

MAG_Ara_19011042_19011093

by Fatih Ecevit

Submission date: 07-Jan-2024 05:19PM (UTC+0300)

Submission ID: 2267320021

File name: MAG_Ara_19011042_19011093.pdf (5.22M)

Word count: 13589

Character count: 84253

**TÜRKİYE CUMHURİYETİ
YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**



**UZUN VE KISA TATİL GÜNLERİ İÇİN TRAFİK TAHMİNİ:
ÖZEL YAKLAŞIMLAR**

19011093 – Fatih ECEVİT

19011042 – Yusuf Mert ÇELİKARSLAN

BİLGİSAYAR PROJESİ

Danışman
Doç. Dr. Mehmet Amaç GÜVENSAN

Aralık, 2023

TEŞEKKÜR

Bir projenin başarılı bir şekilde hayatı geçirilmesi, sadece teknik bilgi ve beceriyle değil, aynı zamanda yürekten verilen destek ve vizyonla mümkündür. Bu nedenle, Sayın Doç.Dr. Mehmet Amaç GÜVENSAN'ın projemizin gerçekleşmesindeki katkılarının büyüklüğünə vurgu yapmadan geçemiyoruz.

Sizler gibi derin bilgi birikimine ve deneyime sahip hocaların varlığı, öğrencilik yıllarımızın en büyük şanslarından biridir. Yol göstericiliğiniz, bize sadece akademik bilgi değil, aynı zamanda sektörde nasıl başarılı olabileceğimiz konusunda da kıymetli ipuçları sundu. Her sorumuzda, her endişemizde yanınızda oluşunuz ve bizi motive eden sözleriniz için minnettarız.

Projemizin her aşamasında bize verdığınız değerli geri dönüşler, vizyonumuzu şekillendirdi ve bizi daha da ileri taşıdı. Sizlerin desteği olmadan bu yolculukta bu kadar ilerleyemeyeceğimizi bilmek, bize gelecekteki projelerimiz için de büyük bir umut ve güven veriyor.

Bu vesileyle, sadece projemize katkılarınız için değil, bize kattığınız değerler, öğretikleriniz ve gösterdiğiniz yol için de teşekkür etmek istiyoruz. Sizin gibi değerli bir hocala sahip olduğumuz için şanslıyız.

Umarız ki gelecekte de sizinle çalışma fırsatı bulabiliriz. Sizlere olan minnettarlığımızı ifade etmekte kelimelerin yetersiz kaldığı bir gerçek, ama yine de dilimiz döndüğünde ifade etmek istedik.

Fatih ECEVİT
Yusuf Mert ÇELİKARSLAN

15 İÇİNDEKİLER

KISALTMA LİSTESİ	13 v
ŞEKİL LİSTESİ	vi
TABLO LİSTESİ	xi
ÖZET	xii
ABSTRACT	xiii
1 Giriş	1
1.1 Trafik Tahmin Probleminin Tarihçesi	1
1.2 Trafik Tahmin Probleminde Kullanılan Veri Türleri	1
1.3 Trafik Tahmin Modelleri ve Hata Metrikleri	2
1.4 Karşılaşılan Sorunlar	2
2 Ön İnceleme	3
2.1 Zaman Serileri	3
2.2 Probleme Farklı Yaklaşımlar	3
2.3 Veri	4
3 Fizibilite	5
3.1 Teknik Fizibilite	5
3.1.1 Yazılım Fizibilitisi	5
3.1.2 Donanım Fizibilitesi	5
3.2 Ekonomik Fizibilite	6
3.3 Legal Fizibilite	6
3.4 İşgücü ve Zaman Fizibilitesi	6
4 Sistem Analizi	7
4.1 Veri Seti Analizi	7
4.1.1 Seasonal Decomposition	7
4.1.2 Geçmiş Verilerle Korelasyon	9
4.1.3 Özellik Seçimi ve Farklı Bayramlardaki Kullanımları	9

4.1.4	Test Seti Olarak Kullanılan Segmentler ve Özellikleri	10
4.2	Uzun Tatiller	15
4.2.1	Kurban Bayramı	15
4.2.2	Ramazan Bayramı	16
4.3	Kısa Tatiller	17
5	Sistem Tasarımı	23
5.1	Algoritmalar ve Model Eğitimi	23
5.2	Karar Ağacı Modelleri	23
5.3	Derin Öğrenme Modelleri	25
6	Uygulama	26
6.1	Fast Fourier Transform Ön İşlemesi	26
6.2	Kayıp Değerleri Doldurma	29
6.3	Random Forest Regressor	29
6.3.1	Grafikler	31
6.4	XGBoost	32
6.4.1	Grafikler	34
6.5	Gradient Boosting	35
6.5.1	Grafikler	36
6.6	CatBoost	37
6.6.1	Grafikler	39
6.7	LGBMRegressor	40
6.7.1	Grafikler	42
6.8	DNNs	43
6.8.1	Grafikler	44
6.9	Ensemble	46
6.9.1	Grafikler	46
6.10	Hata Oranları	47
6.10.1	Kısa Bayramlar	47
6.10.2	Uzun Bayramlar	50
6.11	Deneysel Sonuçlar	50
6.12	Performans Analizi	56
6.13	Sonuç	58
Referanslar		60
Özgeçmiş		61

KISALTMA LİSTESİ

ANN	Artifical Neural Networks
CNN	Convolutional Neural Networks
DFT	Discrete Fourier Transform
GBM	Gradient Boosting Model
LGBM	Light Gradient Boosting Model ²³
LSTM	Long-short Term Memory
MA	Moving Average
RFR	Random Forest Regressor
RNN	Artifical Neural Networks
SVR	Support Vector Regression
MDI	Mean Decrease in Impurity

ŞEKİL LİSTESİ

Şekil 4.1 Sistem Tasarım Modeli	8
Şekil 4.2 5 dk. aralıklarla	8
Şekil 4.3 Günlük	8
Şekil 4.4 Bağdat Caddesi Korelasyon eğrisi	8
Şekil 4.5 Korelasyon	9
Şekil 4.6 Kısmı Korelasyon	10
Şekil 4.7 Değişken Önem Yatay Bar Grafiği	11
Şekil 4.8 MDI Özellik Önemi	11
Şekil 4.9 Permütasyon Özellik Önemi	12
Şekil 4.10 E-5 Ana Arter	13
Şekil 4.11 Kurbağalıdere Ana Arter [6]	13
Şekil 4.12 Kurbağalıdere Yol Çalışması Sonucu Etkilenen 431 Segmenti . .	14
Şekil 4.13 Şehir İçi Yol	14
Şekil 4.14 Bağdat Caddesi 2018 Kurban Bayramı	16
Şekil 4.15 Bağdat Caddesi 2019 Kurban Bayramı	16
Şekil 4.16 Bağdat Caddesi 2018 Ramazan Bayramı	17
Şekil 4.17 2019	17
Şekil 4.18 Bağdat Caddesi 2019 Ramazan Bayramı	17
Şekil 4.19 1 Ocak 2018	22
Şekil 4.20 1 Ocak 2019	22
Şekil 4.21 2018 ve 2019 Yıllarında 1 Ocak Hız Grafikleri	22
Şekil 6.1 Düşük Bant Aralıklı FFT	27
Şekil 6.2 Yüksek Bant Aralıklı FFT	27
Şekil 6.3 Düşük Frekans Filtreleme Uygulanmış Veri ve Gerçek Veri . . .	28
Şekil 6.4 Yüksek Frekanslı Veri	28
Şekil 6.5 1 Mayıs 2019 Random Forest	31
Şekil 6.6 30 Ağustos 2019 Random Forest	31
Şekil 6.7 Avcılar 414 Nolu Segment Haftaiçine Denk Gelen Bayramlar . .	31
Şekil 6.8 23 Nisan 2019 Random Forest	31
Şekil 6.9 29 Ekim 2019 Random Forest	31
Şekil 6.10 Ataşehir 316 Nolu Segment Haftasonu ile Birleşen Bayramlar . .	31

Şekil 6.11 19 Mayıs 2019 Random Forest	31
Şekil 6.12 15 Temmuz 2019 Random Forest	31
Şekil 6.13 Kadıköy 2102 Nolu Segment Farklı Karakteristik İzleyen Bayramlar	31
Şekil 6.14 Ramazan Bayramı 2019 Random Forest	32
Şekil 6.15 Kurban Bayramı 2019 Random Forest	32
Şekil 6.16 Kadıköy 2102 nolu Uzun Bayramlar	32
Şekil 6.17 1 Mayıs 2019 XGBoost	34
Şekil 6.18 30 Ağustos 2019 XGBoost	34
Şekil 6.19 Avcılar 414'nolu Segment Haftaiçine Denk Gelen Bayramlar	34
Şekil 6.20 23 Nisan 2019 XGBoost	34
Şekil 6.21 29 Ekim 2019 XGBoost	34
Şekil 6.22 Ataşehir 316'nolu Segment Haftasonu İle Birleşen Bayramlar	34
Şekil 6.23 19 Mayıs 2019 XGBoost	34
Şekil 6.24 15 Temmuz 2019 XGBoost	34
Şekil 6.25 Kadıköy Ankara Asfaltı 2102'nolu Segment Farklı Karakteristik İzleyen Bayramlar	34
Şekil 6.26 Ramazan Bayramı 2019 XGBoost	35
Şekil 6.27 Kurban Bayramı 2019 XGBoost	35
Şekil 6.28 Kadıköy Ankara Asfaltı Uzun Bayramlar	35
Şekil 6.29 1 Mayıs 2019 Gradient Boosted	36
Şekil 6.30 30 Ağustos 2019 Gradient Boosted	36
Şekil 6.31 Avcılar 414'nolu Segment Haftaiçine Denk Gelen Bayramlar	36
Şekil 6.32 23 Nisan 2019 Gradient Boosted	36
Şekil 6.33 29 Ekim 2019 Gradient Boosted	36
Şekil 6.34 Ataşehir 316'nolu Segment Haftasonu İle Birleşen Bayramlar	36
Şekil 6.35 19 Mayıs 2019 Gradient Boosted	37
Şekil 6.36 15 Temmuz 2019 Gradient Boosted	37
Şekil 6.37 Kadıköy Ankara Asfaltı 2102'nolu Segment Farklı Karakteristik İzleyen Bayramlar	37
Şekil 6.38 Ramazan Bayramı 2019 Gradient Boosted	37
Şekil 6.39 Kurban Bayramı 2019 Gradient Boosted	37
Şekil 6.40 Kadıköy Ankara Asfaltı 2102'nolu Segment Uzun Bayramlar	37
Şekil 6.41 1 Mayıs 2019 Catboost	39
Şekil 6.42 30 Ağustos 2019 Catboost	39
Şekil 6.43 Avcılar 414'nolu Haftaiçine Denk Gelen Bayramlar	39
Şekil 6.44 23 Nisan 2019 Catboost	39
Şekil 6.45 29 Ekim 2019 Catboost	39
Şekil 6.46 Ataşehir 316'nolu Segment Haftasonu İle Birleşen Bayramlar	39
Şekil 6.47 19 Mayıs 2019 Catboost	40

Şekil 6.48 15 Temmuz 2019 Catboost	40
Şekil 6.49 Kadıköy Ankara Asfaltı 2102'nolu Segment Farklı Karakteristik İzleyen Bayramlar	40
Şekil 6.50 Ramazan Bayramı 2019 Catboost	40
Şekil 6.51 Kurban Bayramı 2019 Catboost	40
Şekil 6.52 Kadıköy Ankara Asfaltı 2102'nolu Segment Uzun Bayramlar . . .	40
Şekil 6.53 1 Mayıs 2019 LightGBM	42
Şekil 6.54 30 Ağustos 2019 LightGBM	42
Şekil 6.55 Avcılar 414'nolu Segment Haftaiçine Denk Gelen Bayramlar . . .	42
Şekil 6.56 23 Nisan 2019 LightGBM	42
Şekil 6.57 29 Ekim 2019 LightGBM	42
Şekil 6.58 Ataşehir 316'Nolu Segment Haftasonu İle Birleşen Bayramlar . .	42
Şekil 6.59 19 Mayıs 2019 LightGBM	43
Şekil 6.60 15 Temmuz 2019 LightGBM	43
Şekil 6.61 Farklı Karakteristik İzleyen Bayramlar	43
Şekil 6.62 Ramazan Bayramı 2019 LightGBM	43
Şekil 6.63 Kurban Bayramı 2019 LightGBM	43
Şekil 6.64 Kadıköy Ankara Asfaltı Uzun Bayramlar	43
Şekil 6.65 1 Mayıs 2019 DNNs	45
Şekil 6.66 30 Ağustos 2019 DNNs	45
Şekil 6.67 Avcılar 414'nolu Haftaiçine Denk Gelen Bayramlar	45
Şekil 6.68 23 Nisan 2019 DNNs	45
Şekil 6.69 29 Ekim 2019 DNNs	45
Şekil 6.70 Ataşehir 316'nolu Haftasonu İle Birleşen Bayramlar	45
Şekil 6.71 19 Mayıs 2019 DNNs	45
Şekil 6.72 15 Temmuz 2019 DNNs	45
Şekil 6.73 Kadıköy Ankara Asfaltı Farklı Karakteristik İzleyen Bayramlar .	45
Şekil 6.74 1 Mayıs 2019 Ensemble	46
Şekil 6.75 30 Ağustos 2019 Ensemble	46
Şekil 6.76 Avcılar 414'nolu Segment Haftaiçine Denk Gelen Bayramlar . .	46
Şekil 6.77 23 Nisan 2019 Ensemble	46
Şekil 6.78 29 Ekim 2019 Ensemble	46
Şekil 6.79 Ataşehir 316'nolu Haftasonu İle Birleşen Bayramlar	46
Şekil 6.80 19 Mayıs 2019 Ensemble	46
Şekil 6.81 15 Temmuz 2019 Ensemble	46
Şekil 6.82 Farklı Karakteristik İzleyen Bayramlar	46
Şekil 6.83 Ramazan Bayramı 2019 Ensemble	47
Şekil 6.84 Kurban Bayramı 2019 Ensemble	47
Şekil 6.85 Kadıköy Ankara Asfaltı 2102'nolu Segment Uzun Bayramlar . .	47

Şekil 6.86 Eğitim Setinde MDI kullanılması	51
Şekil 6.87 Test Setinde Permüstasyon Kullanılması	51
Şekil 6.88 Farklı Yöntemlerle Hesaplanan Değişken Önemi	51
Şekil 6.89 Eğitim Setinde MDI kullanılması	51
Şekil 6.90 Test Setinde Permüstasyon Kullanılması	51
Şekil 6.91 Ardışık Tatil Değişkeninin Model Performansına Etkisi	51
Şekil 6.92 23 Nisan 2019	52
Şekil 6.93 1 Mayıs	52
Şekil 6.94 19 Mayıs	52
Şekil 6.95 15 Temmuz	52
Şekil 6.96 30 Ağustos	52
Şekil 6.97 29 Ekim	52
Şekil 6.98 Tatil Günleri İçin 106 Segment'in 'Gerçek Ve Tahmin' Hız Değer Grafikleri	52
Şekil 6.99 Eğitim Seti Nesnesi	53
Şekil 6.10 Kadıköy (312 Numaralı Segment)	53
Şekil 6.101 Maltepe (321 Numaralı Segment)	53
Şekil 6.102 Bakırköy (420 Numaralı Segment)	53
Şekil 6.103 Üsküdar (614 Numaralı Segment)	53
Şekil 6.104 Beşiktaş (64 Numaralı Segment)	53
Şekil 6.105 15 Temmuz 2019 Tatilinde 5 Farklı Segmentin 'Gerçek Ve Tahmin' Hız Değer Grafikleri	53
Şekil 6.106 Kadıköy (312 Numaralı Segment)	54
Şekil 6.107 Maltepe (321 Numaralı Segment)	54
Şekil 6.108 Bakırköy (420 Numaralı Segment)	54
Şekil 6.109 Üsküdar (614 Numaralı Segment)	54
Şekil 6.110 Beşiktaş (64 Numaralı Segment)	54
Şekil 6.111 1 Mayıs 2019 Tatilinde 5 Farklı Segmentin 'Gerçek Ve Tahmin' Hız Değer Grafikleri	54
Şekil 6.112 Random Forest	55
Şekil 6.113 XGB Regressor	55
Şekil 6.114 CatBoost	55
Şekil 6.115 Ensemble	55
Şekil 6.116 Kurban Bayramı(2019) Tatilinde Kadıköye Farklı Modellerin 'Gerçek Ve Tahmin' Hız Değer Grafikleri	55
Şekil 6.117 Random Forest	55
Şekil 6.118 XGB Regressor	55
Şekil 6.119 CatBoost	55
Şekil 6.120 Ensemble	55

Şekil 6.12 Ramazan Bayramı(2019) Tatilinde Kadıköyde Farklı Modellerin
'Gerçek Ve Tahmin' Hız Değer Grafikleri 55

TABLO LİSTESİ

Tablo 4.1 Tatil Günlerinin Türlerine Özel Tahmin Setleri	18
Tablo 6.1 1 Mayıs 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Bayramlar ,Set-2: Set1 + Aynı Sene Geçen Hafta Set-3: Set2 + 23 Nisan 2019, Set-4: Set3 + FFT	48
Tablo 6.2 23 Nisan 2019 Hata Oranları, Set-1: Aynı Yıl Geçen Hafta ,Set-2: Set1 + Aynı Sene Geçen Hafta Set-3: Set1 + FFT, Set-4:,Set2 + FFT	48
Tablo 6.3 29 Ekim 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Tatiller ve önceki günleri ,Set-2: Set1 + Aynı Sene Geçen Hafta Set-3: 2 Hafta 2018 + Önceki 2 Hafta 2019, Set-4:,Set3 + FFT	48
Tablo 6.4 19 Mayıs 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Tatiller ve önceki bir gün, Set-2: Aynı Yıl Bir Hafta Önce Aynı Gün Set-3: Set1 + FFT, Set-4: Set2 + FFT	49
Tablo 6.5 15 Temmuz 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Tatiller ve önceki bir gün, Set-2: Set1 + FFT, Set-3: Set1 + Aynı yıl önceki haftasonu ve önceki hafta aynı gün + FFT	49
Tablo 6.6 30 Ağustos 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Tatiller ve önceki bir gün + Geçen Hafta, Set-2: Set1 + FFT, Set-3: Set2 + Aynı Sene Tek Günlük Bayramlar(23 Nisan, 1 Mayıs)	49
Tablo 6.7 Kurban Bayramı 2019 Hata Oranları, Set-1: Geçen Sene Kurban Bayramı 2 Hafta 13 - 27 Ağustos, Set-2: Set1 + Ramazan Bayramı, Set-3: Set1 + Aynı Sene Önceki Hafta sonu + FFT	50
Tablo 6.8 Ramazan Bayramı 2019 Hata Oranları, Set-1: Geçen Sene Ramazan Bayramı 2 Hafta 11 - 25 Haziran, Set-2: Set1 + Geçen Sene Haftasonu + Aynı Sene Bir Önceki Haftasonu, Set-3: Set1 + Kurban Bayramı	50
Tablo 6.9 İstanbul Geneli Segmentlerin MAPE Ortalaması	57
Tablo 6.10 Örnek Segmentlere Göre MAPE Hata Oranı	57

ÖZET

UZUN VE KISA TATİL GÜNLERİ İÇİN TRAFİK TAHMİNİ: ÖZEL YAKLAŞIMLAR

Fatih ECEVİT

Yusuf Mert ÇELİKARSLAN

Bilgisayar Mühendisliği Bölümü

Bilgisayar Projesi

Danışman: Doç. Dr. Mehmet Amaç GÜVENSAN

Trafik analizi ve araç hız tahmini günümüzde bireylerin günlük yaşantısına önemli ölçüde etki etmektedir. 21. Yüzyılın başlangıcından günümüze dünya nüfusundaki yaklaşık %33'lük artış ve araç sayılarındaki yaklaşık %43'lük artış hem insan hem araç sayılarındaki artış hızının yüksek olduğunu, insan ve araç sayılarındaki bu artışın beraberinde trafik yoğunluğu getirdiğini de görüyoruz. Analiz ve tahmin süreçlerinin başarılı bir şekilde ilerlemesi ve sonuçlanması için trafikteki araçların hız verilerinin eksiksiz ve doğru olması gerekmektedir. Bu projede İstanbul sınırları içerisinde hız tahmini yapılması planlanıp, İstanbul Büyükşehir Belediyesi tarafından tutulan veriler kullanılarak sayısal verilerin görselleştirilmesi, analiz edilmesi ve yorumlanması adımları sonrasında uzun ve kısa tatil günleri için trafik tahmini yapan bir araç geliştirilmiştir. Analiz adımda İstanbul'da birçok lokasyondan alınan verilerde harita üzerindeki konumu, zaman-hız verileri esas alınarak grafiksel olarak görselleştirme işlemi gerçekleştirılmıştır. Görselleştirme sonrası yorumlama kısmı ile belirli günlerin diğer günler ile benzerlikleri tespit edilmiş Random Forest, GradientBoosting, CatBoost, LGBM, XGBRegressor modelleri kullanılarak ortalama %12.63 hata oranı ile kısa ve uzun tatil günleri hız tahminini gerçekleştirebilmekteyiz.

Anahtar Kelimeler: Trafik hız tahmini, uzun süreli, tatil, XGBoost, denetimli, zaman serisi

ABSTRACT

TRAFFIC FORECASTING FOR LONG AND SHORT HOLIDAYS: SPECIAL APPROACHES

Fatih ECEVİT
Yusuf Mert ÇELİKARSLAN

⁷
Department of Computer Engineering
Computer Project

Advisor: Assoc. Prof. Dr. Mehmet Amaç GÜVENSAN

Traffic analysis and vehicle speed estimation have a significant impact on the daily lives of individuals today. Since the beginning of the 21st century, we see that the increase in the world population by approximately 33% and the increase in the number of vehicles by approximately 43%, the rate of increase in the number of both people and vehicles is high, and this increase in the number of people and vehicles brings traffic density. In order for the analysis and prediction processes to proceed and result successfully, the speed data of the vehicles in traffic must be complete and accurate. In this project, we plan to predict the speed within the borders of Istanbul, and after the visualization, analysis and interpretation of the numerical data using the data kept by the Istanbul Metropolitan Municipality, a tool that predicts traffic for long and short holidays has been developed. In the analysis step, graphical visualization was performed based on the location on the map, time-speed data in the data taken from many locations in Istanbul. After visualization, with the interpretation part, the similarities of certain days with other days are determined and by using Random Forest, GradientBoosting, CatBoost, LGBM, XGBRegressor models, we are able to predict the speed of short and long holidays with an average error rate of 12.63%.

Keywords: Traffic speed prediction, long term, holiday, XGBoost, supervised, time-series

1

Giriş

Trafik Tahmin Problemi

1.1 Trafik Tahmin Probleminin Tarihçesi

Varolan makaleler tarandığında trafik tahmini üzerine 1960'lardan beri birçok model geliştirilmiştir. İlk olarak istatistiksel modeller olan ARIMA ve onun geliştirilmiş modelleri olan SARIMA ve SARIMAX modelleri trafik tahmin problemlerinde kullanılmıştır. Daha sonraları makine öğrenmesi modellerinin regresyon verileri üzerindeki başarısı bu modellerin zaman serileri içinde kullanılmasına neden olmuştur. Makine öğrenme modelleri istatistiksel modellere oranla veride oluşan anlık değişimleri yakalamada daha başarılıdır fakat mevsimsel değişimleri tahmin etmekte daha başarısız bir performans gösterir. Günümüzde ise derin öğrenme modellerinin gelişmesiyle zaman seri problemlerinde de çeşitli uygulamaları denenmiştir. Nöral ağların makine öğrenme modellerine göre avantajı uzun süreli tahminlerde geçmiş verileri saklama becerisidir.

1.2 Trafik Tahmin Probleminde Kullanılan Veri Türleri

Her zaman serisinde olduğu gibi trafik verisinde de univariate ve multivariate olarak veri çeşitleri ikiye ayrılmaktadır. Trafik verisinin hava durumu, sıcaklık, yol durumu vs. gibi farklı özelliklerle desteklenmesiyle elimizdeki modelin doğruluk oranı artırlabilir. Trafik verisi kendi içinde de farklılaşmaktadır. Bakıldığından literatürdeki tahmin modelleri traffic akışını, trafik yoğunluğunu, yoldaki araç sayısını ve trafik hızını kullanmaktadır. Bu projede İstanbul Büyükşehir Belediyesinde 5 dakika aralıklarla örneklenen 2018 ve 2019 yılları trafik hızı verisi kullanılmıştır.

1.3 Trafik Tahmin Modelleri ve Hata Metrikleri

Yapılan literatür taramasında [1]DFT-SVR, [2]XGBoost ve derin öğrenme modelleri [3]LSTM, CNN, ANN gibi modellerin uzun zamanlı trafik tahmininde daha başarılı olduğu gözlemlenmiştir. SVR modeli ve lineer regresyon modelleri üzerinde çeşitli testler yapılmış beklenilen sonuç alınamamıştır. Kısa süreli tahminlerde kabul edilebilir sonuçlar bulunmasına rağmen hem uzun süreli ve de tatil günlerinin değişken karakteristiği göz önüne alındığında SVR modeli bu araştırma konusu için yeterli olmadığı kanısına varıldı. Bu nedenle XGBoost ve Derin öğrenme algoritmaları üzerinden ilerlenmiştir. Hata ölçme metriği olarak ise MAPE modellerin sonuçlanma hmasını hesaplamak ve MAE ise yapılan tahminlerdeki en kötü ve en iyi tahminleri bularak buna göre parametreler belirlemek için tercih edilmiştir.

1.4 Karşılaşılan Sorunlar

Grafikler incelendiğinde en sorunlu ve yoğun günler uzun süreli tatil günleridir. Bir ve iki günlük resmi tatiller trafik verisinin olağan akışına pek bir etki etmezken haftasonuyla birleşen resmi tatiller veya kurban ve ramazan bayramı gibi uzun süreli tatiller birkaç günlük trafik yoğunluğuna sebep olduğu görülmüştür. Bu gözlem ışığında iki konu göze çarpmaktadır yılbaşı gibi anlık yoğunluktan oluşan çok anı hız düşüşleri veya uzun tatillerde oluşan kademeli düşüşler.

2 Ön İnceleme

Zaman serisi tahmin problemleri doğası gereği diğer makine öğrenme problemlerinden daha zorludur ve doğru sonuçlar elde edilebilmesi için eldeki verinin analizin doğru yapılması, veriye uygun bir model seçilmesi ve verinin kullanılmadan önce işlemden geçilmesi gereklidir.

2.1 Zaman Serileri

Trafik hız zaman serisi finans, tüketim ve diğer zaman serileri arasında birkaç fark vardır. Bunlardan ilki trafik zaman serisi belirgin bir trend eğrisi izlemez. Bunu İstanbul için düşünürsek TÜİK'in 2022 Mart ayında[4] açıkladığı verilere göre İstanbul trafiğine günlük 585 toplamda 213.486 yeni araç katıldı. Buna rağmen bu araçların trafik hız verisine 1 yıllık bir değişimde etkisi büyük ölçüde değil ve grafiklerde yıllık olarak trafik hız verisinde dikkate değer biçimde bir artış veya azalış gözlemlenmiyor. Zaman serilerindeki durağanlığın bir ölçüsü olan ADF testi elimizdeki veriye uygulandığında da bu gözlemi destekliyor. Eşik p değeri 0.05 olan test veriye uygulandığında 0.0 sonucunu vermiştir.

İkinci fark ise dış etkenlerin diğer verilere oranla trafik verisinde bariz biçimde etkilerinin görülmemesidir. Hava durumu, sıcaklık, tatiller, yol çalışmaları, insanların o anki tutumları gibi birçok etken trafik verisini doğrudan etkilemektedir. Bu nedenle projenin ilerleyen aşamalarında bu dış etkenler ekstra değişkenler olarak veriye eklenebilir. Literatürde de [5] hava durumu üzerinden trafik tahmini yapılan çalışmalara rastlanmıştır.

2.2 Probleme Farklı Yaklaşımlar

Literatür incelediğinde uzun süreli tatil günleri konusunda yapılan çalışma sayısının makine öğrenmesi ve derin öğrenme konularının geniş yelpazesi düşünüldüğünde az olduğu görülmüştür. Bu nedenle [1] DFT-SVR modeli MAPE değerleri incelendiğinde

ARIMA ve diğer istatistiksel modellerden daha tutarlı sonuçlar verdiği görülmektedir. Ayrıyetten SVR modelinin basit doğası nedeniyle makalenin ortaya çıkardığı sonuçları yorumlamak kolaydır. Model Çin'in Jiangsu şehrinde National günü ve Tomb-sweeping gününde test edilmiştir. İstanbul'un nüfusu ve bu tatil günlerinin Türkiye'deki resmi tatillerin uzunluklara benzerliği düşünüldüğünde çalışmamızdaki model seçimimizde etkisi olmuştur.

Bir diğer makalede [3] ise LSTM kullanılarak bir günlük araç yoğunluk tahmini yapılmıştır geçmiş günlerin tahmine olan etkisi grafiklendirilmiştir.

2.3 Veri

Projede İstanbul şehrinde farklı yol segmentlerinden toplanarak elde edilen trafik hızı verisi kullanılacaktır.

3

Fizibilite

3.1 Teknik Fizibilite

Projenin amacına uygun olarak gereksinimlerin belirlenmesi ve planlanması yapıldı ayrıca veri çeşitliliği ve miktarı, incelenerek bellek gereksinimleri ve işlem süreleri gibi teknik detaylar belirlendi.

3.1.1 Yazılım Fizibiletisi

Veri önişleme, veri analizi, makine öğrenimi modeli geliştirme ve modelin test edilmesi için uygun teknolojiler belirlendi (Python programlama dili, Scikit-learn, XGBoost, LightGBM, CatBoost, TensorFlow, Keras, vb.). Belirlenen algoritma doğrultusunda kullanılacak programlar. Süreç içerisinde kullanilan sistemin performans olarak yeterli olduğu görüldü. Projenin işletim sistemi olarak Windows üzerinde hazırlanması ve çalıştırılması yapıldı. Windowsa kurulan Anaconda ile spyder üzerinden Python(3.11) ile algoritma çalıştırılmaları yapıldı

3.1.2 Donanım Fizibilitesi

2 senelik veriler İstanbul için 2000'e yakın farklı lokasyon için 5 dakikada bir alınan verilerin 9 GB yer kaplaması ve bu verileri işleyerek çözüm üretme sürecinin çalışabilir, hızlı ve yüksek performanlı olması gereklidir. Bu gerekliliği karşılamak için donanımsal gereksinimlerin karşılanması gereklidir. Mevcut donanım olarak 16 GB Ram kapasiteli , 4 GB GTX 1650 ekran kartına sahip,500 GB SSD 2.60 GHz CPU işlemcili sahip bir dizüstü bilgisayar kullanılması planlandı ve veri incelemesi, veri işleme ve algoritma geliştirmeye süreçlerinde gerektiği zaman Google Colab'da kullanılarak donanımsal gereksinimler karşılandı.

3.2 Ekonomik Fizibilite

Donanım imkanları örnek veri setleri için şuanlık yeterli olduğundan yeni bir donanıma veya bulut geliştirme ortamına ücret ödenmemiştir.

3.3 Legal Fizibilite

Veri toplama ve işleme faaliyetleri için yasal gerekliliklerin belirlenmesi ve buna uygun hareket edilmesi (örneğin, veri gizliliği, kişisel verilerin korunması vb.) trafik yönetiminde kullanılabilecek verilerin uygunluk ve güvenilirlik açısından incelenmesi ve proje sürecinde elde edilen sonuçların kullanımı veya paylaşımı için yasal gerekliliklerin belirlenmesi adımlarının takip edilmesi sonucu projemizin herhangi bir yasal sorun teşkil etmediği tespit edildi.

3.4 İşgücü ve Zaman Fizibilitesi

Projenin gereksinimlerine uygun olarak bir proje yönetim planı hazırlanması, ekip üyelerinin görevlerinin ve sorumluluklarının tanımlanması, proje süresinin belirlenmesi ve proje süresince yapılacak olan işlerin ayrıntılı bir şekilde proje yürütütürü ile ve ekip içinde planlanması, gereksinimlerin değişimleri durumlarda, ekibin esneklik sağlama ve projeyi başarıyla tamamlaması için gerekli planlamaların yapılması adımlarından geçilmiştir. Süreç içerisinde literatür taraması ve ön çalışma süreci yapılmıştır. Daha sonra gerekli modeller belirlendi ve bu alandaki çalışmalarımız ilerledi.

4

Sistem Analizi

4.1 Veri Seti Analizi

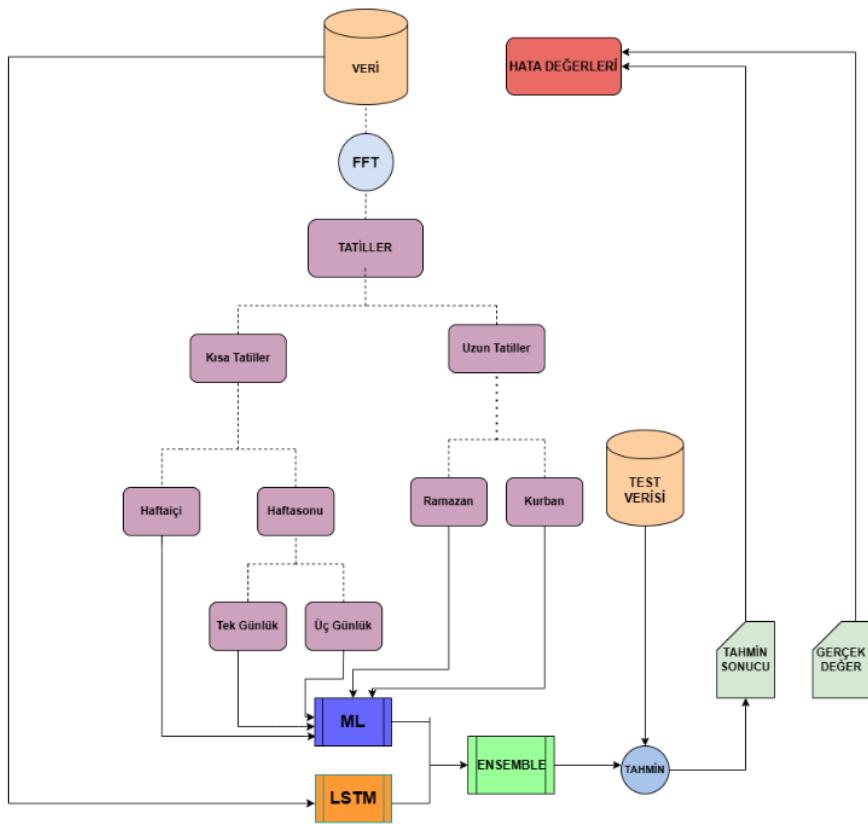
Trafik tahmini için gerekli olan verileri İstanbul Büyükşehir Belediyesi (İBB) tarafından temin edilmiştir. Temin edilen veriler .csv formatında olup Google Drive'da saklanmaktadır. CSV dosyaları 2018 ve 2019 yılına ait en az 2000 farklı lokasyondan gidiş ve dönüş yolları için geçen araçların ortalama hızlarını tutmaktadır. Ortalama hız verilerini her 5 dakikada bir almaktadır. Trafik akış verileri .csv formatında, her lokasyon için farklı csv dosyası ve numarası olacak ve ilgili yolun gidiş yönü için segment0 dönüş yolu için segment1 isimlendirilmesi ile, zaman bilgisi ve ortalama hız bilgisi ile birlikte tutulmaktadır. Hangi segmentin hangi konumu belirttiği bir Json dosyası içerisinde enlem ve boylam bilgisi olarak tutulmaktadır. Veri incelendiğinde bazı aralıklarda verilerin -1 değeri aldığı görülmüştür. Bu verilerin eğitim ve test adımlarında sorun teşkil etmesi nedeni ile bir fonksiyon çağrısı ile ortalama değerler kullanılarak -1 değerler doldurulmuştur. Fonksiyon içeriği daha sonra 'Kayıp değerleri doldurma' başlığı altında detaylı açıklanacaktır.

4.1.1 Seasonal Decomposition

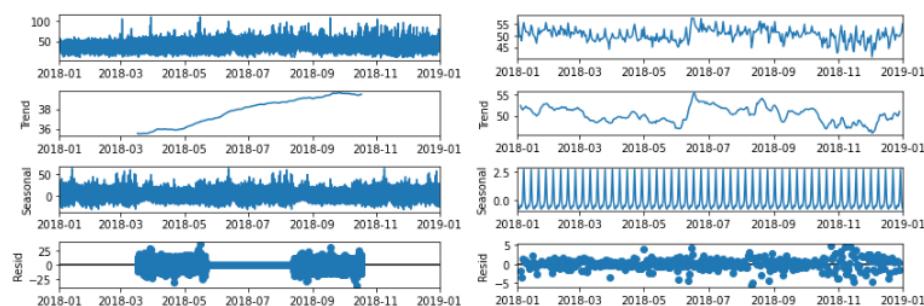
Veriyi mevsimsel olarak ayırttığımızda trend, seasonal ve residual eğrilerini görmemiz mümkün oldu. Bu sayede veri konusunda daha iyi bir anlayış geliştirdik. Şekil 4.4 'de görüldüğü gibi belirgin bir trend eğrisi yok fakat seasonal eğri belirgin biçimde kendini gösteriyor. Yapılan ADF testi de gözlemlerimizin sonucunu doğrulamıştır.

Veri günlük olarak yeniden örneklediğinde veri gürültüsü azaldığı için daha belirgin mevsimsel eğriler gözlemlenmiştir. Şekil 4.3'de eğrideki mevsimsel değişim görülebilir.

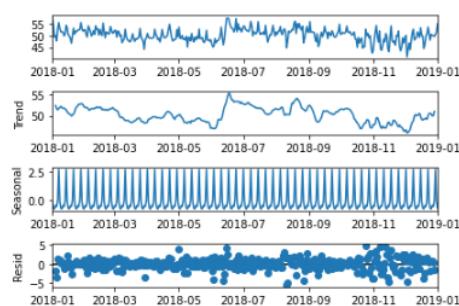
Sonuç olarak bu iki grafik bize uzun tatil günlerinde eğer mevsimsel tekrar zamanlarındaki kışım eğitim veri seti olarak alınırsa tahmin sonucunun daha doğru



Şekil 4.1 Sistem Tasarım Modeli



Şekil 4.2 5 dk. aralıklarla



Şekil 4.3 Günlük Korelasyon Eğrisi

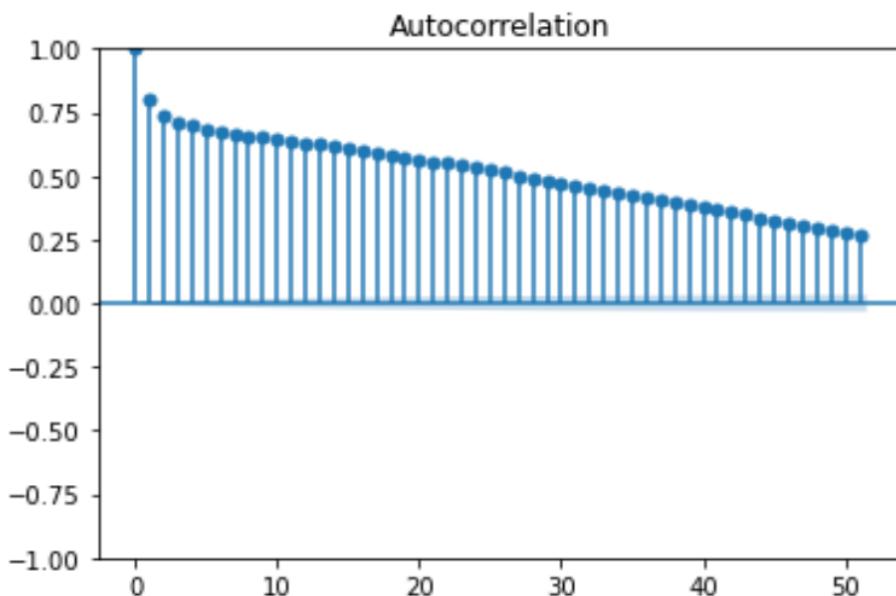
Şekil 4.4 Bağdat Caddesi Korelasyon eğrisi

çıkacağı öngörüsüne ulaştırmıştır.

4.1.2 Geçmiş Verilerle Korelasyon

Bir zaman serisi probleminde geçmiş verilerin modelin yaptığı tahmine etkisi en önemli konulardan birisidir. Elimizdeki veri 5 dk. aralıklarla olduğu için 5dk. önceki bir artış ya da düşüş eğilimini korumaktadır. Bu nedenle korelasyon ve kısmi korelasyon eğrilerini incelememiz gereklidir.

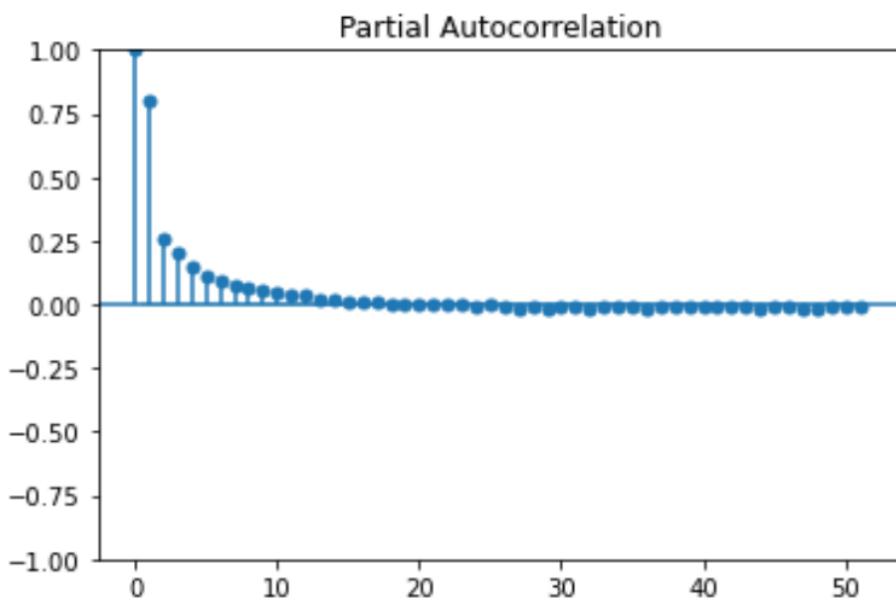
Şekil 4.5 korelasyon eğrisine baktığımızda geçmiş örneklemelere doğru azalarak devam eden bir korelasyon görmekteyiz. Geçmişteki hangi örneklemelerin sonuca daha çok etki ettiğini anlamak için Şekil 4.6'deki kısmi korelasyon eğrisini incelemek gereklidir. Bu grafikte görüldüğü üzere 0 ve 10 adım önceki verilerin korelasyonu göze çarpmaktadır. Bu nedenle tatil verileri incelenirken tatil gününden 1-10 örneklem öncesi ve sonrası göz önüne alınmıştır.



Şekil 4.5 Korelasyon

4.1.3 Özellik Seçimi ve Farklı Bayramlardaki Kullanımları

XGBoost kullanılarak 1 hafta öncesinden alınan örneklemelerin 1 hafta sonrasına tahmin etmekte olan etkisini gözlelemek amacıyla hız verisi, hız verisi arasındaki saatlik fark, hız verisi arasındaki yüzdesel fark ve hız versinin MA metoduyla smoothing işlemi yapılmasıyla oluşan değişkenler model eğitiminde kullanılmıştır.



Şekil 4.6 Kısımlı Korelasyon

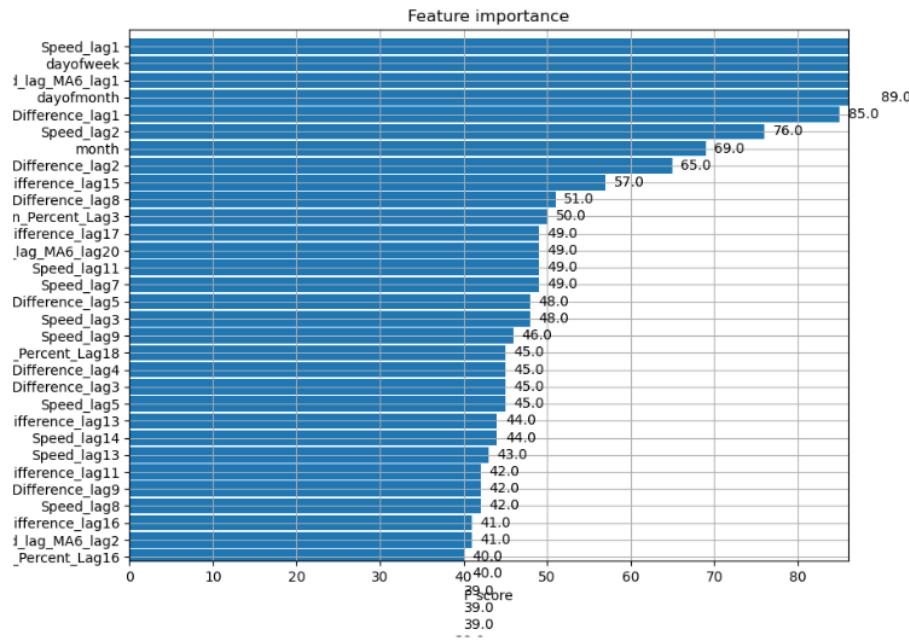
Model Bağdat caddesi segmenti 2018 verisi kullanılarak eğitilmiştir. Bu model tatil tahmini yapmak için değil veri üzerinde geçmiş verilerin ve zaman değişkenlerinin etkisini görmek için geliştirilmiştir.

Şekil??'te sonuçlar Bağdat Caddesi hariç beş farklı segment genelinde incelendiğinde 1 hafta önceki verinin, haftanın günlerinin ve ayın günlerinin tahmin sonucunda en fazla etkiye sahip olduğu gözlemlenmiştir.

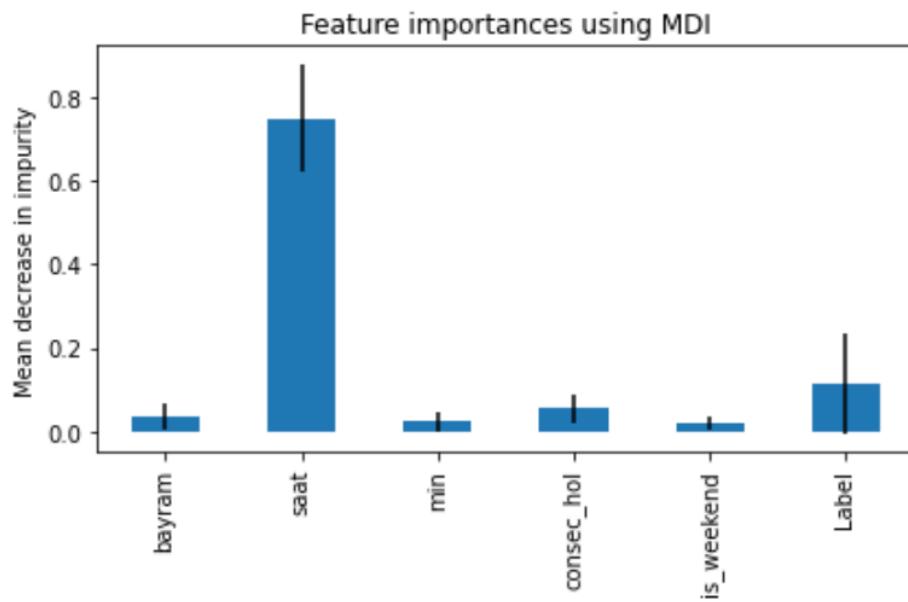
XGBoost ile yaptığımız 1 hafta öncesinin 1 hafta sonrasında yaptığı etkiyi incelemeye ek olarak problemimiz özelinde modellerimizin ve veri setlerimizin uygunluğunu incelemek amacıyla eğitim setini önceleyerek özelliklerin önemlerini çıkarttığımız 4.8 MDI modelini kullandık. Fakat MDI modeli eğitim setini baz aldığı için yaniltıcı sonuçlar verebiliyor. Bu nedenle aynı zamanda 4.9 permütasyon modelini kullanarak da özellik önemlerini çıkardık Şekil 4.8 ve Şekil 4.9'de bu modellerin nasıl farklı sonuçlar verdiğiğini görebilirsiniz. Eğitim setimizin doğasına göre bazı parametrelerin sonucu kötü etkilediği ve bazı parametlerin hiç işe yaramadığını görmek mümkün.

4.1.4 Test Seti Olarak Kullanılan Segmentler ve Özellikleri

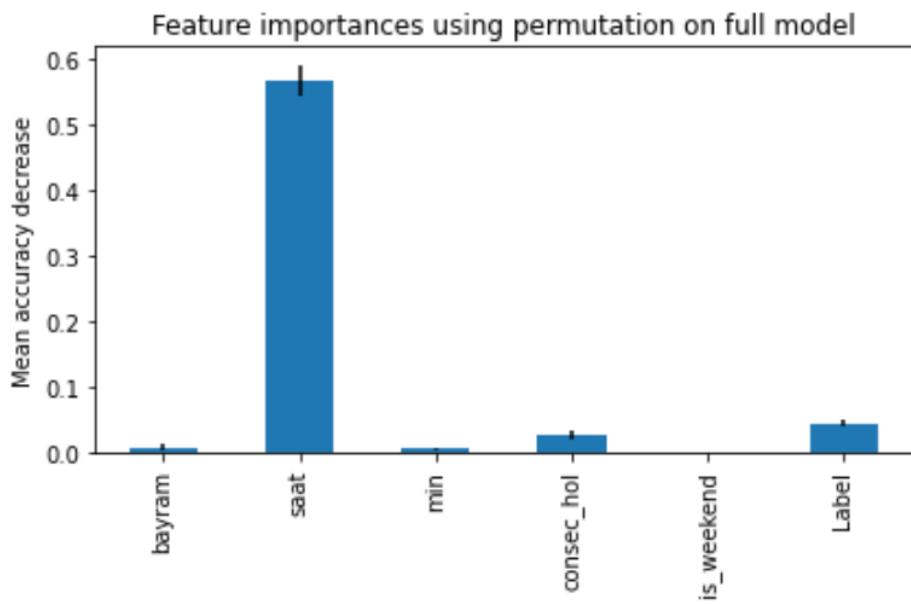
İstanbul genelinde 7500'e yakın segment ve bu segmentlerin gidiş ve geliş şeritleri olduğu için testleri bu 7500 segmentte yapmak yerine seçilen 106 segmentte yapmayı



Şekil 4.7 Değişken Önem Yatay Bar Grafiği



Şekil 4.8 MDI Özellik Önemi

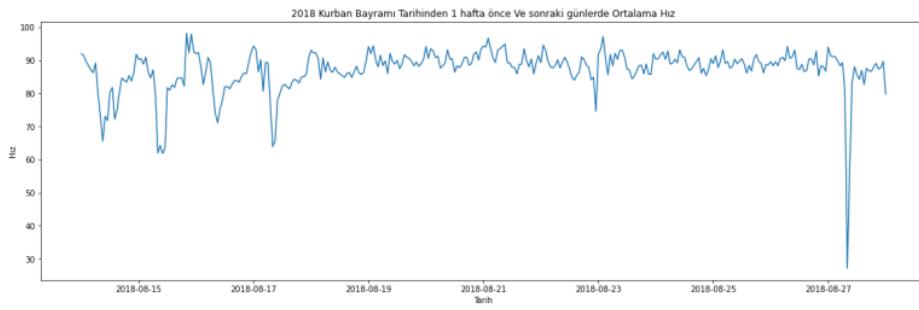


Şekil 4.9 Permüstasyon Özellik Önemi

tercih ettiğimiz. Bu tamamen zaman ve işlem gücü kısıtı nedeniyle yapılmasına rağmen bu 106 segmentin İstanbul'daki trafiğin genel karakteristğini yansıttığını düşünüyoruz. İleride bu hipotezi kanıtlamak için 7500 segment ile 106 segmentin sonuçlarını karşılaştıracağız.

4.1.4.1 Ana Arterler

İstanbul'un genel yol yapısına bakıldığında Avrasya Yolu, E5 gibi çok şeritli ana yollar bulunduğu görüyorum. Bu yolların 4.11 hız grafiklerinin bakıldığından yerel kaynaklı olaylardan kaza, etkinlik vb. daha az etkilendiğini görüyoruz. Çünkü bu yollar anlık değişimleri şerit çöküğünden kaynaklı daha rahat kaldırabiliyor. Aynı zamanda tatil günlerinde de bu yolların şehir içindeki yollara göre daha az etkilendiğini görebiliyoruz. Şekil 4.11 ve Şekil 4.13 farka bakıldığından ana arterlerin o yolun maksimum hız limitine ne kadar yakın ve düz bir grafik izlediğini diğer dar yolun ise ne kadar değişken bir grafik izlediğini görebiliyoruz. Fakat ana arterlerin en çok etkilendiği dış etmenin yol çalışmaları olduğunu görüyoruz. Kadıköy E-5 no'lu segmente bakıldığından bir ana yol bulunuyor. Bu segmentte yaptığımız tahminler ortalama hatanın çok üstünde bir hata oranını verdiği için 2018 ve 2019 hız verilerini bu segmente özel olarak inceledik. Grafiklere bakıldığından Şekil 4.12 2019'da hızın ortalama olarak düşüğünü gördük. Bu düşüş Kurbagaldere yenileme nedeniyle yapılan altyapı çalışmasının yola etki etmesi sonucu olduğunu fark ettik. Yol çalışması



Şekil 4.10 E-5 Ana Arter

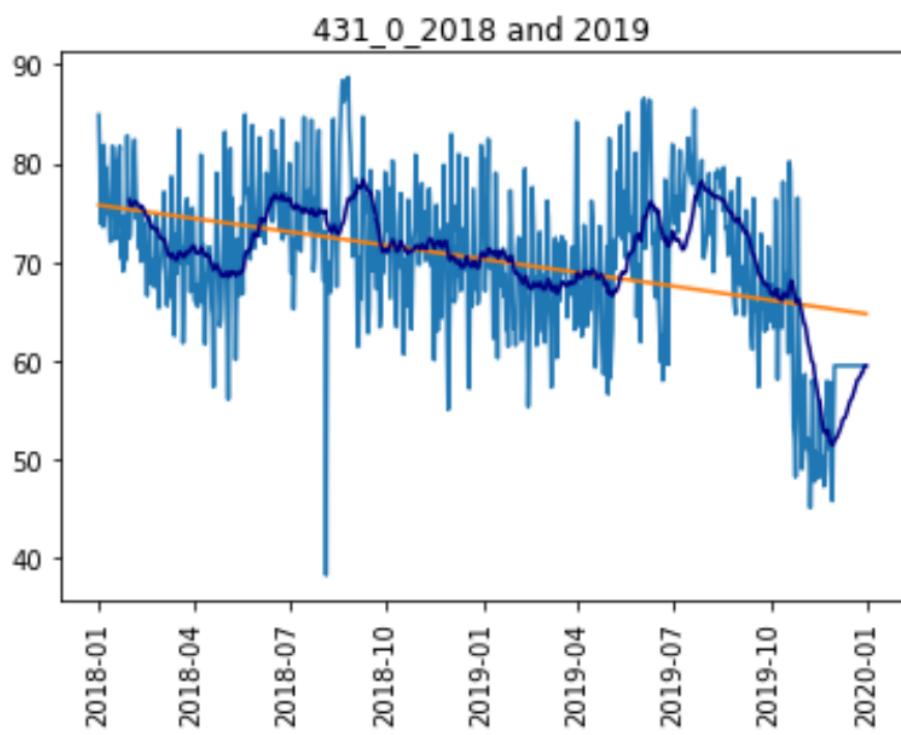


Şekil 4.11 Kurbağalıdere Ana Arter [6]

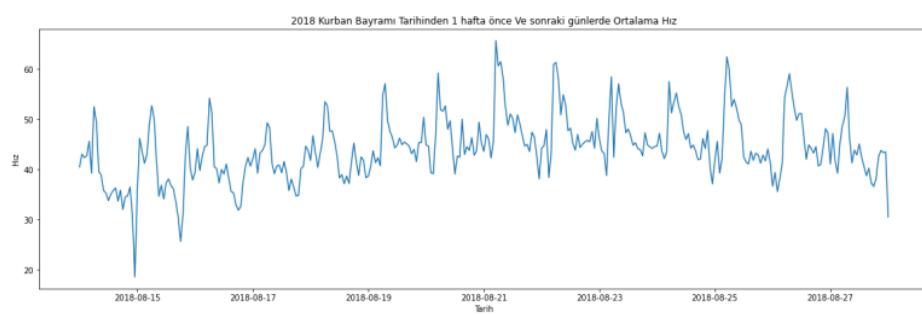
tam olarak 7 Ekim 2019 tarihinde başladığı için 29 EKİM için yapacağımız tahminde 2018 verisini kullanmamızı ve 7 ekimden öncesini kullanmamızı engelliyor. Çünkü trafik karakteristiği bu çalışmayla beraber tamamen değişiyor. Problemi çözmek için yol bakım ve kaza verilerini modelimize entegre etmeyi planlıyoruz.

4.1.4.2 Şehir İçi Yollar

Lokal faktörlerin asıl etkili olduğu yollar ilçelerde mahalleleri bağlayan yollar. Bu yolların tahliye yolu olmadığı için kazalar ve yerel etkinlikler anlık trafik sıkışıklıkları yaratabiliyor. Bu nedenle bu yolları ana arterlerden farklı şekilde inceledik.



Şekil 4.12 Kurbağalıdere Yol Çalışması Sonucu Etkilenen 431 Segmenti



Şekil 4.13 Şehir İçi Yol

4.2 Uzun Tatiller

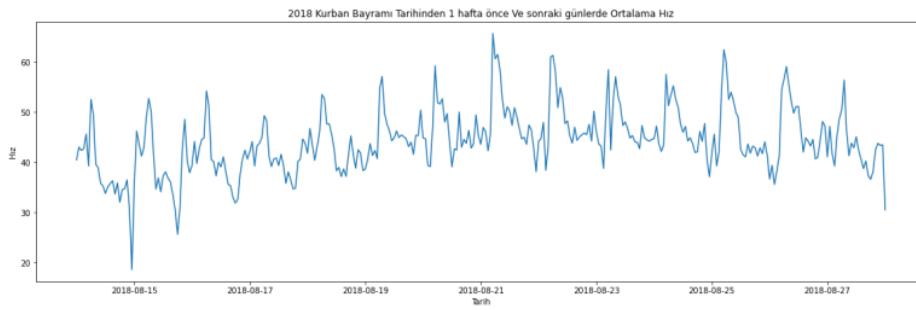
Resmi tatil günlerindeki değişimi incelemek için Ramazan ve Kurban Bayramı tatil günleri 2018 ve 2019 yıllarında ayrı ayrı görselleştirilmiş ve karşılaştırmalar yapılmıştır. Konum seçimlerinde İstanbul trafiğinin genel karakteristiğini yansitan 106 segment teste tabi tutulmuştur. Test sonuçları İstanbulun geneli için uygun bir sonuç vermektedir.

Gözlemlenen konumlar

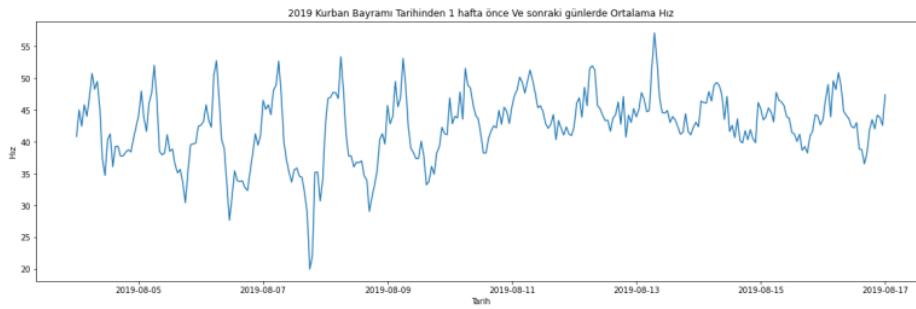
- Bağdat Caddesi
- Londra Asfaltı(Büyükçekmece)
- Haliç
- Fatih Sultan Mehmet Köprüsü
- Dolmabahçe Sarayı
- Üsküdar-Ümraniye TEM
- Kuzey Marmara Otoyolu
- E5

4.2.1 Kurban Bayramı

Kurban bayramlarında genelde insanlar uzun süreli şehir dışına çıktıkları için özellikle haftasonuyla birleştiğinde İstanbul trafiğinde bir yoğunluk yaşanıyor. Şekil 4.14 görüldüğü gibi 2018 yılında bayram haftasonuyla birleştiğinden 15 Ağustos gününde trafik hızında bir düşüş gerçekleşmiştir. Sonrasında İstanbul ilindeki insan sayısının azalması nedeniyle trafik hızı normal değerlerin üstüne çıkmıştır. Bir sonraki yılda ise arife günü 10 Ağustosa denk gelmiştir bu nedenle 7-9 Ağustos tarihleri arasında trafik yoğunlığında bir artış gözlemlenmiştir. 2018'de de olduğu gibi uzun süreli tatil İstanbul trafik yoğunlığında azalmaya sebep olmuştur. Kurban bayramı için eğitim günleri olarak 18 - 26 Ağustos 2018 arasını ve 3, 4 Ağustos 2019 günlerini seçtik. Cumartesi ve pazar günleri tatil günlerine benzer bir karakteristik gösterdiği için 2019 yılında test seti olarak kullandığımız 10 - 14 ağustosun bir hafta önceki haftasonunu da eğitim setine dahil ettik.



Şekil 4.14 Bağdat Caddesi 2018 Kurban Bayramı



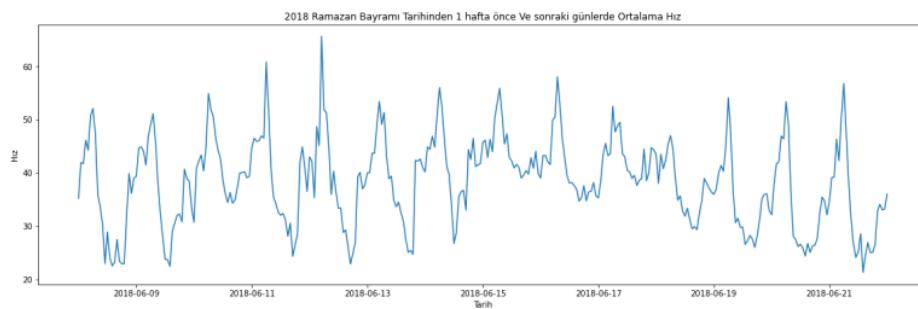
Şekil 4.15 Bağdat Caddesi 2019 Kurban Bayramı

4.2.2 Ramazan Bayramı

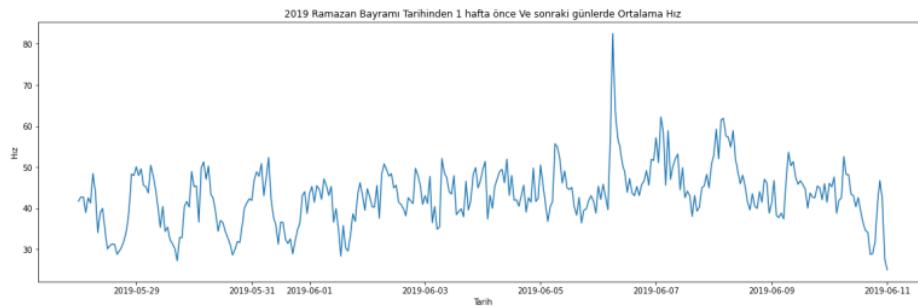
Ramazan Bayramı incelediğinde trafik yoğunlığında kayda değer bir artış gözlemlenmemiştir fakat 2018'de Şekil 4.16 bakıldığından tatilin olduğu 15-17 Haziran tarihleri arasında trafik yoğunlığında azalma gözlemlenmekdir. Ek olarak Şekil 4.18 2019 yılında tatil süresi haftasonuyla birleştiği için 1-9 Haziran tarihleri arasında trafik yoğunlığında bir miktar düşüş gözlemlenebilir. Ramazan bayramında genelde uzun tatiller olduğunda insanlar ilk günlerde şehiri terk etme eğilimi gösterirken, kurban bayramında insanların tatil gidişleri bayramın geneline yayıldığı gözlemlenmiştir bu nedenle Kurban bayramı grafiklerinde birden fazla düşüş ve yükseliş gözlemlenirken Ramazan bayramı grafiklerinde genelde bir düşüş ve yükseliş gözlemlenmiştir.

Ramazan bayramı elimizdeki 2018 ve 2019 verilerine bakıldığında 2018'de haftasonuna denk geldiği için 15-17 Haziran arasında 3 günlük bir tatilken 2019'da 1 - 9 Haziran arasında 9 günlük bir tatil oluyor. Eğitim setimizde 2018 verisini kullandığımızda %18 gibi kötü bir sonuç aldık. Bundan dolayı kurban bayramından farklı bir yaklaşım izlememiz gerekti. Kurban bayramı 2018 yılında 2019 yılındaki ramazan gibi uzun bir dönemi kapsadığı için kurban bayramını kullanarak tahmin etmeyi denedigimizde daha kötü sonuçlar alındığını fark ettik. Bunun bir nedeni de yılın gününü ve ayını da parametre olarak kullanmaktı. Hangi parametrenin

daha iyi sonuç verdiğine dair yaptığımız testlerde bundan daha detaylı bahsedeceğiz. Ancak şimdilik bu parametreleri atarak bu sorunu çözdüğümüzden bahsetmekte yarar var. Kurban bayramı ve ramazan bayramı arasında bayram günlerinin korelasyonuna baktığımızda yüksek bir korelasyon olduğu için kurban bayramının bayram günlerini eğitim setine dahil ettik fakat arife ve ağustos ayının diğer günlerini hazırlan ve ağustos aylarındaki karakteristik farkından dolayı eklemeyi uygun görmedik. Sonuç olarak eğitim setinde bir hafta öncesinin 2019 haftasonu, ramazan bayramı ve kurban bayramı günleri yer aldı. Bu eğitim setiyle %13'e kadar hata oranını indirebildik. Hata oranlarından ?? bölümünde daha detaylı bahsedeceğiz.



Şekil 4.16 Bağdat Caddesi 2018 Ramazan Bayramı



Şekil 4.17 2019

Şekil 4.18 Bağdat Caddesi 2019 Ramazan Bayramı

4.3 Kısa Tatiller

Günlük resmi tatiller uzun süreli tatillere nazaran 1-2 günlük tatiller olduğu için trafik karakteristiğinde uzun süreli tatiller kadar etki göstermez. Burada yola ve güne özel karakteristiklerin etkisinin daha yüksek olduğu görülebilir. Resmi tatillerin ülkelere özgü ve genelgeçer olmak üzere farklı grplara bölünebilir. Genelgeçer olarak bakıldığından günlük tatillerin haftasonuna denk gelmesi ve haftaiçine denk

Tablo 4.1 Tatil Günlerinin Türlerine Özel Tahmin Setleri

Tatil (2019)	Tatil Türü (2019)	Tatil Türü (2018)	Tahmin Seti
23 Nisan Ulusal Egemenlik ve Çocuk Bayramı	Kısa Tatil - Hafta İç-Tek Günlük	Kısa Tatil - Hafta İç-Üç Günlük	1 Mayıs 2018 + 30 Ağustos 2018 + 28-29 Ekim 2018 + 21 Nisan 2019 (Bir önceki pazar)
1 Mayıs Emek ve Dayanışma Günü	Kısa Tatil - Hafta İç-Tek Günlük	Kısa Tatil - Hafta İç-Tek Günlük	30 Nisan-1 Mayıs 2018+ 18-19 Mayıs 2018+ 29-30 Ağustos 2018 + 28-29 Ekim 2018 + Son 1 hafta 2019
19 Mayıs Atatürk'ü Anma, Gençlik ve Spor Bayramı	Kısa Tatil - Hafta Sonu	Kısa Tatil - Hafta Sonu	12 Mayıs 2019 (Geçen hafta aynı gün)
15 Temmuz Demokrasi ve Millî Birlik Günü	Kısa Tatil - Farklı Karakteristik	Kısa Tatil - Farklı Karakteristik	2018 Milli(2) + 2019 15 Temmuza kadarki milli bayramlar + 8 Temmuz 2019 (1 Hafta önce aynı gün) + Bir önceki Cumartesi-Pazar
30 Ağustos Zafer Bayramı	Kısa Tatil - Hafta İç-Üç Günlük	Kısa Tatil - Hafta İç-Tek Günlük	30 Nisan-1 Mayıs 2018+ 18-19 Mayıs 2018+ 29-30 Ağustos 2018 + 28-29 Ekim 2018 + 23 Nisan 2019 + 1 Mayıs 2019 + Son 1 hafta 2019
30 Ağustos Zafer Bayramı	Kısa Tatil - Hafta İç-Tek Günlük	Kısa Tatil - Hafta İç-Üç Günlük	22 Ekim-4 Kasım 2018 + 23 Nisan 2019 + Son 1 hafta 2019
Ramazan Bayramı	Uzun Tatil - 9 Günlük	Uzun Tatil - 3 Günlük	2 Haziran 2018 + 1 Haziran 2019 (Önceki cumartesi) + 14-15-16-17 Haziran 2018 (Ramazan 2018) + Kurban 2018
Kurban Bayramı	Uzun Tatil - 4 Günlük	Uzun Tatil - 9 Günlük	2019 bayramdan önceki haftasonu + 18 Ağustos-26 Ağustos 2018

gelmesi arasında bariz bir fark gözleniyor. Haftasonu ve haftaiçi tatillerinin farklı şekilde incelenmesi gerekiyor. Şekil ?? bakıldığından bu gruplara denk gelen günleri görebiliriz. Ülkelere göre değişen durumlar ise insanların tatillere gösterdiği ilgi çerçevesinde gerçekleşiyor. Örnek vermek gerekirse 29 Ekim günü her sene belirli bir yoğunluk gözlenebilir. Ayrıca resmi tatillerin denk olduğu mevsimler de bu durumda etkili oluyor. Bu nedenle farklı karakteristik gösteren bu tatilleri ayrı bir grupta incelemeyi uygun gördük.

4.3.0.1 Haftaiçine Denk Gelen Tatiller

2018 yılında 23 Nisan pazartesi, 1 Mayıs salı, 30 Ağustos perşembe, 29 Ekim pazartesi gününe denk geliyor. Bir sonraki yılda tek günlük bu tatiller bir sonraki yıl bir gün sonrasına denk geliyor. Eğer veri setimizde bir önceki yılda cuma gününe denk gelen bir tatil olsaydı tahminlerimize göre bu günü eğitim setinde kullandığımızda kötü bir sonuç elde edecektik. Çünkü haftaiçi ve haftasonuna denk gelen tatil günlerinde farklılık gözleniyor. Buna ek olarak cuma ve pazartesine denk gelen günlük tatiller haftasonuyla birleştiği için insanların davranışlarına daha fazla etki ediyor. Yukarıda bahsedildiği gibi 23 Nisan tatili pazartesi gününde. İnsanlar 3 günlük tatil nedeniyle İstanbul'a yakın şehirlere veya aile ziyaretlerini 3 günlük tatil nedeniyle tercih edebilir. İncelediğimiz sonuçlarda da bunu gördük. İki yıl arasındaki bu fark 23 Nisan tahminlerinin diğer haftaiçine denk gelen tatillerden daha kötü olması sonucunu doğurdu. Çünkü haftaiçine denk gelen tatillerde aynı yaklaşımı izledik. Bu grubu 3 günlük ve tek günlük tatiller olarak incelesek de hata bazında 3 günlük tatilleri bu gruptaki tek günlük tatiller seviyesine indirmeyi şuan da gerçekleştiremedik. Şekil ?? bu durumun nedeni şekil ?? karşılaştırıldığında açıkça görülebilir.

4.3.0.1.1 Tek Günlük Tatiller Verisetimizde 2018 yılında yukarıda bahsettiğimiz günlerden 1 Mayıs ve 30 Ağustos günleri bu kategoriye giriyor. Bir sonraki yılda ise sadece 30 Ağustos Zafer bayramı harici diğer bayramlar bu kategoride. İki yıl arasında 1 Mayısın belirlediğimiz grupta göre karakteristik değiştirmedigini görüyoruz bu nedenle 1 Mayıs için kullandığımız eğitim setinde ?? tabloda görüldüğü gibi tek günlük bayramları ve bir önceki günü, 1 Mayıs 2019'dan önceki son bir haftayı ve 1 Mayıs 2018 haftasının haftasonunu kullandık. Zafer bayramında ise 1 Mayıs eğitim setindeki aynı mantığı uygulamakla beraber 2019 senesinde 30 Ağustos'tan önce gelen tek günlük bayramlarında eğitim setine ekledik. Aynı seneden önceki bayramları ekleme yaklaşımını genel olarak bütün bayramlar için yaptık. Daha detaylı sonuçlar ?? tabloda görülebilir. 30 Ağustos tahminimizdeki hata oranlarının ortalaması 1 Mayistan yüzde 1 oranında daha yüksek. Bunun nedeni yukarıda bahsettiğimiz gibi 30 Ağustosun 2019 yılında karakteristik değiştirmesi ve 3 günlük tatil kategorisine girmesi.

4.3.0.1.2 Üç Günlük Tatiller Cumhuriyet Bayramı ve 23 Nisan Ulusal Egemenlik Çocuk Bayramı 2018 yılında bu kategoriye giriyor ve 2019 yılında karakteristik değiştirerek tek günlük tatil kategorisine dahil oluyor. Ancak hata oranlarına baktığımızda 23 Nisanın ortalama olarak yaklaşık yüzde 2 oranında daha fazla hata verdienenğini görüyoruz. Aslında her ne kadar iki yılda da aynı karakteristiği izleyen tatiller olarak gözükseler de 22 Nisan 2019 pazartesi günü tatil değilken 28 Ekim

pazartesi günü yarım gün tatil ediliyor. Bu nedenle 29 ekimin 3 günlük tatil durumunu bir yandan devam ettirdiğini söyleyebiliriz.

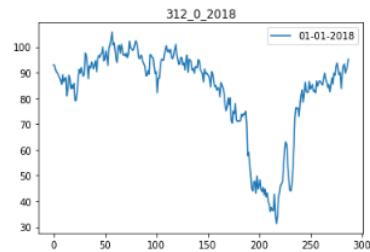
4.3.0.2 Haftasonuna Denk Gelen Tatiller

Gençlik ve Spor Bayramı bu kategoriye giriyor. Veri setinde bulunan iki yılda da haftasonuna denk geldiği için tek günlük tatillere uygulanan yaklaşım burada %15 gibi kötü bir sonuç veriyor. Önemli olan şey insanların haftasonuna denk gelen tatillerde nasıl bir davranışta bulundukları. Herhangi bir haftasonundan farkı olmayan bu günlerde haftaiçinde oluşan sabah 8 akşam 5 trafiği görülmüyor. Ancak trafikte her ne kadar bir azalma olsada tek günlük tatillerde yaşanan belirli bölgelerdeki yoğunlukta gözlenmiyor. Bunun sonucunda haftasonuna denk gelen tatilleri eğitirken aynı yılın bir önceki haftasonu verisini kullanmanın daha isabetli olduğu sonucuna vardık. Bu yaklaşımı izleyerek hata oranını yüzde 2 oranında düşürerek %13 seviyesine indirebildik.

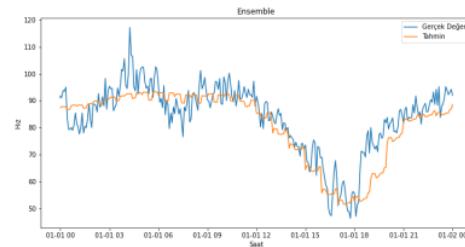
4.3.0.3 Farklı Karakteristik İzleyen Tatiller

Yılbaşı ve 15 Temmuz Bayramı diğer kategorilere girmeyen bir davranış izliyor. Yılbaşı için bunun nedeni 31 Aralık gecesi artan yoğunluk ve 1 Ocak gününün bundan etkilenmesi. Yılbaşı günü incelendiğinde Şekil 4.19 ve 4.20 grafiklerinde görüleceği üzere gece saatlerinde aşırı trafik yoğunluğu gözlemlenmiştir. Demokrasi bayramına baktığımızda ise 2018 yılında haftasonuna 2019 yılında da pazartesi gününe denk gelerek 3 günlük bir tatil oluşturuyor. Bakıldığına 2018 yılındaki 23 Nisan ve 29 Ekim tatilleriyle aynı kategoriye girse de aynı karakteristiği göstermediğini belirtmek isteriz.

15 temmuz eğitim setinde 2018 tek günlük bayramları ve önceki günü, bir hafta önce aynı gün, aynı yılın bir önceki haftasonunu ve aynı yılda o güne kadar olan bayramları kullandık. Tek günlük tatillerde izlediğimiz 2018 verisi kullanma yaklaşımında %15 olan hata oranı bu yeni yaklaşımın sayesinde %13 seviyesine kadar geriledi.



Şekil 4.19 1 Ocak 2018



Şekil 4.20 1 Ocak 2019

Şekil 4.21 2018 ve 2019 Yıllarında 1 Ocak Hız Grafikleri

5

Sistem Tasarımı

5.1 Algoritmalar ve Model Eğitimi

Çeşitli makaleler ve modeller incelendiğinde karar ağaçlarının başarısı görülmüştür. Özellikle ani değişim gösteren verilerde XGBoost modeli klasik makine öğrenimi modellerine göre başarı sağlar. Modele verilecek parametler parametre istatistiksel testleri kullanılarak çıkarılmıştır. Veri üzerindeki mevsimsel değişimleri algılamak için parametrelerin sin ve cos değerleri hesaplanmıştır ve parametre olarak modele verilmiştir. Ayriyetten modeller 2018 ve 2019 resmi tatil günleri veriden ayrılarak bu ayırtırılan veri üzerinde eğitilmiştir gözlemlerimize dayanarak modelin tatil günlerinde normal günlere oranla çok daha iyi performans verdiği söylenebilir.

5.2 Karar AĞacı Modelleri

Karar ağaç modelleri, birçok makine öğrenimi probleminde kullanılan ve yüksek başarı oranlarına sahip bir sınıflandırma ve regresyon yöntemidir. Karar ağaçları, verileri işlemek ve sonuçları tahmin etmek için bir ağaç yapısı kullanır.

1. Sınıflandırma Ağaçları: Sınıflandırma ağaçları, verileri sınıflandırmak için kullanılır. Girdi verileri, ağaç yapısının kök düğümüne girer ve her düğümde, bir karar kuralı ile dallara ayrılır. Bu karar kuralı, bir özellik değeri ile bir eşik değeri arasındaki bir karşılaştırmaya dayanabilir. Bu işlem yaprak düğümüne ulaşınca kadar devam eder. Sonuçta, girdi verilerinin sınıflandırılması, yaprak düğümlerinde tanımlanmış olan sınıflara göre belirlenir.
2. Regresyon Ağaçları: Regresyon ağaçları, verileri tahmin etmek için kullanılır. Girdi verileri, kök düğümünden başlayarak ağaç yapısında ilerler. Her düğümde, bir özellik değeri ile bir eşik değeri arasındaki bir karşılaştırmaya dayanan bir karar kuralı kullanılır. Bu işlem, yaprak düğümüne ulaşınca kadar devam eder. Sonuçta, girdi verilerinin tahmini, yaprak düğümlerinde tanımlanmış olan sayısal değerlere göre belirlenir.

Karar ağaç modelleri, ayrik veya sürekli özellikleri olan verileri işleyebilirler. Ayrıca, çoklu sınıflandırma veya çoklu regresyon problemlerine de uygulanabilirler. Karar ağaç modelleri, yüksek doğruluk, hızlı tahmin ve kolay yorumlanabilirlik gibi avantajlara sahiptirler. Projemiz karar ağaç modellerinde regresyon modelleri kısmi ile ilgili olduğu için o kısmı daha detaylı açıklayabiliyoruz; Regresyon ağaçları, veri setindeki özellikleri kullanarak bağımlı değişkenin (hedef değişken) sürekli bir değerini tahmin etmek için kullanılan bir makine öğrenmesi algoritmasıdır. Bu modeller, karar ağaç algoritmasının bir türüdür ve ağaç yapısı, veri setindeki özelliklerin belirli değerleri tarafından belirli bir hedef değişken değeriley sonuçlanan karar düğümleri tarafından temsil edilir. Bu karar düğümleri, veri setindeki özelliklerin değerlerine göre sıralanır ve hangi özelliklerin hedef değişkeni üzerinde en büyük etkiye sahip olduğunu belirlemeye yardımcı olur. Regresyon ağaçları, diğer makine öğrenimi modellerine göre bazı avantajlara sahiptir. Özellikle, bu modeller, veri setindeki özellikler arasındaki etkileşimleri ve doğrusal olmayan ilişkileri belirlemekte iyidirler. Ayrıca, bu modeller, çok sayıda özellik içeren veri setlerinde çalışabilir ve genellikle sonuçları yorumlamak kolaydır. Regresyon ağaçları aynı zamanda birçok farklı varyasyona sahiptir. Bazı örnekler, CART (Classification and Regression Trees), Random Forest ve Gradient Boosting Trees, XGBRegressor gibi algoritmaları içerir. Projemizde model olarak regresyon ağaçları varyasyonlarından RandomForestRegressor, Gradient Boosted Trees ve XGBRegressor kullanılmıştır. Random ForestRegressor, Gradient Boosted Trees (GBM) ve XGBRegressor, üç farklı makine öğrenimi modelidir. Tüm ağaç tabanlı modellerdir ve çeşitli özellikleri ve kullanımları vardır. Random ForestRegressor, birden fazla karar ağacını eğiterek çalışır. Bu ağaçlar, birbirinden bağımsız olarak oluşturulur ve ardından tahminleri bir araya getirerek son bir tahmin yapılır. Random ForestRegressor, yüksek varyanslı modellerle başa çıkmak için kullanılır.

GBM, Gradient Boosted Trees, tek bir karar ağaç üzerinde çalışır ve ardışık ağaçlar ekler. Bu nedenle, Random ForestRegressor'dan farklı olarak, GBM aşamalı bir öğrenme modelidir. GBM, düşük biaslı modellerle başa çıkmak için kullanılır. XGBRegressor, GBM'ye benzer şekilde çalışır, ancak öğrenme sürecinde farklı bir teknik kullanır. XGBRegressor, GBM'nin aksine, eğitim örneğine ağırlıklar ekleyerek öğrenir. Bu, modelin daha az hata yapmasına ve daha az ağırlığa sahip örnekleri göz ardi etmesine neden olur. XGBRegressor, GBM'ye kıyasla daha hızlı ve daha doğru sonuçlar verir. Tüm bu modeller ağaç tabanlı olmakla birlikte, eğitim yöntemleri ve özellikleri farklıdır. Seçim, veri setine ve probleme bağlıdır. Random ForestRegressor yüksek varyanslı verilerle çalışmak için uygunken, GBM düşük biaslı verilerle çalışmak için uygun olabilir ve XGBRegressor daha hızlı ve doğru sonuçlar verir.

5.3 Derin Öğrenme Modelleri

Derin öğrenme, yapay sinir ağlarının daha karmaşık yapılarını kullanarak, büyük veri setlerindeki kompleks yapıları algılama, sınıflandırma, desen tanıma, tahminleme gibi problemleri çözmek için kullanılan bir makine öğrenmesi yöntemidir. Derin öğrenme, sinir ağı mimarileri aracılığıyla, veri setindeki özelliklerini otomatik olarak öğrenerek, daha önceki örneklerden öğrendiği bilgiyi yeni veriler üzerinde uygulayarak sonuç üretir. Derin öğrenme yöntemleri, resim ve video işleme, doğal dil işleme, otomatik araba sürüsü, ses tanıma ve makine çevirisi gibi alanlarda başarılı sonuçlar vermektedir. Derin öğrenme modelleri arasında Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Autoencoder, ve Generative Adversarial Networks (GAN) , Deep Neural Networks (DNNs) gibi popüler mimariler yer almaktadır. DNN'ler, diğer derin öğrenme modelleri arasında en genel ve temel yapılardan biridir. Diğer modeller, DNN'lerin farklı varyasyonları olarak kabul edilebilir. DNN'ler, girdi katmanı, birden fazla gizli katman ve çıkış katmanı gibi temel yapıya sahip sinir ağı modelleridir. Bu katmanlar, birbirleriyle bağlı olan çok sayıda nöron içerir ve birbirlerine ağırlıklı bağlarla bağlanır. Bu ağırlıklar, modelin öğrenme sürecinde belirlenir. DNN'lerin faydaları, özellikle büyük veri kümeleri üzerinde çalışırken etkileyicidir. Derin sinir ağları, daha önce insanlar tarafından tanımlanmamış özellikleri algılayabilen ve bu özellikleri yorumlayabilme yetisine sahiptir. Derin öğrenme modelleri genel olarak elimizdeki verinin az olmasından kaynaklı olarak klasik makine öğrenimi modellerinden daha kötü bir sonuç vermiştir. Klasik makine öğreniminde verileri özel olarak günlere göre seçme yaklaşımını izlerken derin öğrenme modellerinin çok veriyi daha iyi işleyebilme yeteneğini göz önünde bulundurarak eğitim veri setini daha geniş tutma yaklaşımını izledik. Sonuç olarak klasik ve derin öğrenme modellerinin farklı şekilde eğitilmeleri gerektiğini düşünüyoruz. Bu bilgiler doğrultusunda projemizde derin öğrenme modeli olarak DNN'i tercih ettik.

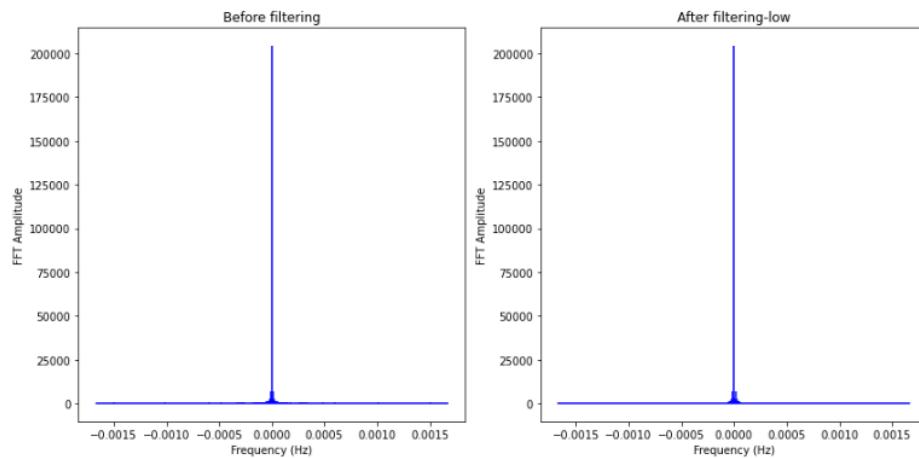
6

Uygulama

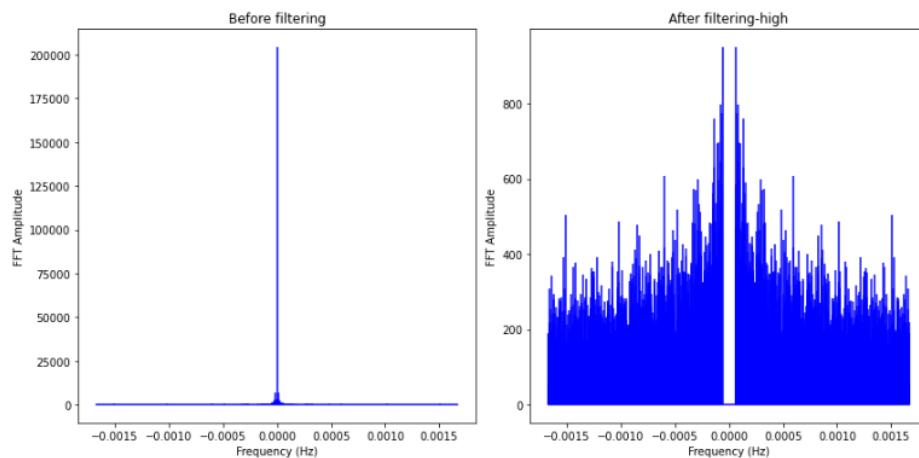
6.1 Fast Fourier Transform Ön İşlemesi

İncelenen DFT-SVR [1] görüldüğü gibi SVR modelini kullanırken eğitim verisine DFT ön işleminin uygulanması eğitim veri setindeki gürültüyü ortadan kaldırıyor. Makalede görüldüğü üzere bu işlem SVR modelindeki % 13 oranındaki MAPE hata değerini veriyor, veri setine DFT uygulandığında ise bu oran % 10 seviyesine kadar düşüyor. Bu iyileştirmeden yola çıkarak ensemble modelini DFT uygulamış veri setiyle eğitmenin iyi sonuçlar vereceğini düşündük. Ancak DFT modelini kullanmak yerine DFT'nin daha hızlı bir versiyonu olan FFT ön işlem metodu bu ensemble modeline uygulandı. Ayırık zamanlı filtrelemede düşük bant aralıklı, yüksek bant aralıklı ve orta bant aralıklı filtreleme yöntemleri literatüre bakıldığından görülebiliyor. Trafik verisi incelendiğinde düşük bant aralıklı filtrelemenin belirli bir kesme frekansı seçilerek (Şekil 6.1) verideki gürültüyü çıkarabiliyoruz. Belirli bir bayram aralığını incelediğimiz için hız verisinin belirgin bir trend eğrisi bulunmuyor. Bu nedenle yaptığımız bu düşük bantlı filtreleme yüksek frekanslı veriyi asıl veriden ayırmamızı sağlıyor.

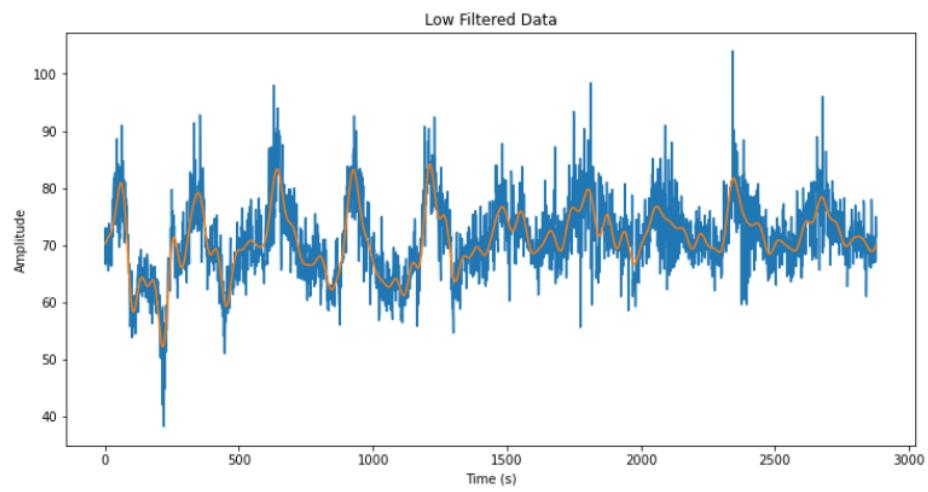
Bu yüksek frekanslı veriye bakıldığından 6.4 istenilen veri setinin ortalamasının 0'a olabildiğince yakın olması gereklidir. Eğer yaptığımız filtreleme sonucunda yüksek frekanslı veride 0'dan farklı bir ortalama bulunursa bu ortalama anlamlı veriyi filtreleme sonucunda kaybetmiş olduğumuzu gösteriyor. Kaybı en aza indirmek için yüksek frekanslı veriyi de ayrı bir SVR modelinde eğiterek ana ensemble modelimize eklemeyi planlıyoruz. Düşük frekanslı veriye baktığımızda Şekil 6.3 verinin ana hatlarını yakaladığını görebiliyoruz.



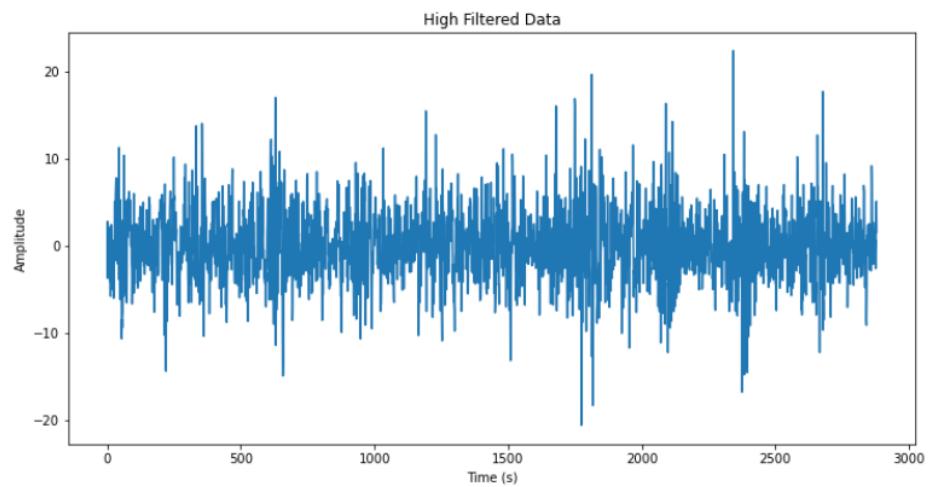
Şekil 6.1 Düşük Bant Aralıklı FFT



Şekil 6.2 Yüksek Bant Aralıklı FFT



Şekil 6.3 Düşük Frekans Filtreleme Uygulanmış Veri ve Gerçek Veri



Şekil 6.4 Yüksek Frekanslı Veri

6.2 Kayıp Değerleri Doldurma

İstanbulun farklı lokasyonlarından alınan hız verileri tablolarında eksik değerler bulunmaktadır. Tabloda '-1' olarak ifade edilen bu eksik değerler tahminleme sürecinde hata oranının artmasına sebep olmaktadır. Eğitim setinde olan eksik değerler modelin yanlış öğrenme sürecine girmesine ve tahminleme sırasında gerçek değerlerden uzak sonuç üretmesine neden olmaktadır. Test setinde olan eksik değerler ise modelimizin ne kadar iyi çalıştığını anlamamıza engel olmaktadır. Tahminleme sonucunda oluşan değerleri test setindeki veriler ile karşılaştırıldığımızda hata oranı yüksek çıkmaktadır. Bu sorunların giderilmesi adına test setinden eksik değerlerin çıkarılması, eğitim setinde ise eksik değerlerin doldurulması gerektiği düşünülmüştür. Eğitim setindeki eksik değerlerin doldurulması için ön işleme kısmına fonksiyon yazılmıştır. Fonksiyonunun yazılma süreci ve son adımda çalışması şu şekildedir;

1-Öncelikle train veri setinde -1 ile karşılaşıldığında son 5 günün aynı saatindeki hızların ortalamasını alarak -1 olan kısma set etmemi denedik. Fakat son 5 gün içerisinde veya tamamında yine aynı saatte -1 değer olma durumunda da hata ile karşılaşıldı

2-Önceki adımda alınan (-1) değer olmayan son 5 günü bulana kadar while ile geriye giderek değer bulma ve ortalama alarak set etme işlemi yaptık. Fakat (-1) değerinin olduğu gün yılın ilk 5 günü olması veya geriye gitme işlemi yaparken yılın ilk günüğe giderek daha geriye gidememe sorunu iel karşılaştık

3-Son adımda çözüm olarak (-1) değeri bulunan zaman yılın ilk 5 günü ise ileri doğru giderek (-1) olmayan 5 günün ortalamasını aldık, yılın ilk 5 günü değil ise geriye giderek (-1) olmayan 5 gün bulunamadığında ileri doğru giderek (-1) olmayan 5 günün ortalamasını aldık

6.3 Random Forest Regressor

RandomForestRegressor, bir makine öğrenimi algoritması olan Random Forest'un bir varyasyonudur ve regresyon problemlerinde kullanılır. Bu algoritma, birden fazla karar ağacını (decision tree) eğiterek tahmin yapar. Random Forest algoritması, öncelikle birbirinden bağımsız ve rastgele özellikler kümesi oluşturarak her bir karar ağacını eğitir. Bu rastgele özellikler kümesi, tüm özellikler kümesinden örneklem alınarak oluşturulur. Daha sonra, her bir karar ağacı, rastgele seçilen bir alt kümesi ile örneklem alınmış veri kümesini kullanarak eğitilir. Bu sayede, farklı alt örneklem ve özellik kümeleri kullanarak eğitilmiş birden fazla karar ağacı elde edilir. Random Forest'un tahmin yapmak için kullandığı yöntem, tüm karar ağaçlarının

ürettiği tahminleri bir araya getirerek (ensemble learning) ortalama bir tahmin değeri oluşturmaktadır. Bu şekilde, tek bir ağacın hatalarından etkilenmeyi engelleyerek daha doğru bir tahmin elde edilir. RandomForestRegressor, regresyon problemlerinde kullanılan bir varyasyondur. Bu algoritma, aynı şekilde birden fazla karar ağacını kullanarak tahmin yapar ve tüm ağaçların tahminlerini bir araya getirerek ortalama bir tahmin değeri oluşturur. Ancak bu tahminler artık sınıflar değil, gerçek sayısal değerlerdir.

Biz Random Forest ana model olarak kullandık. Bu kararı almamızdaki asıl neden random forestin parametre değişimine olan toleransı ve verdiği hata değerlerinin güvenilirliği oldu. Asıl amacımız tatil günlerini etkileyen günleri belirlemek olduğu için random forest temel model olarak iyi bir sonuç verdi.

Randem Forest için en uygun parametre seçimi yapılarak mevcut model içim daha iyi sonuçların elde edilmesi hedeflenmiş, bu doğrultuda 106 segment için sırası ile Rastgele Arama Çapraz Doğrulama yöntemi kullanılmış, belirtilen hiperparametre dağılımlarından rastgele seçilen 10 farklı kombinasyonu değerlendirilmiştir. Bu, 10 farklı hiperparametre setini deneyerek en iyi yapılandırma bulmunya çalışılmıştır. Belirtilen hipermateler aşağıdaki gibidir ;

```
16 n_estimators: [100, 500, 1000]
max_features: ['sqrt', 'log2', 1.0]
max_depth: randint(2, 10)
min_samples_split: randint(2, 20)
min_samples_leaf: randint(1, 20)
bootstrap: [True, False]
```

Sonuç olarak en iyi parametre değerleri aşağıdaki şekilde belirlenmiştir;

```
17 n_estimators: [500]
max_features: [1.0]
max_depth: [19]
min_samples_split: [11]
min_samples_leaf: [19]
bootstrap: [True]
```

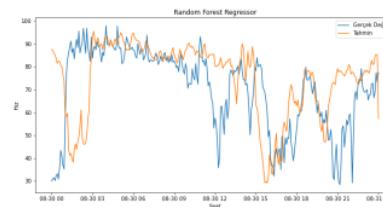
6.3.1 Grafikler

6.3.1.1 Kısa Bayramlar

Kısa bayram olarak sınıflandırılan 23 Nisan, 1 Mayıs, 19 Mayıs, 15 Temmuz, 30 Ağustos, 29 Ekim tarihlerindeki tatillerin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;

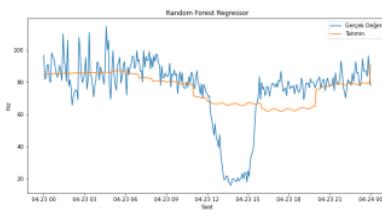


Şekil 6.5 1 Mayıs 2019 Random Forest

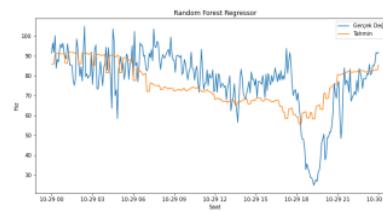


Şekil 6.6 30 Ağustos 2019 Random Forest

Şekil 6.7 Avcılar 414 Nolu Segment Haftaiçine Denk Gelen Bayramlar

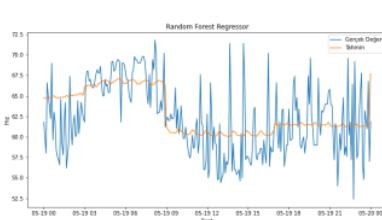


Şekil 6.8 23 Nisan 2019 Random Forest

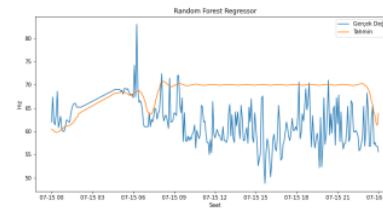


Şekil 6.9 29 Ekim 2019 Random Forest

Şekil 6.10 Ataşehir 316 Nolu Segment Haftasonu İle Birleşen Bayramlar



Şekil 6.11 19 Mayıs 2019 Random Forest

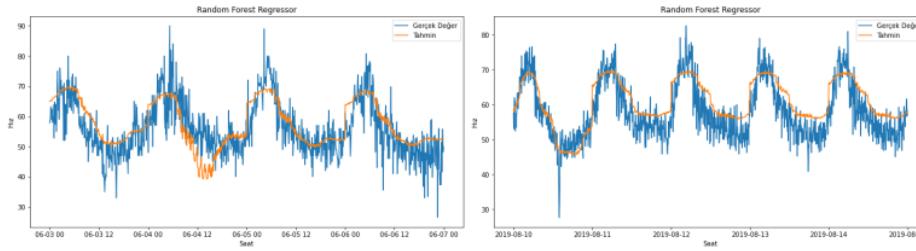


Şekil 6.12 15 Temmuz 2019 Random Forest

Şekil 6.13 Kadıköy 2102 Nolu Segment Farklı Karakteristik İzleyen Bayramlar

6.3.1.2 Uzun Bayramlar

Uzun bayram olarak sınıflandırılan ramazan bayramı ve kurban bayramı tatillerinin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;



Şekil 6.14 Ramazan Bayramı 2019
Random Forest

Şekil 6.15 Kurban Bayramı 2019 Random
Forest

Şekil 6.16 Kadıköy 2102 nolu Uzun Bayramlar

6.4 XGBoost

XGBoost (eXtreme Gradient Boosting), açık kaynak kodlu bir makine öğrenimi kütüphanesidir ve Gradient Boosting yöntemini kullanarak yüksek performanslı tahmin modelleri oluşturmayı hedefler. XGBoost, düzenlileştirme (regularization), paralelleştirme ve hızlı ölçeklenebilirlik özellikleri sayesinde, özellikle yapısal olarak büyük veri kümeleri üzerinde yüksek performanslı modelleme yapmak için tercih edilir.

En uygun parametre değerlerini belirlemek için genellikle çapraz doğrulama (cross-validation) kullanılır. Bu işlem, veri kümesini birden fazla parçaaya ayırarak (genellikle 5 veya 10 parçaaya bölünür), her parçanın sırayla test seti olarak kullanılması ve diğer parçaların eğitim seti olarak kullanılması yoluyla yapılır. Bu yöntem, modelin performansını değerlendirmek ve aynı zamanda en uygun parametre değerlerini belirlemek için kullanılır. Bu amaçla, GridSearchCV veya RandomizedSearchCV gibi bir hiperparametre ayarlama kütüphanesi kullanılabilir. Örneğin, aşağıdaki kod, n-estimators için farklı değerlerin denendiği ve en iyi değerin belirlendiği bir GridSearchCV örneği göstermektedir: Bizim projemizde param-grid aşağıdaki gibi tanıplanıp, param-grid = 'n-estimators': [50, 100, 150, 200], Burada param-grid değişkeninde belirtilen n-estimators değerleri denenecek ve en iyi değer GridSearchCV(grid-search = GridSearchCV(xgb-model, param-grid, cv=5, verbose=2)) yöntemi kullanılarak belirlenmiştir. cv parametresi, çapraz doğrulama için kaç parçaaya bölüneceğini belirtir. Bu örnekte, 5 kullanılmıştır. verbose parametresi, çıktıının ayrıntı düzeyini belirtir. Bu örnekte, 2 kullanılmıştır.

XGBRegressor, XGBoost adı verilen açık kaynaklı bir makine öğrenimi kütüphanesi tarafından sağlanan bir algoritmadır. XGBoost, ağaç tabanlı bir öğrenme algoritmasıdır ve Gradient Boosting yöntemini uygular. Bu yöntem, zayıf öğrenicileri (ağaçları) birden fazla kez eğiterek güçlü bir öğrenici oluşturur. XGBRegressor,

verilerin özelliklerini ve hedef değişkenlerini kullanarak ağaç tabanlı bir model oluşturur. Model, verilerdeki özelliklerin birleşimlerine dayalı olarak kararlar alır. XGBRegressor, Gradient Boosting yöntemini uygularken aynı zamanda Regularization ve Shrinkage gibi teknikleri de kullanır. Bu teknikler, modelin overfitting yapmasını önler ve daha iyi genelleştirme yapmasını sağlar. XGBRegressor, birçok hiperparametreye sahiptir. Bu parametreler arasında ağacın derinliği, ağaç sayısı, öğrenme oranı ve daha birçok parametre bulunur. Bu parametreler, modelin eğitimi ve performansı üzerinde etkilidir. En uygun hiperparametrelerin belirlenmesi, modelin performansını artırabilir. Kod örneğinde XGBRegressor kullanılarak, 2018 ve 2019 yıllarına ait veriler kullanılarak bir hız tahmini yapılmıştır. Model, verilerin özelliklerini (saat ve gün) kullanarak, hedef değişkeni (hız) tahmin etmiştir. Model, diğer makine öğrenimi algoritmaları gibi, eğitim verileri üzerinde eğitilmiş ve ardından test verilerinde performansını ölçmek için kullanılmıştır.

XGBRegressor için en uygun parametre seçimi yapılarak mevcut model içim daha iyi sonuçların elde edilmesi hedeflenmiş, bu doğrultuda 106 segment için sırası ile Rastgele Arama Çapraz Doğrulama yöntemi kullanılmış, belirtilen hiperparametre dağılımlarından rastgele seçilen 10 farklı kombinasyonu değerlendirilmiştir. Bu, 10 farklı hiperparametre setini deneyerek en iyi yapılandırma bulmunya çalışılmıştır. Belirtilen hipermateler aşağıdaki gibidir ;

```
n_estimators: [100, 500, 1000]
learning_rate : [0.01, 0.1, 0.2, 0.3]
max_depth : randint(2, 10)
min_child_weight : randint(1,10)
subsample : [0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
colsample_bytree : [0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
```

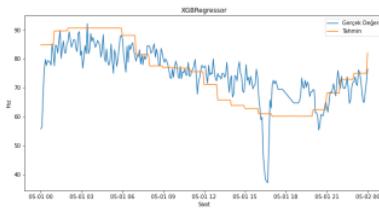
Sonuç olarak en iyi parametre değerleri aşağıdaki şekilde belirlenmiştir;

```
n_estimators: [500]
learning_rate: [0.01]
max_depth: [2]
min_child_weight: [8]
subsample: [0.5]
colsample_bytree: [0.7]
```

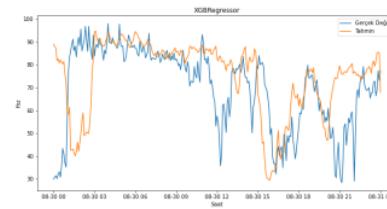
6.4.1 Grafikler

6.4.1.1 Kısa Bayramlar

Kısa bayram olarak sınıflandırılan 23 Nisan, 1 Mayıs, 19 Mayıs, 15 Temmuz, 30 Ağustos, 29 Ekim tarihlerindeki tatillerin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;

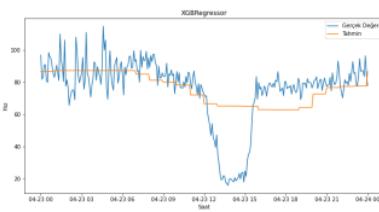


Şekil 6.17 1 Mayıs 2019 XGBoost

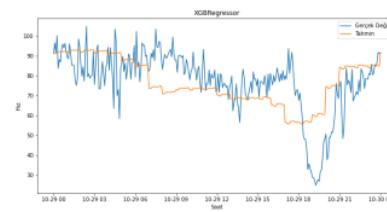


Şekil 6.18 30 Ağustos 2019 XGBoost

Şekil 6.19 Avilar 414'nolu Segment Haftaiçine Denk Gelen Bayramlar

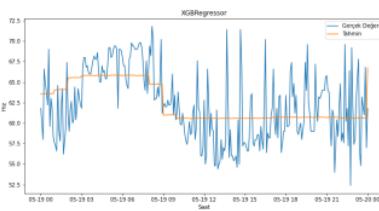


Şekil 6.20 23 Nisan 2019 XGBoost

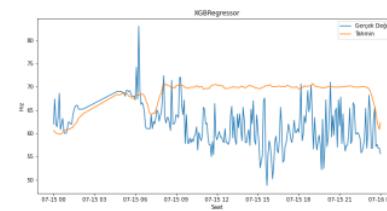


Şekil 6.21 29 Ekim 2019 XGBoost

Şekil 6.22 Ataşehir 316'nolu Segment Haftasonu İle Birleşen Bayramlar



Şekil 6.23 19 Mayıs 2019 XGBoost

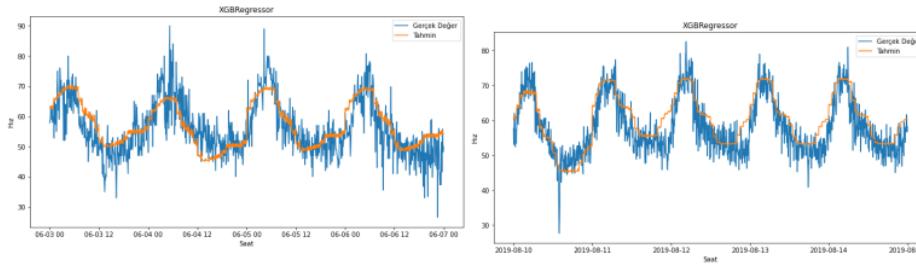


Şekil 6.24 15 Temmuz 2019 XGBoost

Şekil 6.25 Kadıköy Ankara Asfaltı 2102'nolu Segment Farklı Karakteristik İzleyen Bayramlar

6.4.1.2 Uzun Bayramlar

Uzun bayram olarak sınıflandırılan ramazan bayramı ve kurban bayramı tatillerinin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;



Şekil 6.26 Ramazan Bayramı 2019
XGBoost

Şekil 6.27 Kurban Bayramı 2019 XGBoost

Şekil 6.28 Kadıköy Ankara Asfaltı Uzun Bayramlar

6.5 Gradient Boosting

Gradient Boosted Trees (GBT) bir makine öğrenimi algoritmasıdır ve ağaç tabanlı bir öğrenme yöntemidir. GBT, birden fazla zayıf öğrenicinin bir araya gelerek güçlü bir öğrenici oluşturmasına dayanır. GBT, aşamalı bir şekilde, bir önceki aşamadaki hatayı en azı indirmeye çalışarak ağaçlar oluşturur. Her aşamada, ağaçlar modelin zayıf yönlerini düzeltmek için oluşturulur. Her ağaç, önceki ağaçların tahminlerinin hatalarını hedefler. Bu hataların daha sonra diğer ağaçların oluşturulması için kullanılacak yeni bir hedef değişkeni olarak kullanılır.

GBT için en uygun parametre seçimi yapılarak mevcut model içim daha iyi sonuçların elde edilmesi hedeflenmiş, bu doğrultuda 106 segment için sırası ile Rastgele Arama Çapraz Doğrulama yöntemi kullanılmış, belirtilen hiperparametre dağılımlarından rastgele seçilen 10 farklı kombinasyonu değerlendirilmiştir. Bu, 10 farklı hiperparametre setini deneyerek en iyi yapılandırma bulmuna çalışılmıştır. Belirtilen hipermateler aşağıdaki gibidir :

```
4
n_estimators: [100, 500, 1000]
learning_rate: [0.01, 0.1, 0.2, 0.3]
max_depth: randint(2, 10)
min_samples_split : randint(2,20)
min_samples_leaf : randint(1,20)
```

Sonuç olarak en iyi parametre değerleri aşağıdaki şekilde belirlenmiştir;

```
20
n_estimators: [500]
learning_rate: [0.01]
max_depth: [2]
min_samples_split : [5]
```

min_samples_leaf : [3]

6.5.1 Grafikler

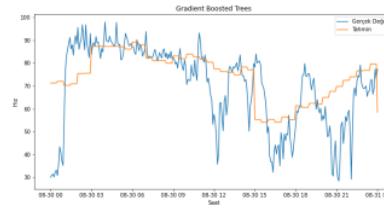
23 Nisan 2018 resmi tatil günü ve 18 -26 Ağustos 2018 Kurban Bayramı haftasında Dolmabahçe Sarayı önünün ve E5 otoyolunun ortalama araç hızı tahmini ‘GBT’ modeli ile yapıldığında ortaya çıkan gerçek değer-tahmin değeri grafiği aşağıdaki gibidir;

6.5.1.1 Kısa Bayramlar

Kısa bayram olarak sınıflandırılan 23 Nisan, 1 Mayıs, 19 Mayıs, 15 Temmuz, 30 Ağustos, 29 Ekim tarihlerindeki tatillerin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;

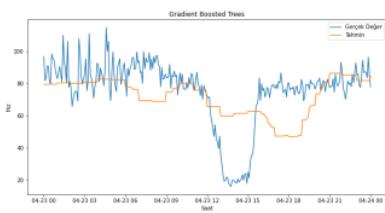


Şekil 6.29 1 Mayıs 2019 Gradient Boosted

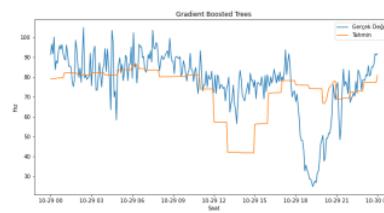


Şekil 6.30 30 Ağustos 2019 Gradient Boosted

Şekil 6.31 Avcılar 414'nolu Segment Haftaiçine Denk Gelen Bayramlar



Şekil 6.32 23 Nisan 2019 Gradient Boosted

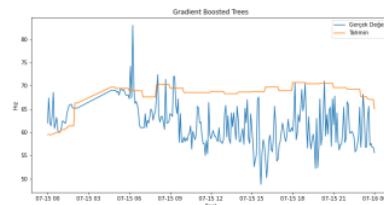


Şekil 6.33 29 Ekim 2019 Gradient Boosted

Şekil 6.34 Ataşehir 316'nolu Segment Haftasonu İle Birleşen Bayramlar



Şekil 6.35 19 Mayıs 2019 Gradient Boosted

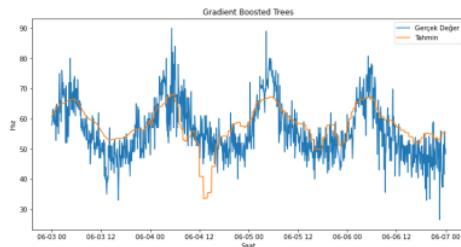


Şekil 6.36 15 Temmuz 2019 Gradient Boosted

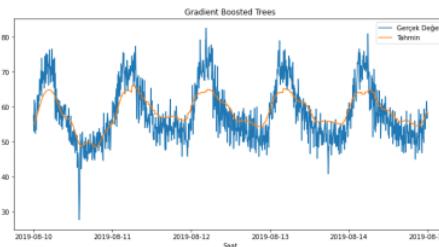
Şekil 6.37 Kadıköy Ankara Asfaltı 2102'nolu Segment Farklı Karakteristik İzleyen Bayramlar

6.5.1.2 Uzun Bayramlar

Uzun bayram olarak sınıflandırılan ramazan bayramı ve kurban bayramı tatillerinin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;



Şekil 6.38 Ramazan Bayramı 2019 Gradient Boosted



Şekil 6.39 Kurban Bayramı 2019 Gradient Boosted

Şekil 6.40 Kadıköy Ankara Asfaltı 2102'nolu Segment Uzun Bayramlar

6.6 CatBoost

CatBoost, bir ağaç tabanlı bir öğrenme algoritmasıdır ve Gradient Boosting Framework (GBM) ailesinin bir üyesidir. Çalışma prensibi, Boosting yöntemi ile verileri modellemek ve tahmin etmek için bir dizi karar ağacı (decision tree) kullanmaktadır.

CatBoost için en uygun parametre seçimi yapılarak mevcut model içim daha iyi sonuçların elde edilmesi hedeflenmiştir, bu doğrultuda 106 segment için sırası ile Rastgele Arama Çapraz Doğrulama yöntemi kullanılmış, belirtilen hiperparametre dağılımlarından rastgele seçilen 10 farklı kombinasyonu değerlendirilmiştir. Bu, 10 farklı hiperparametre setini deneyerek en iyi yapılandırma bulmuna çalışılmıştır. Belirtilen hiperparametreler aşağıdaki gibidir ;

```
iterations: [100, 500, 1000]
depth: [4, 6, 8, 10] learning_rate: [0.01, 0.1, 0.2]
l2_lead,eg : [1,3,5,7]
border_count : [32,64,128]
bagging_temperature : [0.5,0.8,1.0]
```

Sonuç olarak en iyi parametre değerleri aşağıdaki şekilde belirlenmiştir; iterations: [500]
depth: [4]
learning_rate: [0.01]
l2_lead,eg : [3]
border_count : [128]
bagging_temperature :: [0.5]

- iterations: Modelin toplam tekrar sayısını belirtir.
- learning-rate: Her bir adımda model tarafından kullanılan öğrenme oranını belirtir.
- depth: Her bir ağacın maksimum derinliğini belirler.
- l2-leaf-reg: L2 düzenleme parametresi.
- bagging-temperature: Bagging sıcaklığı parametresi.
- random-strength: Randomness kuvveti, özellikle bootstrap sırasında etkilidir.
- border-count: Belirli bir özelliğin ayrıştırılması için kullanılan en az özellik sayısı.
- thread-count: Modelin çalışması için kullanılan iş parçacığı sayısı.
- early-stopping-rounds: Eğitim işlemi sırasında overfitting'i engellemek için kullanılan erken durdurma sayısıdır.

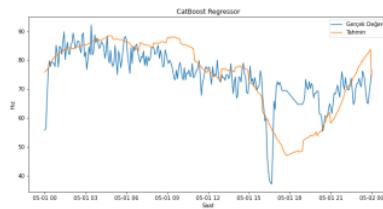
Bu parametrelerin her biri modelin farklı yönlerini kontrol eder ve modelin performansını artırmak için ayarlanabilir. Bununla birlikte, optimal parametre ayarlaması, genellikle deneme yanılma yöntemleri kullanılarak gerçekleştirilir. Projemizde (iterations=1000, learning-rate=0.1, depth=6) değişikleri yapılarak model parametreleri değiştirilmiştir

6.6.1 Grafikler

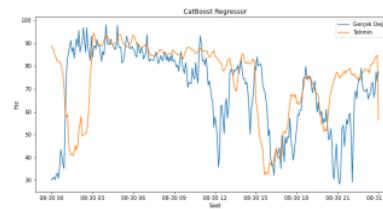
23 Nisan 2018 resmî tatil günü ve 18 -26 Ağustos 2018 Kurban Bayramı haftasında Dolmabahçe Sarayı önünün ve E5 otoyolunun ortalama araç hızı tahmini ‘CatBoost’ modeli ile yapıldığında ortaya çıkan gerçek değer-tahmin değeri grafiği aşağıdaki gibidir

6.6.1.1 Kısa Bayramlar

Kısa bayram olarak sınıflandırılan 23 Nisan, 1 Mayıs, 19 Mayıs, 15 Temmuz, 30 Ağustos, 29 Ekim tarihlerindeki tatillerin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;

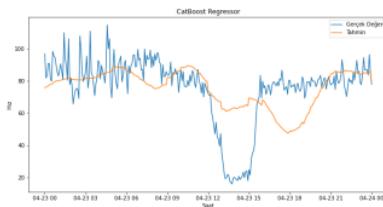


Şekil 6.41 1 Mayıs 2019 Catboost

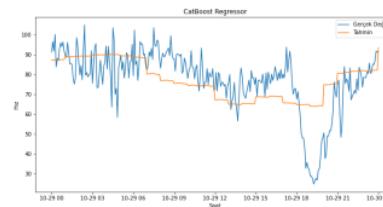


Şekil 6.42 30 Ağustos 2019 Catboost

Şekil 6.43 Avcılar 414'nolu Haftaiçine Denk Gelen Bayramlar



Şekil 6.44 23 Nisan 2019 Catboost

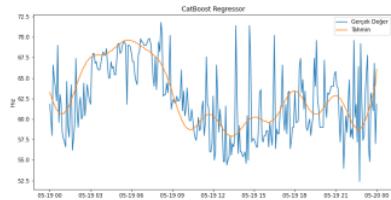


Şekil 6.45 29 Ekim 2019 Catboost

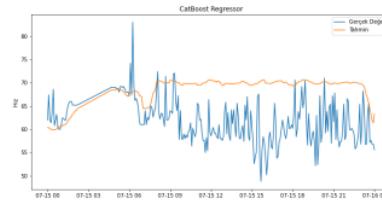
Şekil 6.46 Ataşehir 316'nolu Segment Haftasonu İle Birleşen Bayramlar

6.6.1.2 Uzun Bayramlar

Uzun bayram olarak sınıflandırılan ramazan bayramı ve kurban bayramı tatillerinin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;

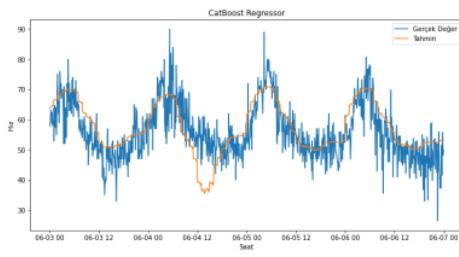


Şekil 6.47 19 Mayıs 2019 Catboost

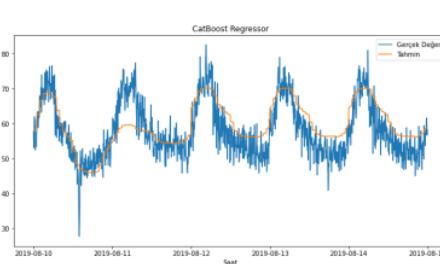


Şekil 6.48 15 Temmuz 2019 Catboost

Şekil 6.49 Kadıköy Ankara Asfaltı 2102'nolu Segment Farklı Karakteristik İzleyen Bayramlar



Şekil 6.50 Ramazan Bayramı 2019 Catboost



Şekil 6.51 Kurban Bayramı 2019 Catboost

Şekil 6.52 Kadıköy Ankara Asfaltı 2102'nolu Segment Uzun Bayramlar

6.7 LGBMRegressor

LGBMRegressor, LightGBM kütüphanesi tarafından sağlanan bir sınıflandırma ve regresyon algoritmasıdır. LightGBM, yüksek performanslı ve dağıtılmış bir makine öğrenimi kütüphanesidir ve Gradient Boosting Framework’ünü kullanarak veri kümesi üzerindeki tahminleri geliştirir. LGBMRegressor, GBM’deki ağaç temelli modelleme stratejisini kullanır ve bu modelleme stratejisi, verilerin özelliklerine ve hedefine dayalı olarak ağaçları büyütmek için özel bir eğitim işlemi kullanır. Bu, LightGBM’ın diğer GBM modellerine göre daha hızlı olmasını sağlar

LGBM için en uygun parametre seçimi yapılarak mevcut model içim daha iyi sonuçların elde edilmesi hedeflenmiş, bu doğrultuda 106 segment için sırası ile Rastgele Arama Çapraz Doğrulama yöntemi kullanılmış, belirtilen hiperparametre dağılımlarından rastgele seçilen 10 farklı kombinasyonu değerlendirilmiştir. Bu, 10 farklı hiperparametre setini deneyerek en iyi yapılandırma bulmuna çalışılmıştır. Belirtilen hiperparametreler aşağıdaki gibidir ;

```
4
n_estimators: [100, 500, 1000]
learning_rate: [0.01, 0.1, 0.2, 0.3]
```

```
max_depth: randint(2, 10)
minchildsamples : randint(1, 10)
3subsample : [0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
colsample_bytree : [0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
reg_alpha : [0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
reg_lambda : [0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
```

Sonuç olarak en iyi parametre değerleri aşağıdaki şekilde belirlenmiştir;

```
n_estimators: [500]
learning_rate': [0.01]
max_d26epth : [6]
min_child_samples : [5]
subsample : [1]
colsample_bytree : [0.6]
reg_alpha : [0.01]
reg_lambda : [0]
```

- objective: Modelin optimize edeceği amaç fonksiyonu.
- boosting-type: Gradient boosting tipi.
- n-estimators: Boosting iterasyonları sayısı.
- learning-rate: Boosting algoritmasının her iterasyonunda kullanacağı öğrenme oranı.
- max-depth: Karar ağacı derinliği sınırı. Burada, 5 olarak belirtilmiştir.
- min-child-samples: Her bir yaprak düğümünde en az kaç veri örneğinin bulunması gerektiği.
- subsample: Her bir iterasyonda kullanılacak alt örneklem oranı.
- colsample-bytree: Her bir iterasyonda kullanılacak özelliklerin oranı.
- n-jobs: Eğitim işleminin paralel olarak çalıştırılacağı işlemci sayısı.
- random-state: Modelin tekrarlanabilir sonuçlar vermesi için kullanılan rastgele sayı üretici için tohum değeri.

Projemizde bu metod kullanılırken parametre değerleri aşağıdaki gibi verilmiştir; 'objective': 'regression', 'boosting-type': 'gbdt', 'n-estimators': 100, 'learning-rate':

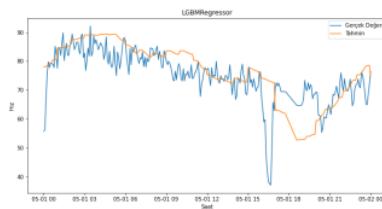
```
0.1, 'max-depth': 5, 'min-child-samples': 20, 'subsample': 0.8, 'colsample-bytree': 0.8,  
'n-jobs': -1, 'random-state': 0
```

6.7.1 Grafikler

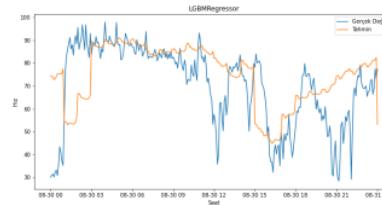
23 Nisan 2018 resmî tatil günü ve 18 -26 Ağustos 2018 Kurban Bayramı haftasında Dolmabahçe Sarayı önünün ve E5 otoyolunun ortalama araç hızı tahmini 'LightGBM' modeli ile yapıldığında ortaya çıkan gerçek değer-tahmin değeri grafiği aşağıdaki gibidir;

6.7.1.1 Kısa Bayramlar

Kısa bayram olarak sınıflandırılan 23 Nisan, 1 Mayıs, 19 Mayıs, 15 Temmuz, 30 Ağustos, 29 Ekim tarihlerindeki tatillerin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;

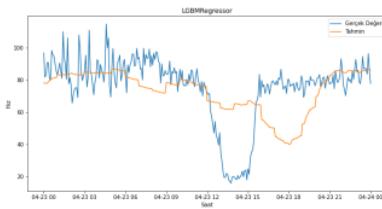


Şekil 6.53 1 Mayıs 2019 LightGBM



Şekil 6.54 30 Ağustos 2019 LightGBM

Şekil 6.55 Avcılar 414'nolu Segment Haftaiçine Denk Gelen Bayramlar



Şekil 6.56 23 Nisan 2019 LightGBM



Şekil 6.57 29 Ekim 2019 LightGBM

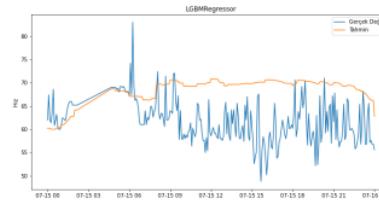
Şekil 6.58 Ataşehir 316'Nolu Segment Haftasonu İle Birleşen Bayramlar

6.7.1.2 Uzun Bayramlar

Uzun bayram olarak sınıflandırılan ramazan bayramı ve kurban bayramı tatillerinin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;

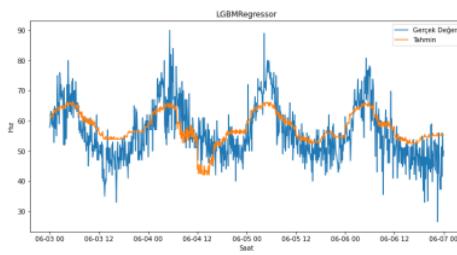


Şekil 6.59 19 Mayıs 2019 LightGBM

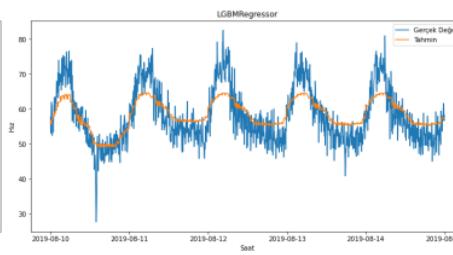


Şekil 6.60 15 Temmuz 2019 LightGBM

Şekil 6.61 Farklı Karakteristik İzleyen Bayramlar



Şekil 6.62 Ramazan Bayramı 2019
LightGBM



Şekil 6.63 Kurban Bayramı 2019
LightGBM

Şekil 6.64 Kadıköy Ankara Asfaltı Uzun Bayramlar

6.8 DNNs

Deep Neural Networks (DNNs) veya derin sinir ağları, yapay sinir ağı mimarisinin bir türüdür ve birçok katmandan oluşur. Giriş katmanı verileri alır, çıktı katmanı sonuçları verir ve ara katmanlar arasında bilgi işlemeyi gerçekleştirir. Her bir katmanın nöronları, girdi değerleriyle ağırlıkların çarpılmasını ve bir aktivasyon fonksiyonuna sokulmasını içeren matematiksel işlemler yapar. Bu işlemler sonucunda çıktılar, bir sonraki katmana veya çıktı katmanına aktarılır. DNN'ler, yüksek ölçekli, karmaşık veri kümelerinde yapısal özelliklerini bulmak ve öğrenmek için kullanılır. Örneğin, bir görüntü tanıma modeli için, ilk katmanlar genellikle basit özellikleri (çizgiler, kenarlar, renkler) algılamak için tasarlanırken, daha sonraki katmanlar bu özelliklerin daha karmaşık kombinasyonlarını öğrenerek daha yüksek seviyeli özellikleri (objeler, yüzler) algılamak için kullanılır. DNN'ler, gerçek dünya problemlerinde başarılı sonuçlar vermek için geniş bir veri kümesi ve yeterli hesaplama gücü gerektirir. Günümüzde DNN'ler, görüntü ve dil işleme, nesne tespiti, doğal dil işleme, otomatik sürüsüz oyun oynama ve daha birçok alanda kullanılır.

DNN'lerin temel parametreleri şunlardır

- Topoloji: DNN'lerin birçok katmanı vardır ve her bir katmanın sayısı, boyutu ve

türü belirlenmelidir. Ayrıca, her katmanın ne kadar öğrenme yapması gerektiği de belirlenir.

- Aktivasyon fonksiyonu: Aktivasyon fonksiyonları, her katmanın çıktısını hesaplamak için kullanılır. ReLU, sigmoid, tanh ve softmax gibi çeşitli aktivasyon fonksiyonları mevcuttur.
- Öğrenme hızı: Bu, ağın öğrenme hızını kontrol etmek için kullanılan bir parametredir. Öğrenme hızı ne kadar yüksek olursa, ağ daha hızlı öğrenir ancak daha yüksek bir hata oranına sahip olabilir.
- Mini-batch boyutu: Bu, ağın eğitim verilerini kaç parçaya ayıracığını belirler. Mini-batch boyutu arttıkça, ağ daha hızlı öğrenir ancak daha az hassas bir sonuç verebilir.
- Epoch sayısı: Bir epoch, tüm verilerin ağıda bir kez eğitilmesini ifade eder. Epoch sayısı, ağın ne kadar süre eğitleceğini belirler.
- Optimizasyon algoritması: Ağın eğitilmesi sırasında kullanılan optimizasyon algoritması, ağın hızını ve doğruluğunu etkiler. Stokastik gradyan inişi (SGD), ADAM ve RMSProp gibi çeşitli optimizasyon algoritmaları mevcuttur.
- Dropout: Dropout, ağın aşırı öğrenmesini önlemek için kullanılan bir tekniktir. Bu, rastgele birimlerin devre dışı bırakıldığı bir yöntemdir ve ağın daha iyi genellemeye yapmasına yardımcı olabilir.

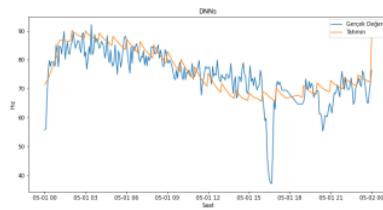
```
Projemizde kullanılan parametreler şu şekildedir;      6  
activation="relu",      input_dim=X-train.shape[1])      model.add(Dense(128,  
model.add(Dense(64,          activation="relu"))      model.add(Dropout(0.3))  
model.add(Dense(1, activation="linear"))  model.compile(loss="mean-absolute-error",  
optimizer=Adam(lr=0.001))  history = model.fit(X-train, y-train, epochs=100,  
batch_size=32, verbose=0)
```

6.8.1 Grafikler

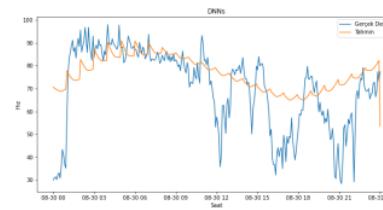
23 Nisan 2018 resmî tatil günü ve 18 -26 Ağustos 2018 Kurban Bayramı haftasında Dolmabahçe Sarayı'ni önenin ve E5 otoyolunun ortalama araç hızı tahmini 'DNNs 'modeli ile yapıldığında ortaya çıkan gerçek değer-tahmin değeri grafiği aşağıdaki gibidir

6.8.1.1 Kısa Bayramlar

Kısa bayram olarak sınıflandırılan 23 Nisan, 1 Mayıs, 19 Mayıs, 15 Temmuz, 30 Ağustos, 29 Ekim tarihlerindeki tatillerin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;

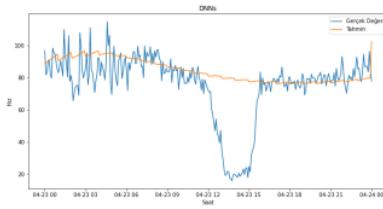


Şekil 6.65 1 Mayıs 2019 DNNs

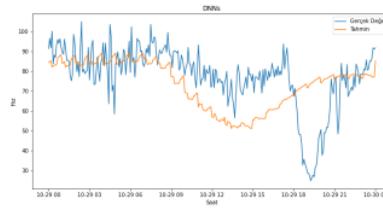


Şekil 6.66 30 Ağustos 2019 DNNs

Şekil 6.67 Avcılar 414'nolu Haftaiçine Denk Gelen Bayramlar

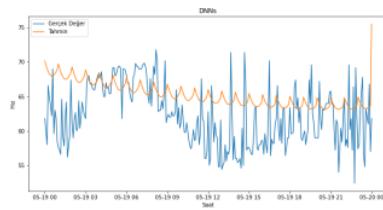


Şekil 6.68 23 Nisan 2019 DNNs

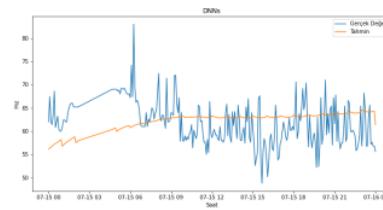


Şekil 6.69 29 Ekim 2019 DNNs

Şekil 6.70 Ataşehir 316'nolu Haftasonu İle Birleşen Bayramlar



Şekil 6.71 19 Mayıs 2019 DNNs



Şekil 6.72 15 Temmuz 2019 DNNs

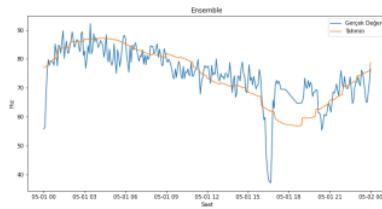
Şekil 6.73 Kadıköy Ankara Asfaltı Farklı Karakteristik İzleyen Bayramlar

6.9 Ensemble

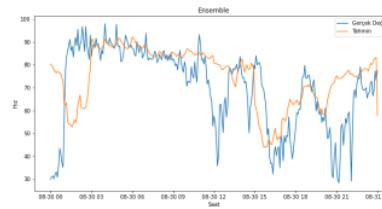
6.9.1 Grafikler

6.9.1.1 Kısa Bayramlar

Kısa bayram olarak sınıflandırılan 23 Nisan, 1 Mayıs, 19 Mayıs, 15 Temmuz, 30 Ağustos, 29 Ekim tarihlerindeki tatillerin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;

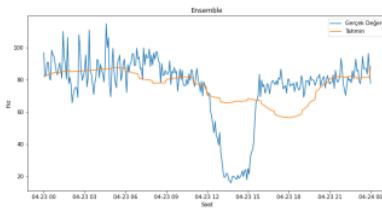


Şekil 6.74 1 Mayıs 2019 Ensemble

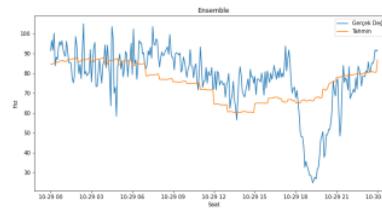


Şekil 6.75 30 Ağustos 2019 Ensemble

Şekil 6.76 Avcılar 414'nolu Segment Haftaiçine Denk Gelen Bayramlar

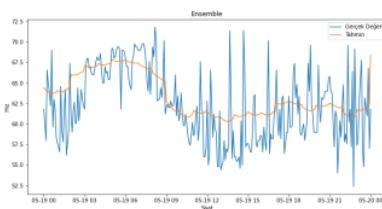


Şekil 6.77 23 Nisan 2019 Ensemble

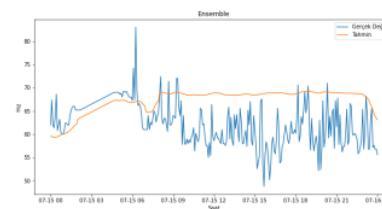


Şekil 6.78 29 Ekim 2019 Ensemble

Şekil 6.79 Ataşehir 316'nolu Haftasonu ile Birleşen Bayramlar



Şekil 6.80 19 Mayıs 2019 Ensemble

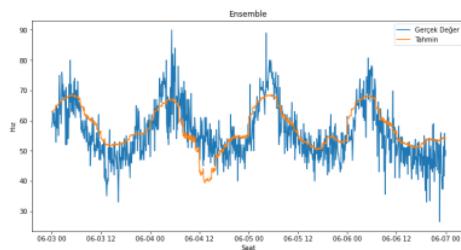


Şekil 6.81 15 Temmuz 2019 Ensemble

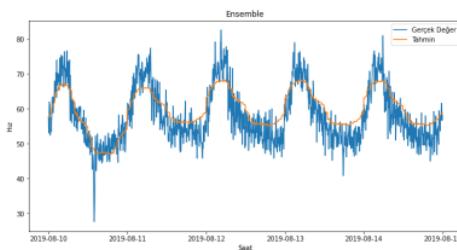
Şekil 6.82 Farklı Karakteristik İzleyen Bayramlar

6.9.1.2 Uzun Bayramlar

Uzun bayram olarak sınıflandırılan ramazan bayramı ve kurban bayramı tatillerinin gerçek değer ve tahminleme grafikleri aşağıdaki gibidir ;



Şekil 6.83 Ramazan Bayramı 2019
Ensemble



Şekil 6.84 Kurban Bayramı 2019
Ensemble

Şekil 6.85 Kadıköy Ankara Asfaltı 2102'nolu Segment Uzun Bayramlar

6.10 Hata Oranları

Hata oranları kısa ve uzun bayramlar olmak üzere ikiye ayrılmıştır. Test seti olarak kısa bayramlar için 1 günlük test seti kullanılırken uzun bayramlar için 4 ve 5 günlük test setleri kullanılmıştır. Ramazan bayramı için test seti arife ve 3 günlük bayramı içerisinde kurban bayramı için arife ve 4 günlük bayramı içermektedir. Bunu yapmaktadır sebep tatil günlerinin birbirine olan benzerliği tatilden önceki veya sonraki günlerin bu benzerliğe sahip olmamasıdır.

6.10.1 Kısa Bayramlar

Kısa bayramları 3 farklı başlık altında (Haftasonuna gelmesi, hafta içine gelerek 3 günlük bayram oluşturmaları, hafta içi tek günlük oluşturmaları) inceleme işlemlerinde her bir bayram günü ayrı ayrı değerlendirilerek set set tahminleme işlemleri yapılmıştır. Aşağıdaki tablolarda her bir bayram günü için hangi zaman aralıkları ile ne kadar hata alındığı bilgileri yer almaktadır.

Tablo 6.1 1 Mayıs 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Bayramlar ,Set-2: Set1 + Aynı Sene Geçen Hafta Set-3: Set2 + 23 Nisan 2019, Set-4: Set3 + FFT

	Set-1	Set-2	Set-3	Set-4
Random Forest	14.00%	13.99%	10.43%	9.53%
GBRegressor	12.03%	11.83%	11.44%	10.26%
XGBRegressor	13.96%	13.88%	11.69%	10.05%
LGBMRegressor	13.07%	12.94%	11.14%	10.34%
Catboost	13.78%	13.70%	10.25%	9.63%
DNNs	11.88%	11.62%	12.76%	11.28%
Ensemble	12.68%	12.57%	10.11%	9.53%

Tablo 6.2 23 Nisan 2019 Hata Oranları, Set-1: Aynı Yıl Geçen Hafta ,Set-2: Set1 + Aynı Sene Geçen Hafta Set-3: Set1 + FFT, Set-4:,Set2 + FFT

	Set-1	Set-2	Set-3	Set-4
Random Forest	15.41%	15.37%	12.24%	11.69%
GBRegressor	14.49%	13.87%	13.35%	12.31%
XGBRegressor	15.46%	15.29%	12.41%	11.69%
LGBMRegressor	14.70%	14.35%	13.27%	12.01%
Catboost	15.28%	15.10%	13.21%	11.60%
DNNs	14.24%	14.28%	13.95%	14.28%
Ensemble	14.34%	14.16%	12.19%	11.56%

Tablo 6.3 29 Ekim 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Tatiller ve önceki günleri ,Set-2: Set1 + Aynı Sene Geçen Hafta Set-3: 2 Hafta 2018 + Önceki 2 Hafta 2019, Set-4:,Set3 + FFT

	Set-1	Set-2	Set-3	Set-4
Random Forest	13.00%	13.00%	12.36%	9.89%
GBRegressor	11.36%	11.27%	12.60%	10.35%
XGBRegressor	12.98%	12.89%	13.29%	11.61%
LGBMRegressor	12.11%	11.90%	12.14%	10.55%
Catboost	12.81%	12.69%	12.27%	10.53%
DNNs	11.60%	11.28%	11.85%	11.42%
Ensemble	11.76%	11.64%	11.68%	9.75%

Tablo 6.4 19 Mayıs 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Tatiller ve önceki bir gün, Set-2: Aynı Yıl Bir Hafta Önce Aynı Gün Set-3: Set1 + FFT, Set-4: Set2 + FFT

	Set-1	Set-2	Set-3	Set-4
Random Forest	15.12%	13.59%	15.27%	14.10%
GBRegressor	15.42%	13.27%	14.11%	13.73%
XGBRegressor	15.70%	14.07%	14.40%	13.98%
LGBMRegressor	15.13%	13.45%	13.29%	13.25%
Catboost	14.65%	13.91%	13.42%	13.73%
DNNs	20.56%	17.56%	18.19%	15.50%
Ensemble	14.72%	13.65%	14.10%	13.32%

Tablo 6.5 15 Temmuz 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Tatiller ve önceki bir gün, Set-2: Set1 + FFT, Set-3: Set1 + Aynı yıl önceki haftasonu ve önceki hafta aynı gün + FFT

	Set-1	Set-2	Set-3
Random Forest	15.20%	15.00%	13.47%
GBRegressor	13.99%	13.90%	12.90%
XGBRegressor	15.20%	15.02%	13.20%
LGBMRegressor	14.42%	14.38%	12.83%
Catboost	15.02%	14.93%	13.15%
DNNs	13.44%	13.53%	13.42%
Ensemble	14.04%	14.01%	12.76%

Tablo 6.6 30 Ağustos 2019 Hata Oranları, Set-1: Geçen Sene Tek Günlük Tatiller ve önceki bir gün + Geçen Hafta, Set-2: Set1 + FFT, Set-3: Set2 + Aynı Sene Tek Günlük Bayramlar(23 Nisan, 1 Mayıs)

	Set-1	Set-2	Set-3
Random Forest	13.43%	10.84%	10.84%
GBRegressor	10.88%	10.64%	10.64%
XGBRegressor	13.33%	9.92%	9.93%
LGBMRegressor	12.14%	11.98%	10.23%
Catboost	13.13%	12.48%	10.10%
DNNs	10.67%	10.72%	10.77%
Ensemble	11.79%	10.50%	9.87%

6.10.2 Uzun Bayramlar

Uzun bayramlar olan ramazan ve kurban bayramları kendine özgü davranışları sergiler. Elimizdeki veri setinde 2018 ramazan bayramı haftasonuna denk gelmesi nedeni ile 3 günlük tatil olarak ilan edilmiş, 2019 senesinde ise ramazan bayramı 9 gün olması nedeni ile eğitim aşamasında 2018 ramazan bayramı günlerini kullanmamız hata sonucunu artırmaktadır. Şekil 6.7 ve Şekil 6.8 tablolarına bakıldığında ramazan bayramında oluşturduğumuz modelin daha kötü sonuç verdiği görülebilir. Bu hata sonucunu düşürmek için Set-1 olarak adlandırılan eğitim setimizi ramazan bayramında kullanmak yerine bu eğitim setine kurban bayramının tatil günlerini de eklemeli düşündük. Kurban bayramında ise elimizdeki veri seti yeterli olduğu ve birbirini takip eden senelerde aynı karakteristiği izlediği için Set-3 6.7 verisetinde Set-1 verisine haftasonu verilerini eklemek dışında herhangi bir iyileştirmeye gitmedik.

Tablo 6.7 Kurban Bayramı 2019 Hata Oranları, Set-1: Geçen Sene Kurban Bayramı 2 Hafta 13 - 27 Ağustos, Set-2: Set1 + Ramazan Bayramı, Set-3: Set1 + Aynı Sene Önceki Hafta sonu + FFT

	Set-1	Set-2	Set-3
Random Forest	11.44%	13.85%	9.14%
GBRegressor	11.44%	14.00%	8.92%
XGBRegressor	11.97%	14.92%	9.58%
LGBMRegressor	11.40%	14.45%	9.23%
Catboost	11.34%	13.75%	9.20%
DNNs	13.34%	15.38%	10.71%
Ensemble	11.28%	13.83%	8.97%

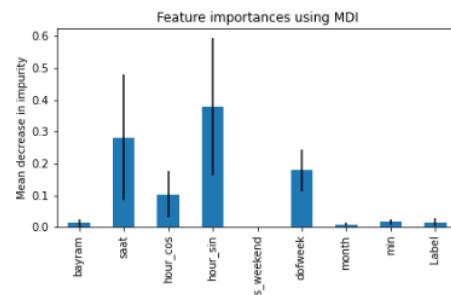
Tablo 6.8 Ramazan Bayramı 2019 Hata Oranları, Set-1: Geçen Sene Ramazan Bayramı 2 Hafta 11 - 25 Haziran, Set-2: Set1 + Geçen Sene Haftasonu + Aynı Sene Bir Önceki Haftasonu, Set-3: Set1 + Kurban Bayramı

	Set-1	Set-2	Set-3
Random Forest	18.05%	15.29%	13.85%
GBRegressor	17.88%	17.01%	14.00%
XGBRegressor	19.16%	16.66%	14.92%
LGBMRegressor	18.31%	18.85%	14.45%
Catboost	17.82%	14.22%	13.75%
DNNs	19.99%	14.88%	15.38%
Ensemble	17.63%	15.61%	13.83%

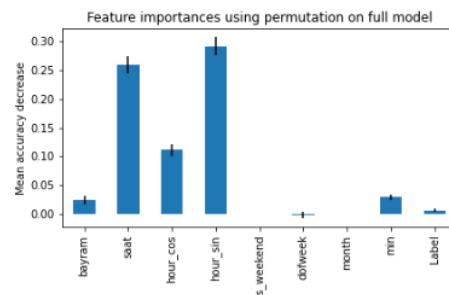
6.11 Deneysel Sonuçlar

Uygulamada kısa bayramlar ve uzun bayramlar için farklı deney ortamları hazırlandı. Modeli eğitirken kullanılan değişkenler kısa bayramlar için (saat, saat_cos, saat_sin,

bayram, hafta_günü) olurken uzun bayramlar için (saat, saat_cos, saat_sin, bayram, hafta_günü, month, ardışık_tatil) olarak seçildi. Şekil 6.88'e bakıldığından bu değişkenlerin modelin başarı oranına etkisinin grafikleştirilmiş hali hem eğitim setinde MDI kullanılarak hem de test setinde permütasyon kullanılarak hesaplanmıştır. Sadece uzun tatil günlerinde kullanılan ardışık tatil değişkeninin model performansına etkisini ise Şekil 6.91'da görülebilir. Aynı şekilde haftasonuna denk gelen veya haftasonuyla birleşen kısa tatil günlerinde de is_weekend değişkeni model performansında etkili olmuştur.

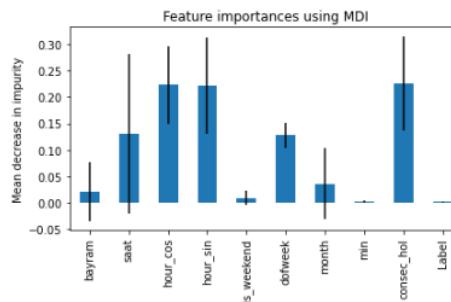


Şekil 6.86 Eğitim Setinde MDI kullanılması

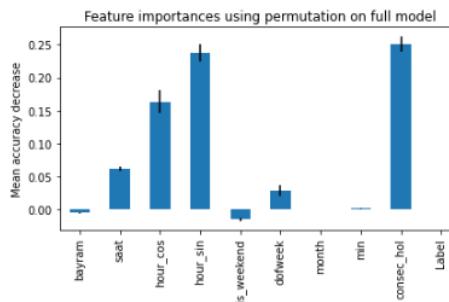


Şekil 6.87 Test Setinde Permütasyon Kullanılması

Şekil 6.88 Farklı Yöntemlerle Hesaplanan Değişken Önemi



Şekil 6.89 Eğitim Setinde MDI kullanılması

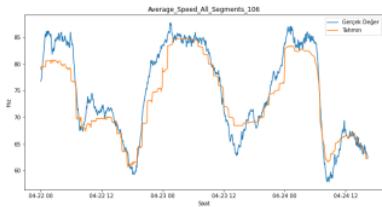


Şekil 6.90 Test Setinde Permütasyon Kullanılması

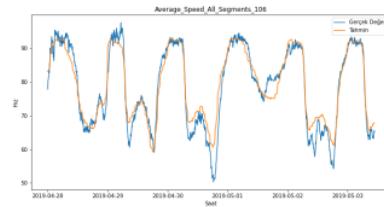
Şekil 6.91 Ardışık Tatil Değişkeninin Model Performansına Etkisi

Değişken seçiminin yanısıra yapılan hız tahmininde tatil öncesi, tatil ve tatil sonrası için farklı eğitim seti yapıları oluşturuldu. Bu eğitim seti yapısının bir örneğini Şekil 6.99'da görebilirsiniz. Kısa bayramlar için bu yapı bir günlük tatillerde 1 gün öncesi ve olmasını kapsayan 3 günlük bir test seti oluşturulurken uzun bayramlar için bütün tatil günleriyle beraber 2 gün öncesi ve olmasını kapsayan bir test seti oluşturuldu. Ayrıca tatil öncesi, tatil ve tatil sonrasında seçilen eğitim seti yapısının değişmesine

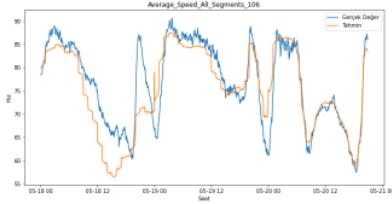
ek olarak her bir model için kullanılan parametrelerde daha önce optimize edilmiş bir şekilde yapının içine dahil edildi. Daha sonra bu yapı .pkl uzantılı pickle objeleri olarak kaydedildi. Bu sayede her bir gün ve model için en optimum parametre ve gün seçimi sağlanmış oldu.



Şekil 6.92 23 Nisan 2019



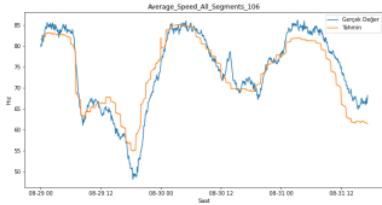
Şekil 6.93 1 Mayıs



Şekil 6.94 19 Mayıs



Şekil 6.95 15 Temmuz



Şekil 6.96 30 Ağustos



Şekil 6.97 29 Ekim

Şekil 6.98 Tatil Günleri İçin 106 Segment'in 'Gerçek Ve Tahmin' Hız Değer Grafikleri

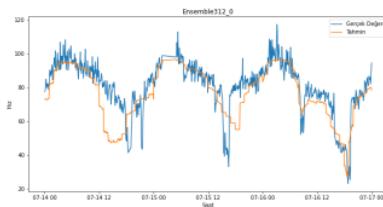
Segmentlere göre hız tahmini yapmak o segmenti kullanan araçların veya trafiği yönlendiren unsurların planlaması için iyi olsa da buna ek olarak İstanbul genelinde bir yoğunluk tahmini yapmak tatil günlerinin karakteristiklerini anlamak ve son kullanıcı için bir fikir vermesi nedeniyle önemli olmakta. Bu nedenle genel bir hız tahmini için Grafik 6.98'de görüldüğü üzere geliştirilen model İstanbul genelinde oldukça başarılı sonuçlar vermektedir. Performans analizi kısmında hata metrikleri olarak bu sonuçlar verilmiştir.

```

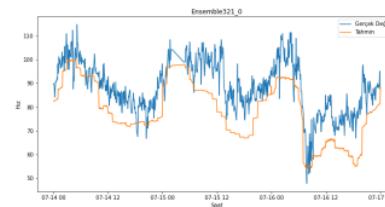
diff_day_sets = [
    (pd.date_range(start="08-16-2018", end="08-16-2018 23:55", freq="5T"),
     ["05-30-2019", "05-31-2019", "07-26-2019", "06-13-2018"], 
    [], 
    [], 
    pd.date_range(start="08-08-2019", end="08-10-2019 00:00", freq="5T")), # Tatil Öncesi
    (pd.date_range(start="08-18-2018", end="08-26-2018 23:55", freq="5T"),
     ["08-03-2019", "08-04-2019"], 
     ["08-21-2018", "08-22-2018", "08-23-2018", "08-24-2018"], 
     ["08-11-2019", "08-12-2019", "08-13-2019", "08-14-2019"], 
     pd.date_range(start="08-10-2019", end="08-19-2019 00:00", freq="5T")), # Tatil
    (pd.date_range(start="07-22-2019", end="07-22-2019 23:55", freq="5T"), # 04-24-2019
     ["08-16-2018", "08-27-2018", "08-28-2018"], 
    [], 
    [], 
    pd.date_range(start="08-19-2019", end="08-21-2019 00:00", freq="5T")) # Tatil Sonrası
]

```

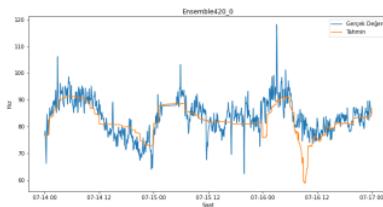
Şekil 6.99 Eğitim Seti Nesnesi



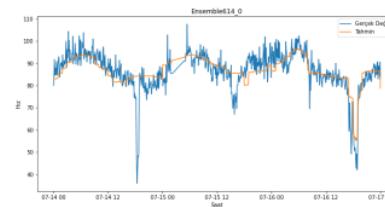
Şekil 6.100 Kadıköy (312 Numaralı Segment)



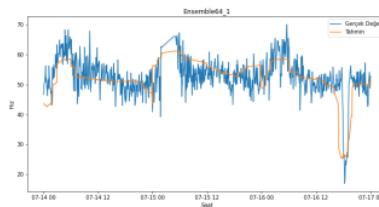
Şekil 6.101 Maltepe (321 Numaralı Segment)



Şekil 6.102 Bakırköy (420 Numaralı Segment)

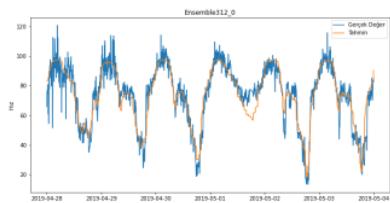


Şekil 6.103 Üsküdar (614 Numaralı Segment)

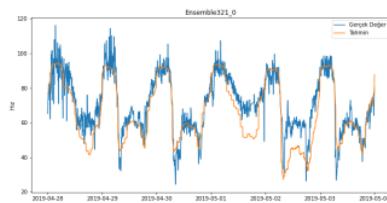


Şekil 6.104 Beşiktaş (64 Numaralı Segment)

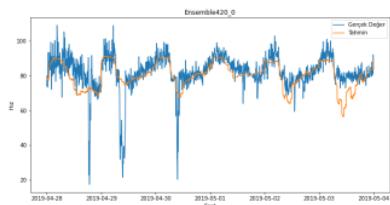
Şekil 6.105 15 Temmuz 2019 Tatilinde 5 Farklı Segmentin 'Gerçek Ve Tahmin' Hiz Değer Grafikleri



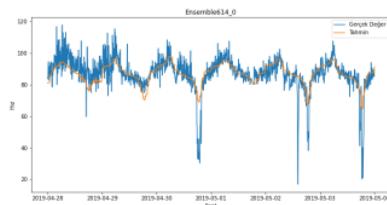
Şekil 6.106 Kadıköy (312 Numaralı Segment)



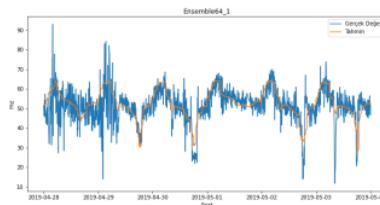
Şekil 6.107 Maltepe (321 Numaralı Segment)



Şekil 6.108 Bakırköy (420 Numaralı Segment)

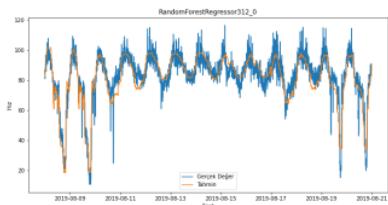


Şekil 6.109 Üsküdar (614 Numaralı Segment)



Şekil 6.110 Beşiktaş (64 Numaralı Segment)

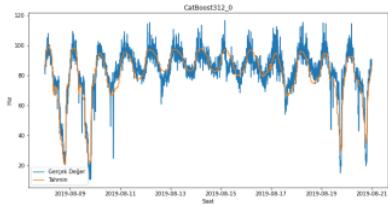
Şekil 6.111 1 Mayıs 2019 Tatilinde 5 Farklı Segmentin 'Gerçek Ve Tahmin' Hız Değer Grafikleri



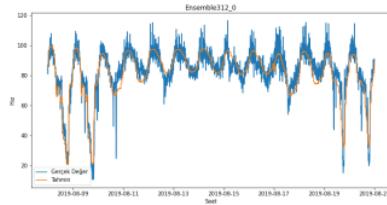
Şekil 6.112 Random Forest



Şekil 6.113 XGB Regressor

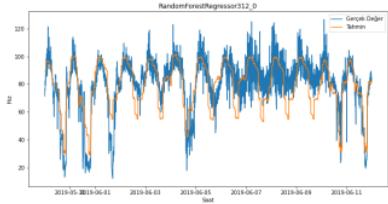


Şekil 6.114 CatBoost

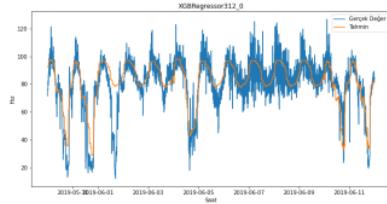


Şekil 6.115 Ensemble

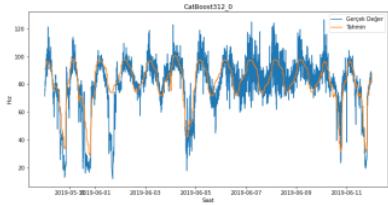
Şekil 6.116 Kurban Bayramı(2019) Tatilinde Kadıköye Farklı Modellerin 'Gerçek Ve Tahmin' Hız Değer Grafikleri



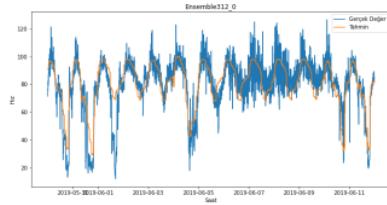
Şekil 6.117 Random Forest



Şekil 6.118 XGB Regressor



Şekil 6.119 CatBoost



Şekil 6.120 Ensemble

Şekil 6.121 Ramazan Bayramı(2019) Tatilinde Kadıköye Farklı Modellerin 'Gerçek Ve Tahmin' Hız Değer Grafikleri

6.12 Performans Analizi

Denenen farklı modeller arasında Gradient Boosting, LGBM modellerinden istenilen verim alınamadığı ve derin öğrenme modelleri için yeterli tatil verisi olmadığı için Random Forest, XGBRegressor ve Catboost modellerini FFT ön işlemesiyle birleştirilip FFT-Ensemble modeli kullanmaya karar verildi. Tablo 6.12 göreceğiniz gibi farklı modeller farklı günler için iyi sonuçlar vermektedir. Bunun nedenin parametre optimizasyonu olabileceğini düşünülebilir fakat bunu en aza indirmek için her bir tatil gününe öncesi ve sonrası farklı olmak üzere parametre optimizasyonu yapılmıştır. Ortalama olarak ensemble modelimiz daha iyi sonuç verdiği görülebilir. Buna ek olarak tablo 6.9 bakıldığından genel segmentlerin karakteristiğini yansitan 106 segmentin hız tahminleri ve gerçek hız değerleri karşılaştırıldığında İstanbul genelinde ortalama %3.67 hata vermektedir. Uzun tatil günlerinde ramazan bayramının kurban bayramından daha fazla hata vermesinin nedeni 2018 yılında ramazan bayramının haftasonuna denk gelerek 3 günlük bir tatil olması aksine 2019 yılında ise haftaiçine denk gelerek 9 günlük bir tatil oluşturulmasıdır. Verideki bu problem 2019 ramazan bayramını tahmin etmede büyük sorun teşkil etmektedir. Bunu çözmek için nispeten daha uzun tatil olan 2018 kurban bayramı verisinide eğitim setine dahil ederek bu hata oranını minimuma indirmeyi denedik. Tablo 6.10 bakarsanız örnek olarak seçilen segmentlerin hata oranlarını modelimiz özeline görebilirsiniz. Bu segmentleri seçerken segmentlerin İstanbulun farklı bölgelerini yansıtmasına dikkat edilmiştir. Üsküdar, Beşiktaş, Bakırköy, Kadıköy ve Maltepe seçilen segmentlerin bulunduğu illerdir. Göründüğü üzere bu segmentlerin ikisi Avrupa yakasında diğer üçü ise Anadolu yakasında bulunmaktadır. Kadıköy ve Beşiktaş gibi nüfusun yanı sıra turistik olarak da yoğun olan bölgelerde de modelimiz başarılı sonuçlar vermektedir. Modelin eksik kaldığı nokta ise milli bayramlarda anlık olarak kutlamalardan dolayı oluşan trafik yoğunluklarını tahmin edememesidir. Ancak bu yoğunluk sadece milli bayramlar nedenli değil aynı zamanda konser, maç, trafik kazası vb. farklı dış etkenler kaynaklı olabilir. Bu projede genel olarak tatil karakteristikleri incelenmesi nedeniyle bu etkenler gözardı edilmiştir. Bunun etkileri milli bayramlarda Kadıköy ve Beşiktaş gibi merkez ilçelerin MAPE oranlarında görülebilir.

Tablo 6.9 İstanbul Geneli Segmentlerin MAPE Ortalaması

	23 Nisan	1 Mayıs	19 Mayıs	15 Temmuz	30 Ağustos	29 Ekim	Ramazan	Kurban	Ortalama
Genel MAPE	3.04%	3.47%	4.32%	3.55%	3.48%	3.41%	4.18%	3.79%	3.65%

Tablo 6.10 Örnek Segmentlere Göre MAPE Hata Oranı

Segment No	23 Nisan	1 Mayıs	19 Mayıs	15 Temmuz	30 Ağustos	29 Ekim	Ramazan	Kurban	Ortalama
614 Üsküdar	5.84%	7.18%	5.66%	5.23%	12.80%	4.95%	7.08%	4.69%	6.68%
420 Bakırköy	6.29%	8.55%	6.19%	4.44%	13.63%	9.41%	9.34%	8.28%	8.27%
312 Kadıköy	13.68%	9.40%	18.10%	11.89%	15.77%	11.32%	14.51%	7.81%	12.81%
64 Beşiktaş	7.87%	11.46%	13.36%	7.92%	11.38%	10.04%	11.68%	7.55%	10.16%
321 Maltepe	9.83%	11.40%	8.15%	11.89%	16.60%	11.41%	11.19%	11.47%	11.49%

Tablo 6.11 Farklı Modellerin 106 Segment Genelinde Ortalama MAPE Hata Oranı

Modeller	23 Nisan	1 Mayıs	19 Mayıs	15 Temmuz	30 Ağustos	29 Ekim	Ramazan	Kurban	Ortalama
RandomForestRegressor	12.26%	14.67%	12.97%	11.28%	13.61%	12.10%	13.66%	11.20%	12.72%
XGBRegressor	11.50%	16.17%	13.15%	11.36%	13.88%	12.32%	13.51%	10.70%	12.82%
CatBoost	14.07%	14.86%	12.99%	11.26%	14.04%	12.17%	13.54%	10.89%	12.98%
Ensemble	12.25%	14.97%	12.97%	11.09%	13.59%	11.99%	13.36%	10.82%	12.63%

6.13 Sonuç

Bu projede İstanbul sınırları içerisinde trafik hızının kısa ve uzun vadeli tahminini gerçekleştiren model geliştirilmiştir. Mevcut verilerde tatil günleri için zamana ve hızla bağlı grafikler ile görselleştirme sağlanmış, grafikler günlük ve haftalık şekilde ayrı ayrı incelenerek benzer bayramlar tespit edilmeye çalışılmış, benzerlik gösteren tatil günlerinin ortak özellikleri belirlenmiş ve gruplama işlemi gerçekleştirilmiştir. Tatil günleri olan milli bayram ve dini bayram günlerini trafik benzerliği olarak bizim de milli-dini olarak ayırabileceğimiz veriler incelenerek belirlenmiştir. Milli bayramların cuma veya pazartesi gününe denk gelerek 3 günlük tatil oluşturma, haftanın diğer herhangi gününe denk gelmesine göre farklı karakteristik göstermesi milli bayram tatillerini de kendi içinde iki farklı grubu ayırmamız gerektiğini göstermiştir. Dini bayram tatillerinde ise tatil günlerini hafta içine denk gelmesi ve 9 günlük tatil oluşturma diğeri günlere denk gelmesine göre farklı bir karakter oluşturmuş ve kendi içerisinde ikiye ayrılmıştır. Gruplara ayırma süreci sonrasında benzerliğin yalnızca geçmiş milli veya dini bayramlar ile olabileceği yanısıra aynı yıl içerisinde yakın tarihteki günler ille de yüksek olduğu tespit edilmiştir. Bu benzerlikler esas alınarak her bir tatil günü için ayrı eğitim günleri bulunmuştur. Verilerin görselleştirilmesi ve yorumlanması sonucu elde edilen benzer günleri tespiti sonrasında tahminleme için uygun modellerin tespiti yapılmış ve sonuç olarak Random Forest, GradientBoosting, CatBoost, LGBM, XGBRegressor modellerine projede yer verilmiştir. Hız tahmini bu 5 farklı modelin tahmin sonuçlarının ortalaması olarak esas alınmıştır. Seçilen model parametrelerinin tahmin edilen tatil gününe göre değişmesi gerektiği yapılan deneyler sonucunda tespit edilmiş ve bu doğrultuda her bir tatil günü için ayrı parametreler belirlenmiştir. Literatürde hava durumunun trafik üzerindeki etkileri üzerine birçok çalışma bulunmaktadır. Özellikle yağmurlu havalarda ıslanan zeminin trafiğe etkisi artmaktadır ve yağmurlu havaların trafiğe etkisi ile ilgili yapılan çalışmalar ilgili tarihin 3 hafta öncesine kadar olan verileri esas almaktadır. Fakat, yağmurlu gün sayısının sürekli olmaması ve sınırlandırıcı etki olması sistemin doğruluğunu olumsuz olarak etkileyebilmektedir. Bu nedenle hava durumu bu projede kullanılmamış yalnızca trafiğe olan etkisi yorumlanmıştır. Proje gerçekleştirilmesi sürecinde eksik verilerin doğru tahminlemeye olumsuz etki etmesi nedeni ile ortalama değer ile eksik verilerin doldurulması işlemi gerçekleştirilmiştir. Projenin sonraki adımlarında veri tamlığı sağlanarak ve yıl çeşitliliği artırılarak gerçeğe daha yakın değerlerde tahmin edilebileceği düşünülmüş ve planlanmıştır. Buna ek olarak gerçekleşen kazalar sonucu normalin dışında davranış gösteren trafik yoğunluğu da tahminleme sürecini olumsuz etkilemektedir. Kaza yer ve tarihleri esas alınarak eğitim ve test aşamalarına kaza zamanlarını dahil etmemenin sistemi olumlu etkileyeceği ve hata oranında azaltıcı etki göstereceği düşünülmüş ve sonraki aşamada

projeye bunun da dahil edilebileceği planlanmıştır. Gelecekteki çalışmalarda, sistem performansını artırmak için önceki paragrafta bahsedilen eksik verileri minimal hale getirilmesinin ve kaza saatleri sisteme dahil edilmemesinin yanı sıra hava durumu verisini kullanılması ve aynı lokasyon için daha geniş zamanlı veri setlerinin sisteme dahil edilmesi planlanmaktadır.

Referanslar

- [1] X. Luo, D. Li, and S. Zhang, “Traffic flow prediction during the holidays based on dft and svr,” *Journal of Sensors*, vol. 2019, 2019.
- [2] B. Lartey, A. Homaifar, A. Girma, A. Karimoddini, and D. Opoku, “Xgboost: A tree-based approach for traffic volume prediction,” in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2021, pp. 1280–1286.
- [3] N. C. Petersen, F. Rodrigues, and F. C. Pereira, “Multi-output bus travel time prediction with convolutional lstm neural network,” *Expert Systems with Applications*, vol. 120, pp. 426–435, 2019.
- [4] Ş. Yaprak and A. Akbulut, “Trafik kaza ve denetim istatistikleri,” *Polis Akademisi Yayınları*, vol. 75, pp. 8–13, 2019.
- [5] H. Al-Selwi, A. Aziz, F. Abas, N. Hamzah, and A. Mahmud, “The impact of weather data on traffic flow prediction models,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, p. 1223, Dec. 2022. doi: 10.11591/ijai.v11.i4. pp1223–1231.
- [6] UYM. “Kurbagalidere yol çalışması.” (2019), [Online]. Available: <https://uym.ibb.gov.tr/kurumsal/haberler-ve-duyurular/kurba%C4%9Fal%C4%B1dere-%C4%B1slah-projesi-kapsam%C4%B1nda-e-5-te-%C3%A7a1%C4%B1%C5%9Fma-yap%C4%B1lacak> (visited on 10/07/2019).

Özgeçmiş

BİRİNCİ ÜYE

İsim-Soyisim: Fatih ECEVİT
Doğum Tarihi ve Yeri: 09.08.2000, Yozgat
E-mail: fatih.ecevit@std.yildiz.edu.tr
Telefon: 0553 684 41 65
Staj Tecrübeleri: YazEvi Yazılım

İKİNCİ ÜYE

İsim-Soyisim: Yusuf Mert ÇELİKARSLAN
Doğum Tarihi ve Yeri: 05.09.2000, Trabzon
E-mail: mert.celikarslan@std.yildiz.edut.tr
Telefon: 0546 760 82 52
Staj Tecrübeleri: Logo Yazılım

Proje Sistem Bilgileri

Sistem ve Yazılım: Windows İşletim Sistemi, Python
Gerekli RAM: 16GB
Gerekli Disk: 20GB

ORIGINALITY REPORT



PRIMARY SOURCES

Rank	Source URL	Type	Percentage
1	www.hurriyet.com.tr	Internet Source	1 %
2	www.ncbi.nlm.nih.gov	Internet Source	<1 %
3	savoirs.usherbrooke.ca	Internet Source	<1 %
4	www.geeksforgeeks.org	Internet Source	<1 %
5	www.polmelb.net	Internet Source	<1 %
6	jumba.me	Internet Source	<1 %
7	v1.overleaf.com	Internet Source	<1 %
8	etu.edu.tr	Internet Source	<1 %
9	alphapy.readthedocs.io	Internet Source	<1 %

10	"New Trends in Intelligent Software Methodologies, Tools and Techniques", IOS Press, 2021 Publication	<1 %
11	hdl.handle.net Internet Source	<1 %
12	turkmia.net Internet Source	<1 %
13	icsc.un.org Internet Source	<1 %
14	topbigdata.es Internet Source	<1 %
15	www.dspace.yildiz.edu.tr Internet Source	<1 %
16	qiita.com Internet Source	<1 %
17	Dimitrios Vitsios, Slavé Petrovski. "Mantis-ml: Disease-Agnostic Gene Prioritization from High-Throughput Genomic Screens by Stochastic Semi-supervised Learning", The American Journal of Human Genetics, 2020 Publication	<1 %
18	sciendo.com Internet Source	<1 %

- 19 "Computational Science and Its Applications – ICCSA 2023 Workshops", Springer Science and Business Media LLC, 2023 <1 %
Publication
-
- 20 indico.ihep.ac.cn <1 %
Internet Source
-
- 21 dspace.yildiz.edu.tr <1 %
Internet Source
-
- 22 issuu.com <1 %
Internet Source
-
- 23 Noor Nayyer, Nadeem Javaid, Mariam Akbar, Abdulaziz Aldegheishem, Nabil Alrajeh, Mohsin Jamil. "A New Framework for Fraud Detection in Bitcoin Transactions through Ensemble Stacking Model in Smart Cities", IEEE Access, 2023 <1 %
Publication
-
- 24 acikbilim.yok.gov.tr <1 %
Internet Source
-
- 25 fin-ai-projects-and-resources.readthedocs.io <1 %
Internet Source
-
- 26 opac.ll.chiba-u.jp <1 %
Internet Source
-
- 27 www.hse.ru <1 %
Internet Source

Exclude quotes On

Exclude bibliography On

Exclude matches < 6 words
