleri için artış görülmüştür. Lineer Regresyon ve AdaBoost parametrelerinde ise bir azalma görülmektedir. Destek Vektör Regresörü parametresi için bir değişiklik gözlenmemektedir.

R^2 skoru, 1'e yaklaştıkça, modelin veriyi daha iyi açıklama durumu ortaya çıkmaktadır. MAE, MSE ve RMSE ise tahmin hatalarının büyüklüğünü ölçmektedir. Bu metriklerin değerleri 0'a yaklaştığında, modelin daha doğru tahminler yaptığı ve hataları minimize ettiği gözlenmektedir.

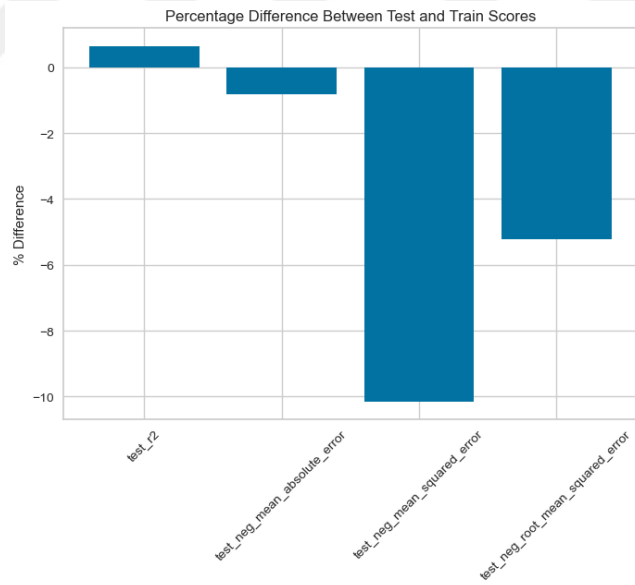
R^2 değerleri farklı bir dağılıma sahipken GridSearchCV sonrası algoritmaların R^2 değerleri birbirine yaklaşılarak 0,89-0,95 aralığında hesaplanmıştır. En yüksek performans değeri Destek Vektör Regresörü olarak tespit edilmiştir. En düşük performans değerlerinin AdaBoost'a ait olduğu görülmüştür.

5.1.2. Test ve eğitim skorları arasındaki yüzdesel farklar

Tablo 5.3. Gecikme süresi parametresi için test ve eğitim skorları arası yüzdelik fark

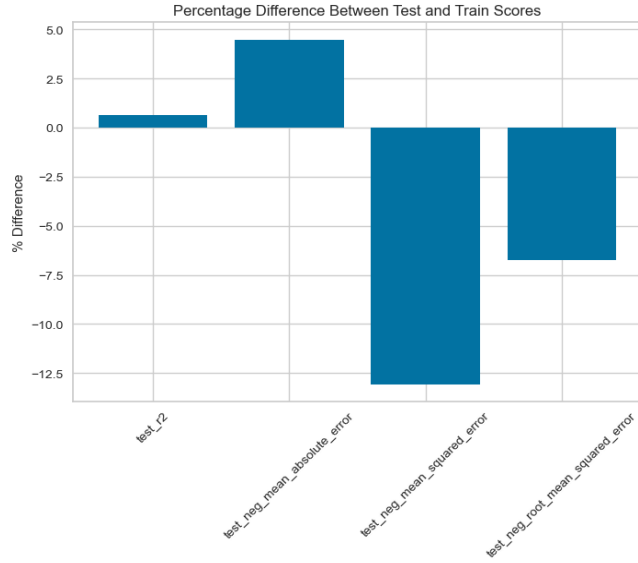
	R ² (%)	MAE (%)	MSE (%)	RMSE (%)
Lineer Regresyon	0,65	-0,83	-10,15	-5,21
SVR	0,63	4,48	-13,07	-6,76
Karar Ağacı	0,74	-1,40	-6,07	-3,08
Rastgele Orman	0,74	-1,42	-6,08	-3,09
AdaBoost	0,45	0,65	-2,02	-1,02
Gradyan Artırma	0,64	1,01	-4,62	-2,34
XGBoost	0,64	0,21	-6,83	-3,47

Test ve Eğitim skorları arasındaki yüzdesel farklar, modelde aşırı öğrenme (overfitting) ya da yetersiz öğrenme (underfitting) olup olmadığını anlamak için önemlidir. Test skorunun, eğitim skoruna göre çok daha düşük olduğu durumlarda yetersiz öğrenme söz konusu olabilmektedir. Ancak eğitim skorunun, test skoruna göre çok daha yüksek olduğu durumlarda aşırı öğrenme görülebilmektedir.



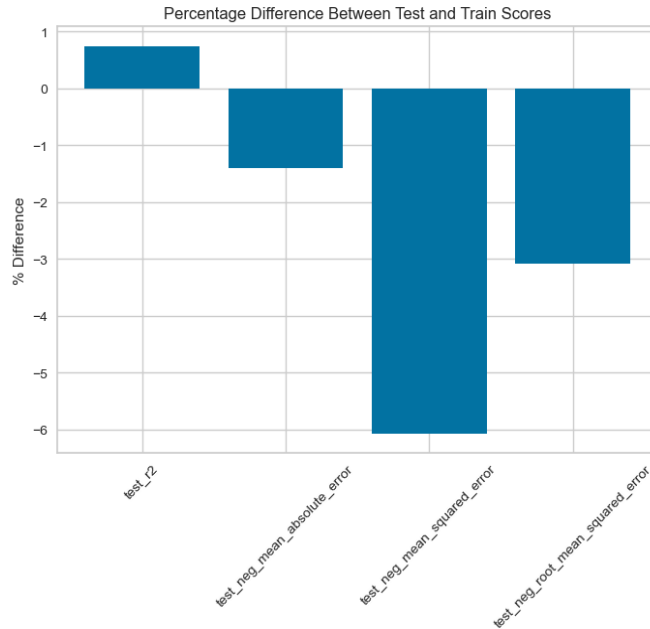
Şekil 5.1. Lineer Regresyon için gecikme süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.1’de Lineer Regresyon modeli R² metriği için test skorunun eğitim skorundan yüksek, MAE, MSE ve RMSE için ise daha düşük olduğu görülmektedir. MAE için bu değer 0’a çok yakın çıkmıştır.



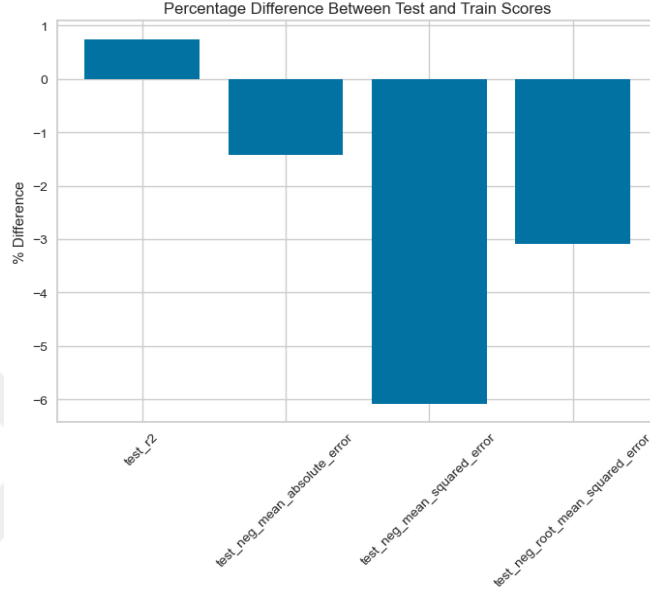
Şekil 5.2. Destek Vektör Regresörü için gecikme süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.2'ye göre R^2 ve MAE metrikleri açısından Destek Vektör Regresörü için test skorunun, eğitim skorundan yüksek olduğu görülmektedir. Ancak MSE ve RMSE için inceleme yapıldığında test skorunun bir miktar düşük olduğu görülmektedir. Ancak tam anlamıyla aşırı uyumun olduğu söylenemez.



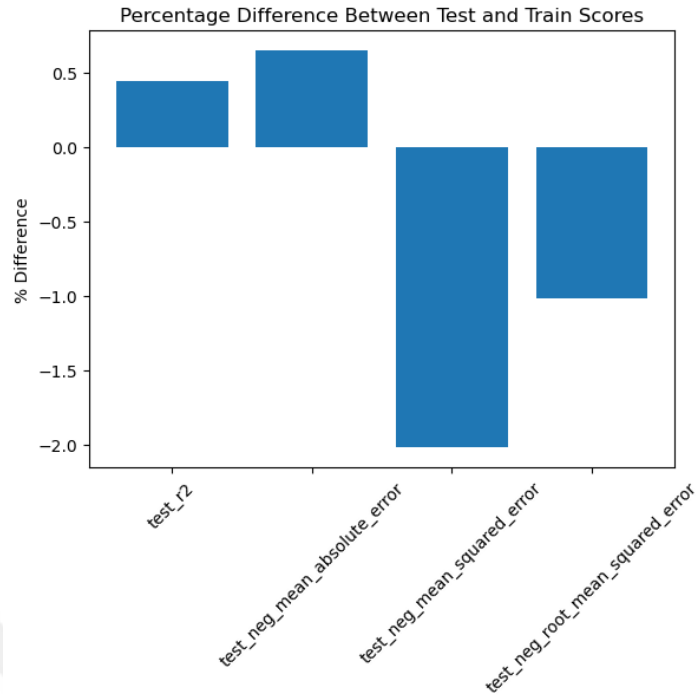
Şekil 5.3. Karar Ağacı için gecikme süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.3'e bakıldığında ise Karar Ağacı modelinde test skorunun R^2 metriği için eğitim skorundan yüksek, MAE için ise diğer modellere göre daha düşük bir değerde olduğu gözlenebilir. Ancak MSE ve RMSE için inceleme yapıldığında test skorunun nispeten düşük olduğu görülmektedir.



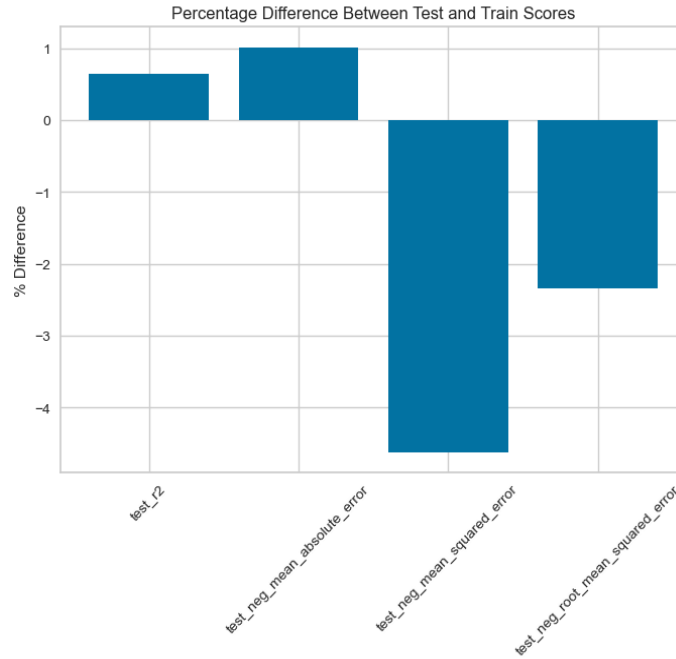
Şekil 5.4. Rastgele Orman için gecikme süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.4'te Rastgele Orman test skoru, R^2 metriği için eğitim skorundan yüksektir. MAE için test skoru, eğitim skorundan düşüktür. Ancak MSE ve RMSE için inceleme yapıldığında test skorunun, eğitim skorundan daha düşük olduğu görülmektedir.



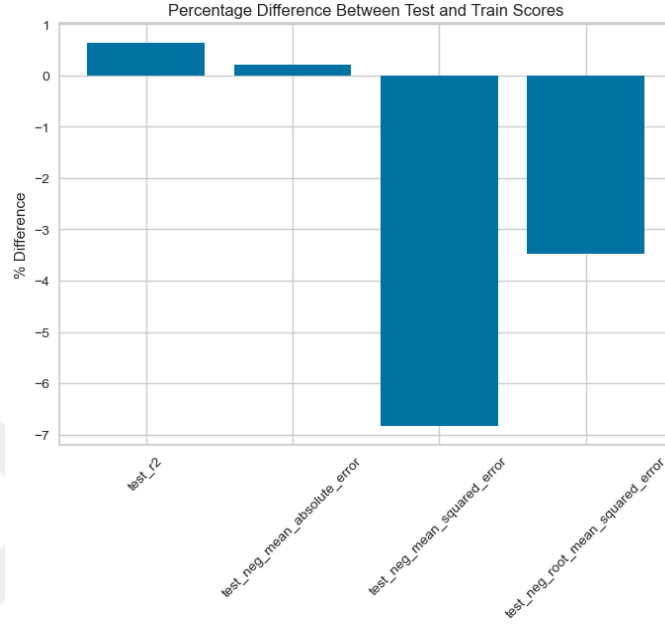
Şekil 5.5. AdaBoost için gecikme süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.5'te R^2 ve MAE metrikleri için AdaBoost için test skorunun, eğitim skorundan yüksek olduğu görülmektedir. Ancak MSE ve RMSE için inceleme yapıldığında test skorunun oldukça düşük olduğu görülmektedir.



Şekil 5.6. Gradyan Artırma için gecikme süresi eğitim ve test verileri arasında yüzdelik fark

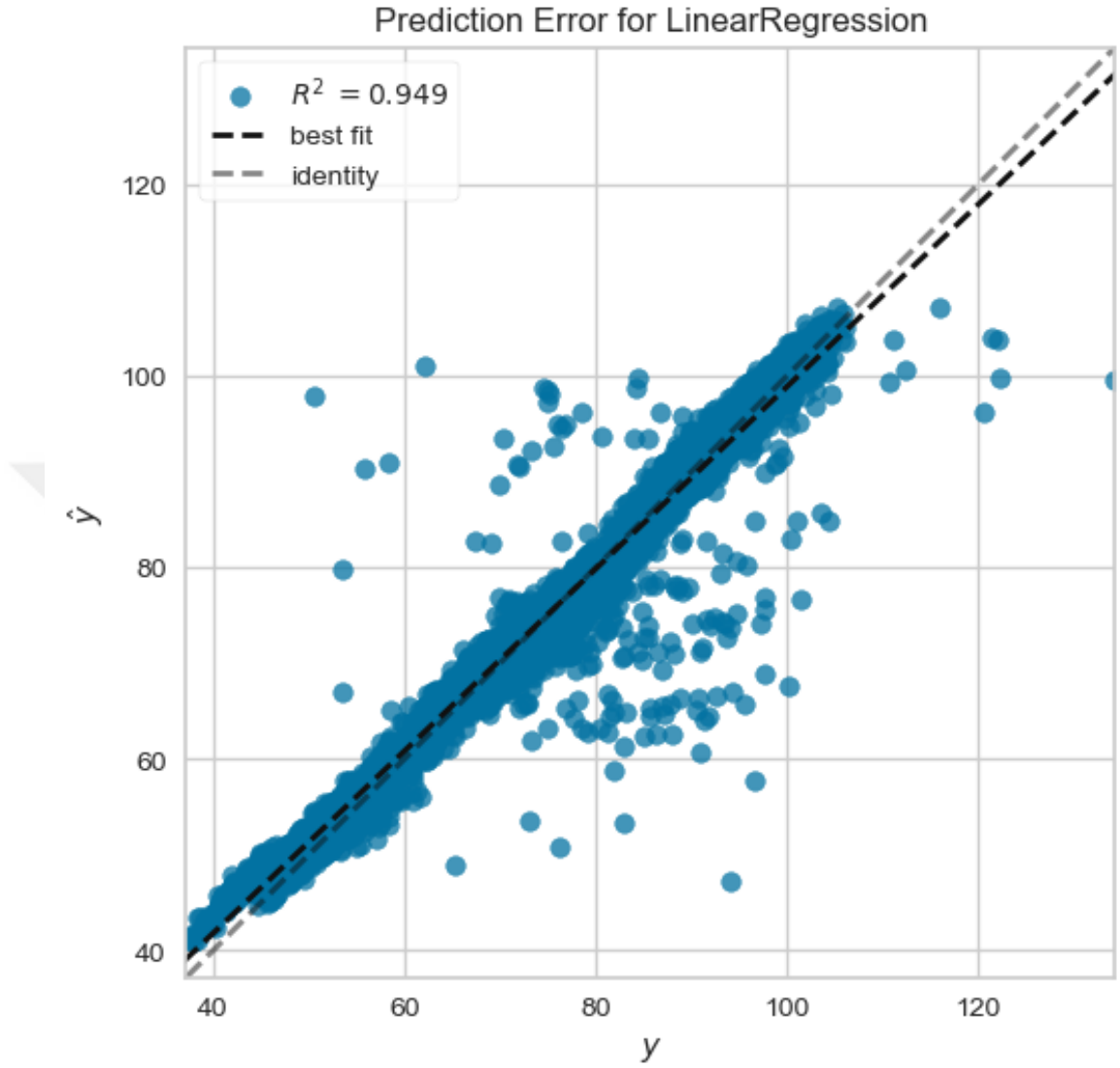
Şekil 5.6'ya göre R^2 ve MAE metrikleri için Gradyan Artırma modelinde test skorunun, eğitim skorundan yüksek olduğu görülmektedir. Ancak MSE ve RMSE için inceleme yapıldığında test skorunun oldukça düşük olduğu görülmektedir.



Şekil 5.7. XGBoost için gecikme süresi eğitim ve test verileri arasında yüzdelik fark

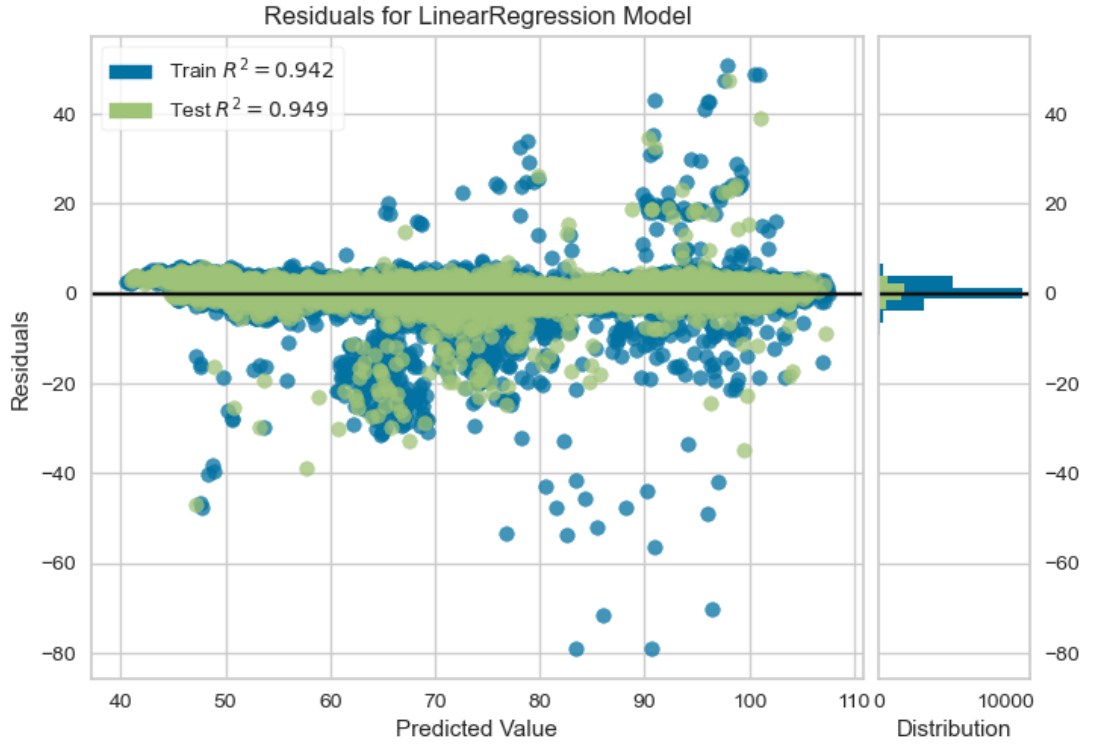
Şekil 5.7'de XGBoost modeli için R^2 ve MAE metriklerine göre test skorunun, eğitim skorundan yüksek olduğu görülmektedir. Ancak MSE ve RMSE için inceleme yapıldığında test skorunun oldukça düşük olduğu görülmektedir.

5.1.3. Final model sonrası değerlendirme



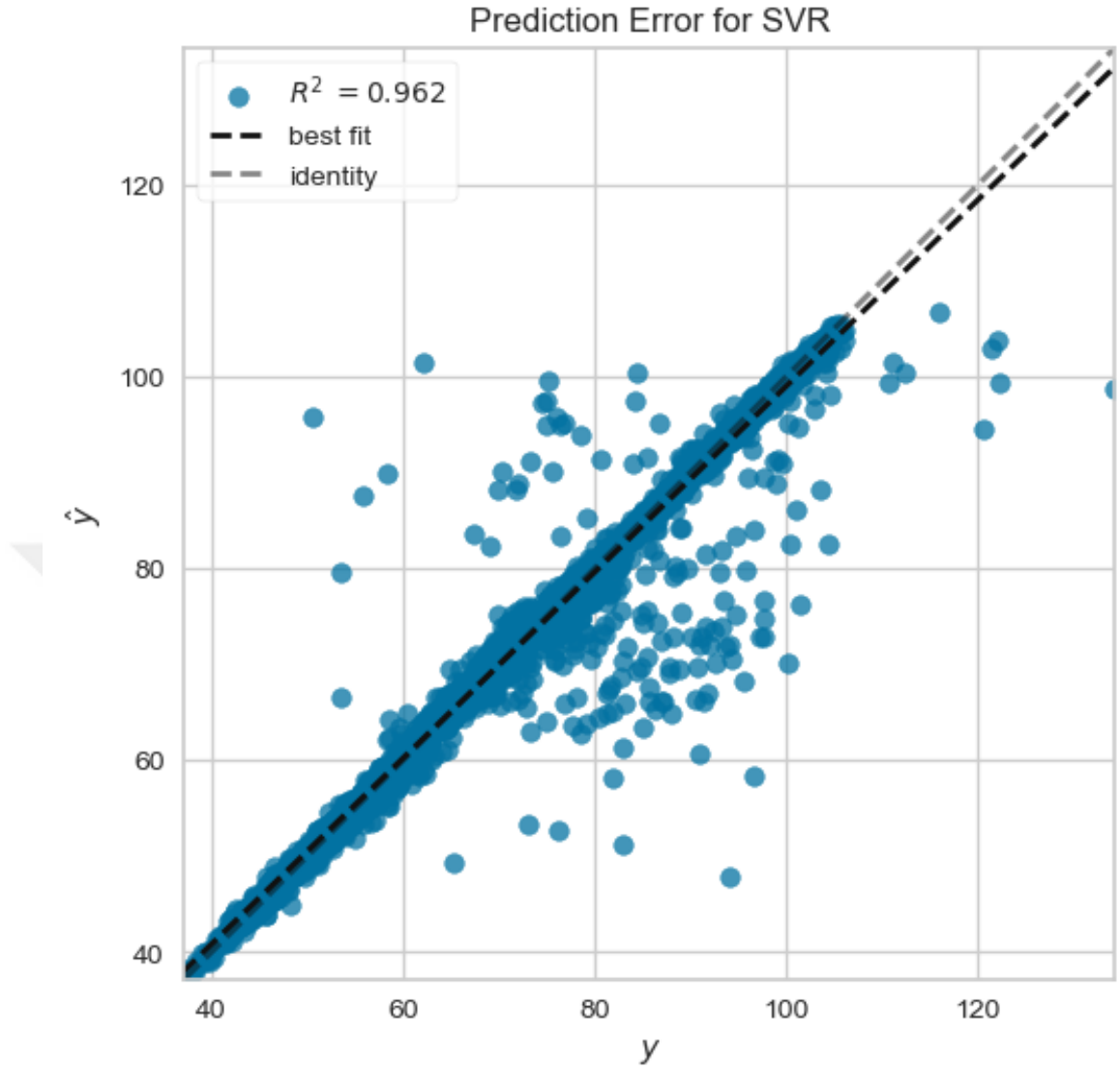
Şekil 5.8. Gecikme süresi için Lineer Regresyon modeli tahmin hatası grafiği

Şekil 5.8’de “Best Fit” eğrisi etrafında toplanan değerler, anlamlı bir Lineer Regresyon modelini ifade etmektedir. “Identity” eğrisi ise tahmin edilmiş eğriyi göstermektedir. Bu egride biraz sapma olmasına rağmen eğrinin etrafında da toplu bir şekilde saçılan değerler, tahmin modelinin %95 doğruluk oranına sahip olduğuna işaret etmektedir. Aşırı uyum durumunun bulunmadığı da söylenebilir.



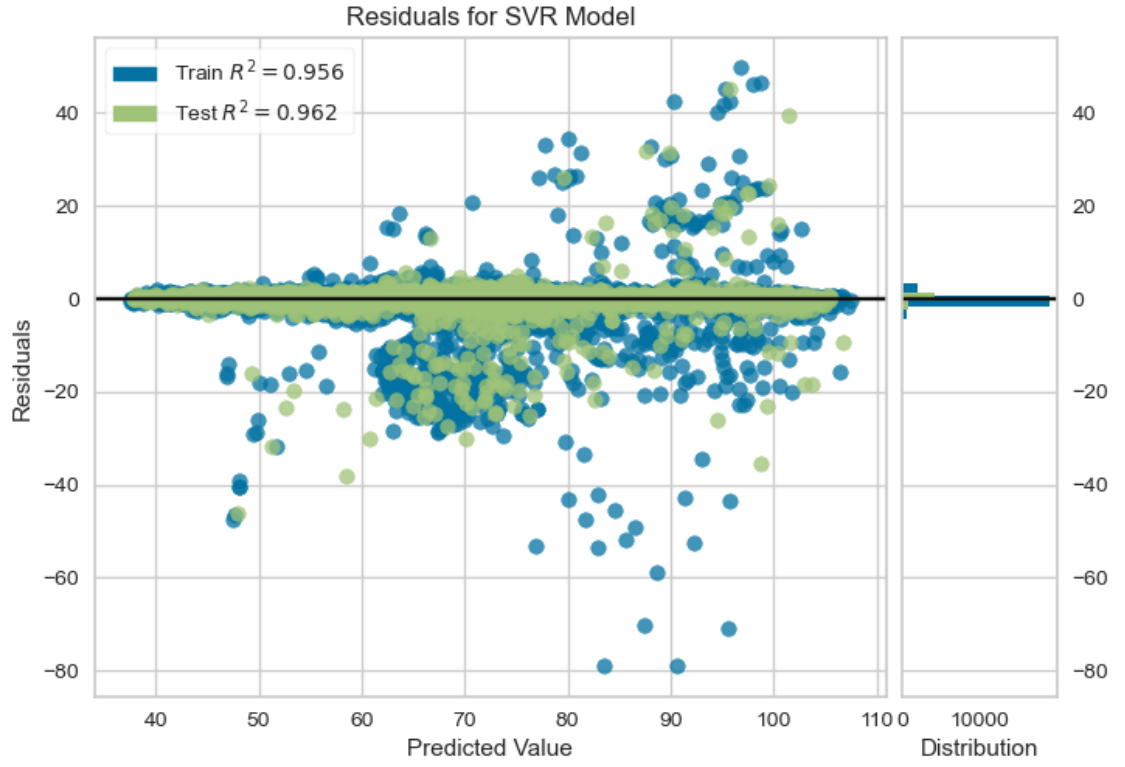
Şekil 5.9. Gecikme süresi için Lineer Regresyon modelinde artık değerler dağılımı

Şekil 5.9’da artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmında herhangi bir “pattern” yani bir desen veya kalıp olmadığı görülmektedir. Sütun dağılımında da normal dağılım bulunduğu için artık değerler açısından bu modelin uygun olduğu tespit edilmiştir.



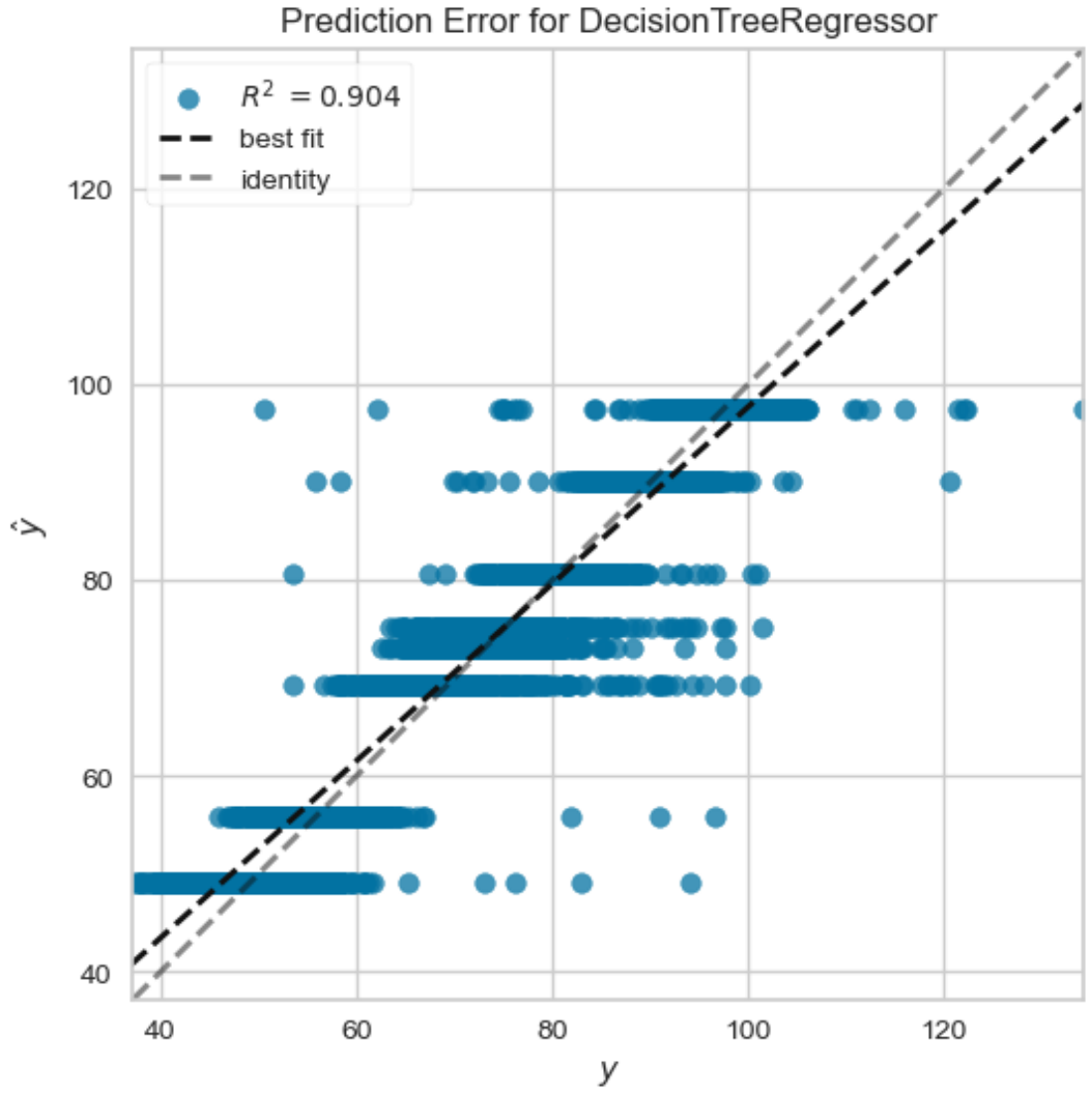
Şekil 5.10. Gecikme süresi için Destek Vektör Regresörü modeli tahmin hatası grafiği

Şekil 5.10’da ise Destek Vektör Regresörü, Lineer Regresyon Modeline göre daha fazla performans gösterdiği görülmektedir. Bu modelde R^2 0,9616, RMSE 3,0892 çıkmıştır. Denklem 5.1’e göre RMSE hata oranı hesaplandığında yaklaşık %4 (0,0407) çıkmıştır. Modelde aşırı uyum bulunmamaktadır.



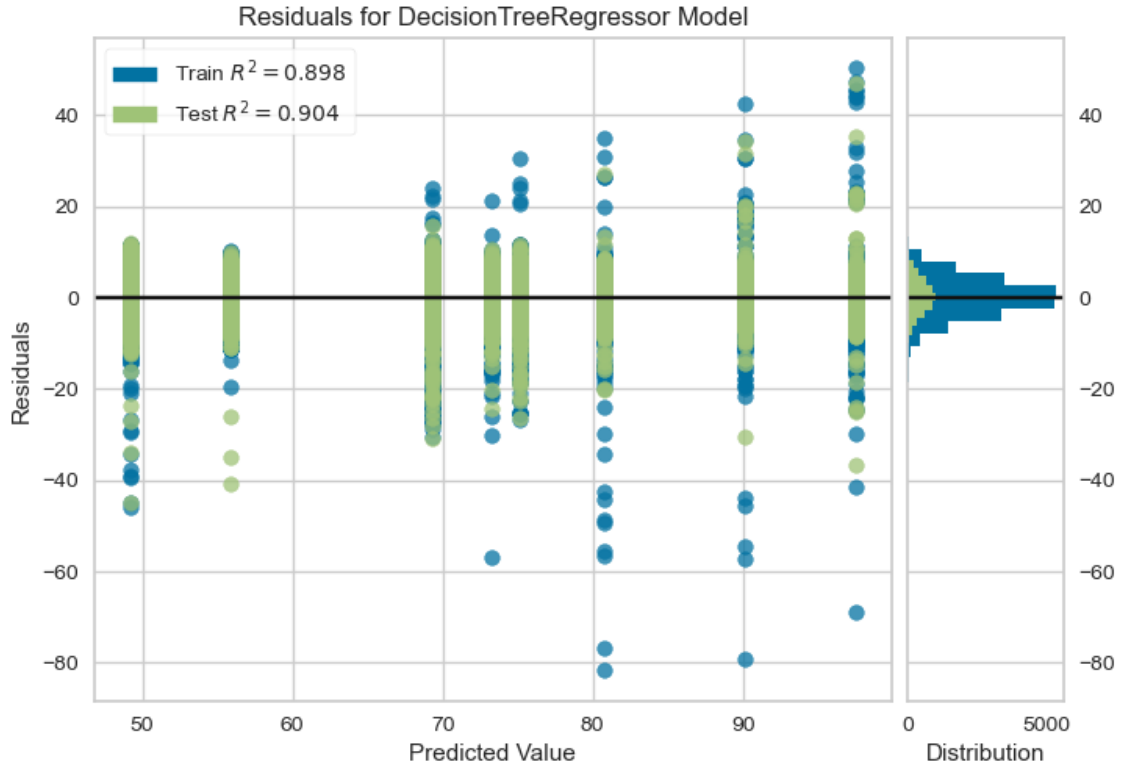
Şekil 5.11. Gecikme süresi için Destek Vektör Regresörü modelinde artık değerler dağılımı

Şekil 5.11’de artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmında herhangi bir desen veya kalıp olmadığı görülmektedir. Sütun dağılımının da normal dağılım göstermesi, artık değerler açısından bu modelin uygun olduğuna işaret etmektedir.



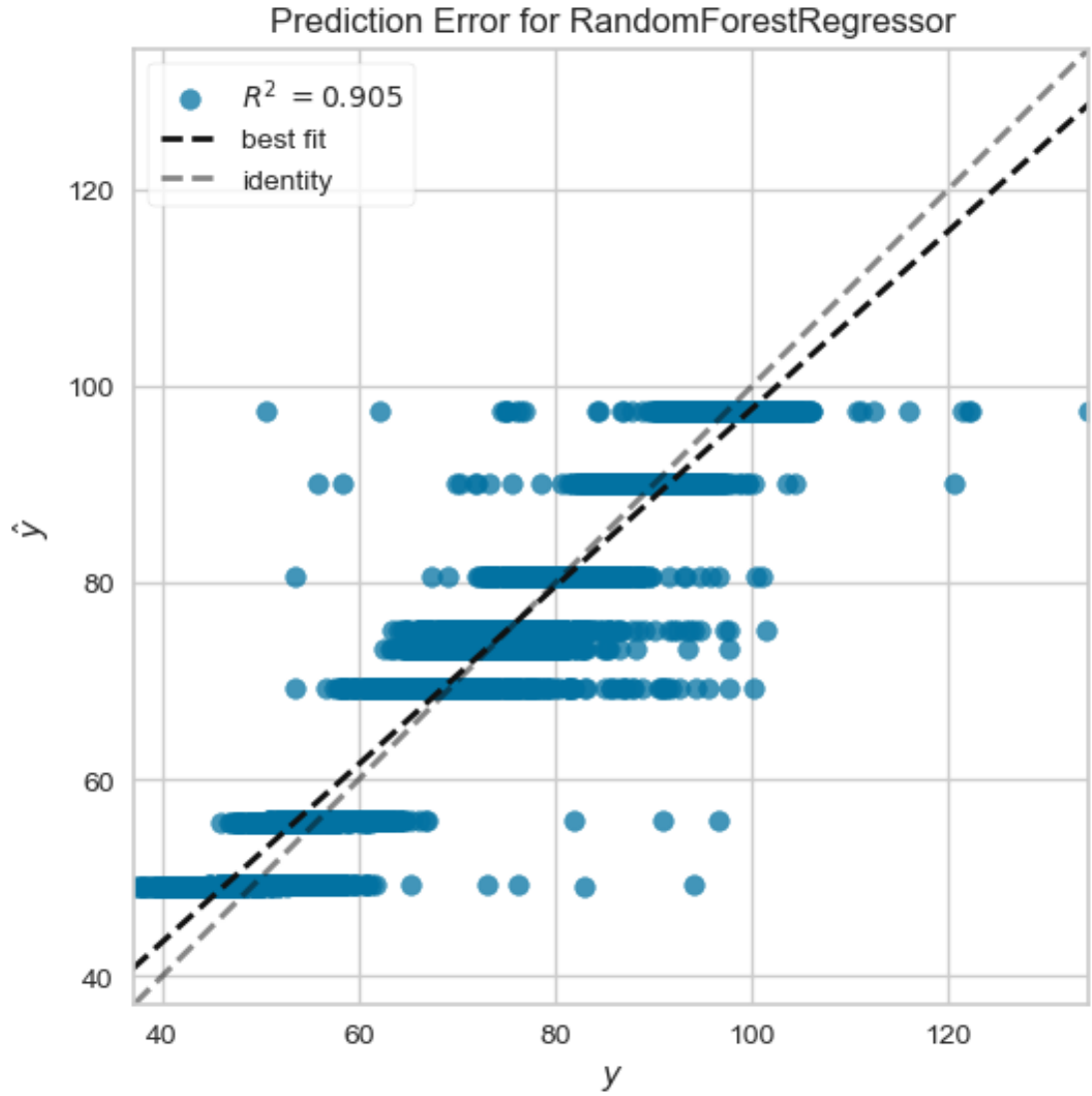
Şekil 5.12. Gecikme süresi için Karar Ağacı modeli tahmin hatası grafiği

Şekil 5.12'e göre modelin belli başlı değerleri tahmin etmesi aşırı uyumun göstergesidir. Bu sebeple, Gecikme süresi parametresi için Karar Ağacı modellemesi başarısız olmuştur.



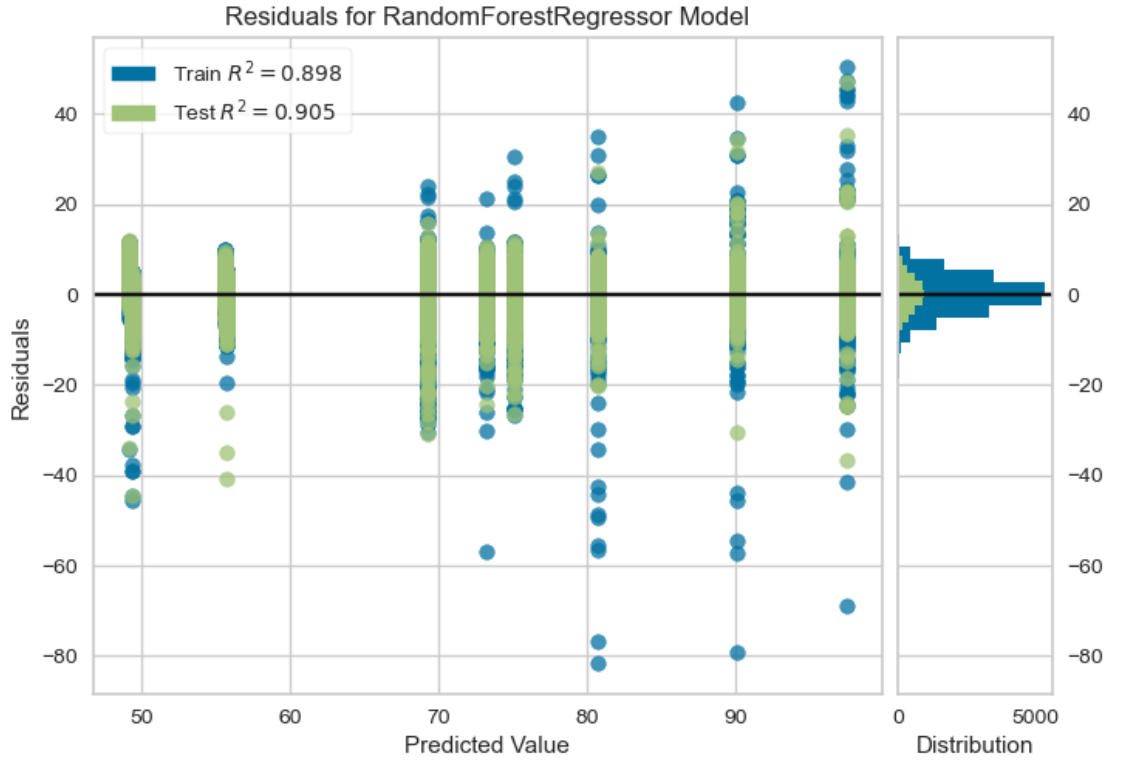
Şekil 5.13. Gecikme süresi için Karar Ağacı modelinde artık değerler dağılımı

Şekil 5.13'te artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmında “pattern” dağılımına sahip olduğu görülmektedir. Bu sebeple gecikme süresi modelinin, artık değerler açısından uygun bir model olmadığı tespit edilmiştir.



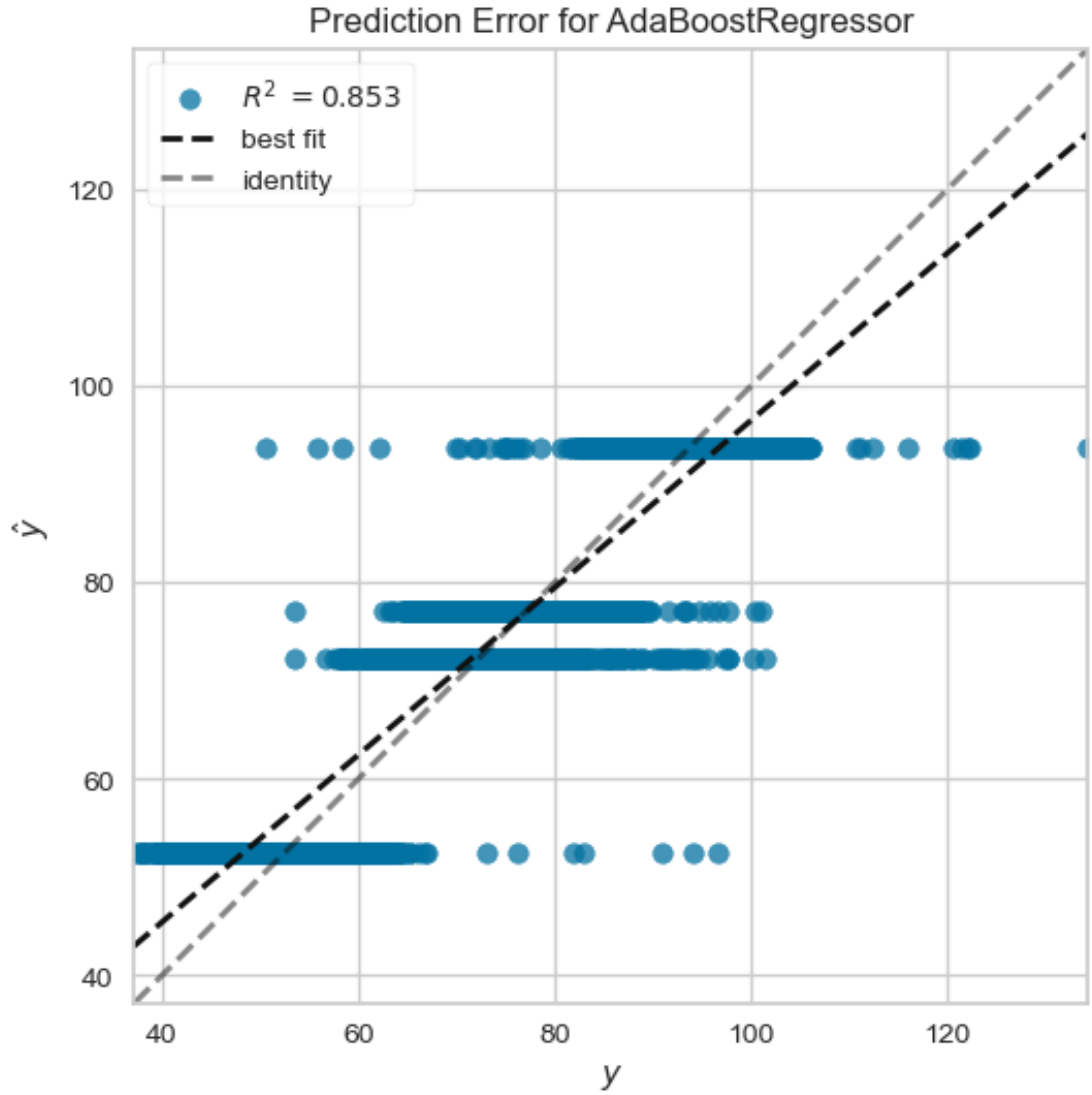
Şekil 5.14. Gecikme süresi için Rastgele Orman modeli tahmin hatası grafiği

Şekil 5.14’de görüldüğü gibi Rastgele Orman için oluşturulan tahmin hatası dağılım grafiği, Karar Ağacı ile benzerlik göstermektedir. Dolayısıyla modelin aşırı uyum gösterdiği ifade edilebilir. Bu sebeple gecikme süresi parametresi için Rastgele Orman modellemesi başarısız olmuştur.



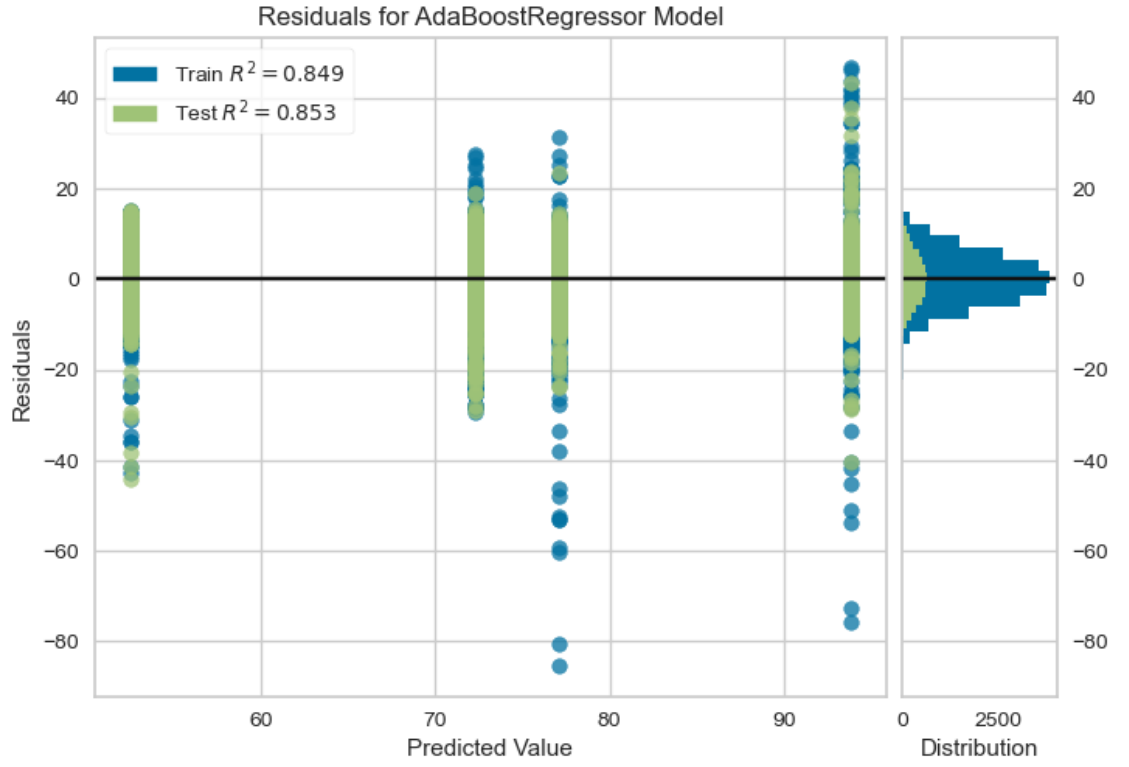
Şekil 5.15. Gecikme süresi için Rastgele Orman modelinde artık değerler dağılımı

Şekil 5.15. artık değerler dağılım grafiğinde ise “pattern” dağılımı oluşmuştur. Yani dağılımda tekrar eden bir desen vardır. Modelin aşırı uyum gösterdiği tespit edilmiştir. R^2 değerleri makul olsa da model, gecikme süresi için başarısız olmuştur.



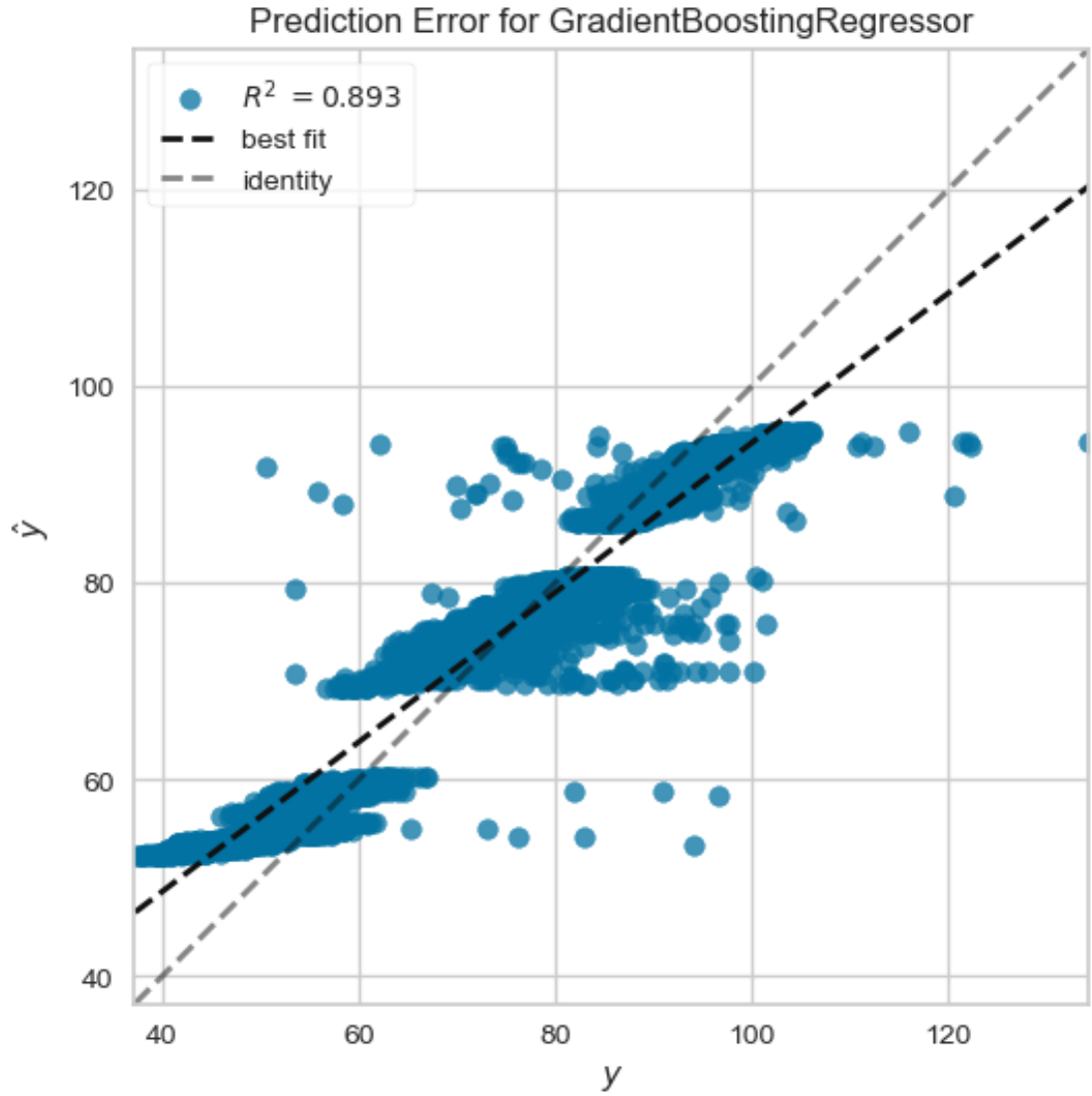
Şekil 5.16. Gecikme süresi için AdaBoost modeli tahmin hatası grafiği

Şekil 5.16’da görüldüğü gibi AdaBoost algoritması, gecikme süresi parametresi için diğer yöntemlerden zayıf tahmin yapmıştır. R^2 0,8525, RMSE ise 6,0557 çıkmıştır. Yani R^2 için %15, RMSE için %8 değerlerinde hata oranı hesaplanmıştır. Bu modelin aşırı uyum gösterdiği tespit edilmiştir. Bu sebeple gecikme süresi parametresi için AdaBoost modellemesi başarısız olmuştur.



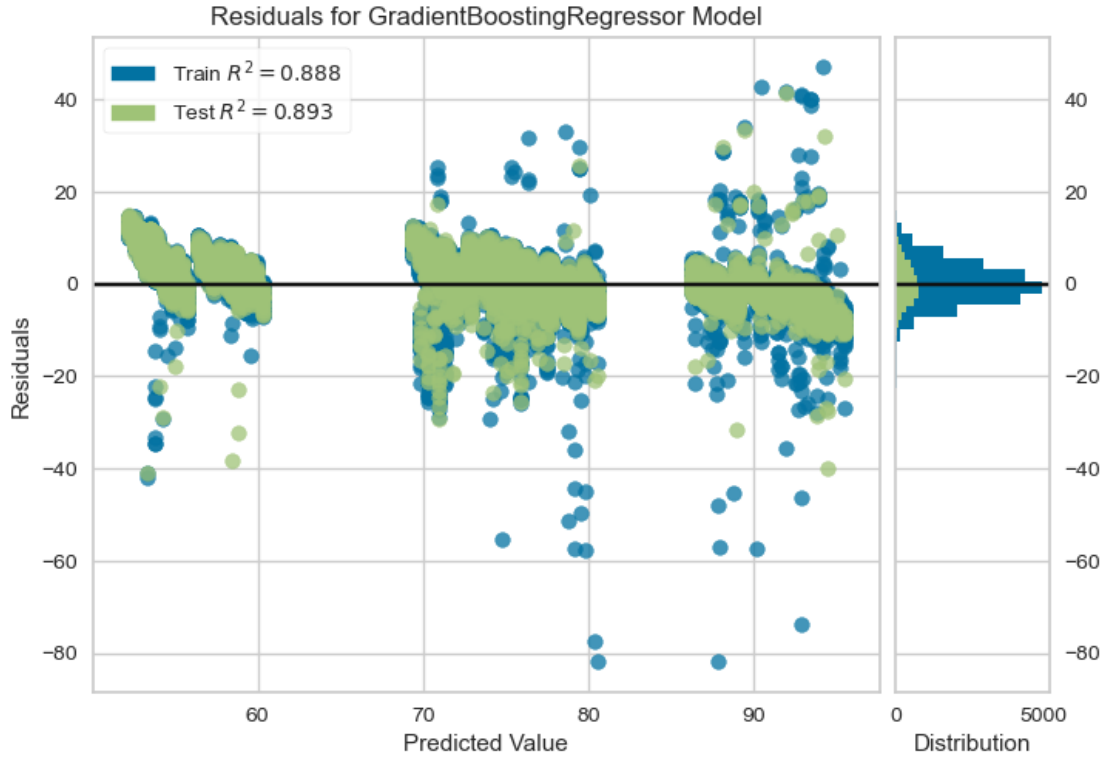
Şekil 5.17. Gecikme süresi için AdaBoost modelinde artık değerler dağılımı

AdaBoost artık değerler dağılımına bakıldığında desenli yani “pattern” bir dağılım bulunmaktadır. Modelin aşırı uyum durumu açıkça görülmektedir. R^2 değerleri de düşük çıkmıştır. Yani model, gecikme süresi parametresi için başarısız olmuştur.



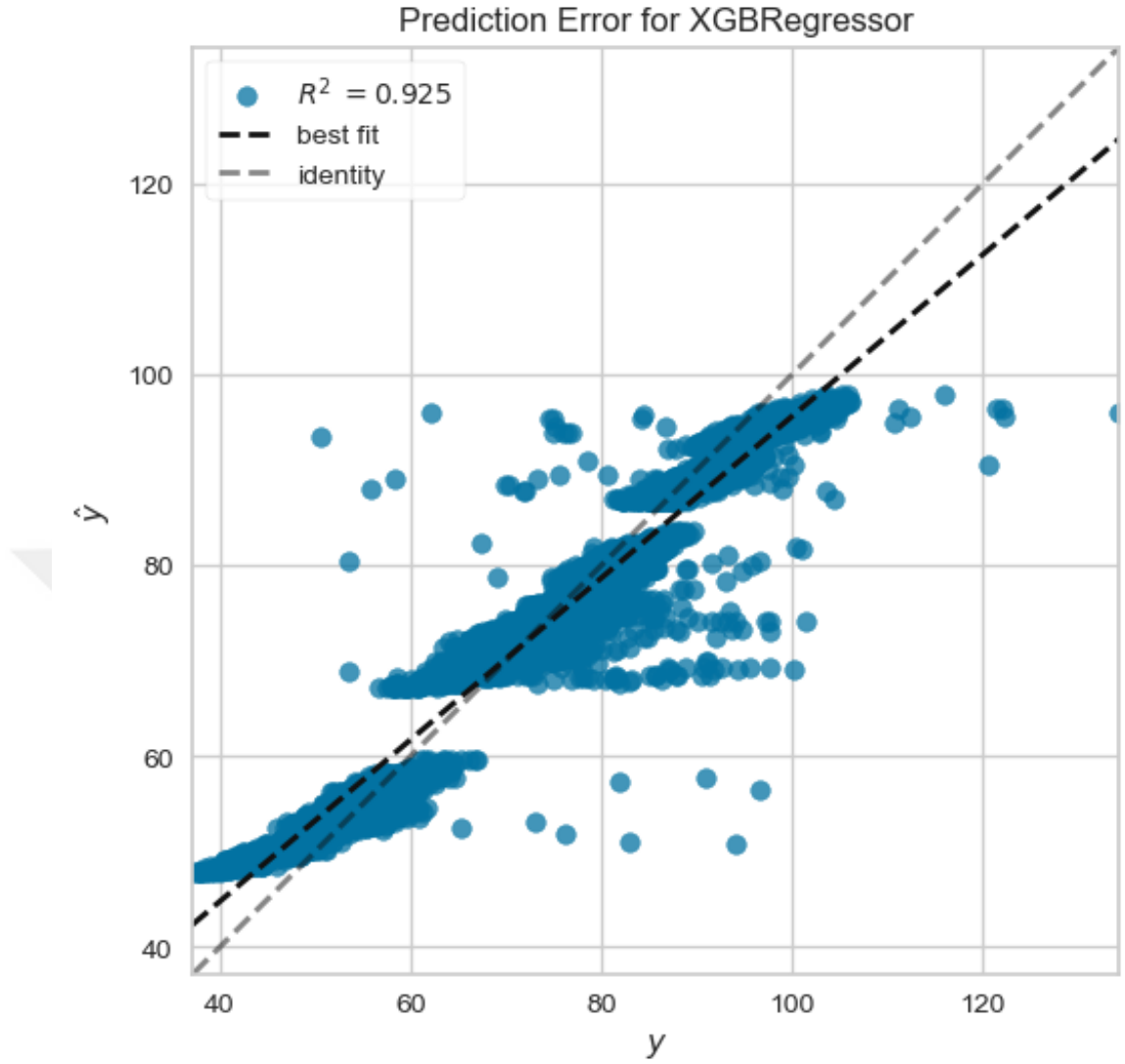
Şekil 5.18. Gecikme süresi için Gradyan Artırma modeli tahmin hatası grafiği

Şekil 5.18’de verilerin hata grafiğinde best fit ve identity eğrileri arasındaki açı yüksek çıkmıştır. R^2 değeri (0,8933) ile %10 hata oranı ve RMSE (5,1510) değeri ile %7 hata oranı hesaplanmıştır. Karar Ağacı, Rastgele Orman ve AdaBoost algoritmalarına göre aşırı uyum durumu daha azdır.



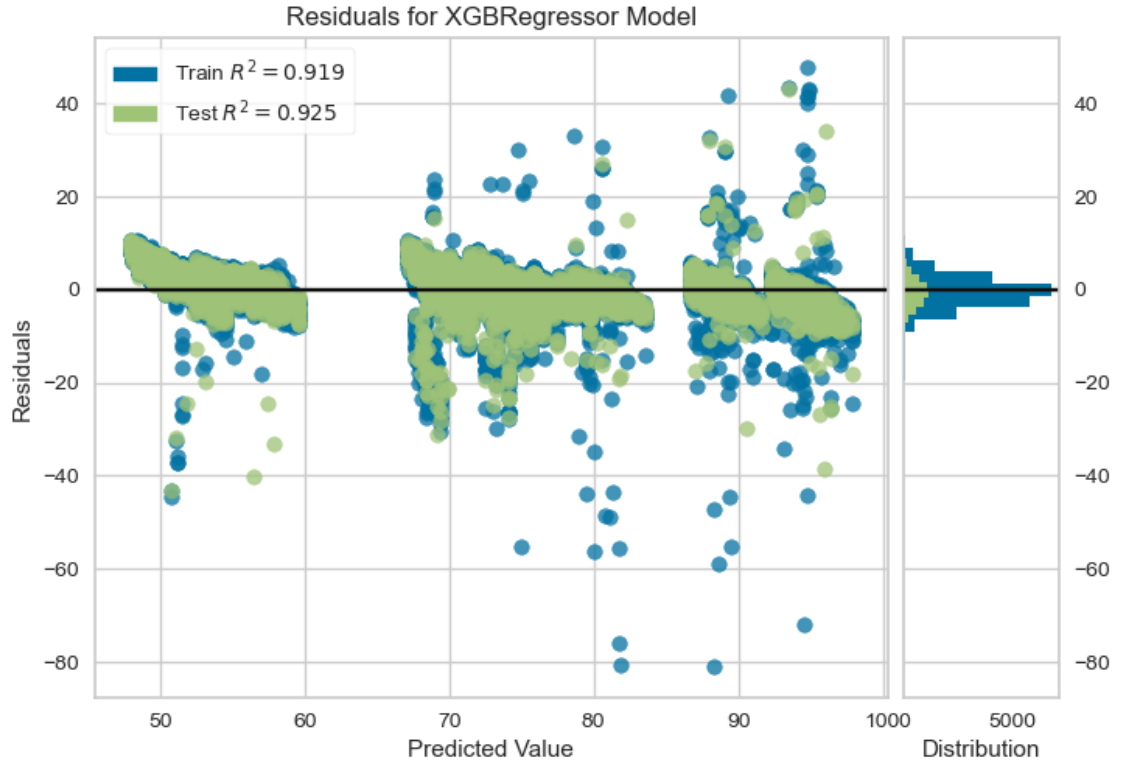
Şekil 5.19. Gecikme süresi için Gradyan Artırma modelinde artık değerler dağılımı

Gradyan Artırma artık değerler dağılımına bakıldığında önceki modellere kıyasla daha az desenli bir dağılım bulunmaktadır. Yine de modelin aşırı uyum gösterdiği söylenebilir. R^2 değerleri de düşük çıktığından diğer modellere kıyasla daha az “aşırı uyum” vardır. Ancak model, gecikme süresi parametresi için başarısız olmuştur.



Şekil 5.20. Gecikme süresi için XGBoost modeli tahmin hatası grafiği

Şekil 5.20’de XGBoost algoritması, Gradyan Artırma algoritmasının gelişmiş hali olarak daha yüksek bir performans gösterdiği ve tanımlanan eğri ile ideal eğri arasındaki açısal farkın daha da az olduğu görülmektedir. R^2 0,9249 ve RMSE 4,3248 çıktığı modelde hata oranları R^2 için yaklaşık %8 ve RMSE için yaklaşık %6 oranında hata çıktığı hesaplanmıştır. Model için aşırı uyum durumuna daha uzak bir dağılım bulunmaktadır.



Şekil 5.21 Gecikme süresi için XGBoost modelinde artık değerler dağılımı

Şekil 5.21 XGBoost artık değerler dağılım grafiğinde test ve eğitim kümelerinin dağılımının örtüştüğü ve kümelerin kendi aralarında yapısal bir ilişki olduğu saptanmıştır. R^2 değerlerinin yakınlığı da bu kanıyı güçlendirmektedir. Gradyan Artırma modeline göre “pattern” yani desenli durum daha azdır. Bu modelleme, bazı başarısız modellere kıyasla kabul edilebilir bir seviyededir.

5.1.4. Modellerin karşılaştırılması

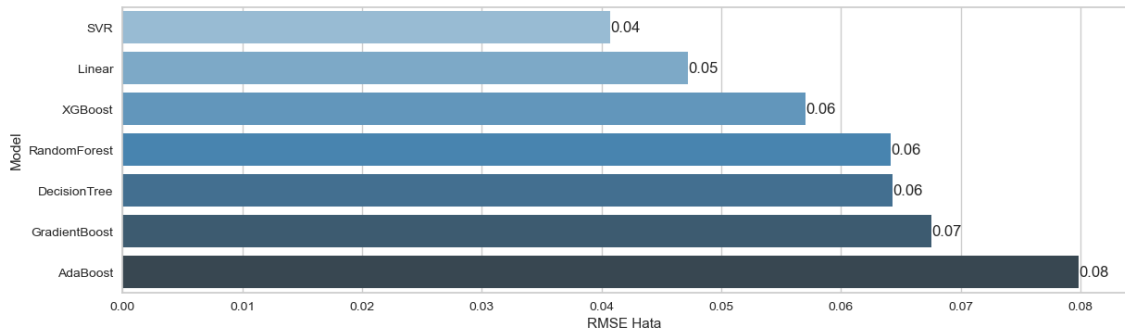
Bu bölümde belirtilen parametrelerin, makine öğrenmesi metotları ile elde edilen modellerin hata metrikleri ile nihai değerlendirilmesi yapılmıştır. Tablo 5.4-5.5’te bu hata metrikleri verilmiştir. Şekil 5.22-5.23’te ise R^2 ve RMSE grafikli karşılaştırmaları verilmiştir.

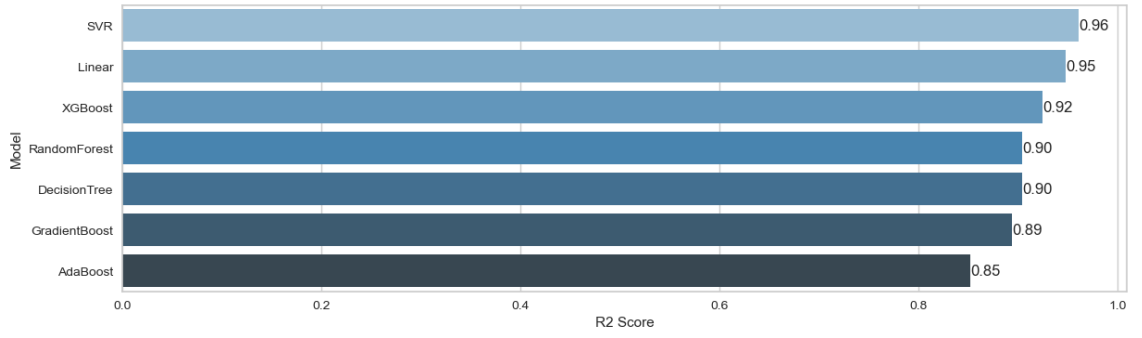
Tablo 5.4. Gecikme süresi için makine öğrenmesi modellerinin hata metrikleri değerleri

	R²	MAE	MSE	RMSE
Lineer Regresyon GridSearchCV Eğitim	0,9424	1,8842	14,2533	3,7754
Lineer Regresyon GridSearchCV Test	0,9485	1,8686	12,8069	3,5787
SVR GridSearchCV Eğitim	0,9556	0,9180	10,9784	3,3134
SVR GridSearchCV Test	0,9616	0,9591	9,5433	3,0892
Karar Ağacı GridSearchCV Eğitim	0,8977	3,5283	25,3279	5,0327
Karar Ağacı GridSearchCV Test	0,9043	3,4790	23,7895	4,8775
Rastgele Orman GridSearchCV Eğitim	0,8979	3,5226	25,2630	5,0262
Rastgele Orman GridSearchCV Test	0,9046	3,4725	23,7272	4,8711
AdaBoost GridSearchCV Eğitim	0,8488	4,6629	37,4268	6,1177
AdaBoost GridSearchCV Test	0,8525	4,6932	36,6705	6,0556
Gradyan Artırma GridSearchCV Eğitim	0,8879	3,7587	27,7336	5,2663
Gradyan Artırma GridSearchCV Test	0,8936	3,7969	26,4519	5,1431
XGBoost GridSearchCV Eğitim	0,9189	2,9161	20,0751	4,4805
XGBoost GridSearchCV Test	0,9248	2,9222	18,7043	4,3248

Tablo 5.5 Gecikme süresi için performansa göre sıralanmış RMSE ve R² skorları

	Model	RMSE Hata	R² Skoru
1	SVR	0,0408	0,9616
2	Lineer Regresyon	0,0472	0,9485
3	XGBoost	0,0571	0,9248
4	Rastgele Orman	0,0642	0,9048
5	Karar Ağacı	0,0643	0,9043
6	Gradyan Artırma	0,0676	0,8945
7	AdaBoost	0,0799	0,8525

**Şekil 5.22.** Gecikme süresi için makine öğrenme modelleri RMSE hata oranlarının karşılaştırılması



Şekil 5.23. Gecikme süresi için makine öğrenme modelleri R^2 skorlarının karşılaştırılması

Tablo 5.4 ve 5.5’ göre Destek Vektör Regresörü eğitilen veriler doğrultusunda gecikme süresi parametresi için en iyi sonucu veren makine öğrenmesi modeli olduğu görülmektedir. Lineer Regresyon modeli de başarılı makine öğrenmesi modeli sayılabilir. En başarısız makine öğrenmesi modeli ise “AdaBoost” olmuştur.

Şekil 5.22’de RMSE metriğine göre en iyi tahmin sonuçlarının 0,04 hata oranıyla %96’lık bir performansla Destek Vektör Regresörü ile elde edildiği görülmektedir. Bu değeri, %5’lik hata oranı ile Lineer Regresyon ve %6’lık hata oranı ile XGBoost takip etmektedir.

Şekil 5.23’te ise R^2 skorlarının karşılaştırıldığı bir grafik verilmiştir. Bu grafiğe göre %96’lık doğruluk oranı ve %4’lük bir hata oranı ile Destek Vektör Regresörü modeli en iyi tahmin sonuçlarını üretmiştir. Bu duruma göre %95 R^2 skoru ile Lineer Regresyon modeli ve %92 R^2 skoru ile XGBoost modeli doğruluk oranları yüksek modeller olarak takip etmektedir.

Edinilen bulgular doğrultusunda Destek Vektör Regresörü ve Lineer Regresyon modelleri gecikme süresi parametresi için başarılı modeller olarak seçilmiştir.

5.1.5. Modellere ait tahmin değerlerinin karşılaştırılması

0, 100 ve 200 numaralı senaryolar örnek senaryo olarak kullanılmıştır. Bu Senaryolara ait kavşakların geometrik ve sinyalizasyon özellikleri, trafik hacim bilgileri ve şerit sayısına ait özellikler Tablo 5.6’da verilmiştir.

Tablo 5.6. Karşılaştırma yapılacak senaryolara ait bilgiler

	Dörtüol Kavşağı	Merplaza Kavşağı	Adliye Kavşağı	Terzibaba Kavşağı	Polis Okulu Kavşağı	Tedaş Kavşağı	Terminal Kavşağı	Trafik Hacmi Değişimi Araba%	Trafik Hacmi Değişimi Otobüs%	Trafik Hacmi Değişimi Kamyon%	Batı-Doğu istikametine (I) şerit	Doğu-Batı istikametine (I) şerit
0	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Trafik	Mevcut Trafik	Mevcut Trafik	Mevcut Şerit	Mevcut Şerit
100	Doğu-Batı yönünde	Mevcut Durum	Mevcut Durum	Alıgeçit	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Trafik	Mevcut Trafik	Mevcut Trafik	Mevcut Şerit	Mevcut Şerit
200	Modem Dönel Kavşak	Doğu-Batı yönünde	Modem Dönel Kavşak	Mevcut Durum	Mevcut Durum	Modem Dönel Kavşak	Mevcut Durum	%5 artış	Mevcut Trafik	Mevcut Trafik	Mevcut Şerit	Mevcut Şerit

Tablo 5.7 Gecikme süresi için model tahmin değerlerinin karşılaştırılması

	Simülasyon Sonucu	Lineer Regresyon	SVR	Karar Ağacı	Rastgele Orman	AdaBoost	Gradyan Artırma	XGBoost
0	98,52	99,60	98,63	97,51	97,51	93,82	95,09	95,55
100	84,90	85,49	84,23	80,77	80,77	76,96	80,46	82,21
200	71,51	71,57	73,44	69,32	69,32	72,29	71,41	70,31

Tablo 5.7'ye göre senaryo 0 için Destek Vektör Regresörü ile senaryo 100 ve senaryo 200 için de Lineer Regresyon modeli ile gerçeğe en yakın tahminler yapılmıştır. Hata oranları ile başarılı olarak tespit edilen Lineer Regresyon ve Destek Vektör Regresörü modellerinin yeterliliği somut bir şekilde teyit edilmiş olmaktadır.

5.1.6. Başarılı Modellerin Etki Katsayılarının Değerlendirilmesi

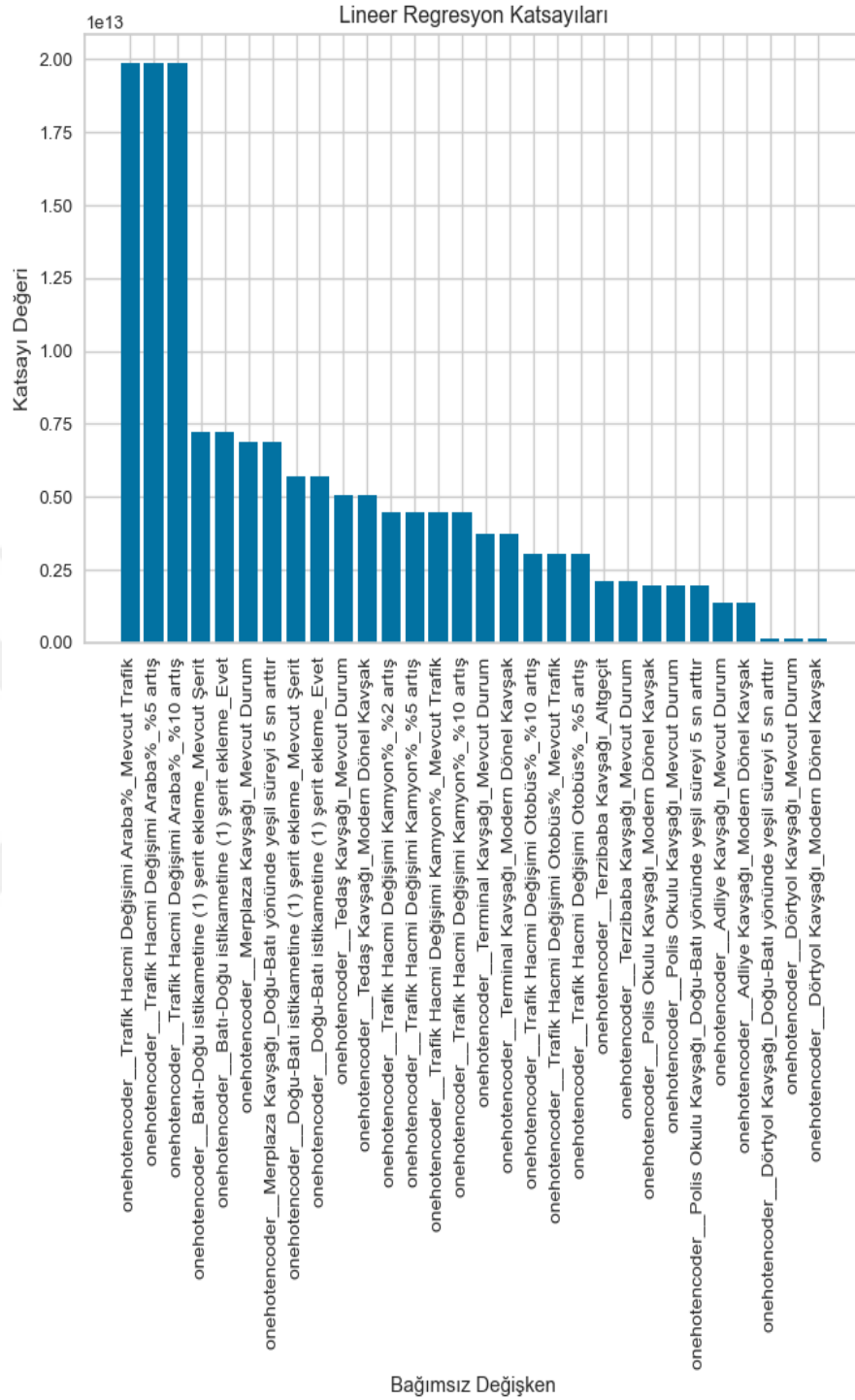
Gecikme süresi için başarılı modeller olarak seçilen Lineer Regresyon'un ve Destek Vektör Regresörü'nün, Denklem 3.6'ya göre katsayı değerleri hesaplanmıştır. Bu değerler Şekil 5.24'te ve 5.26'da, bu değerlere ait grafik ise Şekil 5.25'te ve 5.27'de verilmiştir.

```

onehotencoder__Trafik Hacmi Değişimi Araba%_Mevcut Trafik: -19847709983033.97
onehotencoder__Trafik Hacmi Değişimi Araba%_5 artış: -19847709983032.234
onehotencoder__Trafik Hacmi Değişimi Araba%_10 artış: -19847709983030.566
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Mevcut Şerit: 7225880096120.885
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Evet: 7225880096118.943
onehotencoder__Merplaza Kavşağı_Mevcut Durum: -6853836740034.913
onehotencoder__Merplaza Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: -6853836740032.415
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Mevcut Şerit: 5694873530703.431
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Evet: 5694873530703.234
onehotencoder__Tedaş Kavşağı_Mevcut Durum: 5060923004892.606
onehotencoder__Tedaş Kavşağı_Modern Dönel Kavşak: 5060923004889.648
onehotencoder__Trafik Hacmi Değişimi Kamyon%_2 artış: 4480383160106.762
onehotencoder__Trafik Hacmi Değişimi Kamyon%_5 artış: 4480383160106.697
onehotencoder__Trafik Hacmi Değişimi Kamyon%_Mevcut Trafik: 4480383160106.503
onehotencoder__Trafik Hacmi Değişimi Kamyon%_10 artış: 4480383160106.347
onehotencoder__Terminal Kavşağı_Mevcut Durum: 3750085827154.7983
onehotencoder__Terminal Kavşağı_Modern Dönel Kavşak: 3750085827152.983
onehotencoder__Trafik Hacmi Değişimi Otobüs%_10 artış: 3023728554915.9043
onehotencoder__Trafik Hacmi Değişimi Otobüs%_Mevcut Trafik: 3023728554915.7793
onehotencoder__Trafik Hacmi Değişimi Otobüs%_5 artış: 3023728554915.386
onehotencoder__Terzibaba Kavşağı_Altgeçit: -2114509939806.2952
onehotencoder__Terzibaba Kavşağı_Mevcut Durum: -2114509939788.271
onehotencoder__Polis Okulu Kavşağı_Modern Dönel Kavşak: -1935120073092.2422
onehotencoder__Polis Okulu Kavşağı_Mevcut Durum: -1935120073087.9695
onehotencoder__Polis Okulu Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: -1935120073087.7742
onehotencoder__Adliye Kavşağı_Mevcut Durum: 1382198254038.8538
onehotencoder__Adliye Kavşağı_Modern Dönel Kavşak: 1382198254031.8855

```

Şekil 5.24. Gecikme süresi Lineer Regresyon modelinde değişkenlerin katsayı değerleri



Şekil 5.25 Gecikme süresi Lineer Regresyon modelinde değişkenlerin katsayı değerleri sütun grafiği

Şekil 5.24 ve Şekil 5.25 için gecikme süresi parametresi üzerindeki en fazla etkinin otomobil hacimleri ile ilgili değişkenler olduğu görülmektedir. Bu değişkenleri sinyalizasyon ile ilgili değişiklikler ile şerit sayısı ile ilgili değişiklikler takip etmektedir.

Ancak katsayı değeri 0,75'ten düşük olduğundan bu etki otomobil hacmine bağlı değişkenler kadar yüksek değildir. Kavşak değişkenleri göz önünde bulundurulduğunda ise kavşak geometrisi ile gecikme süresi parametresinin en çok etkilendiği kavşağın Tedaş Kavşağı olduğu gözlenmiştir. En az etkilenen kavşak ise Dört Yol Kavşağı olarak tespit edilmiştir. Bu durum, Tedaş Kavşağı'nın planlanmasıyla gecikme süresi değerini büyük ölçüde azaltacağı görülmektedir.

Sinyalizasyondaki değişiklikler katsayısı 2'ye yakın bir değerle en yüksek olan Merplaza Kavşağı olmuştur. Bu kavşaktaki sinyalizasyonun düzenlenmesi koridordaki gecikme süresi değerinde hayati bir öneme sahip olduğu görülmektedir.

```
onehotencoder__Dört Yol Kavşağı_Modern Dönel Kavşak: -328.63337608183446
onehotencoder__Terzibaba Kavşağı_Altgeçit: -191.68708132081497
onehotencoder__Terzibaba Kavşağı_Mevcut Durum: 191.6870813208149
onehotencoder__Dört Yol Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 178.9342103673047
onehotencoder__Dört Yol Kavşağı_Mevcut Durum: 149.6901657145292
onehotencoder__Adliye Kavşağı_Mevcut Durum: 77.75260791070355
onehotencoder__Adliye Kavşağı_Modern Dönel Kavşak: -77.7526079107035
onehotencoder__Polis Okulu Kavşağı_Modern Dönel Kavşak: -58.255526448138625
onehotencoder__Polis Okulu Kavşağı_Mevcut Durum: 38.20564530228058
onehotencoder__Trafik Hacmi Değişimi Araba%_Mevcut Trafik: -36.36081106327356
onehotencoder__Trafik Hacmi Değişimi Araba%_10 artış: 34.911881527450156
onehotencoder__Tedaş Kavşağı_Modern Dönel Kavşak: -32.76715227582967
onehotencoder__Tedaş Kavşağı_Mevcut Durum: 32.76715227582954
onehotencoder__Polis Okulu Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 20.04988114585818
onehotencoder__Terminal Kavşağı_Modern Dönel Kavşak: -19.530449124650517
onehotencoder__Terminal Kavşağı_Mevcut Durum: 19.530449124650502
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Evet: -18.710365026274214
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Mevcut Şerit: 18.71036502627416
onehotencoder__Merplaza Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 12.663976377787302
onehotencoder__Merplaza Kavşağı_Mevcut Durum: -12.663976377787275
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Evet: -6.171618339413297
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Mevcut Şerit: 6.171618339413238
onehotencoder__Trafik Hacmi Değişimi Otobüs%_5 artış: -4.780642965323663
onehotencoder__Trafik Hacmi Değişimi Otobüs%_10 artış: 4.183139393878909
onehotencoder__Trafik Hacmi Değişimi Kamyon%_10 artış: -3.8087611137435076
onehotencoder__Trafik Hacmi Değişimi Kamyon%_2 artış: 3.3377595256776225
onehotencoder__Trafik Hacmi Değişimi Kamyon%_5 artış: 1.5734217620519413
onehotencoder__Trafik Hacmi Değişimi Araba%_5 artış: 1.4489295358234417
onehotencoder__Trafik Hacmi Değişimi Kamyon%_Mevcut Trafik: -1.1024201739860533
onehotencoder__Trafik Hacmi Değişimi Otobüs%_Mevcut Trafik: 0.5975035714447197
```

Şekil 5.26. Gecikme süresi Destek Vektör Regresörü modelinde değişkenlerin katsayı değerleri

Şekil 5.26 ve Şekil 5.27 incelendiğinde Destek Vektör Regresörü modeli için gecikme süresi parametresi üzerindeki en fazla etkinin kavşaklardaki geometrik değişikliklere ait değişkenlerle olduğu; parametreyi özellikle de Dört Yol Kavşağı modern dönel kavşak değişkeninin etkilediği gözlenmektedir. Bu değerin 300’den fazla olduğu görülmektedir. Bu durum, özellikle Dört Yol Kavşağı’nda yapılan bir iyileştirmenin gecikme süresindeki azalmaya katkı sağlayacağını göstermektedir.

Katsayı değeri yaklaşık 150 çıkan Dört Yol Kavşağı sinyalizasyon süresi ile ilgili değişkenin de gecikme süresi üzerinde büyük etkisi olduğu görülmektedir.

Bu değişkenleri taşıt hacmindeki değişiklikler ile sinyalizasyondaki değişiklikler takip etmektedir. Katsayı değeri 200’den fazla çıkan Terzibaba Kavşağı Mevut Durum ve Modern Dönel Kavşak değişkenlerinin de gecikme süresi üzerinde etkisi olduğu açıkça görülmektedir.

Gecikme süresi üzerinde en az etkiye ise kamyon hacmine ait değişkenler etki etmektedir. Öyle ki bu değişkenlere ait katsayı değerleri 0’a yakın çıkmıştır.

Dört Yol Kavşağı’nda bir modern dönel kavşak ya da doğu-batı yönündeki sinyalizasyon düzenlemesinin, gecikme süresi parametresinin kontrol altına alınabilmesinde hayati bir öneme sahip olduğu anlaşılmaktadır.

Terzibaba Kavşağı’ndaki altgeçit düzenlemesi gibi bir düzenlemenin de gecikme süresi için büyük etkileri olacağı açıkça görülmektedir.

5.2. IEM CO₂ parametresi için bulgular

5.2.1. GridSearchCV öncesi ve sonrası performans değerlendirme

Tablo 5.8. IEM CO₂ parametresi için GridSearchCV öncesi performans değerleri

	R ² Eğitim	R ² Test	MAE Eğitim	MAE Test	MSE Eğitim	MSE Test	RMSE Eğitim	RMSE Test
Lineer Regresyon	7,852733e-01	7,853326e-01	8,105470e+04	8,121438e+04	7,797756e+09	7,792533e+09	8,830490e+04	8,827533e+04
SVR	-6,127469e-02	-5,327391e-02	1,465131e+05	1,461319e+05	3,853997e+10	3,823437e+10	1,963160e+05	1,955361e+05
Karar Ağacı	1,0	9,670559e-01	0,0	1,227330e+04	0,0	1,195886e+09	0,0	3,458158e+04
Rastgele Orman	9,980328e-01	9,875628e-01	3,519779e+03	9,243045e+03	7,143954e+07	4,514750e+08	8,452191e+03	2,124794e+04
AdaBoost	7,507739e-01	7,377046e-01	7,814330e+04	7,999957e+04	9,050593e+09	9,521453e+09	9,513461e+04	9,757793e+04
Gradyan Artırma	9,751582e-01	9,745785e-01	1,503109e+04	1,514054e+04	9,021265e+08	9,228126e+08	3,003542e+04	3,037783e+04
XGBoost	9,748548e-01	9,742947e-01	1,518495e+04	1,531317e+04	9,131437e+08	9,331162e+08	3,021827e+04	3,054695e+04

Tablo 5.8'e bakıldığında R² değerlerinin hem test hem de eğitim verileri için tüm algoritmalarda 1 değerine yakın olduğu görülmektedir. Değerler gayet makul seviyede çıkmıştır. Karar Ağacı modelinde bu değer 1,0 olarak hesaplanmıştır. Bu durum hiçbir hata olmadığını gösterse de normalde çok fazla mümkün olmayan bir durumdur. Bu durum modelin genelleştirme özelliğinin kaybedildiği aşırı uyum durumunu göstermektedir.

AdaBoost R² değerleri ise 0,7 civarında diğer algoritmalarından daha az bir değerde çıkmıştır.

MAE değerinin için veriler, değer büyüklüğü ile kıyaslandığında her modelleme yöntemi için tahmin değerlerinin, nispeten başarılı bir model oluşturulduğunu göstermektedir. Burada Karar Ağacı 0 değeri alarak aşırı uyum olasılığını yükseltmektedir.

Birbirleriyle bağlantılı olan MSE ve RMSE metrikleri MAE gibi tahmin değerlerinin gözlenen değerlere uzaklığını temsil etmesi bağlamında makul seviyede bulunmuştur. Ayrıca Karar Ağacı diğer metriklerde olduğu gibi bu iki metrikte de 0 çıkmıştır. Yani aşırı uyum durumuna göre işlem yapılması gerekmektedir.

Tablo 5.9. IEM CO₂ parametresi için GridSearchCV sonrası performans değerleri

	R ² Eğitim	R ² Test	MAE Eğitim	MAE Test	MSE Eğitim	MSE Test	RMSE Eğitim	RMSE Test
Lineer Regresyon	7,866148e-01	7,869817e-01	8,115230e+04	8,123977e+04	7,749041e+09	7,732670e+09	8,802864e+04	8,793560e+04
SVR	1,762122e-03	1,017963e-02	1,410657e+05	1,406203e+05	3,625080e+10	3,593097e+10	1,903964e+05	1,895547e+05
Karar Ağacı	9,076595e-01	9,061153e-01	4,320539e+04	4,346659e+04	3,353328e+09	3,408062e+09	5,790792e+04	5,837861e+04
Rastgele Orman	8,452191e+03	2,124794e+04	4,317874e+04	4,344424e+04	3,350645e+09	3,405373e+09	5,788475e+04	5,835557e+04
AdaBoost	8,414300e-01	8,373362e-01	5,932572e+04	5,970058e+04	5,758436e+09	5,904778e+09	7,588436e+04	7,684255e+04
Gradyan Artırma	8,908357e-01	8,901941e-01	4,811332e+04	4,804491e+04	3,964278e+09	3,986009e+09	6,296251e+04	6,313485e+04
XGBoost	9,199768e-01	9,184510e-01	4,036919e+04	4,064018e+04	2,906028e+09	2,960268e+09	5,390758e+04	5,440835e+04

GridSearchCV ile fit edilen test ve eğitim R² değerleri Karar Ağacı, Rastgele Orman, Gradyan Artırma ve XGBoost modelleri için azalmıştır. Destek Vektör Regresörü ve AdaBoost modellerinde GridSearchCV kullanımı ile R² değerinde artış gözlenmiştir. Lineer Regresyon parametresi için ise bir değişiklik gözlenmemiştir.

MAE, MSE ve RMSE değerlerinde ise GridSearchCV ile Karar Ağacı, Rastgele Orman, Gradyan Artırma ve XGBoost modelleri için artış görülmüştür. Destek Vektör Regresörü ve AdaBoost modellerinde ise bir azalma görülmekte ve Lineer Regresyon parametresi için ise bir değişiklik gözlenmemektedir.

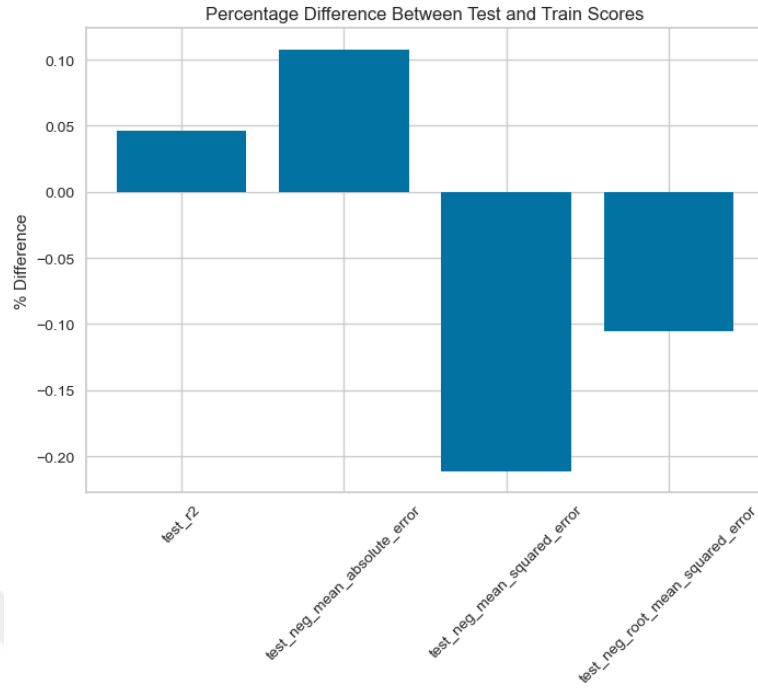
R^2 değerleri farklı bir dağılıma sahipken GridSearchCV sonrası algoritmaların R^2 değerleri birbirine yaklaşılarak 0,89-0,95 aralığında hesaplanmıştır. En yüksek performansa Destek Vektör Regresörü modelinde ve en düşük performansa AdaBoost modelinde erişilmiştir.

5.2.2. Test ve Eğitim Skorları Arasındaki Yüzdesel Farklar

Tablo 5.10 IEM CO₂ parametresi için test ve eğitim skorları arası yüzdelik fark

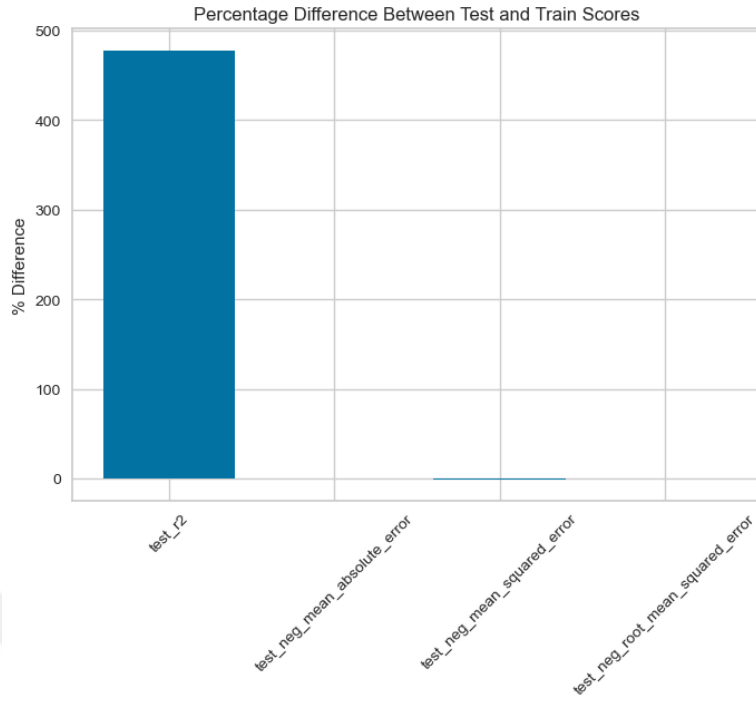
	R^2 (%)	MAE (%)	MSE (%)	RMSE (%)
Lineer Regresyon	0,05	0,11	-0,21	-0,11
SVR	477,69	-0,32	-0,88	-0,44
Karar Ağacı	-0,17	0,60	1,63	0,81
Rastgele Orman	-0,17	0,61	1,63	0,81
AdaBoost	-0,49	0,63	2,54	1,26
Gradyan Artırma	-0,07	-0,14	0,55	0,27
XGBoost	-0,17	-0,637	1,87	0,93

Her bir modele ait test ve eğitim skorları arası yüzdesel farkla ilgili grafikler Şekil 5.28-5.34'te gösterilmiştir.



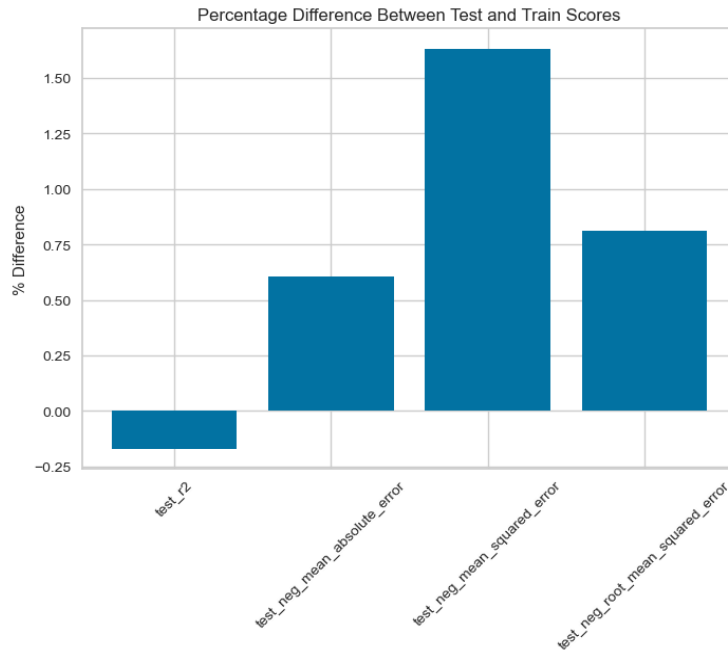
Şekil 5.28. Lineer Regresyon için IEM CO₂ eğitim ve test verileri arasında yüzdelik fark

Şekil 5.28’de R^2 metriği ve MAE metriği için Lineer Regresyon modeline ait test skorunun, eğitim skorundan yüksek bir değerde olduğu kolayca gözlenmektedir. Ancak MSE ve RMSE için inceleme yapıldığında test skorunun eğitim skorundan düşük olduğu görülmektedir. Yine de metriklere ait yüzdesel farkların fazla olmaması sebebiyle herhangi bir aşırı öğrenme ya da yetersiz öğrenme bulunmamaktadır.



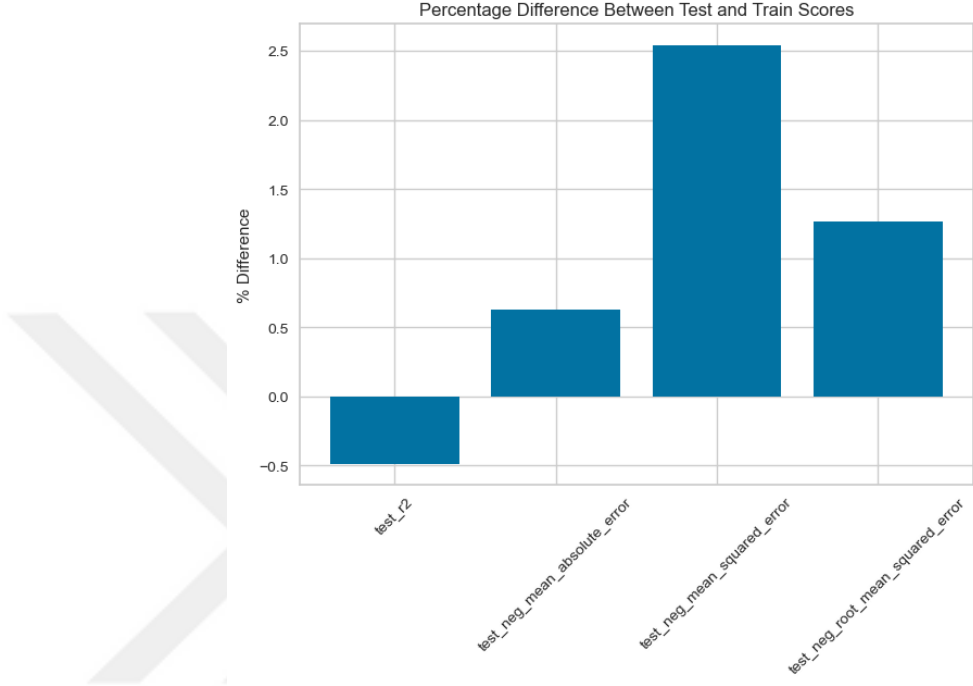
Şekil 5.29. Destek Vektör Regresörü için IEM CO₂ eğitim ve test verileri arasında yüzdelik fark

Şekil 5.29’da ise test ve eğitim skorları arasındaki fark MAE, MSE ve RMSE için fazla değildir. Ancak yukarıda da belirtildiği gibi R^2 metriği için yüzdelik farkın %477,69 olması test ve eğitim verileri arası uyumsuzluğun fazla olduğunu göstermektedir.



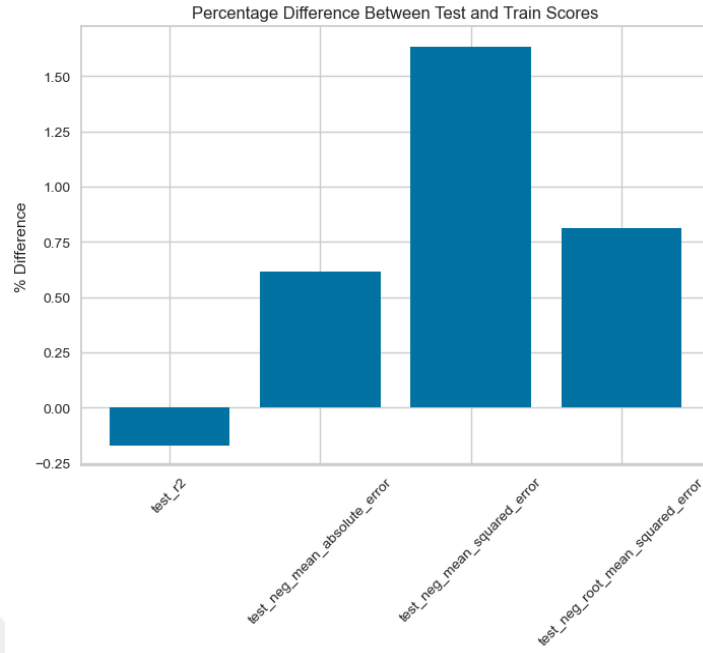
Şekil 5.30. Karar Ağacı için IEM CO₂ eğitim ve test verileri arasında yüzdelik fark

Karar Ağacı için Şekil 5.30 incelendiğinde test ve eğitim skorları arası yüzdelik farkın fazla olmadığı söylenebilir. R^2 için ölçülen değer hariç tüm değerlerde test skoru eğitim skorunu geçmiştir. En çok yüzdelik farka sahip MSE metriği, %1,63'lük bir değer almıştır.



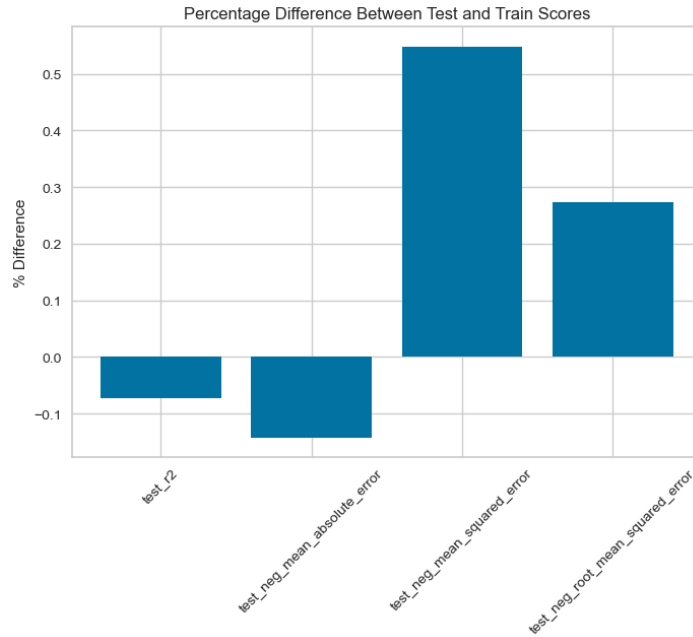
Şekil 5.31. Rastgele Orman için IEM CO₂ eğitim ve test verileri arasında yüzdelik fark

Şekil 5.31’de Rastgele Orman modeli için verilen metriklerle ait yüzdesel fark skorlarına bakıldığında ise Karar Ağacı modelindeki gibi eğitim R^2 test R^2 değerini geçmiş; MAE, MSE ve RMSE için test değeri eğitim değerini geçmiştir. Burada da test ve eğitim skorları arası yüzdelik fark fazla değildir.



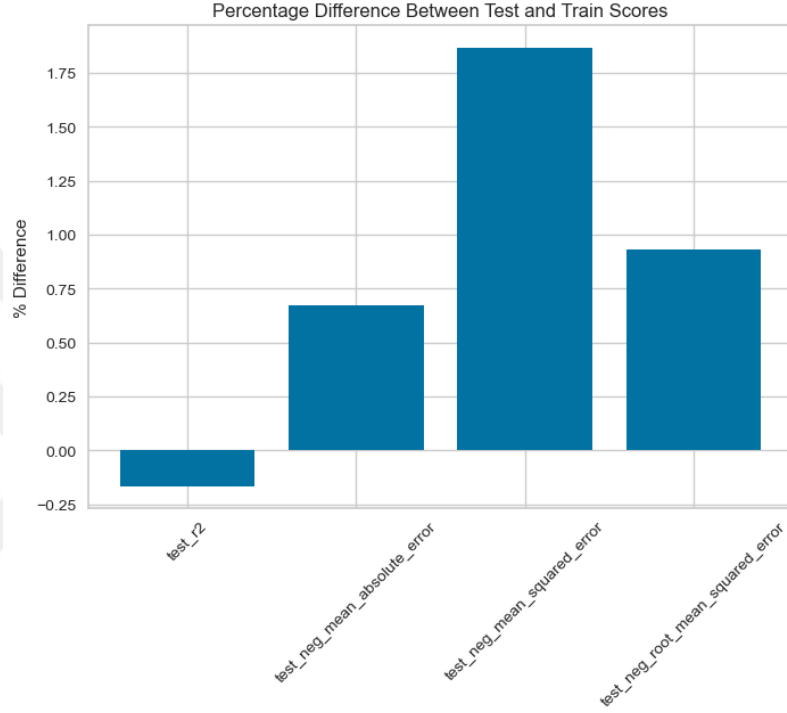
Şekil 5.32. AdaBoost için IEM CO₂ eğitim ve test verileri arasında yüzdelik fark

AdaBoost için Karar Ağacı ve Rastgele Orman modellerindeki bulgulara benzer bulgulara erişilmiştir. Bu modelde eğitim R^2 test R^2 değerini geçmiş, MAE, MSE ve RMSE için ise test değeri eğitim değerini geçmiştir. Burada da test ve eğitim skorları arası yüzdelik farkın fazla olmadığı söylenebilir. En fazla yüzdesel skor farkına sahip değer ise MSE için %2,54 olarak hesaplanmıştır.



Şekil 5.33. Gradyan Artırma için IEM CO₂ eğitim ve test verileri arasında yüzdelik fark

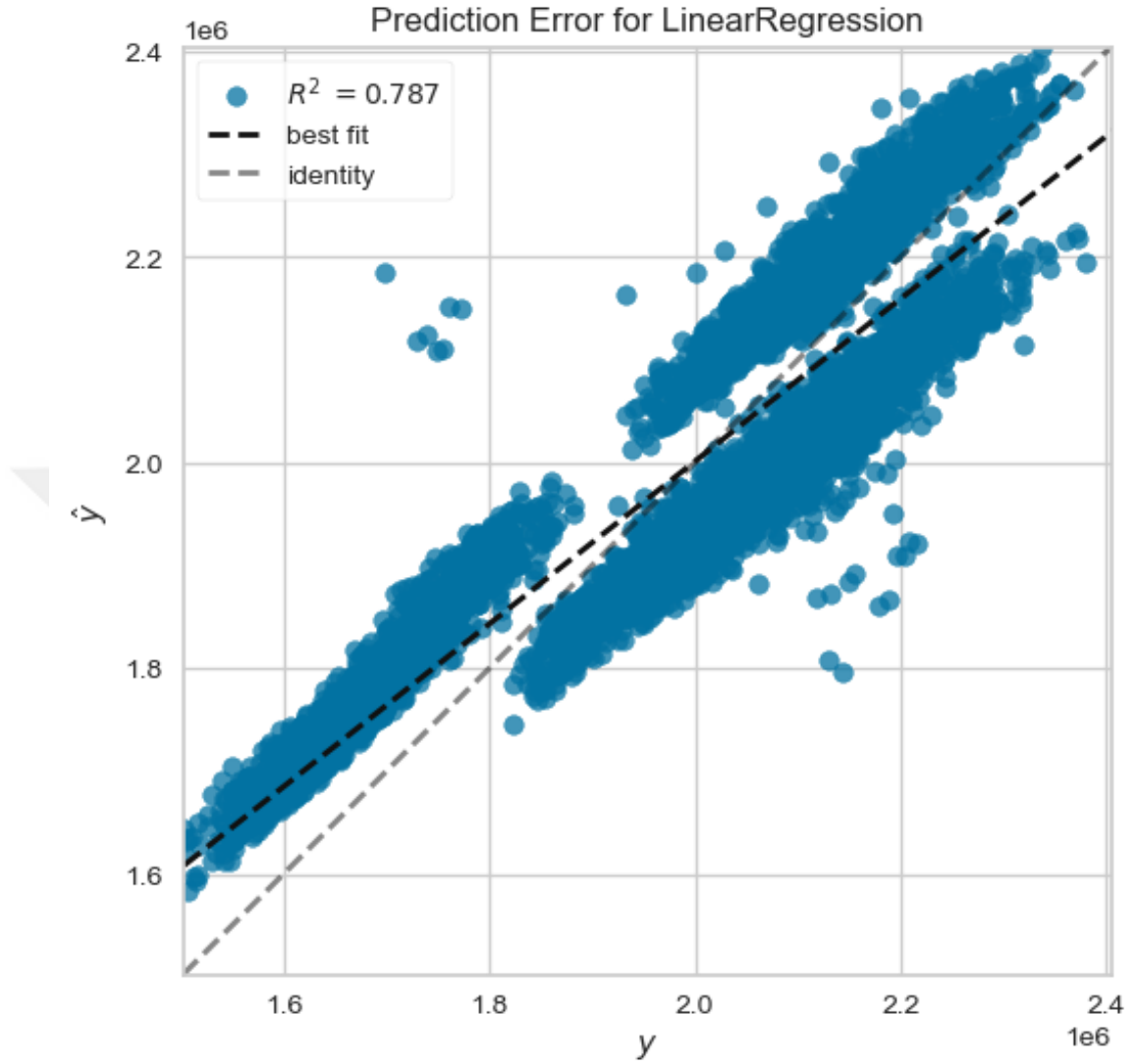
Bu modele ait metriklerde yüzdesel fark değerleri 0'a yakın değer almıştır. Yani test ve eğitim veri kümelerinin birbirleri ile uyumlu olduğu ortaya çıkmıştır. R^2 ve MAE için eğitim skorları test skorlarını geçmiştir; MSE ve RMSE'de ise test skorları eğitim skorlarından fazla değer almıştır. En fazla yüzdesel farka sahip RMSE'dir ve değeri %0,55'tir.



Şekil 5.34. XGBoost için IEM CO₂ eğitim ve test verileri arasında yüzdelik fark

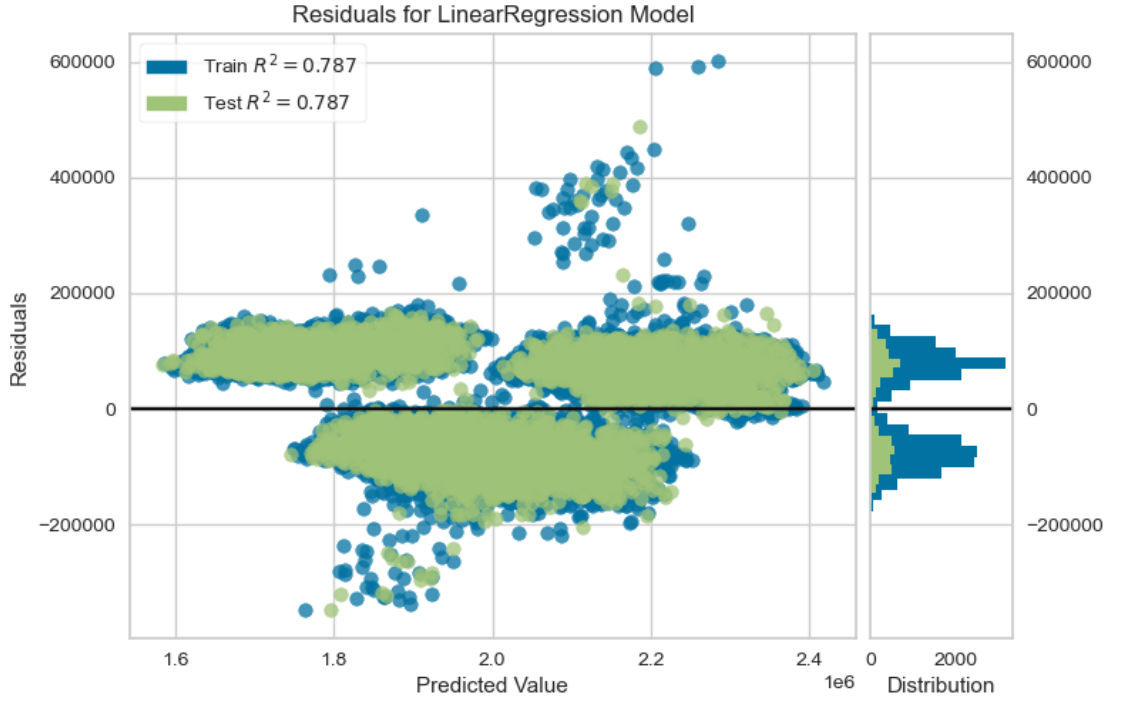
Şekil 5.34'te ise Karar Ağacı, Rastgele Orman ve AdaBoost modellerindeki bulgulara benzer bulgulara erişilmiştir. Bu model için eğitim R^2 test R^2 değerini geçmiştir; MAE, MSE ve RMSE için ise test değeri eğitim değerini geçmiştir. Yani test ve eğitim skorları arası yüzdelik farkın fazla olmadığı söylenebilir. En fazla yüzdesel skor farkına sahip değer ise MSE %1,87 olarak hesaplanmıştır.

5.2.3. Final Model Sonrası Değerlendirme



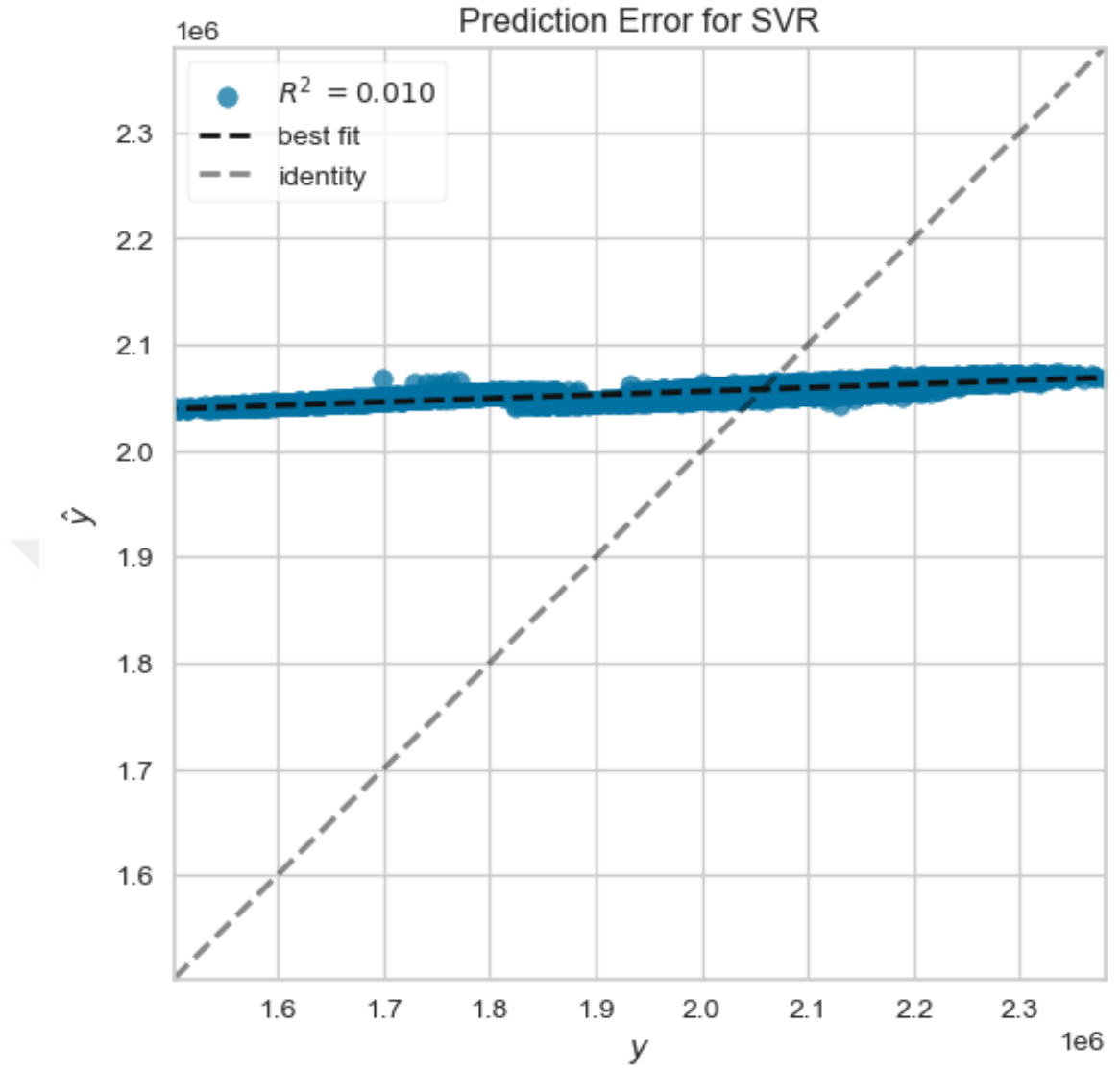
Şekil 5.35. IEM CO₂ için Lineer Regresyon modeli tahmin hatası grafiği

Şekil 5.35'te IEM CO₂ parametresi için Lineer Regresyon modelinin tahmin hatasına ait saçılım grafiği görülmektedir. Bu parametre için R^2 0,7870 ve RMSE 87935,6034 çıkmıştır. Denklem 5.1'e göre hesaplanan RMSE hata oranı ise 0,0438 çıkmıştır. Yani R^2 değerine göre yaklaşık %21, RMSE'ye göre yaklaşık %4'lük bir hata oranı bulunmuştur. Bu durum çok başarılı bir modelleme yapılamadığını göstermektedir.



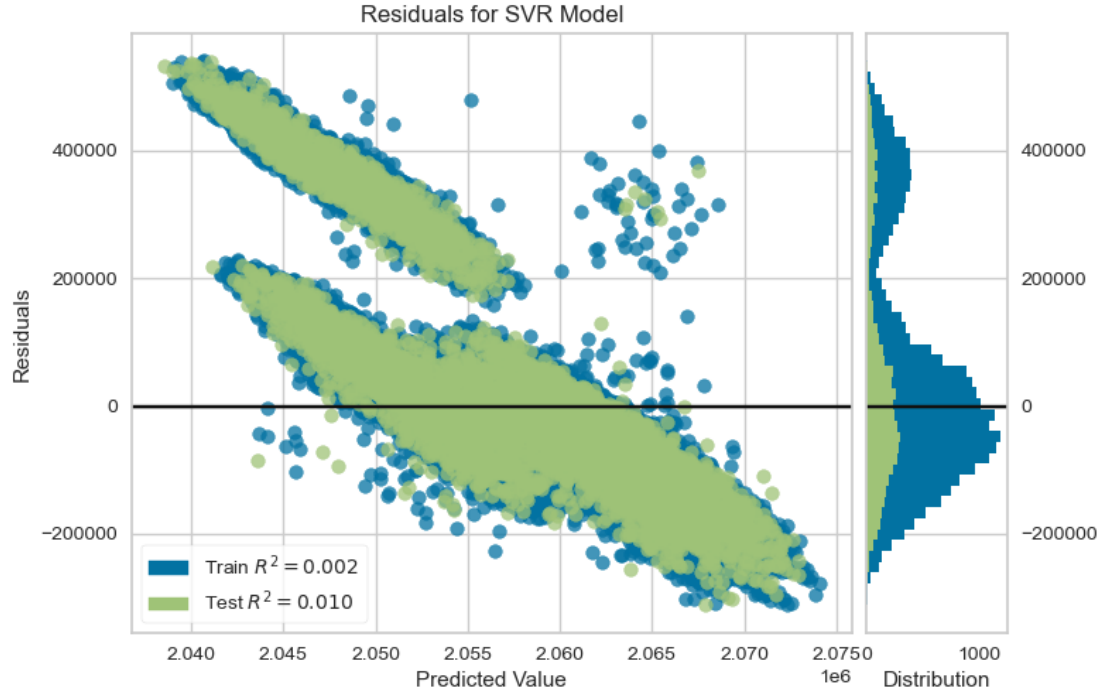
Şekil 5.36. IEM CO₂ için Lineer Regresyon modelinde artık değerler dağılımı

Şekil 5.36’da ise grafik “pattern” yani desenlere sahiptir. Önceki grafikte de belirtilen modelin başarılı olmama durumu burada da çıkmıştır. Bu durumda modelde aşırı öğrenme tespit edilmiştir.



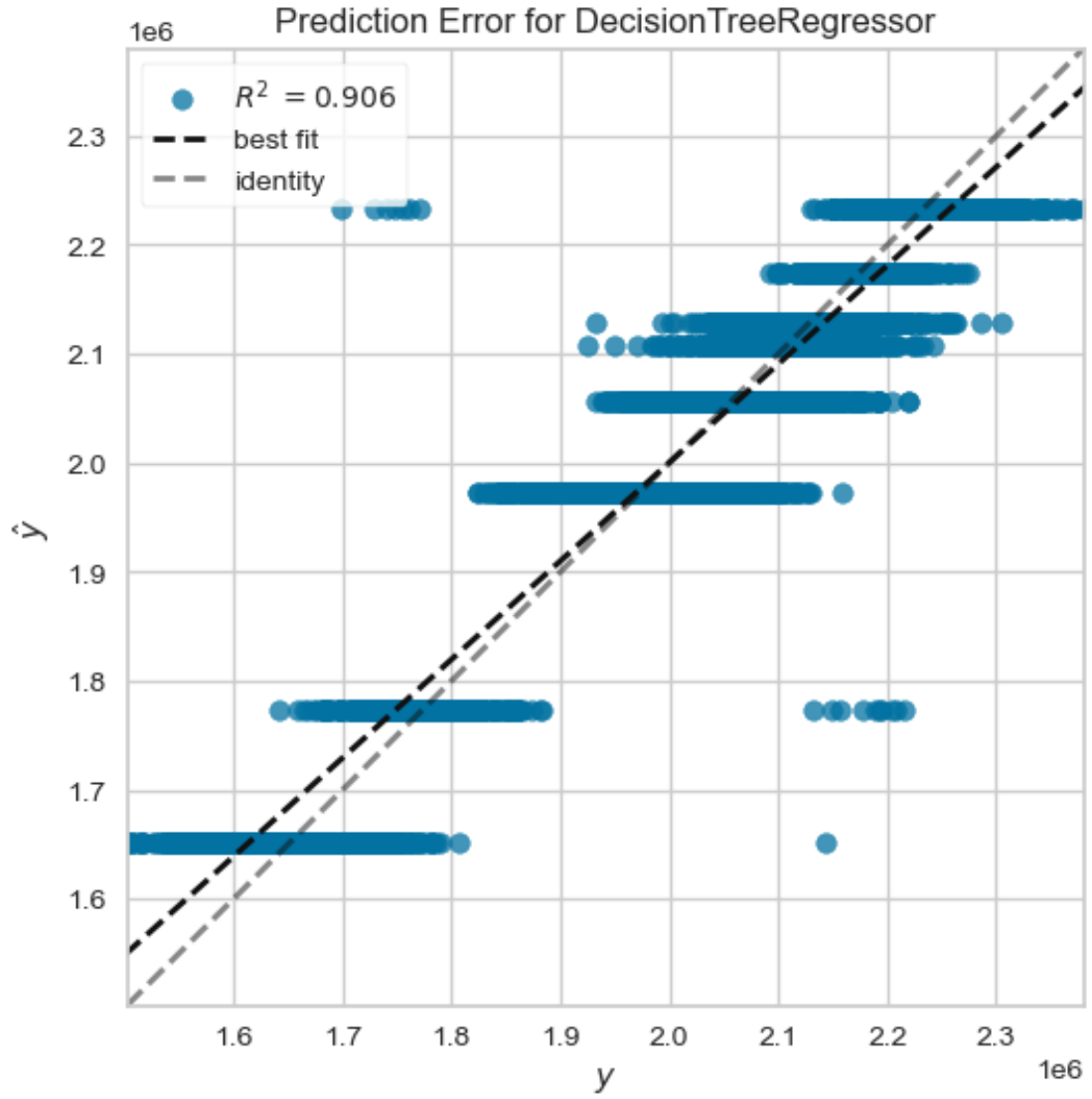
Şekil 5.37. IEM CO₂ için Destek Vektör Regresörü modeli tahmin hatası grafiği

Şekil 5.37’de IEM CO₂ parametresi Destek Vektör Regresörü modelinin tahmin hatasının görselleştirilmiş hali bulunmaktadır. Buna göre, başarısız bir modelleme bulunmaktadır.



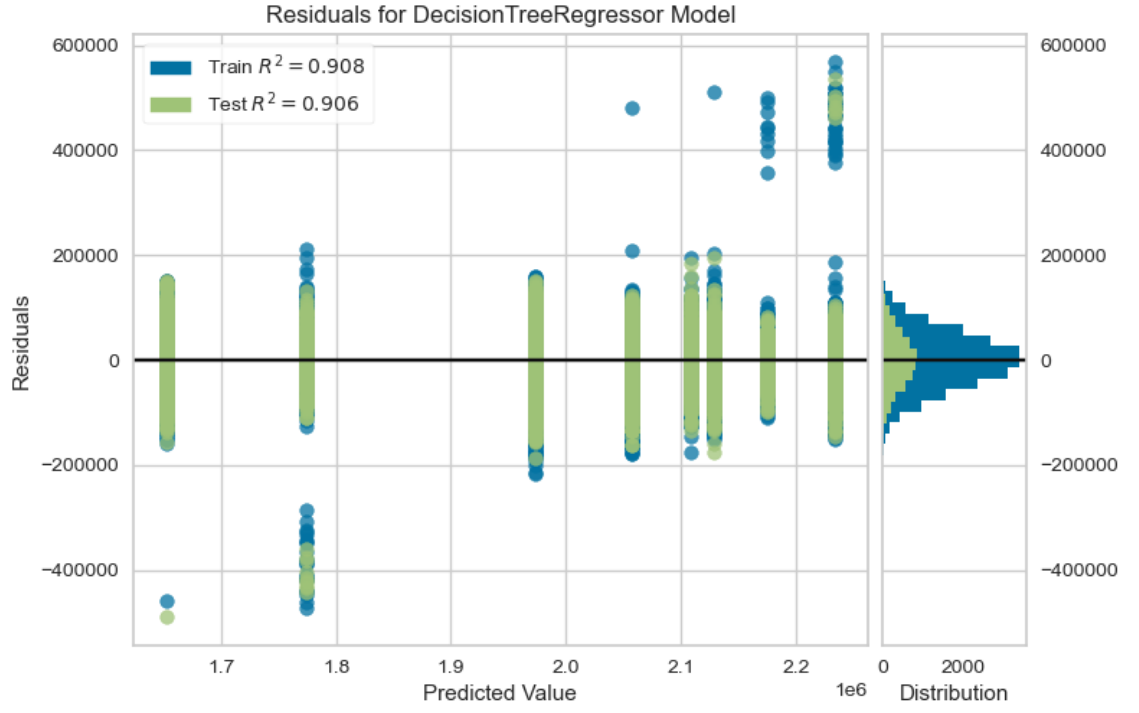
Şekil 5.38. IEM CO₂ için Destek Vektör Regresörü modelinde artık değerler dağılımı

Şekil 5.38'e bakıldığında grafik "pattern" yani desen dağılımına sahiptir. Önceki grafikte görülen bariz aşırı öğrenme durumu, bu grafikte de görülmektedir.



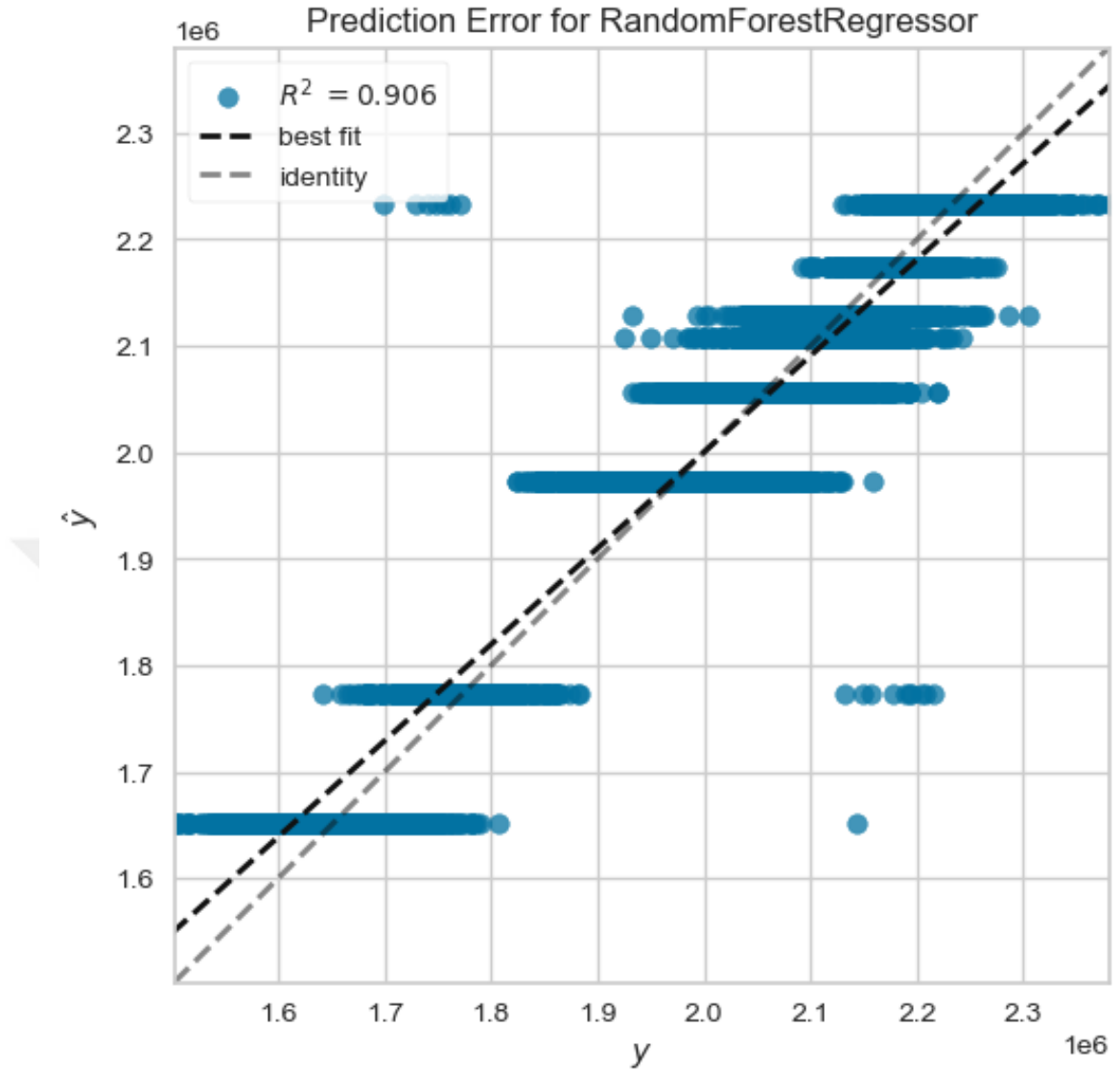
Şekil 5.39. IEM CO₂ için Karar Ağacı modeli tahmin hatası grafiği

Şekil 5.39’da ise modelde aşırı uyum olduğu görülmektedir. R^2 0,9061 ve RMSE 58378,6083 çıkmıştır. Yani R^2 değerime göre yaklaşık %9’luk, RMSE’ye göre yaklaşık %3’lük bir hata oranı bulunmuştur. Metrik değerleri bu parametre için Karar Ağacının başarısız bir model olduğunu gösterir.



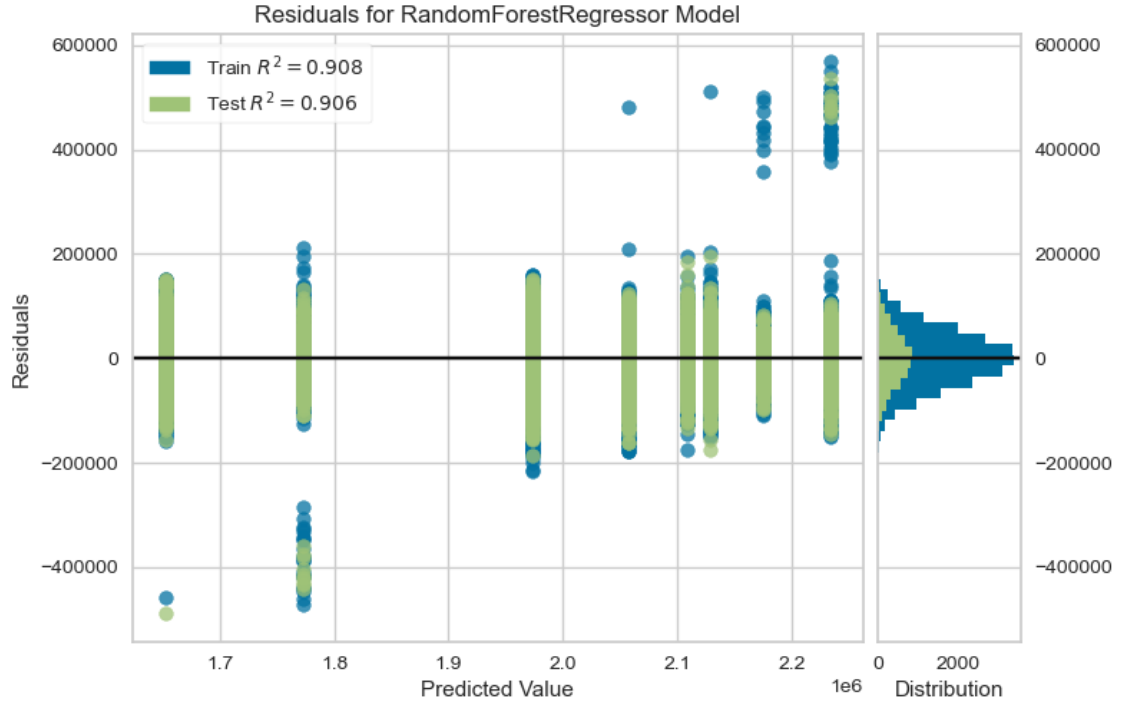
Şekil 5.40. IEM CO₂ için Karar Ağacı modelinde artık değerler dağılımı

Şekil 5.40'a göre grafiğin bir “pattern” yani kalıp/desen yapısına sahip oluşu modelin aşırı öğrenme durumunun ve başarısızlığının göstergesidir.



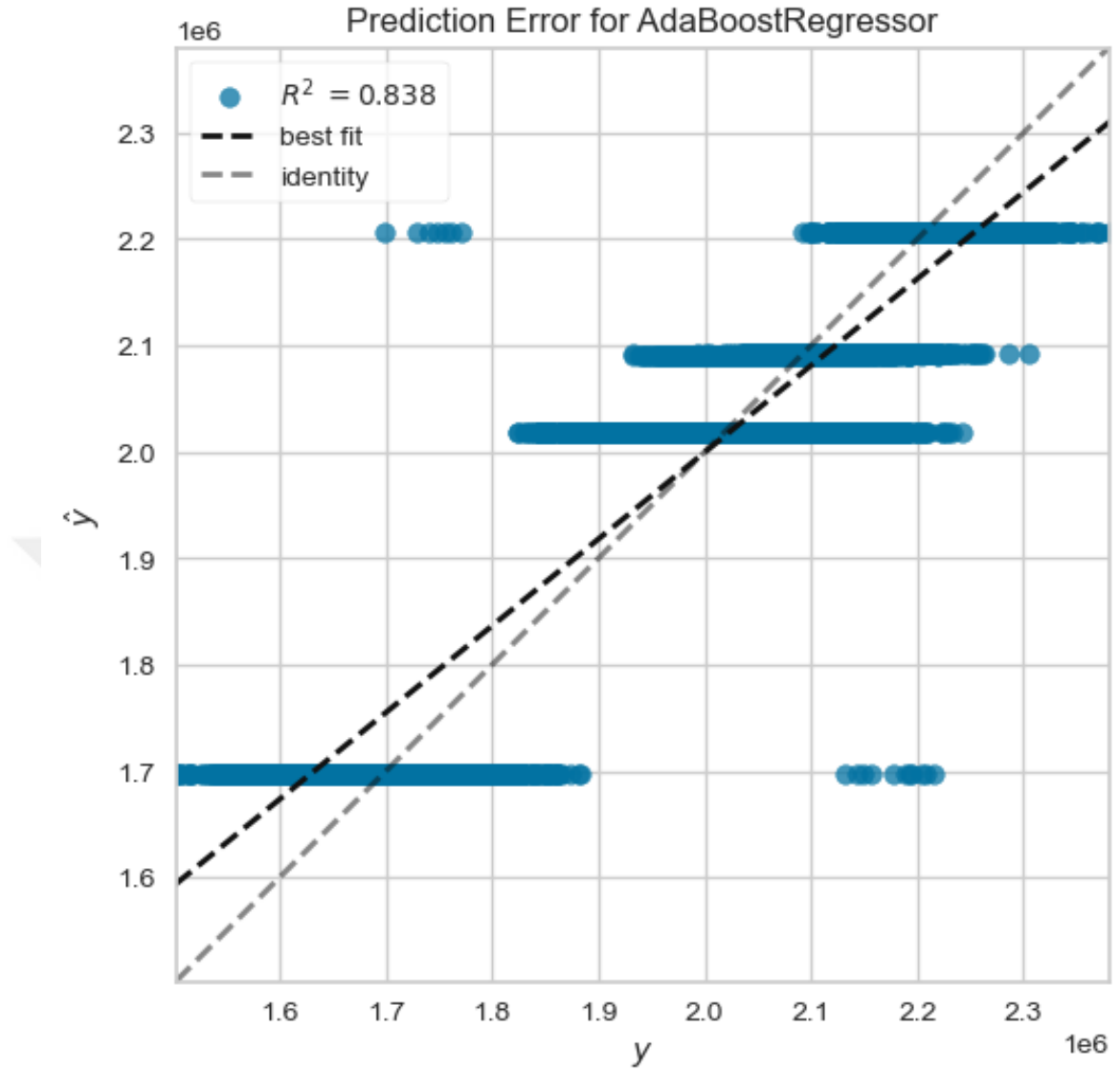
Şekil 5.41. IEM CO₂ için Rastgele Orman modeli tahmin hatası grafiği

Şekil 5.41’de ise Rastgele Orman modeli için tipik bir aşırı uyum modeline ait bir grafik görülmektedir. R^2 0,9061 ve RMSE 58378,6083 çıkmıştır. Yani R^2 değerine göre yaklaşık %9’luk, RMSE’ye göre yaklaşık %3’lük bir hata oranı bulunmuştur. Hata oranları düşük olsa da aşırı uyumun görülmesi, bu modeli başarılı kabul etmememize sebep olmaktadır.



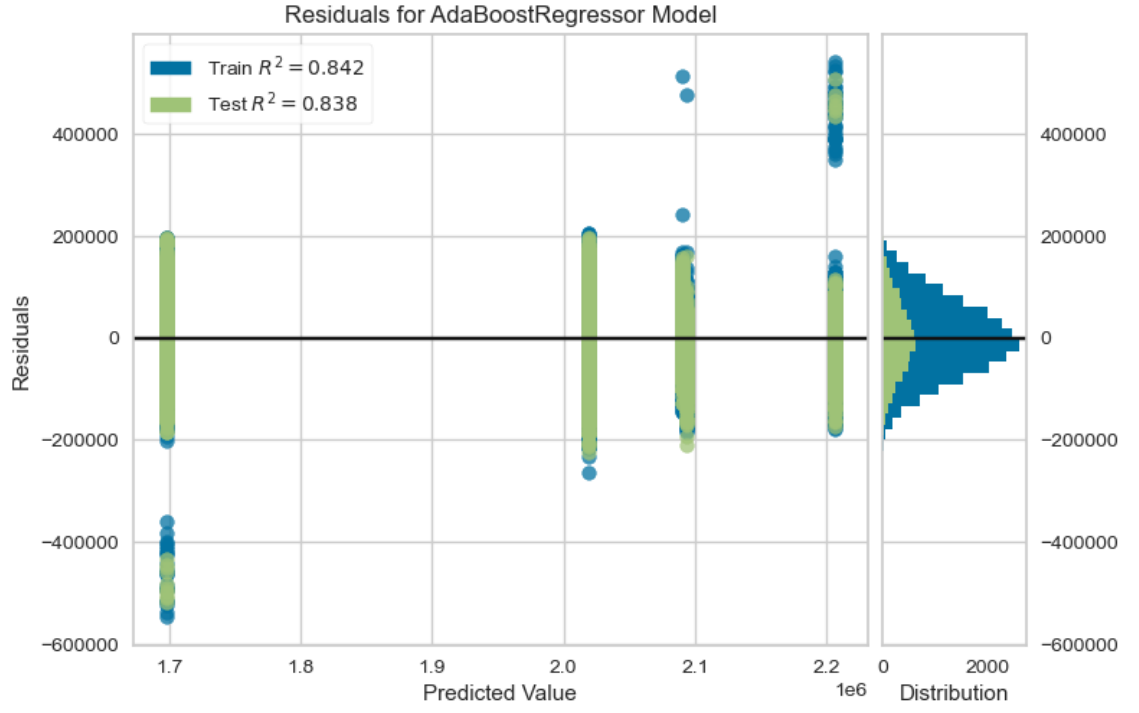
Şekil 5.42 IEM CO₂ için Rastgele Orman modelinde artık değerler dağılımı

Şekil 5.42'ye göre değerler Karar Ağacı ile benzer bir dağılıma sahip olarak aşırı öğrenme sebebiyle desen ya da kalıp şeklinde dağılmıştır. Yani bu parametre için Rastgele Orman modeli başarısız kalmıştır.



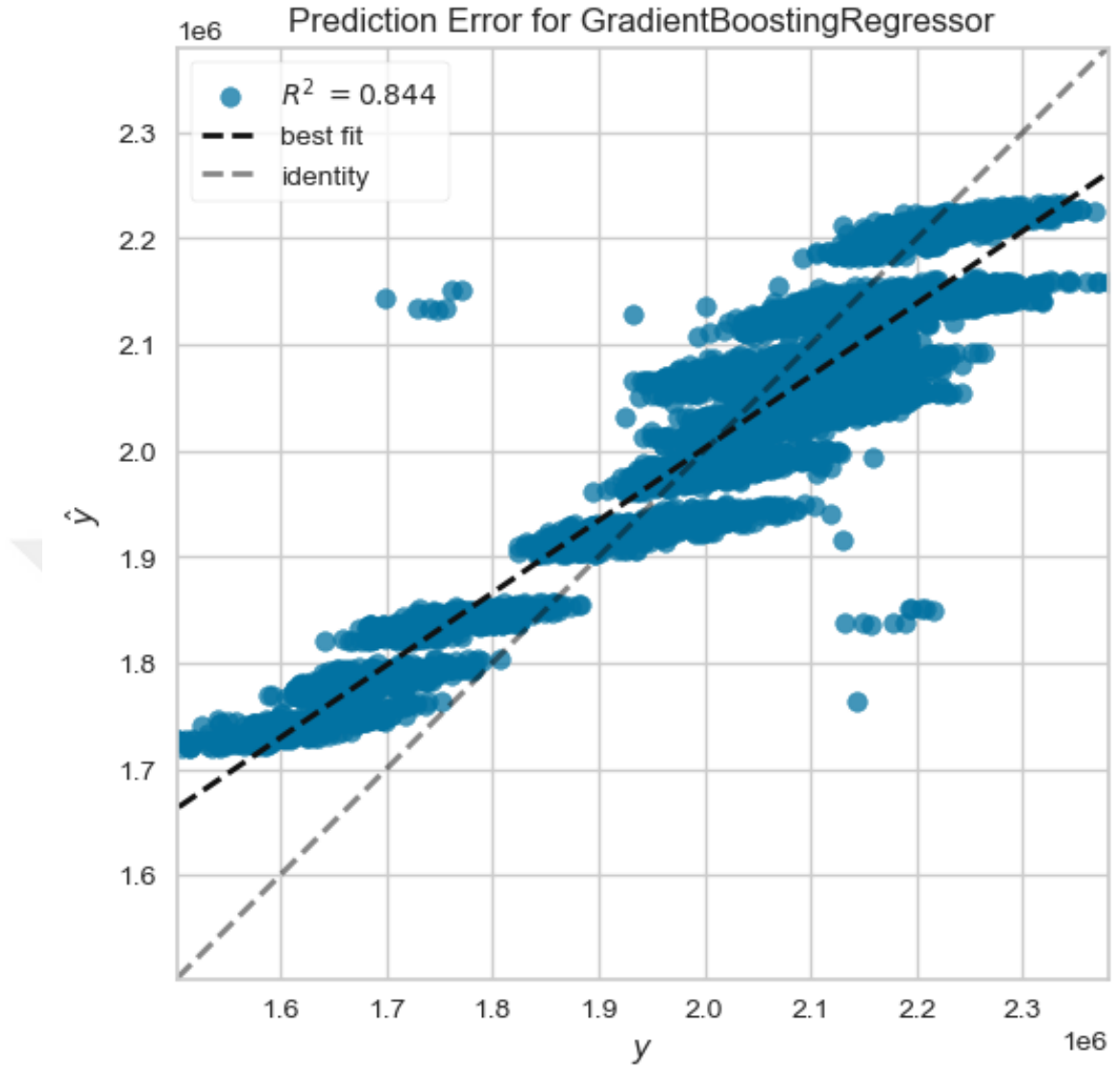
Şekil 5.43. IEM CO₂ için AdaBoost modeli tahmin hatası grafiği

Şekil 5.43'te görüldüğü üzere AdaBoost modelinde bariz bir şekilde aşırı uyum oluşmuştur. R^2 (0,8378) değerine göre yaklaşık %16'lık, RMSE'ye (76729,2963) göre yaklaşık %4'lük bir hata oranı bulunmuştur. Görüldüğü gibi R^2 metriğinin de düşük olması modelin başarısız olduğunu göstermektedir.



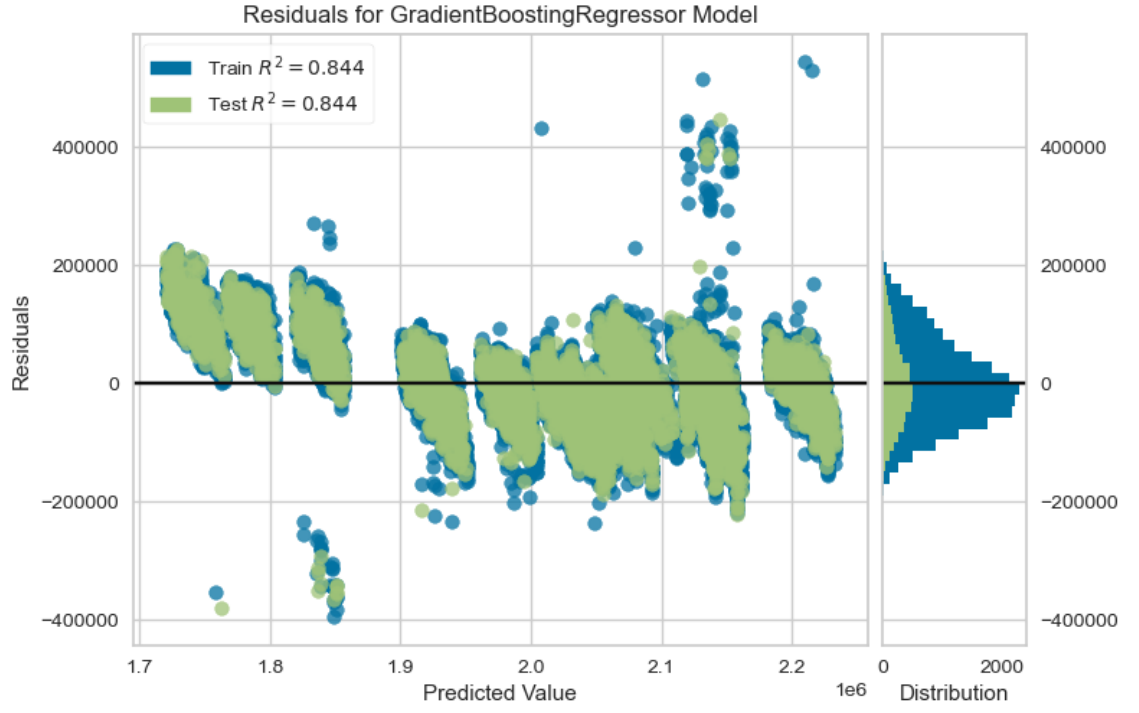
Şekil 5.44. IEM CO₂ için AdaBoost modelinde artık değerler dağılımı

Şekil 5.44'e göre değerler aşırı uyum sebebiyle pattern yani desenli bir yapıda dağılmıştır. Yani bu parametre için AdaBoost modeli başarısız kalmıştır



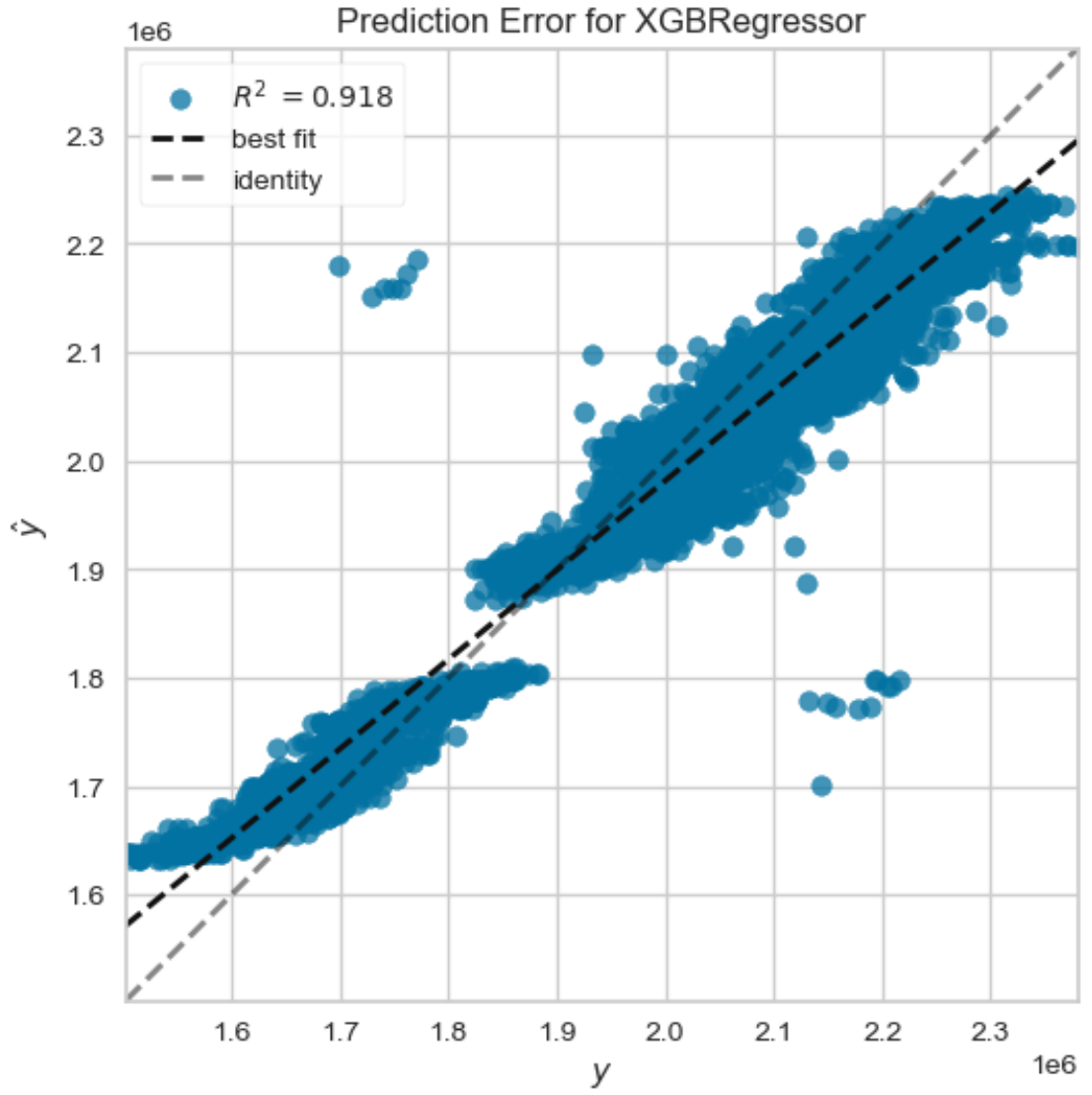
Şekil 5.45. IEM CO₂ için Gradyan Artırma modeli tahmin hatası grafiği

Şekil 5.45'te Gradyan Artırma modeli için tahmin hata grafiğindeki verilerin düzgün şekilde dağıldığı söylenebilir. Modelin aşırı uyum durumu olmadığı varsayılabilir. Ancak R^2 (0,8444) değerine göre yaklaşık %16'lık, RMSE'ye (75150,3890) göre yaklaşık %4'lük bir hata oranı bulunmuştur. R^2 değerinin 0,90'dan düşük olma durumunda modeli başarılı saymamak daha sağlıklı olacaktır.



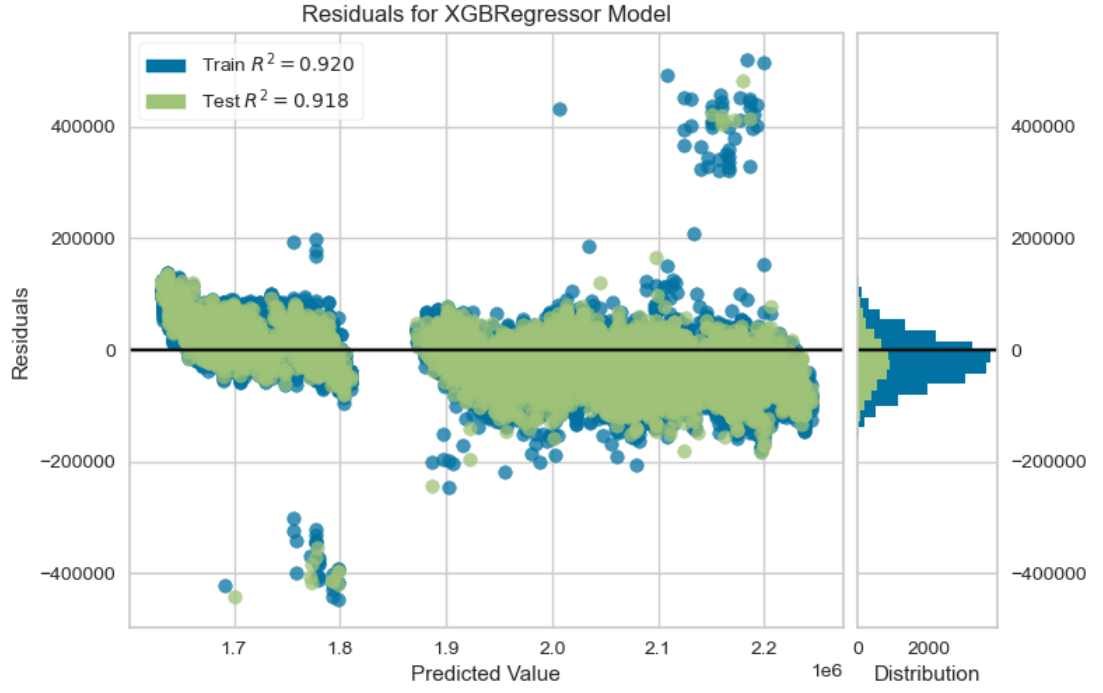
Şekil 5.46. IEM CO₂ için Gradyan Artırma modelinde artık değerler dağılımı

Şekil 5.46'da noktasal dağılımın belli bir kalıp/desen yani pattern yapısında tam anlamıyla dağılmadığı gözlemlenmektedir. Yani modelleme zayıf kalsa da tamamen başarısız olmadığı görülmektedir. Ancak önceden de belirtildiği gibi R² skorunun yeterli olmaması sebebiyle modelin başarısız olduğu söylenebilir.



Şekil 5.47. IEM CO₂ için XGBoost modeli tahmin hatası grafiği

Şekil 5.47’de görüldüğü üzere XGBoost modeli Gradyan Artırma modeline benzer şekilde eğri etrafında toplanmıştır. R^2 (0,9185) değerine göre yaklaşık %8’lik, RMSE’ye (54408,3471) göre yaklaşık %3’lük bir hata oranı bulunmuştur. Aşırı uyum sayılabilecek bir dağılım da bulunmadığından bu model, başarılı olarak değerlendirilmiştir.



Şekil 5.48. IEM CO₂ için XGBoost modelinde artık değerler dağılımı

Şekil 5.48’de artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmına bakıldığında bir desen veya kalıp olmadığı görülmektedir. Sütun dağılımında da normal dağılım bulunduğu için artık değerler açısından bu modelin uygun olduğu tespit edilmiştir.

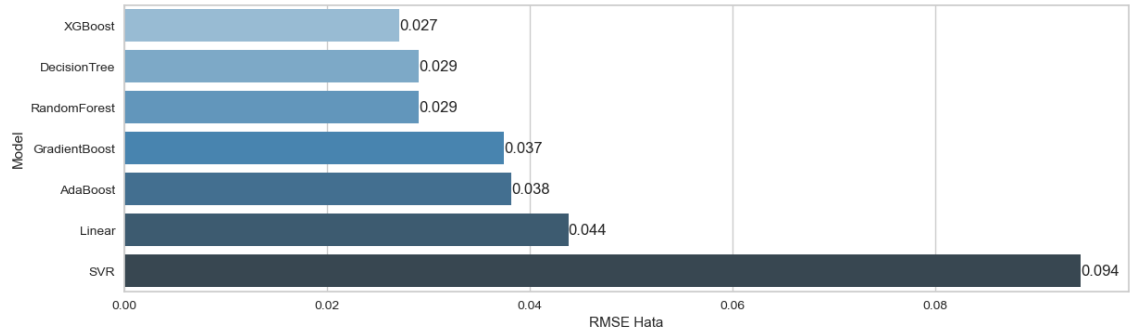
5.2.4. Modellerin Karşılaştırılması

Tablo 5.11 IEM CO₂ için makine öğrenmesi modellerinin hata metrikleri değerleri

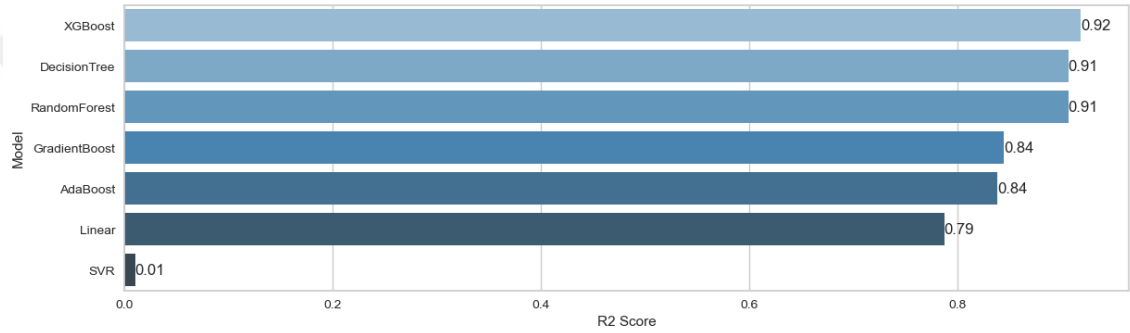
	R ²	MAE	MSE	RMSE
Lineer Regresyon GridSearchCV Eğitim	7,866148e-01	8,115230e+04	7,749041e+09	8,802864e+04
Lineer Regresyon GridSearchCV Test	7,869817e-01	8,123977e+04	7,732670e+09	8,793560e+04
SVR GridSearchCV Eğitim	1,762122e-03	1,410657e+05	3,625080e+10	1,903964e+05
SVR GridSearchCV Test	1,017963e-02	1,406203e+05	3,593097e+10	1,895547e+05
Karar Ağacı GridSearchCV Eğitim	9,076595e-01	4,320539e+04	3,353328e+09	5,790792e+04
Karar Ağacı GridSearchCV Test	9,061153e-01	4,346659e+04	3,408062e+09	5,837861e+04
Rastgele Orman GridSearchCV Eğitim	9,077333e-01	4,317874e+04	3,350645e+09	5,788475e+04
Rastgele Orman GridSearchCV Test	9,061894e-01	4,344424e+04	3,405373e+09	5,835557e+04
AdaBoost GridSearchCV Eğitim	8,414300e-01	5,932572e+04	5,758436e+09	7,588436e+04
AdaBoost GridSearchCV Test	8,373362e-01	5,970058e+04	5,904778e+09	7,684255e+04
Gradyan Artırma GridSearchCV Eğitim	8,908357e-01	4,811332e+04	3,964278e+09	6,296251e+04
Gradyan Artırma GridSearchCV Test	8,901941e-01	4,804491e+04	3,986009e+09	6,313485e+04
XGBoost GridSearchCV Eğitim	9,199768e-01	4,036919e+04	2,906028e+09	5,390758e+04
XGBoost GridSearchCV Test	9,184510e-01	4,064018e+04	2,960268e+09	5,440835e+04

Tablo 5.12 IEM CO₂ için performansa göre sıralanmış RMSE ve R² skorları

	Model	RMSE Hata	R ² Skoru
1	XGBoost	0,0271	0,9185
2	Karar Ağacı	0,0291	0,9061
3	Rastgele Orman	0,0291	0,9061
4	Gradyan Artırma	0,0374	0,8444
5	AdaBoost	0,0382	0,8378
6	Lineer Regresyon	0,0438	0,7870
7	SVR	0,0944	0,0102



Şekil 5.49. IEM CO₂ için makine öğrenme modelleri RMSE hata oranlarının karşılaştırılması



Şekil 5.50. IEM CO₂ için makine öğrenme modelleri R² skorlarının karşılaştırılması

Tablo 5.11 ve 5.12’de ise en az hata oranı ve en iyi tahmin sonuçları XGBoost ile en çok hata oranı ise Destek Vektör Regresörü ile elde edilmiştir.

En başarılı model olarak XGBoost modelinin RMSE’ye göre 0,027’lik bir hata oranı vardır ve R² skoru ise 0,92 çıkmıştır. Bu modeli, 0,029’luk RMSE hata oranı, 0,91 R² skoru ve eşit performansla Karar Ağacı ve Rastgele Orman takip etmektedirler. Ancak aşırı uyum gösterdikleri için bu modeller, başarılı sayılamazlar. En başarısız model ise %9 RMSE hata oranı ve 0,01’lik R² değeri ile Destek Vektör Regresörü’nün olduğu görülmektedir.

Sonuç olarak parametre için XGBoost tek başarılı model olarak seçilmiştir.

5.2.5. Modellere Ait Tahmin Değerlerinin Karşılaştırılması

Tablo 5.13 IEM CO₂ için model tahmin değerlerinin karşılaştırılması

	Simülasyon Sonucu	Lineer Regresyon	SVR	Karar Ağacı	Rastgele Orman	AdaBoost	Gradyan Artırma	XGBoost
0	2081881,95	2144768,	2064227,41	2057132,53	2057189,59	2091779,37	2077185,75	2040712,4
100	1951360,65	1899328,	2048061,41	1973609,20	1973708,49	2018771,01	1937845,81	1935003,2
200	2047506,98	2121472,	2062579,41	2128508,53	2128495,87	2093921,39	2100133,34	2081161,4

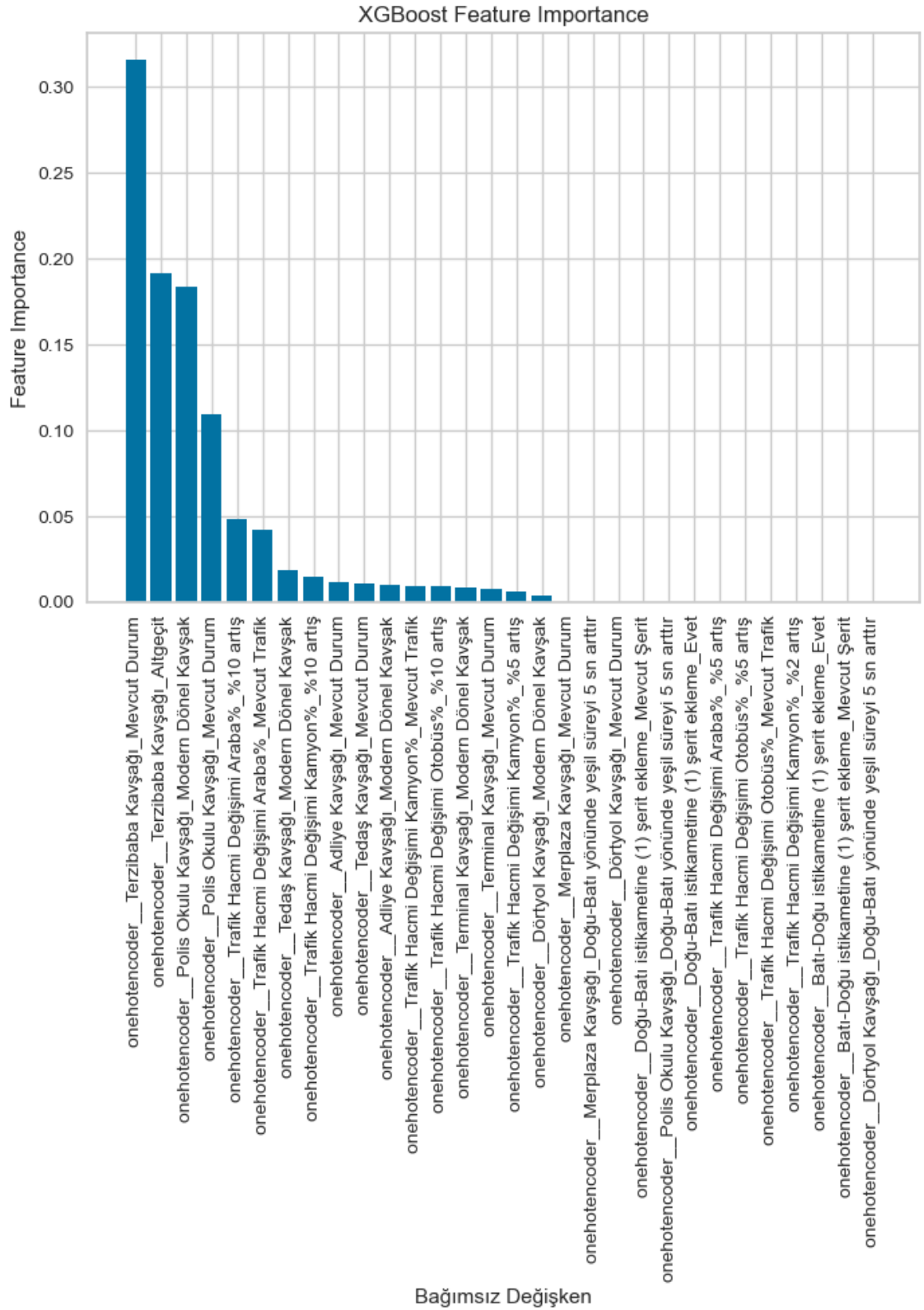
Tablo 5.13'e göre 0. ve 100. senaryoların gerçek değere en yakın tahminleri Gradyan Artırma modeli, en uzak tahminler ise Lineer Regresyon ile yapılmıştır. 200. senaryo için Destek Vektör Regresörü ile en yakın tahmin, Karar Ağacı ile en kötü tahmin modeli yapılmıştır. Bu tabloya göre genel olarak en başarılı tahminleri Destek Vektör Regresörü ve Gradyan Artırma yardımıyla yapılmıştır. En başarısız tahminler ise AdaBoost ve Karar Ağacı tarafından yapılmıştır.

5.2.6. Başarılı Modellerin Etki Katsayılarının Değerlendirilmesi

Nitelik önemi değerleri, modelin eğitimi sırasında kullanılan bağımsız değişkenlerin tahminler üzerinde ne kadar etkili olduğunu gösterir.

```
[21:54:28] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.
onehotencoder__Terzibaba Kavşağı_Mevcut Durum: 0.3157430589199066
onehotencoder__Terzibaba Kavşağı_Altgeçit: 0.19160284101963043
onehotencoder__Polis Okulu Kavşağı_Modern Dönel Kavşak: 0.18383082747459412
onehotencoder__Polis Okulu Kavşağı_Mevcut Durum: 0.10923486948013306
onehotencoder__Trafik Hacmi Değişimi Araba%_10 artış: 0.04832526668906212
onehotencoder__Trafik Hacmi Değişimi Araba%_Mevcut Trafik: 0.04195234924554825
onehotencoder__Tedaş Kavşağı_Modern Dönel Kavşak: 0.01847015880048275
onehotencoder__Trafik Hacmi Değişimi Kamyon%_10 artış: 0.014879802241921425
onehotencoder__Adliye Kavşağı_Mevcut Durum: 0.011239615269005299
onehotencoder__Tedaş Kavşağı_Mevcut Durum: 0.010473242960870266
onehotencoder__Adliye Kavşağı_Modern Dönel Kavşak: 0.009795538149774075
onehotencoder__Trafik Hacmi Değişimi Kamyon%_Mevcut Trafik: 0.009301038458943367
onehotencoder__Trafik Hacmi Değişimi Otobüs%_10 artış: 0.009293211624026299
onehotencoder__Terminal Kavşağı_Modern Dönel Kavşak: 0.008417680859565735
onehotencoder__Terminal Kavşağı_Mevcut Durum: 0.007627058308571577
onehotencoder__Trafik Hacmi Değişimi Kamyon%_5 artış: 0.006150806788355112
onehotencoder__Dört Yol Kavşağı_Modern Dönel Kavşak: 0.0036626183427870274
onehotencoder__Merplaza Kavşağı_Mevcut Durum: 0.0
onehotencoder__Merplaza Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 0.0
onehotencoder__Dört Yol Kavşağı_Mevcut Durum: 0.0
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Mevcut Şerit: 0.0
onehotencoder__Polis Okulu Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 0.0
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Evet: 0.0
onehotencoder__Trafik Hacmi Değişimi Araba%_5 artış: 0.0
onehotencoder__Trafik Hacmi Değişimi Otobüs%_5 artış: 0.0
onehotencoder__Trafik Hacmi Değişimi Otobüs%_Mevcut Trafik: 0.0
onehotencoder__Trafik Hacmi Değişimi Kamyon%_2 artış: 0.0
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Evet: 0.0
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Mevcut Şerit: 0.0
onehotencoder__Dört Yol Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 0.0
```

Şekil 5.51. XGBoost modeli için bağımsız değişkenlerin nitelik önemi değerleri



Şekil 5.52. XGBoost modeli için bağımsız değişkenlerin nitelik önemi değerleri sütun grafiği

Şekil 5.51 ve Şekil 5.52’ e göre, IEM CO₂ parametresi üzerindeki en fazla etkinin Terzibaba kavşağına ait katsayı değeri 0,30’ dan fazla olan mevcut durum ve katsayı değeri 0,20’ ye yakın olan altgeçit değişkenidir. Bu değişkenleri Polis Okulu Kavşağı’ nın Modern Dönel Kavşak olduğu ve otomobil hacmindeki değişiklikler ile ilgili değişkenler takip etmektedir. Bu değerler birbirlerine yakın olduğundan benzer bir etki gösterebilirler.

Sonuç olarak Kavşak değişkenleri göz önünde bulundurulduğunda IEM CO₂ parametresi için Terzibaba Kavşağı’ ndaki değişiklikler en fazla etki oluşturacaktır. Taşıt hacmindeki değişikliklerde ise en fazla otomobil hacmindeki değişiklikler etki edecektir.

Dört yol ve Merplaza Kavşağı ile ilişkili değişkenlerin, kamyon ve otobüs hacmi ile ilişkili değişkenler ve şerit sayıları ile ilişkili değişkenlerin 0’ a yakın çıkmıştır. Dolayısıyla bu değişkenlerin, XGBoost modelinde IEM CO₂ parametresi için önem arz etmediği gözlenmiştir.

5.3. Toplam Seyahat Süresi Parametresi İçin Bulgular

5.3.1. GridSearchCV öncesi ve sonrası performans değerlendirme

Tablo 5.14. Toplam seyahat süresi parametresi için GridSearchCV öncesi performans değerleri

	R ² Eğitim	R ² Test	MAE Eğitim	MAE Test	MSE Eğitim	MSE Test	RMSE Eğitim	RMSE Test
Lineer Regresyon	0,8764	0,8866	6,6377	6,5478	59,4237	53,4820	7,7087	7,3132
SVR	0,9668	0,9734	0,9012	0,9433	15,9793	12,5569	3,9974	3,5436
Karar Ağacı	1,0	0,9545	0,0	1,4726	0,0	21,4682	0,0	4,6334
Rastgele Orman	0,9951	0,9762	0,4634	1,2452	2,3522	11,2182	1,5337	3,3494
AdaBoost	0,7246	0,7337	8,9927	8,7992	132,3690	125,6196	11,5052	11,2080
Gradyan Artırma	0,9660	0,9726	1,4776	1,4713	16,3469	12,9315	4,0431	3,5960
XGBoost	0,9661	0,9727	1,4777	1,4674	16,2861	12,8634	4,0356	3,5866

Tablo 5.14’ te toplam seyahat süresi Parametresi için test ve eğitim veri kümelerinin belirtilen makine öğrenmesi metotları ile elde edilen hata metrikleri bulunmaktadır. Diğer

parametreler gibi burada da yüksek performans değerleri sağlanmıştır. Genel olarak Destek Vektör Regresörü'nün iyi sonuç çıkardığı söylenebilir.

Tablo 5.15. Toplam seyahat süresi parametresi için GridSearchCV sonrası performans değerleri

	R^2 Eğitim	R^2 Test	MAE Eğitim	MAE Test	MSE Eğitim	MSE Test	RMSE Eğitim	RMSE Test
Lineer Regresyon	0,8763	0,8866	6,6393	6,5483	59,4420	53,5015	7,7099	7,3145
SVR	0,9668	0,9734	0,9012	0,9433	15,9793	12,5569	3,9974	3,5436
Karar Ağacı	0,8356	0,8388	6,8526	6,8695	79,0117	76,0608	8,8889	8,7213
Rastgele Orman	0,7718	0,7727	8,1938	8,2600	109,6910	107,2640	10,4733	10,3568
AdaBoost	0,8430	0,8471	6,6549	6,6399	75,4918	72,1570	8,6886	8,4945
Gradyan Artırma	0,8662	0,8742	6,0535	5,9691	64,3259	59,3394	8,0203	7,7032
XGBoost	0,9193	0,9258	4,4300	4,4193	38,7941	35,0202	6,2285	5,9178

Tablo 5.15'te verilmiş olan GridSearchCV ile düzenlenen modellemelerde daha tutarlı modellemeler elde edilmiştir. Diğer parametrelerde olduğu gibi algoritmalara ait hata metrik değerleri birbirine daha da yaklaşmıştır. GridSearchCV sonrası da en iyi performans Destek Vektör Regresörü'ne aittir.

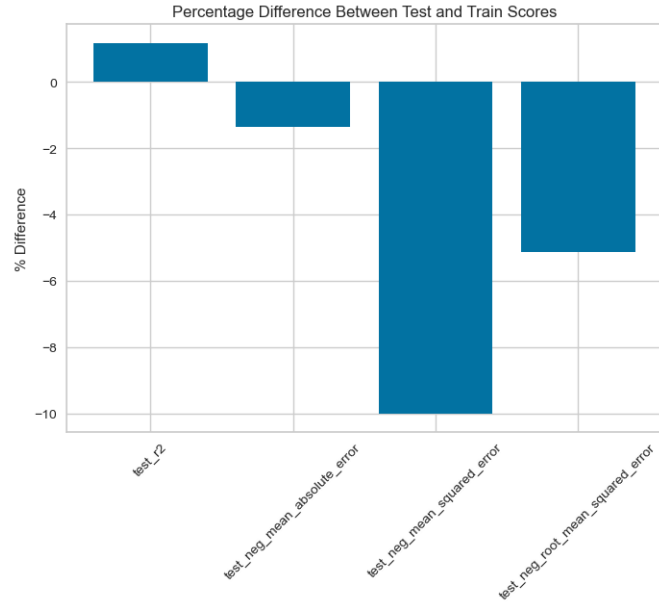
5.3.2. Test ve eğitim skorları arası yüzdesel farklar

Toplam seyahat süresi parametresine ait yüzdelik fark skorları Tablo 5.16'da verilmiştir.

Tablo 5.16. Toplam seyahat süresi parametresi için test ve eğitim skorları arası yüzdelik farklar

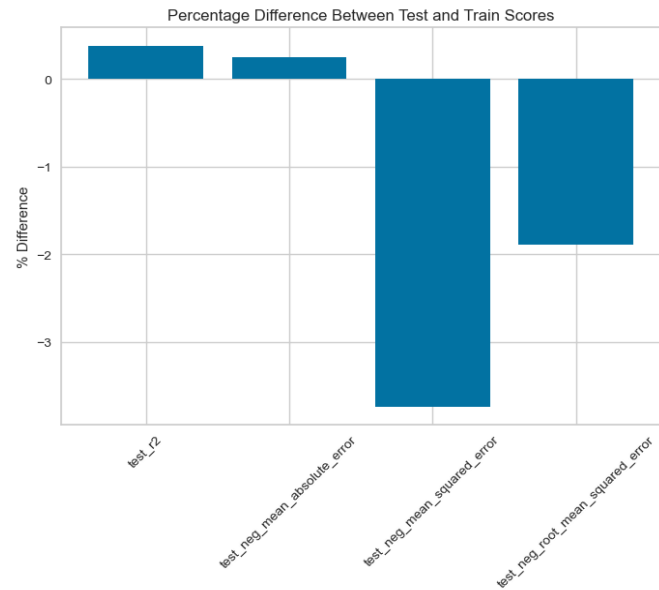
	R^2 (%)	MAE (%)	MSE (%)	RMSE (%)
Lineer Regresyon	1,17	-1,37	-9,99	-5,13
SVR	0,69	4,66	-21,42	-11,35
Karar Ağacı	0,38	0,25	-3,73	-1,89
Rastgele Orman	0,11	0,81	-2,21	-1,11
AdaBoost	0,49	-0,23	-4,42	-2,23
Gradyan Artırma	0,93	-1,39	-7,75	-3,95
XGBoost	-0,70	-0,24	-9,73	-4,99

Tablo 5.16'da birçok parametre için MAE yüzdelik fark değeri 0'a yakınlık göstermektedir. Bu durum başarılı bir modelleme yapıldığını göstermektedir.



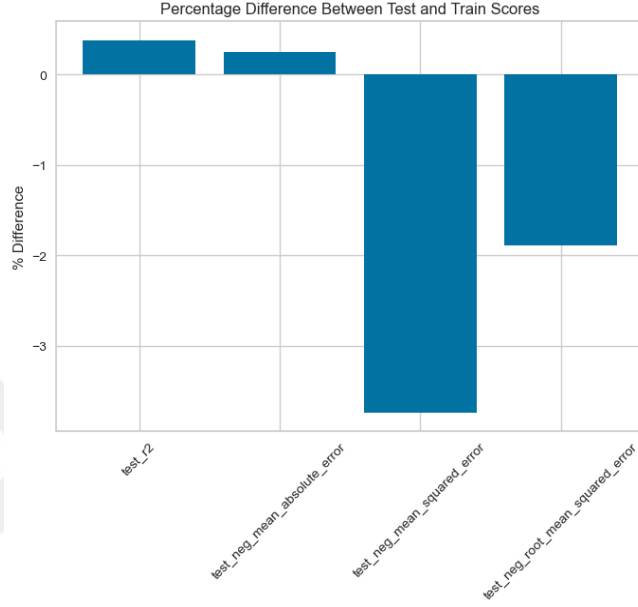
Şekil 5.53. Lineer Regresyon için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.53' göre Lineer Regresyon modeli R^2 metriği için test skoru, eğitim skorundan yüksek bir değerdedir. MAE metriği ise 0'a yakın ve test skoru eğitim skorundan düşük bir değer almıştır. Ancak MAE MSE ve RMSE için inceleme yapıldığında test skorunun eğitim skorundan düşük olduğu görülmektedir. Yine de metriklerle ait yüzdesel farkların fazla olmaması sebebiyle herhangi bir aşırı uyum ya da yetersiz uyum bulunmamaktadır.



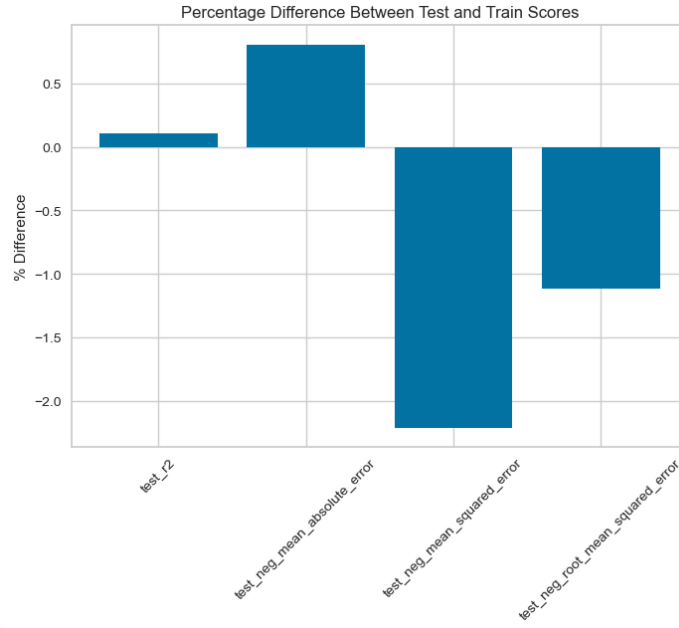
Şekil 5.54. Destek Vektör Regresörü için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.54'e bakıldığında ise R^2 ve MAE metriklerinde test skorunun eğitim skorundan yüksek olduğu görülmektedir. MSE ve RMSE'de ise bu durumun tam tersi durum karşımıza çıkmaktadır. Her 4 metrik için de yüzdesel farklar 0'a yakın olduğundan modellerin başarılı olduğu söylenebilir.



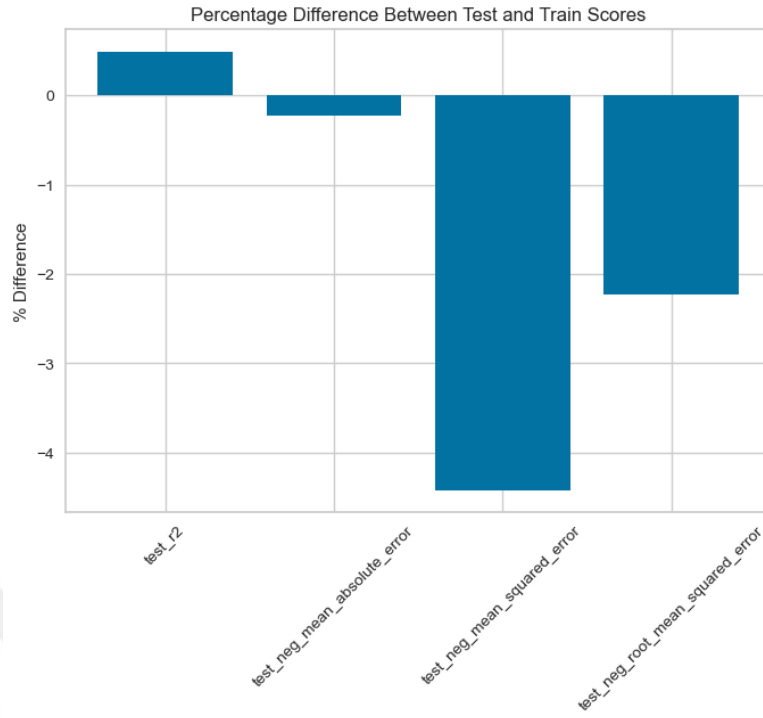
Şekil 5.55. Karar Ağacı için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.55'e bakıldığında ise Karar Ağacı için R^2 ve MAE metriklerinde test skorunun eğitim skorundan yüksek olduğu görülmektedir. MSE ve RMSE'de ise bu durumun tam tersi durum ortaya çıkmaktadır. Her 4 metrik için de yüzdesel farklar 0'a yakın olduğundan modellerin başarılı olduğu varsayımına varılabilir.



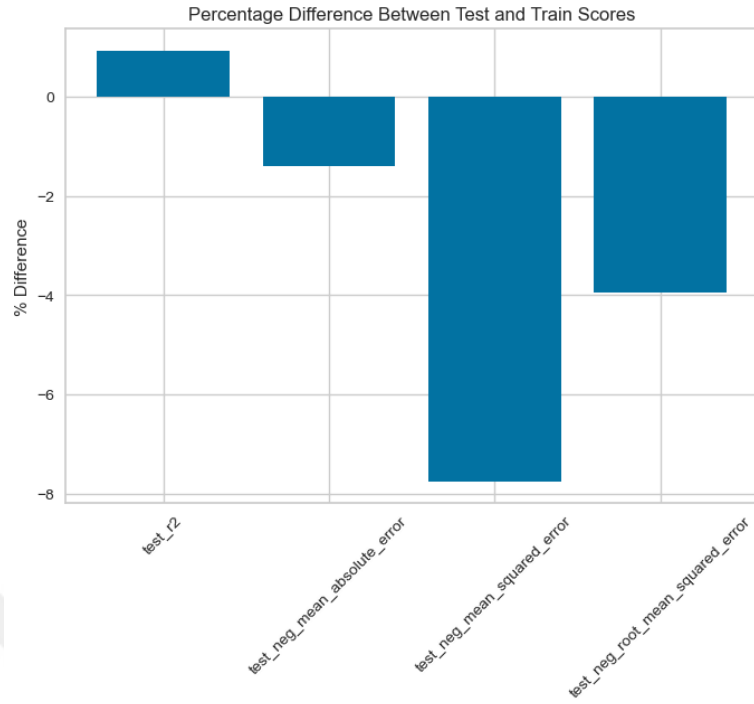
Şekil 5.56. Rastgele Orman için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.56’da ise Rastgele Orman için R^2 ve MAE metriklerinde test skorunun eğitim skorundan yüksek olduğu görülmektedir. Ancak MAE yüzdelik fark değerinin daha yüksek olduğu görülmektedir. MSE ve RMSE’de ise bu durumun tam tersi durum ortaya çıkmaktadır. Her 4 metrik için de yüzdesel farklar 0’a yakın olduğundan modellerin başarılı olduğu ifade edilebilir.



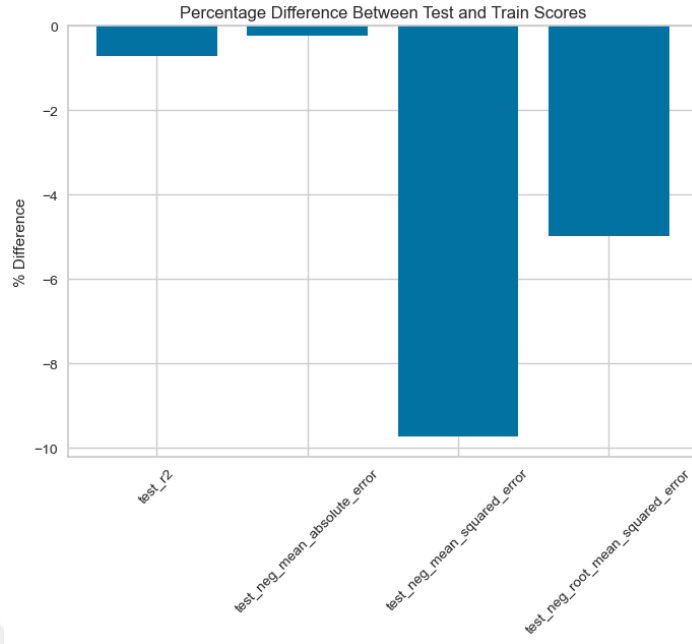
Şekil 5.57. AdaBoost için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark

Şekil 5.57 incelendiğinde AdaBoost modeli R^2 metriği için test skorunun, eğitim skorundan yüksek bir değerde olduğu gözlenebilir. MAE metriği ise 0'a yakın test skoru eğitim skorundan düşük bir değer almıştır. Ancak MAE MSE ve RMSE için inceleme yapıldığında test skorunun eğitim skorundan düşük olduğu görülmektedir. Yine de metriklerle ait bu yüzdesel farkların fazla olmaması sebebiyle herhangi bir aşırı uyum ya da yetersiz uyum bulunmamaktadır.



Şekil 5.58. Gradyan Artırma için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark

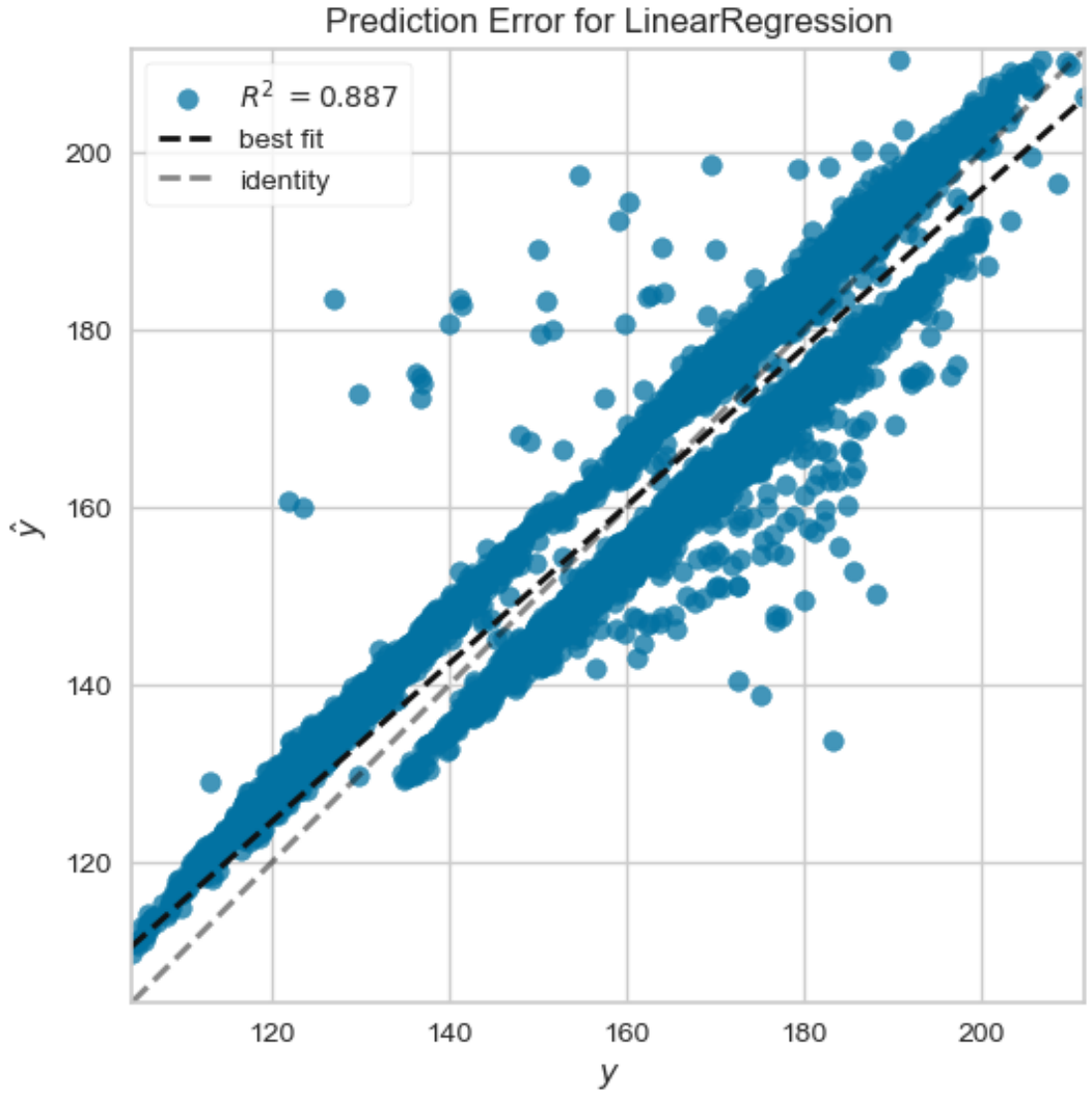
Şekil 5.58 incelendiğinde Gradyan Artırma modeli R^2 metriği için test skorunun, eğitim skorundan yüksek bir değerde olduğu gözlenebilir. MAE metriği ise 0'a yakın ve test skoru, eğitim skorundan düşük bir değer almıştır. Ancak MAE, MSE ve RMSE için inceleme yapıldığında test skorunun eğitim skorundan düşük olduğu görülmektedir. Yine de metriklerle ait bu yüzdesel farkların fazla olmaması sebebiyle herhangi bir aşırı uyum ya da yetersiz uyum bulunmamaktadır.



Şekil 5.59. XGBoost için toplam seyahat süresi eğitim ve test verileri arasında yüzdelik fark

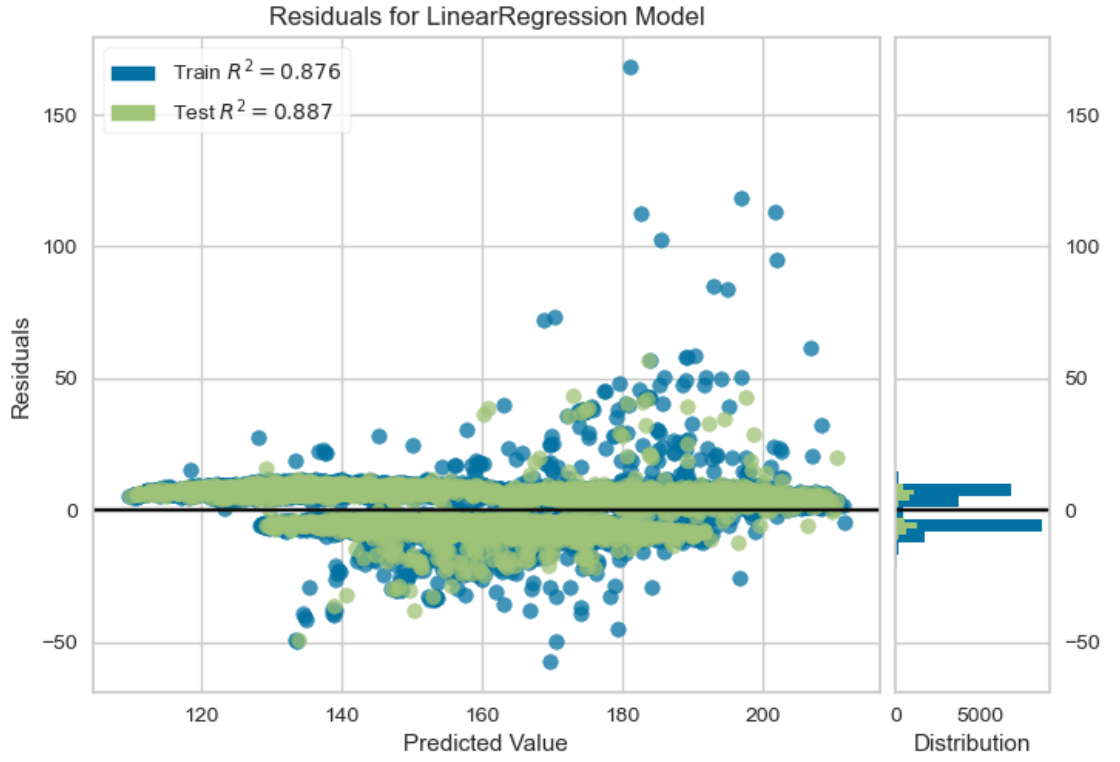
Şekil 5.59 incelendiğinde XGBoost modeli R^2 , MAE, MSE ve RMSE metriklerinin tamamında test skorunun, eğitim skorundan yüksek bir değerde olduğu gözlenebilir. Ancak metriklerle ait bu yüzdesel farkların fazla olmaması sebebiyle aşırı uyum ya da yetersiz uyum bulunmamaktadır.

5.3.3. Final Model Performans Değerlendirme



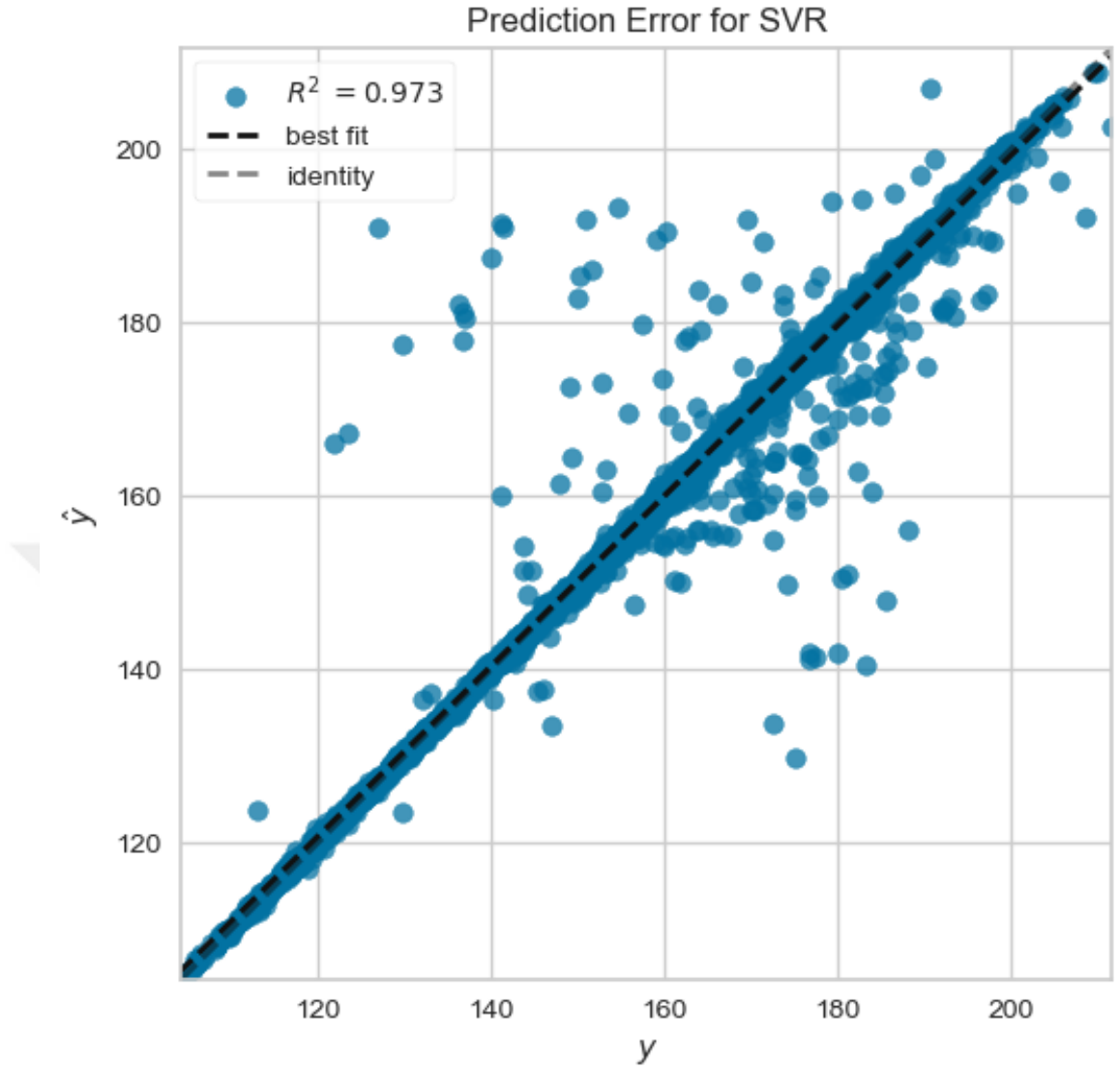
Şekil 5.60. Toplam seyahat süresi için Lineer Regresyon modeli tahmin hatası grafiği

Şekil 5.60’da Lineer Regresyon modelinin doğrusal bir dağılım gösterdiği görülmüştür. R^2 (0,8866) değerine göre yaklaşık %9’luk, RMSE’ye (7,3145) göre yaklaşık %4’lük bir hata oranı bulunmuştur. Bu durumda bu model başarılı sayılabilmektedir. Aşırı uyum durumuna uzak bir modelleme yapılmıştır.



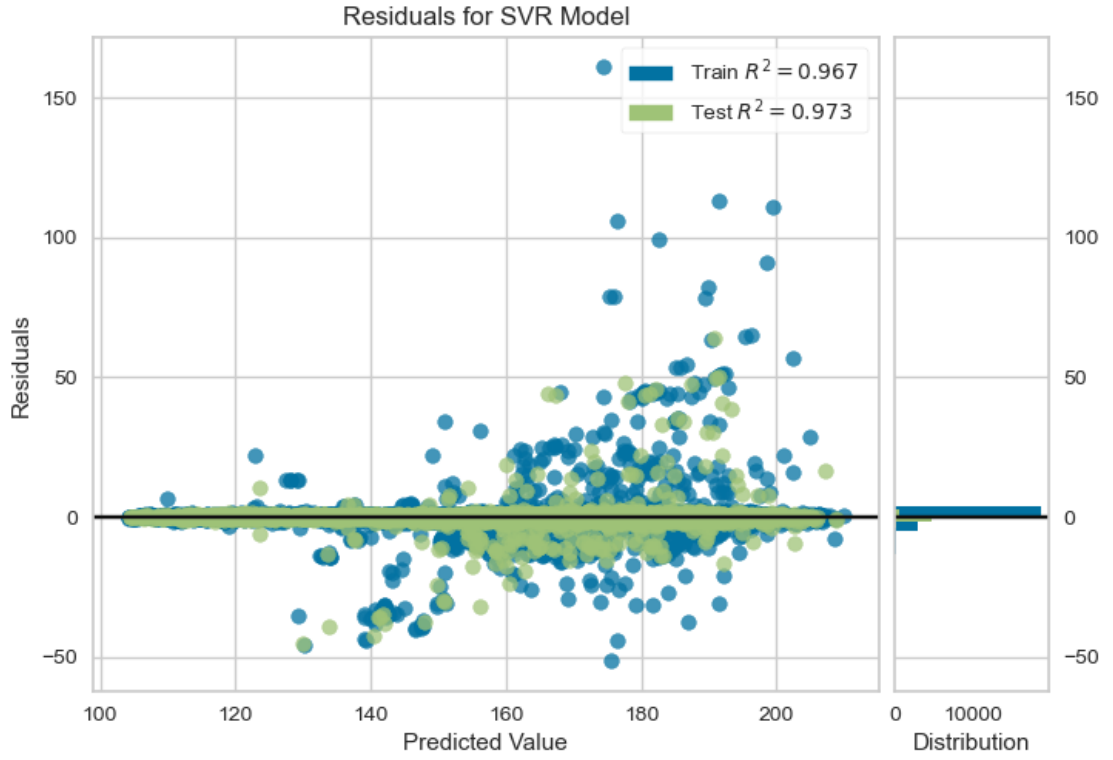
Şekil 5.61. Toplam seyahat süresi için Lineer Regresyon modelinde artık değerler dağılımı

Şekil 5.61’de Lineer Regresyon için artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmında herhangi bir “pattern” yani bir desen veya kalıp olmadığı görülmektedir. Sütun dağılımında da normal dağılım bulunduğu için artık değerler açısından bu modelin uygun olduğu tespit edilmiştir.



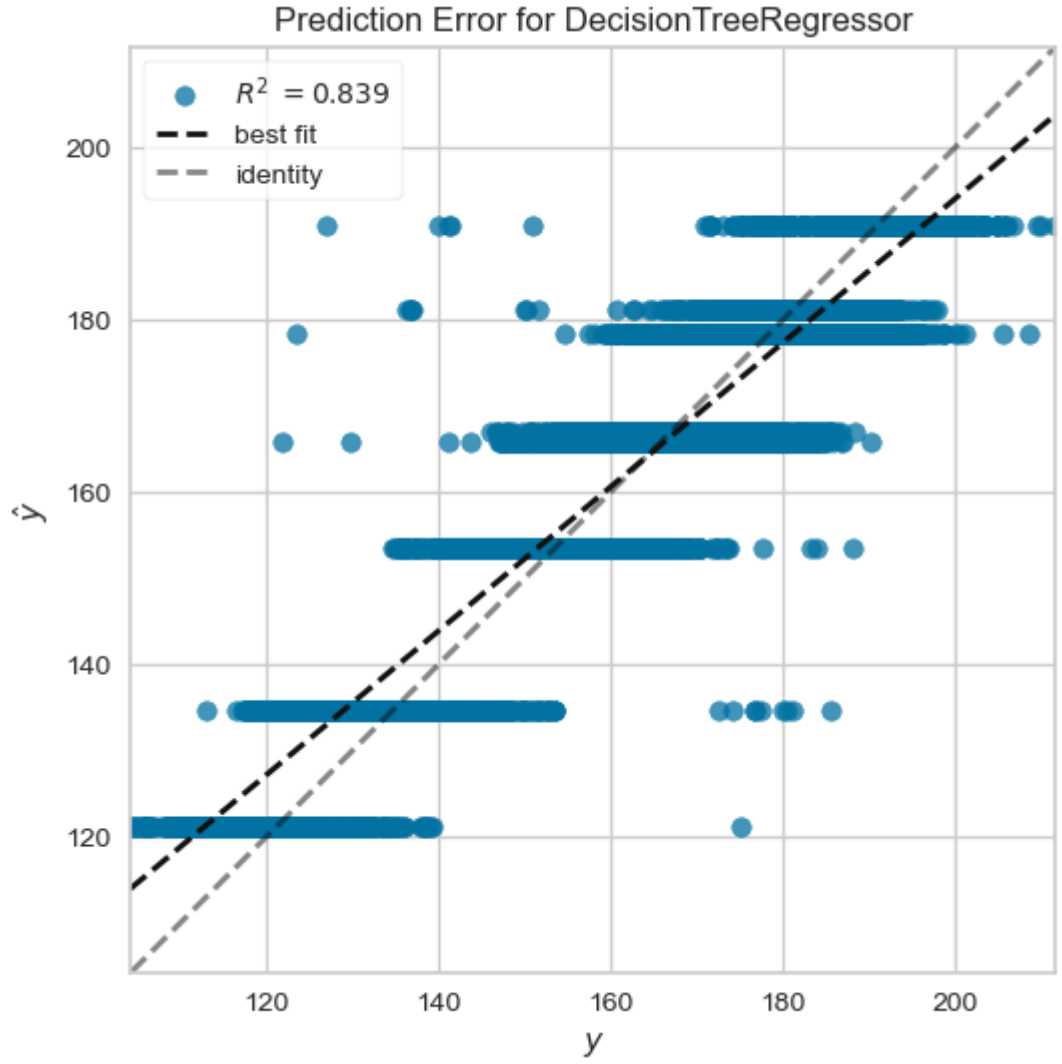
Şekil 5.62. Toplam seyahat süresi için Destek Vektör Regresörü modeli tahmin hatası grafiği

Şekil 5.62’de Destek Vektör Regresörü modelinin doğrusal bir dağılım gösterdiği görülmüştür. R^2 (0,9734) değerine göre yaklaşık %9’luk, RMSE’ye (3,5436) göre yaklaşık %2’lik bir hata oranı bulunmuştur. Bu durum Destek Vektör Regresörü ve parametre için güçlü bir modelleme yapıldığını göstermektedir.



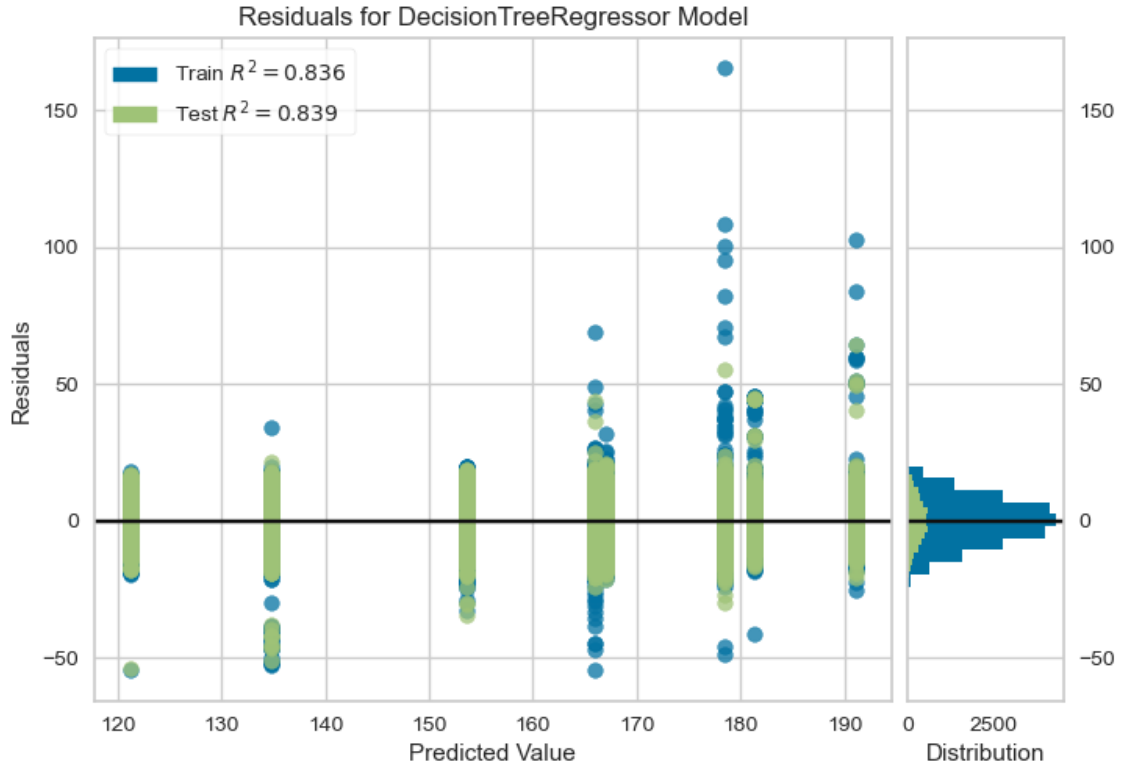
Şekil 5.63. Toplam seyahat süresi için Destek Vektör Regresörü modelinde artık değerler dağılımı

Şekil 5.63'te Destek Vektör Regresörü için artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmına bakıldığında herhangi bir “pattern” yani bir desen veya kalıp olmadığı görülmektedir. Sütun dağılımında da normal dağılım bulunduğu için artık değerler açısından bu modelin uygun olduğu tespit edilmiştir.



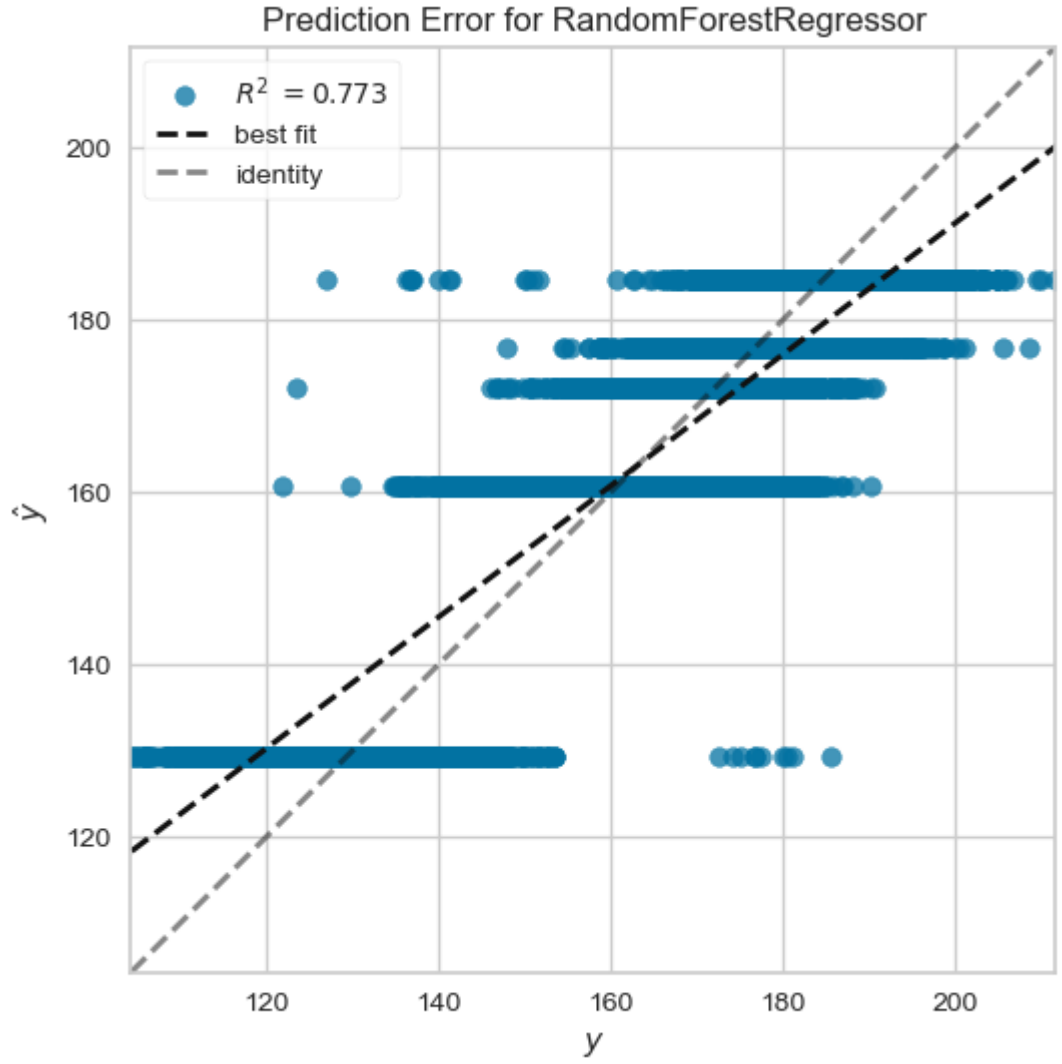
Şekil 5.64. Toplam seyahat süresi için Karar Ağacı modeli tahmin hatası grafiği

Şekil 5.64'te görüldüğü üzere burada aşırı uyum durumu bulunmaktadır. Modelleme, benzer tahminlerde bulunarak yapması gereken tahmin şekli dışında hatalı bir tahminleme yapmıştır. R^2 (0,839) değerinin de düşük olmasından dolayı bu modellemenin başarısız olduğu ifade edilebilir.



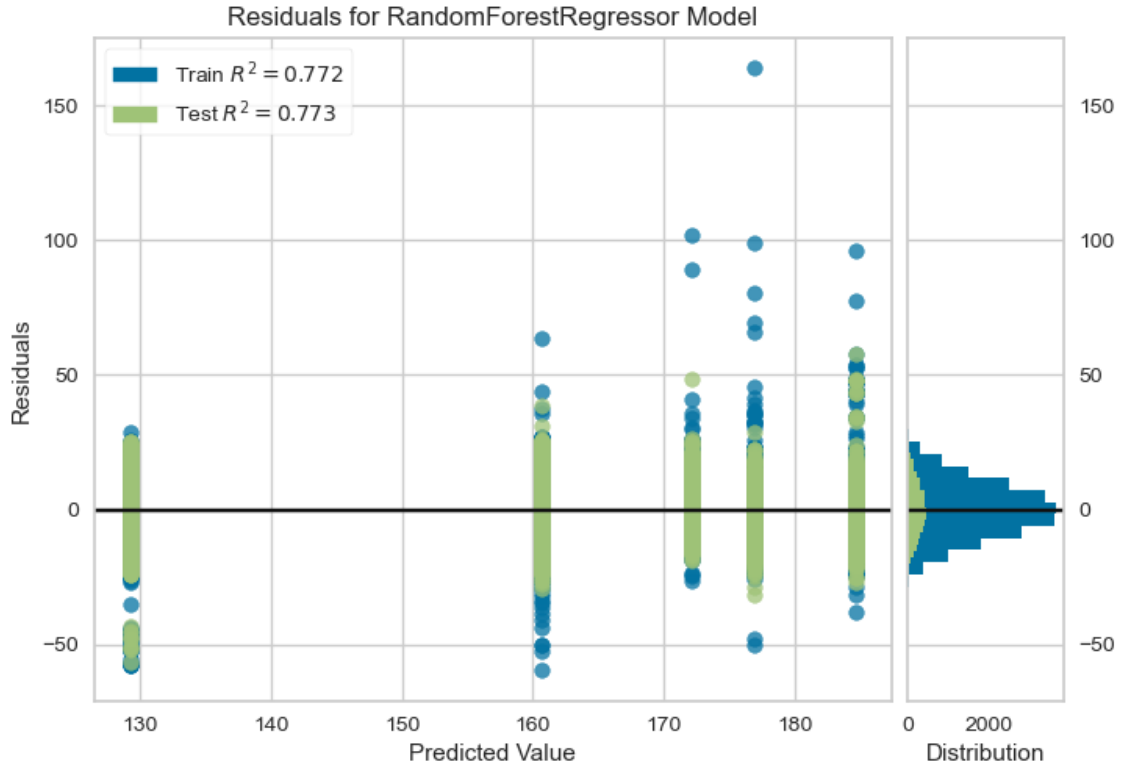
Şekil 5.65 Toplam seyahat süresi için Karar Ağacı modelinde artık değerler dağılımı

Şekil 5.65'te Karar Ağacı için artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmına bakıldığında “pattern” yani bir desen veya kalıp şeklinde bir dağılım söz konusudur. Bu sebeple artık değerler açısından bu modelin uygun olmadığı tespit edilmiştir.



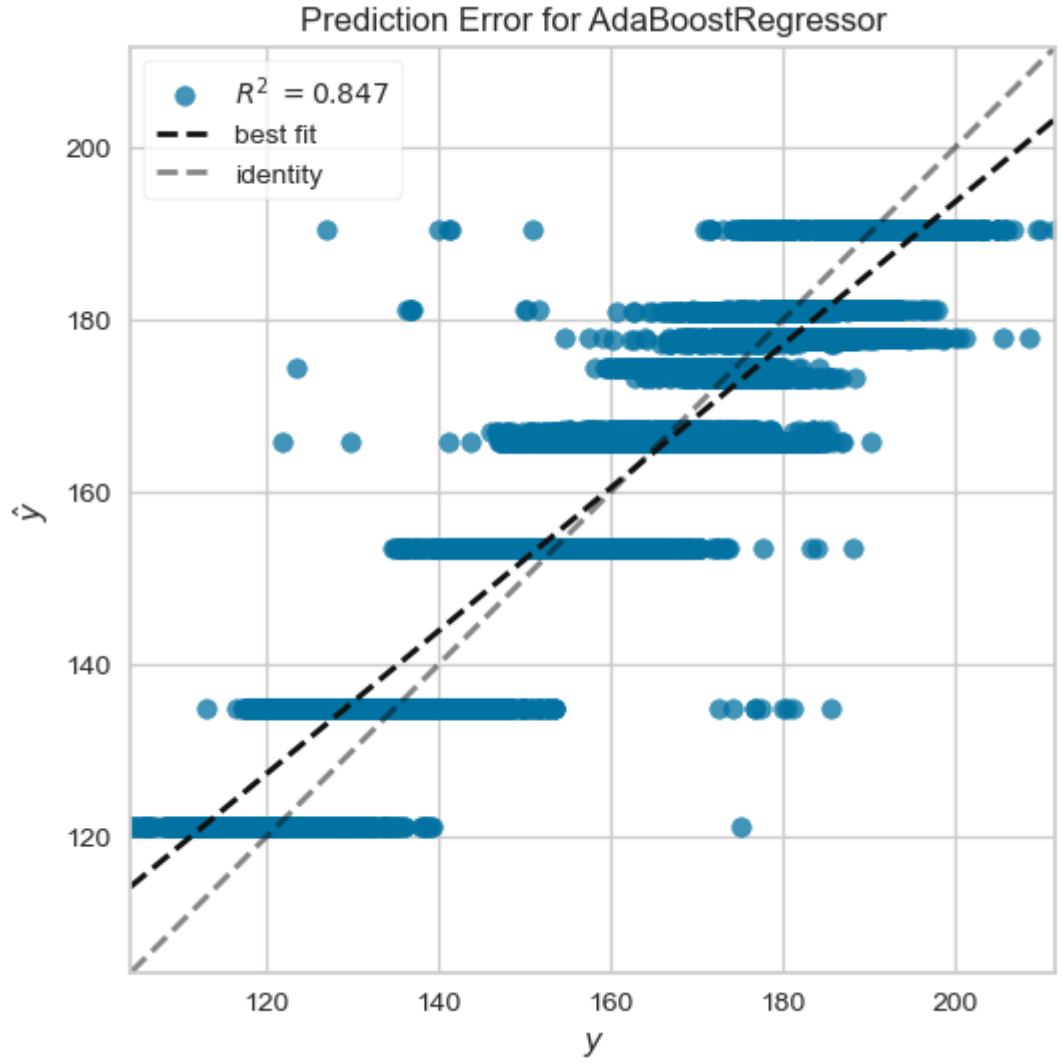
Şekil 5.66. Toplam seyahat süresi için Rastgele Orman modeli tahmin hatası grafiği

Şekil 5.66’da Rastgele Orman modeli, Karar Ağacı’na göre daha da başarısız bir modelleme yapmıştır. R^2 (0,773) değerinin az olması ve daha az değer tahmin ettiğinin grafikte görülmesi, bu durumun kanıtıdır.



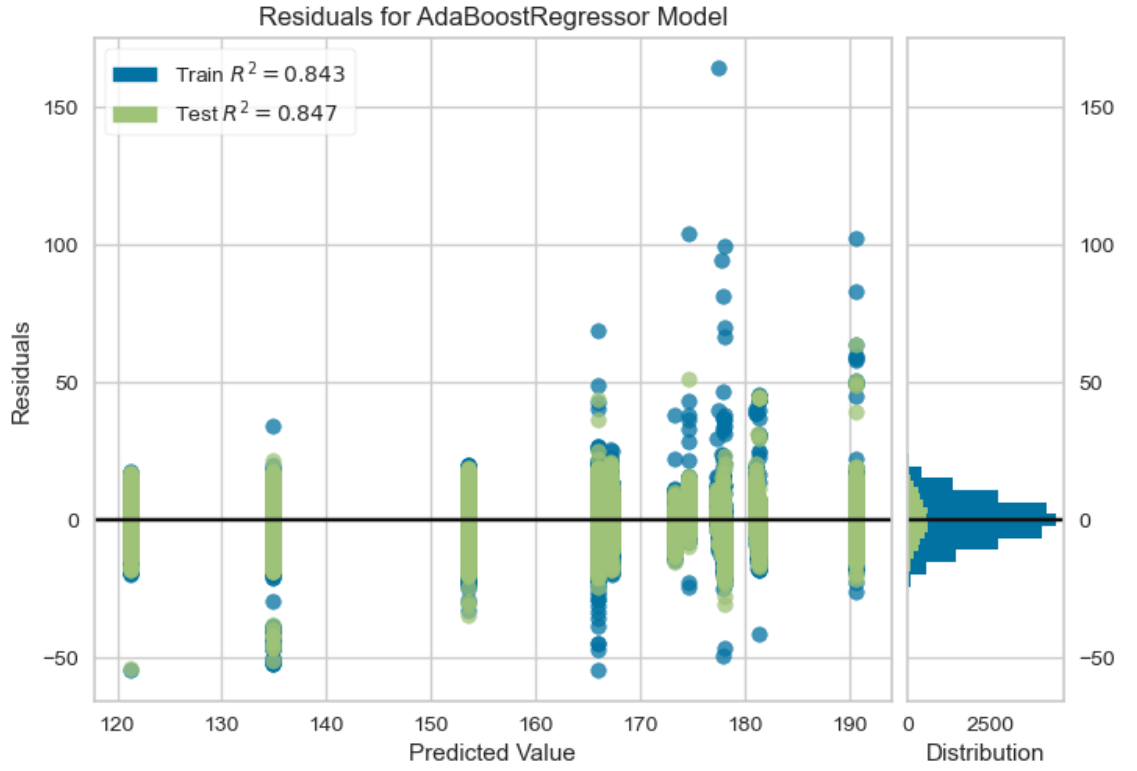
Şekil 5.67. Toplam seyahat süresi için Rastgele Orman modelinde artık değerler dağılımı

Şekil 5.67’de Rastgele Orman için artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmına bakıldığında “pattern” şeklinde bir dağılım bulunmaktadır. Bu sebeple artık değerler açısından bu modelin uygun olmadığı tespit edilmiştir.



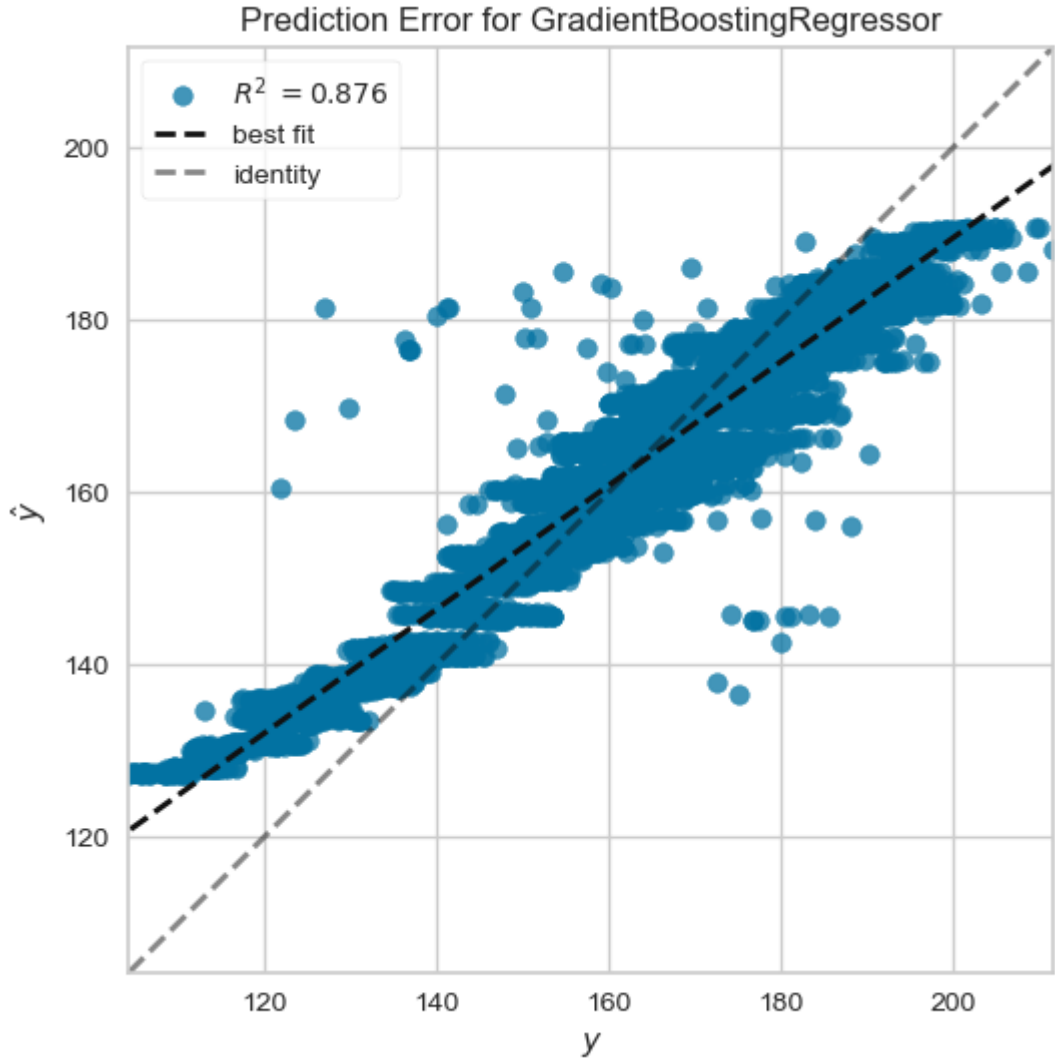
Şekil 5.68. Toplam seyahat süresi için AdaBoost modeli tahmin hatası grafiği

Şekil 5.68’de AdaBoost modeli Karar Ağacı ve Rastgele Orman algoritmalarına göre daha başarılı bir model olduğu görülmektedir. R^2 değerinin 0,847 olması, bu durumun numerik bir kanıtıdır. Ancak söz konusu bulgular, AdaBoost’u başarılı bir model sayılması için yeterli değildir. Çünkü grafikte modelin aşırı uyum gösterdiği çıkasanmaktadır.



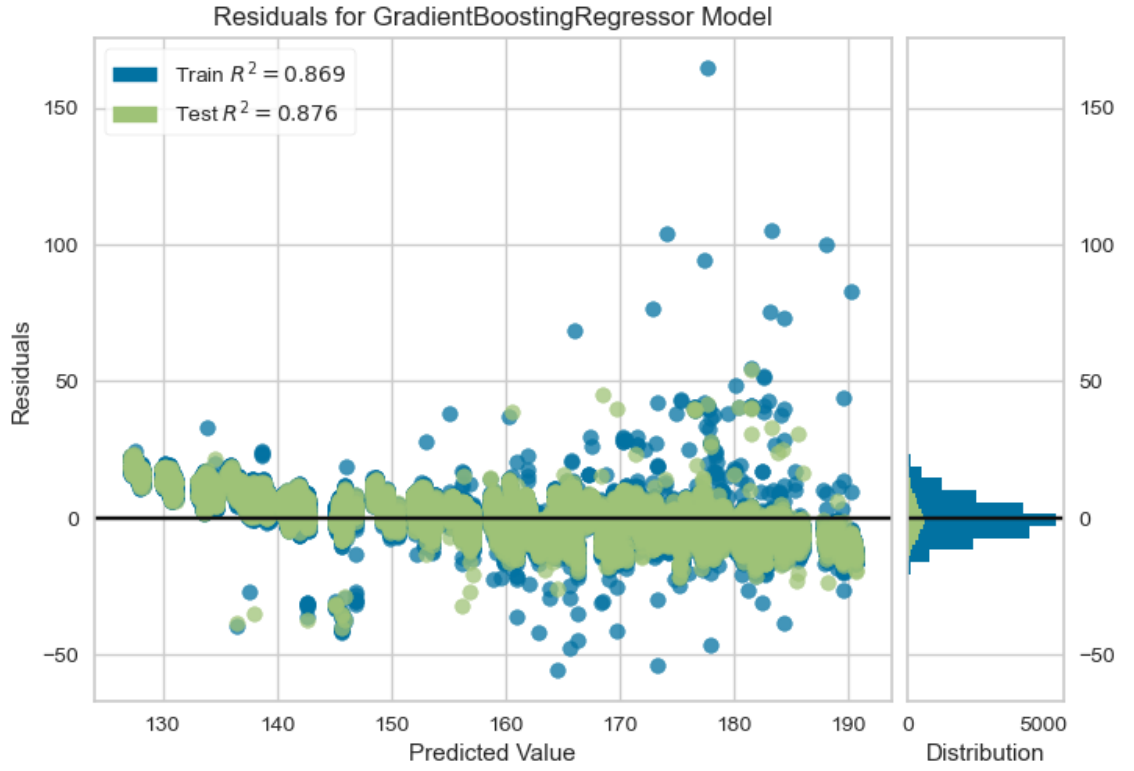
Şekil 5.69. Toplam seyahat süresi için AdaBoost modelinde artık değerler dağılımı

Şekil 5.69’da AdaBoost için artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmına bakıldığında “pattern” yani bir desen veya kalıp şeklinde bir dağılım söz konusudur. Bu sebeple artık değerler açısından bu modelin uygun olmadığı tespit edilmiştir.



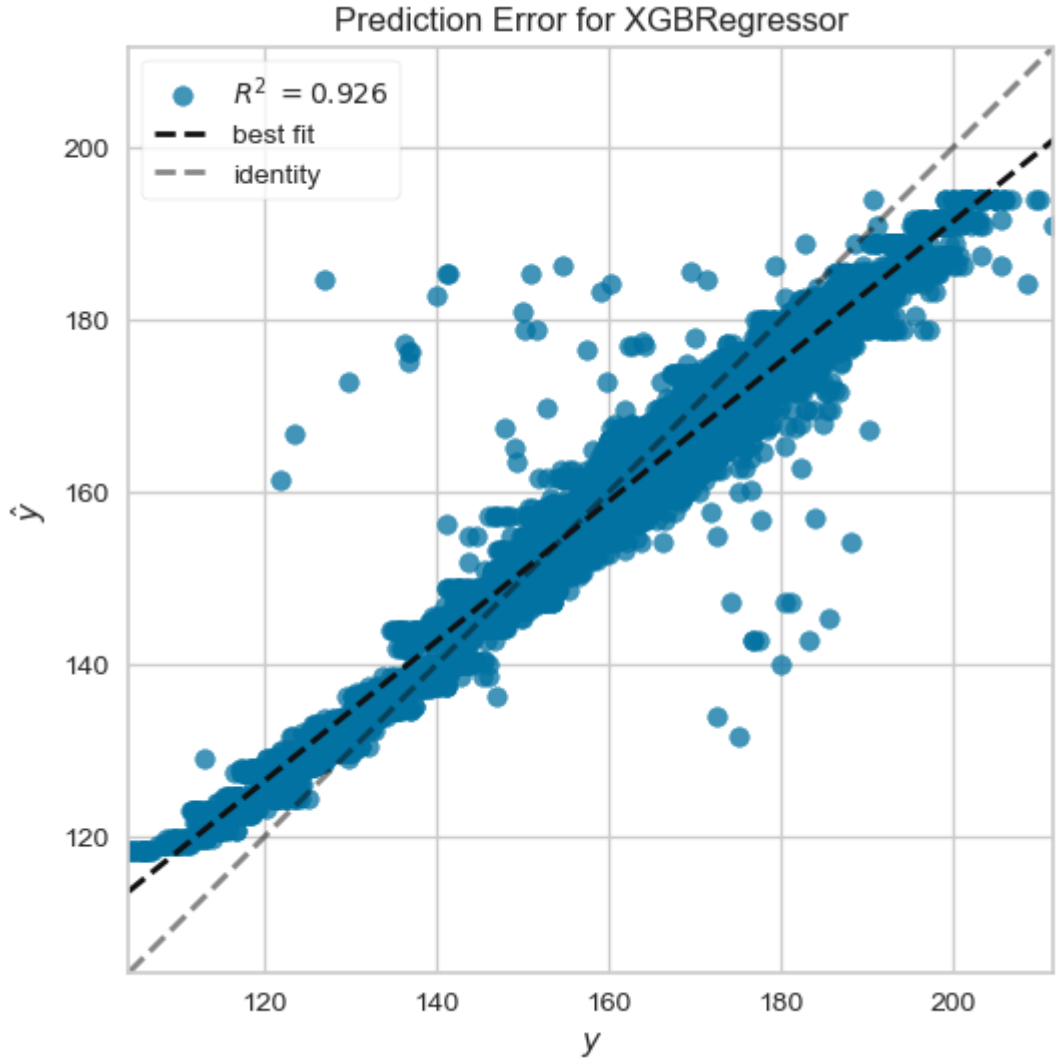
Şekil 5.70. Toplam seyahat süresi için Gradyan Artırma modeli tahmin hatası grafiği

Şekil 5.70’te Gradyan Artırma modelinin doğrusal bir dağılım gösterdiği görülmüştür. R^2 (0,8758) değerine göre yaklaşık %18’lik, RMSE’ye (7,6541) göre yaklaşık %5’lik bir hata oranı bulunmuştur. R^2 ’nin biraz düşük olması, bu modelin çok başarılı olmadığını göstermektedir.



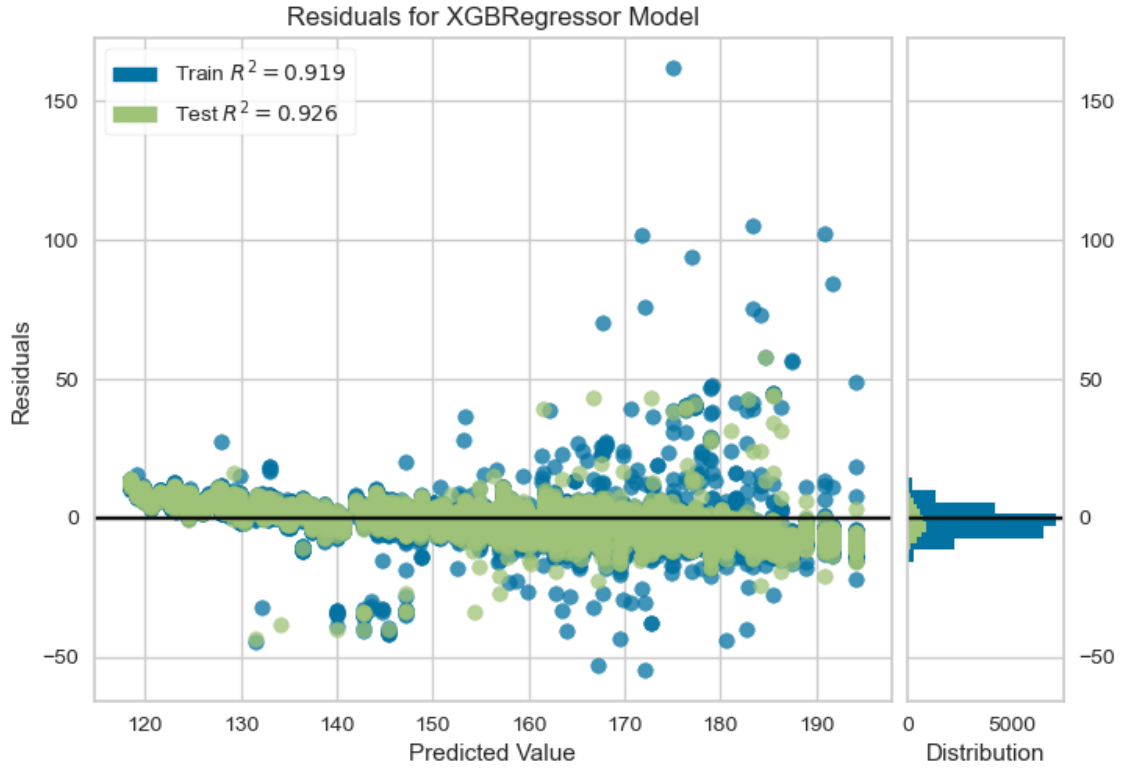
Şekil 5.71 Toplam seyahat süresi için Gradyan Artırma modelinde artık değerler dağılımı

Şekil 5.71'e göre Gradyan Artırma için artık değerler dağılımının tam anlamıyla bir "pattern" gösterdiği söylenemez. Ancak R^2 değerlerinin düşük olması da işin içine katıldığında bu modellemenin toplam seyahat süresi için başarısız kabul edilmelidir.



Şekil 5.72. Toplam seyahat süresi için XGBoost modeli tahmin hatası grafiği

Şekil 5.72’de XGBoost modelinin doğrusal bir dağılım gösterdiği görülmüştür. R^2 (0,9258) değerine göre yaklaşık %7’lik, RMSE’ye (5,9178) göre yaklaşık %4’lük bir hata oranı bulunmuştur. Modelde aşırı uyum durumu da olmadığından başarılı sayılabilmektedir.



Şekil 5.73. Toplam seyahat süresi için XGBoost modelinde artık değerler dağılımı

Şekil 5.73'te XGBoost artık değerler tahmin dağılımı grafiği verilmiştir. Grafiğin nokta saçılım kısmına bakıldığında herhangi bir “pattern” olmadığı görülmektedir. Sütun dağılımında da normal dağılım görüldüğü için artık değerler açısından bu modelin uygun olduğu tespit edilmiştir.

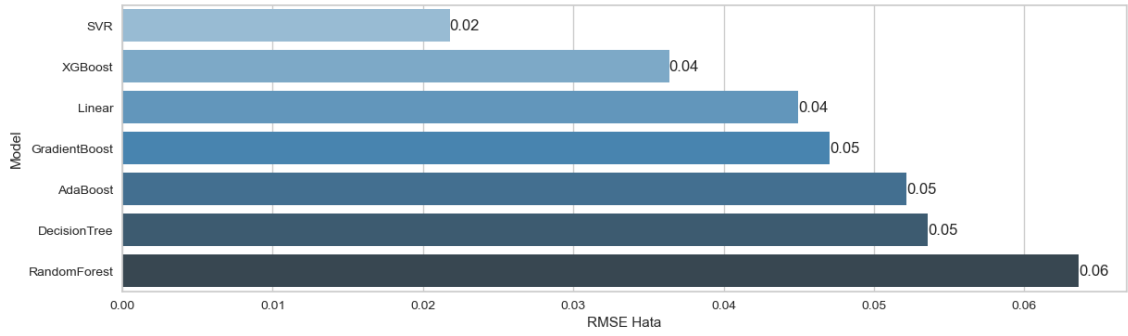
5.3.4. Modellerin Karşılaştırması

Tablo 5.17. Toplam seyahat süresi için makine öğrenmesi modellerinin hata metrikleri değerleri

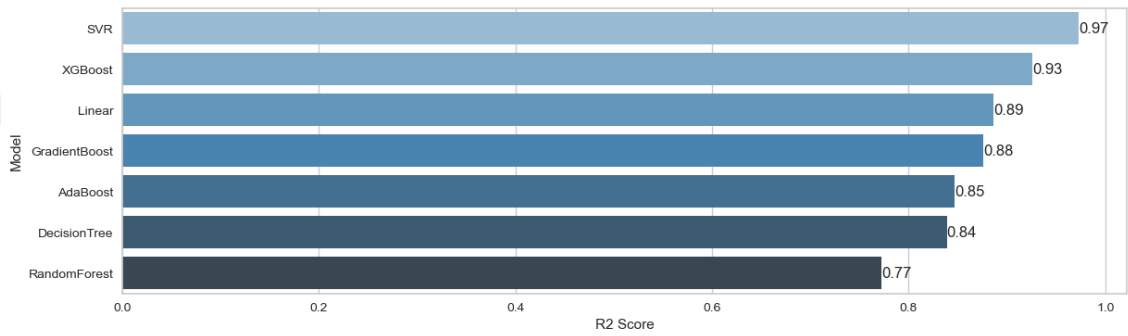
	R^2	MAE	MSE	RMSE
Lineer Regresyon GridSearchCV Eğitim	0,8763	6,6393	59,4420	7,7099
Lineer Regresyon GridSearchCV Test	0,8866	6,5483	53,5015	7,3145
SVR GridSearchCV Eğitim	0,9668	0,9012	15,9793	3,9974
SVR GridSearchCV Test	0,9734	0,9433	12,5569	3,5436
Karar Ağacı GridSearchCV Eğitim	0,8356	6,8526	79,0117	8,8889
Karar Ağacı GridSearchCV Test	0,8388	6,8695	76,0608	8,7213
Rastgele Orman GridSearchCV Eğitim	0,7718	8,1938	109,6908	10,473
Rastgele Orman GridSearchCV Test	0,7727	8,2600	107,2640	10,3568
AdaBoost GridSearchCV Eğitim	0,8430	6,6549	75,4918	8,6886
AdaBoost GridSearchCV Test	0,8471	6,6399	72,1570	8,4945
Gradyan Artırma GridSearchCV Eğitim	0,8662	6,0535	64,3259	8,0203
Gradyan Artırma GridSearchCV Test	0,8742	5,9691	59,3394	7,7032
XGBoost GridSearchCV Eğitim	0,9193	4,4300	38,7941	6,2285
XGBoost GridSearchCV Test	0,9258	4,4193	35,0202	5,9178

Tablo 5.18. Toplam seyahat süresi için performansa göre sıralanmış RMSE ve R^2 skorları

	Model	RMSE Hata	R^2 Skoru
1	SVR	0,0218	0,9734
2	XGBoost	0,0364	0,9258
3	Lineer Regresyon	0,0450	0,8866
4	Gradyan Artırma	0,0470	0,8758
5	AdaBoost	0,0522	0,8471
6	Karar Ağacı	0,0536	0,8388
7	Rastgele Orman	0,0637	0,7727



Şekil 5.74. Toplam seyahat süresi için makine öğrenme modelleri RMSE hata oranlarının karşılaştırılması



Şekil 5.75. Toplam seyahat süresi için makine öğrenme modelleri R^2 skorlarının karşılaştırılması

Tablo 5.17 ve 5.18 incelendiğinde ise eğitilen veriler doğrultusunda Destek Vektör Regresörü, toplam seyahat süresi parametresi için en iyi sonucu veren makine öğrenmesi modeli olmuştur. XGBoost ve Lineer Regresyon da bu modeli takip eden diğer makine öğrenmesi modelidir. En başarısız modelleme ise Rastgele Orman ile elde edilmiştir.

Şekil 5.74'te açıkça görülmektedir ki RMSE metriğine göre 0,02 hata oranıyla %98'lik bir performansla en iyi tahmin sonuçları Destek Vektör Regresörü ile elde edilmiştir. Bu değeri %4'lük hata oranları ile Lineer Regresyon ve XGBoost takip etmektedir.

Şekil 5.75'te ise R^2 skorlarının karşılaştırıldığı bir grafik verilmiştir. Bu grafiğe göre %97'lik doğruluk oranı ile ya da %3'lük bir hata oranı ile Destek Vektör Regresörü modeli, en iyi tahmin sonuçlarını üretmiştir. Bu duruma göre, %93 R^2 skoru ile XGBoost modeli ve %89 R^2 skoru ile Lineer Regresyon modeli, doğruluk oranları yüksek modeller olarak takip etmektedir. Lineer Regresyon modelinin belirtildiği gibi R^2 değerinin Destek Vektör Regresörü ve XGBoost'tan geride kalması sebebiyle başarılı model olarak kabul edilmemiştir.

Edinilen bulgular doğrultusunda toplam seyahat süresi parametresi için Destek Vektör Regresörü ve XGBoost modelleri, başarılı modeller olarak seçilmiştir.

5.3.5. Tahmin Değerleri Karşılaştırılması

Bu karşılaştırma için veri setindeki 0. 100. ve 200. Senaryolar kullanılmıştır. Bu senaryolara ait kavşak geometrik ve sinyalize özellikleri, trafik hacim bilgileri ve şerit sayısına ait özellikler Tablo 5.19’da verilmiştir.

Tablo 5.19. Karşılaştırma yapılacak senaryolara ait bilgiler

	Dörtüyl Kavşağı	Merplaza Kavşağı	Adliye Kavşağı	Terzibaba Kavşağı	Polis Okulu Kavşağı	Tedaş Kavşağı	Terminal Kavşağı	Trafik Hacmi Değişimi Araba %	Trafik Hacmi Değişimi Otobüs %	Trafik Hacmi Değişimi Kamyon %	Batı-Doğu istikametine (1) şerit ekleme	Doğu-Batı istikametine (1) şerit ekleme
0	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Trafik	Mevcut Trafik	Mevcut Trafik	Mevcut Şerit	Mevcut Şerit
100	Doğu-Batı yönünde yeşil süreyi 5 sn arttır	Mevcut Durum	Mevcut Durum	Altgeçit	Mevcut Durum	Mevcut Durum	Mevcut Durum	Mevcut Trafik	Mevcut Trafik	Mevcut Trafik	Mevcut Şerit	Mevcut Şerit
200	Modern Dönel Kavşak	Doğu-Batı yönünde yeşil süreyi 5 sn arttır	Modern Dönel Kavşak	Mevcut Durum	Mevcut Durum	Modern Dönel Kavşak	Mevcut Durum	%5 artış	Mevcut Trafik	Mevcut Trafik	Mevcut Şerit	Mevcut Şerit

Tablo 5.20. Toplam seyahat süresi için model tahmin değerlerinin karşılaştırılması

	Simülasyon Sonucu	Lineer Regresyon	SVR	Karar Ağacı	Rastgele Orman	Ada Boost	Gradyan Artırma	XGBoost
0	186,42	192,00	186,40	174,54	170,62	177,66	180,54	178,77
100	167,50	163,46	167,45	165,91	160,56	165,92	159,81	162,19
200	169,94	175,05	171,65	175,99	179,96	167,24	172,91	168,56

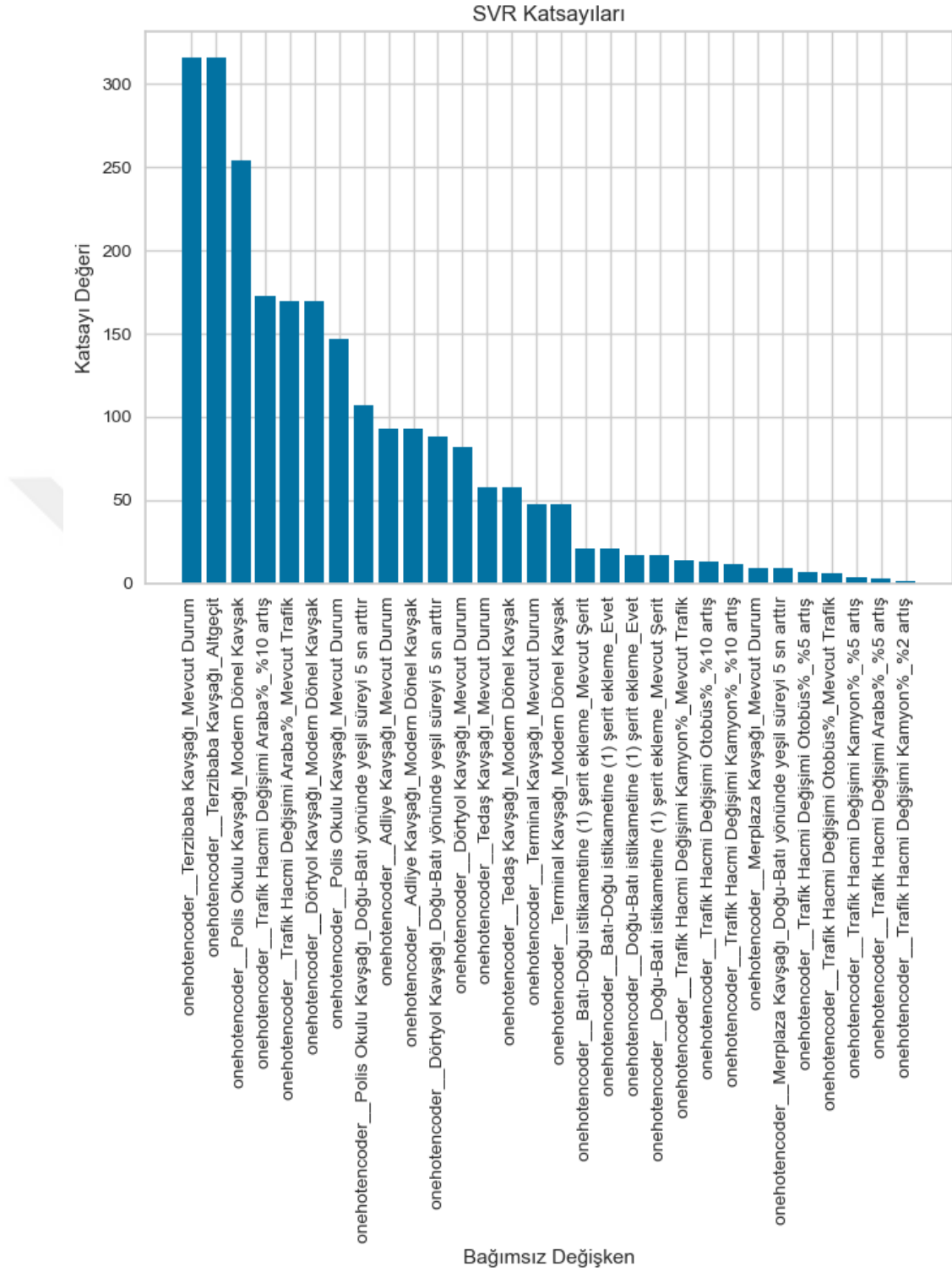
Tablo 5.20’ye göre senaryo 0 ve 100’ün gerçek değere en yakın tahminlerini Destek Vektör Regresörü modeli yapmıştır. En yakın tahminleri, senaryo 200 için ise XGBoost yapmıştır En uzak tahminler ise senaryo 0 ve senaryo 100 için XGBoost modeli, senaryo

200 için ise Rastgele Orman modelleri ile yapılmıştır. Bu tablo için en başarılı model Destek Vektör Regresörü, en başarısız model ise Rastgele Orman modelidir.

5.3.6. Başarılı modellerin etki katsayılarının değerlendirilmesi

```
onehotencoder__Terzibaba Kavşağı_Mevcut Durum: 315.26844105518717
onehotencoder__Terzibaba Kavşağı_Altgeçit: -315.268441055187
onehotencoder__Polis Okulu Kavşağı_Modern Dönel Kavşak: -253.73472301181133
onehotencoder__Trafik Hacmi Değişimi Araba%_10 artış: 172.6839908091503
onehotencoder__Trafik Hacmi Değişimi Araba%_Mevcut Trafik: -169.60268027658614
onehotencoder__Dörtüol Kavşağı_Modern Dönel Kavşak: -169.2043123431711
onehotencoder__Polis Okulu Kavşağı_Mevcut Durum: 146.531617646166
onehotencoder__Polis Okulu Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 107.203105365645
onehotencoder__Adliye Kavşağı_Mevcut Durum: 92.98287562372094
onehotencoder__Adliye Kavşağı_Modern Dönel Kavşak: -92.98287562372079
onehotencoder__Dörtüol Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 87.72188443180474
onehotencoder__Dörtüol Kavşağı_Mevcut Durum: 81.48242791136643
onehotencoder__Tedaş Kavşağı_Mevcut Durum: 57.2829462427529
onehotencoder__Tedaş Kavşağı_Modern Dönel Kavşak: -57.282946242752786
onehotencoder__Terminal Kavşağı_Mevcut Durum: 47.68621460339917
onehotencoder__Terminal Kavşağı_Modern Dönel Kavşak: -47.68621460339905
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Mevcut Şerit: 21.17376658828212
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Evet: -21.173766588281914
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Evet: -16.828002574942296
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Mevcut Şerit: 16.828002574942246
onehotencoder__Trafik Hacmi Değişimi Kamyon%_Mevcut Trafik: -13.723801700385875
onehotencoder__Trafik Hacmi Değişimi Otobüs%_10 artış: 12.78828096960406
onehotencoder__Trafik Hacmi Değişimi Kamyon%_10 artış: 11.420184776882607
onehotencoder__Merplaza Kavşağı_Mevcut Durum: 9.468095396075318
onehotencoder__Merplaza Kavşağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: -9.468095396075178
onehotencoder__Trafik Hacmi Değişimi Otobüs%_5 artış: -6.973553849402129
onehotencoder__Trafik Hacmi Değişimi Otobüs%_Mevcut Trafik: -5.814727120202066
onehotencoder__Trafik Hacmi Değişimi Kamyon%_5 artış: 4.031322635103972
onehotencoder__Trafik Hacmi Değişimi Araba%_5 artış: -3.0813105325642542
onehotencoder__Trafik Hacmi Değişimi Kamyon%_2 artış: -1.727705711600759
```

Şekil 5.76. Destek Vektör Regresörü modeli için bağımsız değişkenlere ait katsayı değerleri



Şekil 5.77. Destek Vektör Regresörü modeli için bağımsız değişkenlere ait katsayı değerleri sütun grafiği

Şekil 5.76 ve Şekil 5.77 incelendiğinde Destek Vektör Regresörü modeli için toplam seyahat süresi üzerindeki en fazla etkinin kavşaklardaki geometrik değişiklikler göze

çarpmaktadır. Özellikle de modelin Terzibaba Kavşağı'nın mevcut durum değişkenini etkilediği söylenebilir. Bu değerin 300'den fazla olduğu görülmektedir. Bu durumun Dört yol Kavşağı'nda yapılan bir değişikliğin gecikme süresindeki azalmaya çok fazla etkisinin olduğu görülmektedir.

Katsayı değeri yaklaşık 150 çıkan Dört yol Kavşağı sinyalizasyon ile ilgili değişkenin de gecikme süresi üzerinde büyük etkisi olduğu görülmektedir.

Bu değişkenleri taşıt hacmindeki değişiklikler ile sinyalizasyondaki değişiklikler takip etmektedir. Katsayı değeri 200'den fazla çıkan Terzibaba Kavşağı'na ait mevcut durum ve Modern Dönel Kavşak değişkenlerinin de toplam seyahat süresi üzerinde etkisi olduğu açıkça görülmektedir.

Toplam seyahat süresi parametresini en az kamyon hacmindeki değişiklikler etki etmektedir. Öyle ki bu değişkenlere ait katsayı değerleri 0'a yakın çıkmıştır.

Dört yol Kavşağı'nda bir modern dönel kavşak düzenlemesi ya da doğu-batı yönündeki sinyalizasyon düzenlemesi, gecikme süresi parametresinin kontrol altına alınabilmesinde önemli olduğu anlaşılmaktadır.

Terzibaba Kavşağı'ndaki altgeçit düzenlemesi gibi bir düzenlemenin de toplam seyahat süresi için büyük etkileri olacağı açıkça görülmektedir.

```
[13:50:39] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.
onehotencoder__Terzibaba Kavsağı_Mevcut Durum: 0.2656928598880768
onehotencoder__Terzibaba Kavsağı_Altgeçit: 0.21627625823020935
onehotencoder__Polis Okulu Kavsağı_Modern Dönel Kavşak: 0.14034870266914368
onehotencoder__Polis Okulu Kavsağı_Mevcut Durum: 0.12589885294437408
onehotencoder__Trafik Hacmi Değişimi Araba%_Mevcut Trafik: 0.044957809150218964
onehotencoder__Trafik Hacmi Değişimi Araba%_10 artış: 0.04406438022851944
onehotencoder__Dört Yol Kavsağı_Modern Dönel Kavşak: 0.04343992471694946
onehotencoder__Adliye Kavsağı_Mevcut Durum: 0.03796129673719406
onehotencoder__Tedaş Kavsağı_Modern Dönel Kavşak: 0.02222219668328762
onehotencoder__Adliye Kavsağı_Modern Dönel Kavşak: 0.016780182719230652
onehotencoder__Terminal Kavsağı_Modern Dönel Kavşak: 0.016405433416366577
onehotencoder__Tedaş Kavsağı_Mevcut Durum: 0.013219332322478294
onehotencoder__Terminal Kavsağı_Mevcut Durum: 0.012732713483273983
onehotencoder__Polis Okulu Kavsağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 0.0
onehotencoder__Merplaza Kavsağı_Mevcut Durum: 0.0
onehotencoder__Merplaza Kavsağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 0.0
onehotencoder__Dört Yol Kavsağı_Mevcut Durum: 0.0
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Mevcut Şerit: 0.0
onehotencoder__Doğu-Batı istikametine (1) şerit ekleme_Evet: 0.0
onehotencoder__Trafik Hacmi Değişimi Araba%_5 artış: 0.0
onehotencoder__Trafik Hacmi Değişimi Otobüs%_10 artış: 0.0
onehotencoder__Trafik Hacmi Değişimi Otobüs%_5 artış: 0.0
onehotencoder__Trafik Hacmi Değişimi Otobüs%_Mevcut Trafik: 0.0
onehotencoder__Trafik Hacmi Değişimi Kamyon%_10 artış: 0.0
onehotencoder__Trafik Hacmi Değişimi Kamyon%_2 artış: 0.0
onehotencoder__Trafik Hacmi Değişimi Kamyon%_5 artış: 0.0
onehotencoder__Trafik Hacmi Değişimi Kamyon%_Mevcut Trafik: 0.0
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Evet: 0.0
onehotencoder__Batı-Doğu istikametine (1) şerit ekleme_Mevcut Şerit: 0.0
onehotencoder__Dört Yol Kavsağı_Doğu-Batı yönünde yeşil süreyi 5 sn arttır: 0.0
```

Şekil 5.78. XGBoost modeli için bağımsız değişkenlerin nitelik önemi değerleri

Şekil 5.78 ve Şekil 5.79 incelendiğinde ise toplam seyahat süresi parametresi üzerindeki en fazla etkiyi, katsayı değeri 0,25'ten fazla olan Terzibaba kavşağına ait mevcut durum ve katsayı değeri 0,15'e yakın olan altgeçit değişkenleri göstermektedir. Bu değişkenleri, modern dönel kavşak olan Polis Okulu Kavşağı ve otomobil hacmindeki değişiklikler ile ilgili değişkenler takip etmektedir. Bu değerler birbirlerine yakın olduğundan benzer bir etki gösterebilirler.

Sonuç olarak Kavşak değişkenleri göz önünde bulundurulduğunda IEM CO₂ parametresi için Terzibaba Kavşağı'ndaki değişiklikler en fazla etkiyi oluşturacaktır. Taşıt hacmindeki değişikliklerde ise en fazla otomobil hacmindeki değişiklikler etkili olacaktır.

Dörtyol ve Merplaza kavşağı ile ilişkili değişkenler, kamyon ve otobüs hacmi ile ilişkili değişkenler ve şerit sayıları ile ilişkili değişkenlerin 0'a yakın çıkması, bu değişkenlerin XGBoost modelinde IEM CO₂ parametresi için önem arz etmediği söylenebilir.

6. SONUÇ ve ÖNERİLER

Gerçekleştirilen araştırma sonucunda elde edilen bulgular ışığında sonuç ve öneriler sunulmuştur.

- Makine öğrenmesi her alanda olduğu gibi Trafik Mühendisliği alanında da planlama, iyileştirme gibi durumlarda başvurulabilecek bilimsel bir metot olduğu görülmüştür.
- Bu metot ile seçilen herhangi bir koridorda geometrik değişiklikler, trafik hacmindeki değişiklikler ya da sinyalizasyon süresindeki birtakım teknik değişiklikler ile koridorda konfor, güvenlik, sürdürülebilirlik ve ekolojik açıdan ne gibi etkileri olacağı hakkında tahminler yapılabileceği anlaşılmıştır.
- Veri üretiminin, özellikle makine öğrenimi modelleri oluşturmak için çok önemli olduğu; simülasyon tekniklerinin ve simülasyon programlarının, veri üretimini kolaylaştırması sebebiyle önemli bir konuma sahip olduğu görülmüştür.
- AIMSUN programı büyük veri analizi, derin öğrenme ve makine öğrenmesi gibi yapay zekâ metotları için veri temini sağlayabilecek yeterlilikte bir trafik simülasyon programı olduğu söylenebilir.
- GridSearchCV işlemi R^2 , MAE, MSE ve RMSE metrikleri incelendiğinde verilerin doğru bir şekilde fit edilmesini sağlayarak başarılı bir eğitim modelini oluşturmayı sağladığı görülmüştür.
- Gecikme süresi parametresi için eğitilen veriler doğrultusunda en iyi sonucu veren makine öğrenmesi modelleri $0,96 R^2$ skoruyla ve RMSE metriğine göre %4 hata oranıyla Destek Vektör Regresörü ve $0,95 R^2$ skoruyla ve RMSE metriğine %5 hata oranı ile Lineer Regresyon modeli olmuştur.
- IEM CO₂ parametresi için eğitilen veriler doğrultusunda en iyi sonucu veren makine öğrenmesi modeli $0,92 R^2$ skoruyla ve RMSE metriğine göre %2,7 hata oranıyla XGBoost bulunmuştur.
- Toplam seyahat süresi parametresi için eğitilen veriler doğrultusunda en iyi sonucu veren makine öğrenmesi modelleri $0,97 R^2$ skoru ve RMSE metriğine göre %2 hata oranıyla Destek Vektör Regresörü ve $0,93 R^2$ skoru ve RMSE metriğine göre %4 hata oranıyla XGBoost olmuştur.

- Gecikme süresi parametresi Lineer Regresyon modellemesinde katsayı değeri 2'ye yakın bir değer olan otomobil hacmi ile ilgili değişkenlerin parametre üzerinde en fazla etkiye sahip değişkenler olduğu saptanmıştır.
- Gecikme süresi parametresi Destek Vektör Regresörü modellemesinde katsayı değerleri 150-350 aralığında değişen Dört Yol Kavşağı'na ve Terzibaba Kavşağı'na ait değişkenler (mevcut durum, sinyalizasyonda artış, modern dönel kavşak, altgeçit) en büyük etkiye sahip değişkenler olarak bulunmuştur.
- Gecikme süresi parametresinin kontrol altına alınmasının istendiği durumlarda, Dört Yol Kavşağı'nda ve Terzibaba Kavşağı'nda geometrik düzenlemeler ve sinyalizasyon düzenlemeleri yapılması gerekliliği ortaya çıkmaktadır.
- IEM CO₂ parametresi XGBoost modellemesinde nitelik önemi değerleri 0,3 çıkan Terzibaba Kavşağı mevcut durum ile ilgili değişkeni ve 0,2 Terzibaba Kavşağı altgeçit durumu ile ilgili değişkeni bu parametrede en fazla etkiye sahip değişkenlerdir. Bu değişkenleri Eski Polis Okulu Kavşağı'na ait değişkenler takip etmektedir.
- IEM CO₂ parametresinin kontrol altına alınmasının istendiği durumlarda, Terzibaba ve Eski Polis Okulu kavşaklarında geometrik düzenlemeler ve sinyalizasyon düzenlemeleri yapılması gerekliliği ortaya çıkmaktadır.
- Toplam seyahat süresi parametresi Destek Vektör Regresörü modellemesinde katsayı değeri 200-300 bandında seyreden Terzibaba Kavşağı'na ait değişkenler (mevcut durum, sinyalizasyonda artış, modern dönel kavşak) en büyük etkiye sahip değişkenler olarak bulunmuştur.
- Toplam seyahat süresi parametresi XGBoost modellemesinde “feature importance” değerleri 0,25'e yakın çıkan Terzibaba Kavşağı mevcut durum ile ilgili değişkeni ile 0,2'ye yakın çıkan Terzibaba Kavşağı altgeçit durumu ile ilgili değişkeni bu parametrede en fazla etkiye sahip değişkenlerdir.
- Toplam seyahat süresi parametresinin kontrol altına alınmasının istendiği durumlarda Terzibaba Kavşağı'nda geometrik düzenlemeler mecburidir.
- Başarılı modellemeler incelendiğinde her üç parametrenin de kontrolü için gerekli kavşak planlamaları yapılması sonucu ortaya çıkmaktadır.
- Her üç parametre için de özellikle otomobil hacminin kontrolünün yapılmaması sonucunda bu parametrelerde artışın gözleneceği görülmektedir.

- IEM modeline göre CO₂ salınımının ciddi oranda kavşak sinyalizasyonundaki değişikliklere bağlı olduğu gözlenmiştir. Bu sebeple kavşak sinyalizasyonunun da önemli planlamalara tabi tutulması gerekliliği ortaya çıkmaktadır.



KAYNAKLAR

- Aboah, A. and Arthur, E. (2021) “Comparative Analysis of Machine Learning Models for Predicting Travel Time”, *arXiv preprint arXiv:2111.08226*.
- AIMSUN. “Calculation Of Traffic Statistics”
<https://docs.aimsun.com/next/22.0.1/UsersManual/CalculationOfTrafficStatistics.html>
Son erişim tarihi: 05.06.2023
- Akkaya, S. ve Engin, T. (2022) “Trafik simülasyon yazılımlarına genel bakış”, *Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi*, 5(2), 157-168.
- Alatawneh, A., Shatanawi, M. and Mészáros, F. (2023) “Analysis of the Emergence of Autonomous Vehicles Using Simulation-based Dynamic Traffic Assignment–The Case of Budapest”, *Periodica Polytechnica Transportation Engineering*, 51(2), 126-132.
- Anh, N. T. N., Thang, T. N. and Solanki, V. K. (2022) “Quantifying the Qualitative Features”, *Artificial Intelligence for Automated Pricing Based on Product Descriptions*, 19-30.
- Aytekin, K. (2021) “Delay prediction using machine learning algorithms for connected autonomous traffic flow in uninterrupted facilities”, MSc. Thesis, *Institute for Graduate Studies in Science and Engineering, Boğaziçi University*, İstanbul.
- Bahuleyan, H. and Vanajakshi, L. D. (2017) “Arterial path-level travel-time estimation using machine-learning techniques”, *Journal of Computing in Civil Engineering*, 31(3), 04016070.
- Bayrak, O. Ü., Bayata, H. F., Hattatoğlu, F. and Çolak, M. A. (2016) “Evolution of a Junction Traffic Management Measures Using Microsimulation Model”, *Journal of the Institute of Science and Technology*, 6(3), 87-96.
- Bisong, E. (2019) “Building machine learning and deep learning models on Google cloud platform”, *Berkeley, CA: Apress*, Berkeley, 151-165.
- Celikoglu, H. B. and M. Dell’Orco (2007) “Mesoscopic Simulation Of A Dynamic Link Loading Process”, *Transportation Research Part C: Emerging Technologies* 15(5): 329-344.
- Chikaraishi, M., Garg, P., Varghese, V., Yoshizoe, K., Urata, J., Shiomi, Y. and Watanabe, R. (2020) “On the possibility of short-term traffic prediction during disaster with machine learning approaches: An exploratory analysis”, *Transport Policy*, 98, 91-104.
- Çoban, F. and Demir, L. (2021) “Yapay Sinir Ağları ve Destek Vektör Regresyonu ile Talep Tahmini: Gıda İşletmesinde Bir Uygulama”, *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 23 (67) , 327-338. DOI: 10.21205/deufmd.2021236729.
- IEA. “CO₂ Emissions in 2022”,
<https://www.iea.org/reports/co2-emissions-in-2022>
Son erişim tarihi: 05.06.2023

- de Villa, A. R., Casas, J., Breen, M. and Perarnau, J. (2014) “Static OD estimation minimizing the relative error and the GEH index”, *Procedia-Social and Behavioral Sciences*, 111, 810-818.
- Demircioğlu, M. S. (2021) “Sinyalize Kavşaklarda Performans Analizi ve Simülasyonu: Erzincan İli Örneği”, Yüksek Lisans Tezi, *Fen Bilimleri Enstitüsü, Erzincan Binali Yıldırım Üniversitesi*, Erzincan.
- Demiriz, A. O. (2019) “Mezoskopik Simülasyon ile Koridor Kapasite Analizi Yapılması: Erzincan İli Örneği”, Yüksek Lisans Tezi, *Fen Bilimleri Enstitüsü, Atatürk Üniversitesi*, Erzurum.
- Deveci, A. and Esen, M. F. (2022) “Medikal Sentetik Veri Üretimiyle Veri Dengelemesi”, *İstatistik ve Uygulamalı Bilimler Dergisi*, (5), 17-28.
- Elhenawy, M. M. Z. (2015) “Applying Machine and Statistical Learning Techniques to Intelligent Transport Systems: Bottleneck Identification and Prediction, Dynamic Travel Time Prediction, Driver Stoprun Behavior Modeling, and Autonomous Vehicle Control at Intersections”, Doctoral dissertation, *Virginia Polytechnic Institute and State University*, Virginia.
- Garmei, S. and Kashi, E. (2019) “Calibration of Aimsun roundabout model: Pedestrian and vehicles flow”, *Cogent Engineering*, 6(1), 1686796.
- Gnat, S. (2021) “Impact of categorical variables encoding on property mass valuation”, *Procedia Computer Science*, 192, 3542-3550.
- Google Earth. “Google Earth”,
<https://earth.google.com/>
Son erişim tarihi: 01.03.2023
- Güldamlaşı, G. (2007) “Tek Yön Sistemlerinin Çift Yöne Dönüştürülmesinin Sonuçları ve Performans Analizi Üzerine Araştırma (Balıkesir ve İzmir Örnekleri)”, Yüksek lisans Tezi, *Fen Bilimleri Enstitüsü, Balıkesir Üniversitesi*, Balıkesir.
- Günel, A. (2003) “Regresyon Denkleminin Başarısını Ölçmede Kullanılan Belirleme Katsayısı ve Kritiği”, *Doğuş Üniversitesi Dergisi*, 4(2), 133-140.
- Hattatoğlu, F., Bayata, H. F., Çolak, M. A. ve Bayrak, O. Ü. (2016) “Erzincan İli N. Muratoğlu Meydanı Kavşağı İçin Farklı Alternatiflerin Mikrosimülasyon Modellerlerinin Değerlendirilmesi”, *21. Sürdürülebilir Ulaşım İçin Yol Trafik Güvenliği Ulusal Kongresi Bildiri Kitabı*, 54.,
- Hedges&Company. “How Many Cars Are There In The World”
<https://hedgescompany.com/blog/2021/06/how-many-cars-are-there-in-the-world/>
Son erişim tarihi: 05.06.2023
- Hou, Q., Zhang, X., Li, B., Zhang, X. and Wang, W. (2019) “Identification of low-carbon travel block based on GIS hotspot analysis using spatial distribution learning algorithm”, *Neural Computing and Applications*, 31, 4703-4713.
- Hou, Y. and Edara, P. (2018) “Network scale travel time prediction using deep learning.”, *Transportation Research Record*, 2672(45), 115-123.

- Huang, J. and Perry, M. (2016) “A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting”, *International Journal of Forecasting*, 32(3), 1081-1086.
- Jaderberg, M., Simonyan, K., Vedaldi, A. and Zisserman, A. (2014) “Synthetic data and artificial neural networks for natural scene text recognition”, *arXiv preprint arXiv:1406.2227*.
- Jiang, D., Lin, W. and Raghavan, N. (2020) “A novel framework for semiconductor manufacturing final test yield classification using machine learning techniques”, *IEEE Access*, 8, 197885-197895.
- KGM, (2016) “Karayolu Tasarım El Kitabı”, *Karayolları Genel Müdürlüğü Yayınları*, Ankara, 100-199.
- Kırımlı, Y. S. (2019) “Eşdüzey kavşaklarda kavşaklar arası mesafenin kavşak performansına etkisinin araştırılması”, Yüksek Lisans Tezi, *Fen Bilimleri Enstitüsü, Karadeniz Teknik Üniversitesi*, Trabzon.
- Korkmaz E. (2019) “İzole Sinyalize Kavşaklarda Yapay Zekâ Teknikleri ile Trafik Sinyal Kontrolü ve Optimizasyonu”, Doktora Tezi, *Fen Bilimleri Enstitüsü, Kırıkkale Üniversitesi*, Kırıkkale.
- Kök, B. V. (2019) Karayolu Mühendisliği ve Tasarımı, *Nobel Akademik Yayıncılık*, Ankara, 137-147.
- Kramer, O. (2016) “Scikit-learn. Machine learning for evolution strategies”, *Springer*, Switzerland, 45-53.
- Kuzey, C. (2012) “Veri madenciliğinde destek vektör makinaları ve karar ağaçları yöntemlerini kullanarak bilgi çalışanlarının kurum performansı üzerine etkisinin ölçülmesi ve bir uygulama”, Doktora Tezi, *İstanbul Üniversitesi Sosyal Bilimler Enstitüsü*, İstanbul.
- Li, C. and Xu, P. (2021) “Application on traffic flow prediction of machine learning in intelligent transportation”, *Neural Computing and Applications*, 33, 613-624
- Jones, A. S. (2020) “A framework for smart traffic management using heterogeneous data sources”, Doctoral dissertation, *University of Wolverhampton*, England.
- Jiber, M., Mbarek, A., Yahyaouy, A., Sabri, M. A. and Boumhidi, J. (2020) “Road traffic prediction model using extreme learning machine: the case study of Tangier, Morocco”, *Information*, 11(12), 542.
- Mayr, A., Binder, H., Gefeller, O. and Schmid, M. (2014) “The evolution of boosting algorithms”, *Methods of information in medicine*, 53(06), 419-427.
- Manibardo, E. L., Laña, I. and Del Ser, J. (2021) “Deep learning for road traffic forecasting: Does it make a difference?”, *IEEE Transactions on Intelligent Transportation Systems*, 23(7), 6164-6188.
- Masiero, L. P., Casanova, M. A. and de Carvalho, M. T. M. (2011) “Travel time prediction using machine learning”, *In Proceedings of the 4th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, 34-38.

- Mecheva, T., Furnadzhiev, R. and Kakanakov, N. (2022) "Modeling driver behavior in road traffic simulation", *Sensors*, 22(24), 9801.
- McKinney, W. (2022) "Python for Data Analysis", *O'Reilly Media, Inc.*, 3-6.
- Natekin, A. and Knoll, A. (2013) "Gradient boosting machines, a tutorial", *Frontiers in neurorobotics*, 7, 21.
- Nguyen, J., Powers, S. T., Urquhart, N., Farrenkopf, T. and Guckert, M. (2021) "An overview of agent-based traffic simulators", *Transportation research interdisciplinary perspectives*, 12, 100486.
- Nikolenko, S. I. (2021) "Synthetic data for deep learning", (Vol. 174), *Springer Nature*, 11-13.
- Özen, N. S., Saraç, S. and Koyuncu, M. (2021) "COVID-19 Vakalarının Makine Öğrenmesi Algoritmaları ile Tahmini: Amerika Birleşik Devletleri Örneği", *Avrupa Bilim ve Teknoloji Dergisi*, *Ejosat 2021 Ocak*, 134-139.
- Özgün, B. (2021) "Trafik verilerinin üretilmesinde makine öğrenmesi yönteminin kullanımı", Yüksek Lisans Tezi, *Lisansüstü Eğitim Enstitüsü, İstanbul Aydın Üniversitesi*, İstanbul.
- Panis, L. I., Broekx, S. and Liu, R. (2006) "Modelling instantaneous traffic emission and the influence of traffic speed limits", *Science of the total environment*, 371(1-3), 270-285.
- Paz, A., Molano, V., Martinez, E., Gaviria, C. and Arteaga, C. (2015) "Calibration of traffic flow models using a memetic algorithm", *Transportation Research Part C: Emerging Technologies*, 55, 432-443.
- Purbarani, S. C., Sanabila, H. R., Wibisono, A., Alfiany, N., Wisesa, H. A. and Jatmiko, W. (2019) "Distance-To-Mean Continuous Conditional Random Fields: Case Study in Traffic Congestion", *Information*, 10(12), 382.MAP
- Python, W. (2021) "Python", *Python Releases for Windows*, 24.
- Rahmati, O., Naghibi, S. A., Shahabi, H., Bui, D. T., Pradhan, B., Azareh, A., Rafiei-Sardooi E., Samani A. N. and Melesse, A. M. (2018) "Groundwater spring potential modelling: Comparing the capability and robustness of three different modeling approaches", *Journal of hydrology*, 565, 248-261.
- Sekhar, C. R., Raj, P., Parida, P. and Gangopadhyay, S. (2013) "Estimation of delay and fuel loss during idling of vehicles at signalised intersection in Ahmedabad", *Procedia-Social and Behavioral Sciences*, 104, 1178-1187.
- Sider, T. M., Alam, A., Farrell, W., Hatzopoulou, M. and Eluru, N. (2014) "Evaluating vehicular emissions with an integrated mesoscopic and microscopic traffic simulation", *Canadian Journal of Civil Engineering*, 41(10), 856-868.
- Sykes, A. O. (1993) "An introduction to regression analysis", *Coase-Sandor Institute for Law & Economics Working Paper*, No. 20.
- ŞİRİN, S. (2020) "Rastgele orman algoritmaları ile otel özellikleri analizi", *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 4(2), 123-132.

- TÜİK. “2022 İl Göstergeleri”,
<https://biruni.tuik.gov.tr/ilgosterge>
Son erişim tarihi: 05.06.2023
- Tunalıoğlu, N. (2005) “Katlı Kavşakların ve Bağlantı Yollarının Örnek Uygulama Olarak Altunizade Katlı Kavşağı”, Yüksek Lisans Tezi, *Fen Bilimleri Enstitüsü, Yıldız Teknik Üniversitesi*, İstanbul.
- Van Rossum, G. (1991) “Python 0”, *Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County*, Baltimore.
- Vanajakshi, L. and Rilett, L. R. (2007) “Support vector machine technique for the short term prediction of travel time”, *In 2007 IEEE Intelligent Vehicles Symposium*, 600-605.
- Varga, B., Doba, D. and Tettamanti, T. (2023) “Optimizing vehicle dynamics co-simulation performance by introducing mesoscopic traffic simulation”, *Simulation Modelling Practice and Theory*, 125, 102739.
- WHO. “Road traffic injuries”,
<https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
Son erişim tarihi: 13.06.2023.
- Yürük, M. F. (2022) “Xgboost (Extreme Gradient Boosting) Tabanlı Algoritma ile Gümüş Fiyatlarının Tahmin Edilmesi”, *10th International Conference On Social Sciences & Humanities*, July 18-19, 2022, Sivas.
- Zahid, M., Chen, Y., Jamal, A. and Memon, M. Q. (2020) “Short term traffic state prediction via hyperparameter optimization based classifiers”, *Sensors*, 20(3), 685.
- Zhu, X. H., Lu, K. F., Peng, Z. R. and Gao, H. O. (2022) “Characterizing carbon emissions from China V and China VI gasoline vehicles based on portable emission measurement systems”, *Journal of Cleaner Production*, 378, 134458.