

**PAMUKKALE ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ**

**MAKİNE ÖĞRENMESİ İLE KISA VADELİ TRAFİK TAHMİNİ
VE SİNYALİZASYON OPTİMİZASYONU**

LİSANS TEZİ

**Osman Doğukan URKAN
(19253502)**

Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Meriç ÇETİN

Mayıs 2022

Pamukkale Üniversitesi Bilgisayar Mühendisliği Bölümü, 19253502 numaralı Lisans Öğrencisi Osman Doğukan URKAN, ilgili yönetmeliklerin belirllediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “MAKİNE ÖĞRENMESİ İLE KISA VADELİ TRAFİK TAHMİNİ VE SİNYALİZASYON OPTİMİZASYONU” başlıklı tezini aşağıdaki imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Doç. Dr. Meriç ÇETİN**

Jüri Üyeleri :
Pamukkale Üniversitesi

.....
Pamukkale Üniversitesi

Teslim Tarihi : **Haziran 2022**
Savunma Tarihi : **Haziran 2022**

Önsöz

Çalışmalarım sırasında bilgi, tecrübe ve bağlantılarını benden esirgemeyen değerli hocam ve tez danışmanım Sayın Doç. Dr. Meriç ÇETİN'e, ulaşırma ve trafik mühendisliği alanındaki bilgileri ve yardımcıları sebebiyle Prof. Dr. Soner HALDENBİ-LEN'e, trafik verilerine ulaşabilme ve sistemi anlamama yardımcı olan Sayın Gökhan Tokmak'a ve Denizli Bel. Ulaşım Daire Başkanlığı'na sonsuz teşekkürlerimi sunarım.

Çalışmamın son halini almاسında katkıları büyük olan Sayın Ramazan ÖLMEZ'e, fikirleri ile ufkumu açan Sayın İbrahim KARAHAN'a ve tabii ki maddi manevi desteklerini esirgemeyen aileme sonsuz teşekkürlerimi sunarım.

Mayıs 2022

Osman Doğukan URKAN

İçindekiler

| | <u>Sayfa</u> |
|---|--------------|
| ÖNSÖZ | v |
| İÇİNDEKİLER | vii |
| KISALTMALAR..... | ix |
| SEMBOLLER | xi |
| ÇİZELGE LİSTESİ..... | xiii |
| ŞEKİL LİSTESİ..... | xv |
| ÖZET | xvii |
| SUMMARY | xix |
| 1. GİRİŞ | 1 |
| 1.1 Tezin Amacı..... | 1 |
| 1.2 Literatür Araştırması | 2 |
| 2. BİLİMSEL ARKAPLAN | 5 |
| 2.1 Kısa Vadeli Trafik Tahmini..... | 5 |
| 2.2 Makine Öğrenmesi | 5 |
| 2.2.1 Doğrusal Regresyon | 6 |
| 2.2.2 Lojistik Regresyon..... | 6 |
| 2.2.3 Destek Vektör Makinesi | 7 |
| 2.2.3.1 Radial Basis Function Kernel | 8 |
| 2.2.3.2 Polynomial Kernel | 9 |
| 2.2.4 Extreme Gradient Boosting | 9 |
| 2.2.4.1 Karar Ağacı..... | 9 |
| 2.2.4.2 Gradyan Artırma | 10 |
| 3. METODOLOJİ | 13 |
| 3.1 Veri Seti | 13 |
| 3.2 Ön İşleme..... | 14 |
| 3.2.1 Eksik değerlerin tamamlanması | 14 |
| 3.2.2 Kavşak ve yön özelliklerinin eklenmesi..... | 14 |
| 3.2.3 Tarih kolonunun parçalanması..... | 14 |
| 3.3 Model seçimi | 14 |
| 4. BAŞARI..... | 15 |
| 4.1 Hata Ölçütleri | 15 |
| 4.1.1 Ortalama Hata Karesi | 15 |
| 4.1.2 Karekök Ortalama Hata..... | 15 |
| 4.1.3 Ortalama Mutlak Hata | 16 |

| | |
|--|-----------|
| 5. SONUÇLAR VE DEĞERLENDİRME | 17 |
| 5.1 Uygulama | 17 |
| 5.2 Sonuçlar | 17 |
| 5.2.1 Makine Öğrenimi Modeli Sonuçları..... | 17 |
| 5.2.1.1 Destek Vektör Makinesi..... | 17 |
| 5.2.1.2 Extreme Gradient Boosting | 18 |
| 5.2.2 Hesaplanan Faz Değerlerinin Simülasyon Sonuçları | 19 |
| 5.3 Genel Değerlendirme..... | 20 |
| 5.4 Tartışma | 20 |
| KAYNAKLAR..... | 31 |
| ÖZGEÇMİŞ | 33 |

KISALTMALAR

| | |
|----------------|--|
| AUS | : Akıllı Ulaşım Sistemi |
| MÖ | : Makine Öğrenmesi |
| GB | : Gradient Boosting |
| MAE | : Mean Absolute Error |
| PMAE | : Mean Absolute Percentage Error |
| SVM | : Support Vektor Machines |
| ARIMA | : Autoregressive integrated moving average |
| SARIMA | : Seasonal Autoregressive Integrated Moving Average |
| BPNN | : Back Propagation Neural Network |
| ANN | : Artificial Neural Network |
| XGBoost | : Extreme Gradient Boosting |
| YSA | : Yapay Sinir Ağları |
| LSTM | : Long-Short Term Memory |
| TVGCN | : Time-Variant Graph Convolutional Network |
| SVR | : Support Vector Regression |
| DCRNN | : Diffusion Convolutional Recurrent Neural Network |
| ASTGCN | : Attention Based Spatial–Temporal Graph Convolutional Network |

SEMBOLLER

| | |
|---------------|--|
| Y | : Bağımlı Değişken |
| X | : Bağımsız Değişken |
| β_0 | : Doğrunun y Eksenini Kestiği Nokta |
| β_1 | : Doğrunun Eğimi |
| ε | : Hata oranı |
| p | : Karakteristik özelliğin var olma olasılığı |
| b | : Lojistik regresyon katsayılarının vektörü |
| w | : SVM Ağırlık Vektörü |
| x | : SVM Girdi Vektörü |
| σ | : RBF Varyans Değeri |
| \hat{y}_i | : Tahmin Edilen Değer |
| y_i | : Gerçek Değer |

Tablo Listesi

| | <u>Sayfa</u> |
|--|--------------|
| Tablo 3.1 : Veri kümesi..... | 13 |
| Tablo 3.2 : Ön işleme sonrası veri kümesi | 14 |
| Tablo 5.1 : SVM Sonuçları..... | 17 |
| Tablo 5.2 : Xgboost Sonuçları | 18 |
| Tablo 5.3 : AUS Verisi Tiyatro Kavşağı Trafik Simülasyonu Sonuçları | 19 |
| Tablo 5.4 : Tahmin Modeli Verisi Tiyatro Kavşağı Trafik Simülasyonu Sonuçları | 20 |

Şekil Listesi

| | <u>Sayfa</u> |
|--|--------------|
| Şekil 2.1 : Destek Vektör Makinesi [1]..... | 7 |
| Şekil 2.2 : SVM Doğru Denklemi..... | 8 |
| Şekil 2.3 : Polinomal Kernel [2] | 8 |
| Şekil 2.4 : Karar ağacı ile regresyon [3] | 10 |
| Şekil 3.1 : Tiyatro Kavşağı 1 Günlük Trafik Verisi..... | 13 |
| Şekil 5.1 : SVM Tahmin Sonuçları | 18 |
| Şekil 5.2 : Xgboost Eğitim ve Test Loss Değerleri..... | 19 |
| Şekil 5.3 : Xgboost Tiyatro Yön 1 Tahmin Değerleri..... | 21 |
| Şekil 5.4 : Xgboost Tiyatro Yön 2 Tahmin Değerleri..... | 22 |
| Şekil 5.5 : Xgboost Tiyatro Yön 3 Tahmin Değerleri..... | 23 |
| Şekil 5.6 : Xgboost Tiyatro Yön 4 Tahmin Değerleri..... | 24 |
| Şekil 5.7 : Xgboost Halley Yön 1 Tahmin Değerleri..... | 25 |
| Şekil 5.8 : Xgboost Halley Yön 2 Tahmin Değerleri..... | 26 |
| Şekil 5.9 : Xgboost Halley Yön 3 Tahmin Değerleri..... | 27 |
| Şekil 5.10 : Xgboost Halley Yön 4 Tahmin Değerleri..... | 28 |
| Şekil 5.11 : AUS Faz Değerleri İle Simülasyon Sonuçları | 29 |
| Şekil 5.12 : Tahmin Modeli Faz Değerleri İle Simülasyon Sonuçları | 30 |

MAKİNE ÖĞRENMESİ İLE KISA VADELİ TRAFİK TAHMİNİ VE SİNYALİZASYON OPTİMİZASYONU

Özet

Çalışmada Denizli Büyükşehir Belediyesi bünyesinde bulunan Ulaşım Daire Başkanlığı'nın akıllı ulaşım sisteminden alınan verilerle kısa vadeli trafik tahmini modeli geliştirilmiştir. Bu modeli geliştirmenin amacı, akıllı ulaşım sistemlerinin oluşturacak trafiği çok daha önceden tahmin edip aksiyon almalarını kolaylaştırmaktır. Sistemlerin çevrim zamanlarının kısa olması nedeniyle model karmaşıklığını ve hesap yükünü en aza indirmek için Makine Öğrenmesi metodlarından Destek Vektör Makinesi (SVM) ve Extreme Gradient Boosting kullanılmıştır(Xgboost). İki farklı kavşak, sekiz farklı yön ve otuz günlük veriler ile eğitilen modellerin hiperparametre optimizasyonları yapılmış ve modeller hız, mutlak ortalama hata(MAE), mutlak ortalama hata yüzdesi (MAPE) vb. kriterlerle karşılaştırılmıştır. Karşılaştırma sonucunda bu problem ve veri seti özelinde Extreme Gradient Boosting modelinin daha doğru olacağını sonucuna varılmıştır.

Daha sonra trafik tahmin modelinden elde edilen trafik verileri Webster yönteminde kullanılarak yeni faz süreleri elde edilmiştir. Bu faz süreleri ile AUS süreleri trafik simülasyon programı olan PTV Vissim üzerinde simüle edilip sonuçlar karşılaştırılmıştır. Simülasyon sonuçları hızlı artan veya azalan anlarda trafik tahmin modelinin kuyruklama gecikmesinde $\sim 10\%$ ile $\sim 15\%$, kavşak yoğunlığında $\sim 3\%$ ile $\sim 8\%$ azalma olduğunu göstermiştir.

Anahtar Kelimeler: Kısa Vadeli Trafik Tahmini, Makine Öğrenmesi, Gradyan Güçlendirme, Destek Vektör Makinesi, Sinyalizasyon Optimizasyonu

SHORT TERM TRAFFIC PREDICTION WITH MACHINE LEARNING AND SIGNALIZATION OPTIMIZATION

SUMMARY

In this study, a short-term traffic forecast model was developed with the data obtained from the smart transportation system of the Transportation Department within the Denizli Metropolitan Municipality. The purpose of developing this model is to make it easier for smart transportation systems to predict the traffic that will occur much earlier and take action. Due to the short cycle times of the systems, Machine Learning methods Support Vector Machine (SVM) and Extreme Gradient Boosting (Xgboost) were used to minimize model complexity and computational load. Hyperparameter optimizations of the models trained with two different intersections, eight different directions and thirty-day data were made, and the models were analyzed for speed, absolute mean error (MAE), absolute mean error percentage (MAPE), etc. compared with the criteria. As a result of the comparison, it was concluded that the Extreme Gradient Boosting model would be more accurate for this problem and the data set.

Then, new phase times were obtained by using the traffic data obtained from the traffic prediction model in the Webster method. These phase times and AUS times were simulated on the traffic simulation program PTV Vissim and the results were compared. The simulation results showed that the queuing delay of the traffic prediction model decreased by $\sim 10\%$ to $\sim 15\%$ and the intersection density was reduced by $\sim \%3$ and $\sim \%8$ at fast increasing or decreasing moments.

Keywords: Short Term Traffic Prediction, Machine Learning, Gradient Boosting, Support Vector Machine, Signalization Optimization

1. GİRİŞ

Dünya nüfusunun hızla artması, gelir düzeylerinin yükselmesi, üretimdeki iyileştirmeler ve teknolojinin hızla yaygınlaşmasının sonucunda trafikteki araç sayısı hızla artmaktadır [4]. Özellikle şehir merkezlerindeki trafik yoğunluğu problemi oldukça ciddi bir konu haline gelmiştir. Seyahat süresinin artması kişilerin psikolojisini kötü yönde etkilemeye ve zaman kaybını da artırmaktadır. Bunların yanı sıra araçların trafikte geçirdikleri zaman, kirliliği ve küresel ısınmayı da beslemektedir [5]. Bu sorunların tümü göz önüne alındığında Akıllı Ulaşım Sistemleri (AUS) her zamankinden daha önemli bir hal almaktadır.

Akıllı ulaşım sistemleri; trafik güvenliğinin artması, enerji ve zamanın verimli kullanılması, trafikte geçirilen zamanın azaltılması ve benzeri amaçlar için geliştirilmiş sistemlerdir. Bu sistemler; trafik verilerini inceleme, analiz etme ve müdahale etme gibi görevleri yerine getirirler. Dünyada iyi örnekleri bulunan bu sistemler trafik akışında önemli iyileştirmeler yapabilmektedir [6]. Fakat birçok akıllı ulaşım sistemi oluşacak trafiği, trafik olmadan ancak birkaç dakika önceden öngörebilmektedir. Bu sebeple bazı durumlarda refleks göstermekte yetersiz kalabilmektedirler. Böylece kısa vadeli trafik tahmini çok daha önemli bir hal almaktadır. Başarılı bir tahmin ve gerçek verilerin hibrit bir şekilde kullanılması akıllı ulaşım sistemlerinin verimini de artıracaktır.

1.1 Tezin Amacı

Bu tez çalışmasının amacı, Denizli Büyükşehir Belediyesi bünyesinde bulunan Ulaşım Daire Başkanlığı tarafından sağlanan trafik verilerini kullanarak tahmin modeli geliştirmek ve bu tahmin sonuçları üzerinden sinyalizasyon iyileştirmesi yapmaktadır. Tez çalışması süresince izlenecek adımlar şu şekildedir:

- Veri setinin ön işlenmesi(eksik veriyi tamamlama, temizleme, ayristırma vb.)
- Seçilen Makine Öğrenmesi (MÖ) modellerinin oluşturulması
- Modellerin test edilmesi ve iyileştirilmesi (tuning)

- Sonuçların görselleştirilmesi
- Sinyalizasyon iyileştirmesi için model seçimi
- İyileştirme sonrası trafiğin simülasyonu
- Sonuçların değerlendirilmesi

1.2 Literatür Araştırması

Kısa vadeli trafik tahmini alanında Makine Öğrenmesi (SVM, Doğrusal Regresyon vb.), Zaman serileri (ARIMA, Kalman Filtreleme vb.), Derin Öğrenme (KNN vb.), Sinir Ağları (BPNN, ANN vb.) gibi çeşitli yöntemlerle günümüze kadar birçok çalışma yapılmış olup, günümüzde de çalışmalar devam etmektedir. Kullanılan veri setlerinin çeşitlilikleri ve problemin karmaşıklığı nedeniyle en iyi yöntemden bahsetmek oldukça zordur. Son yıllarda Destek Vektör Makinesi (SVM), Gradyan Güçlendirme (Gradient Boosting) gibi çeşitli Makine Öğrenmesi teknikleri başarılı sonuçlar üretmektedir. Geçmişten günümüze yapılan araştırmaların zamanla klasik yöntemler ve zaman serisi yöntemlerinden Yapay Sinir Ağları ve Makine Öğrenmesi gibi yöntemlere evrildiği görülmüştür. Literatürde Yapay Sinir Ağları ve Derin Öğrenme metodlarının sıkça kullanıldığı görülmüştür. Çalışmada kullanılacak Denizli Ulaşım Daire Başkanlığı verilerinin çevrim süresinin kısa olması sebebiyle karmaşık modellerden kaçınmak, hesap yükünü azaltmak ve literatüre katkı sağlamak için Makine Öğrenmesi modelleri kullanılmasına karar verilmiştir. 1980 yılında Nihan ve Nancy L. [7] tarafından yazılan bir makalede dokuz yıllık trafik verisi üzerinde Box and Jenkins zaman serisi modeli kullanılmıştır. Ortalamada %5 gibi bir hata oranıyla gayet başarılı sonuçlar elde edilmiştir. 1991 yılında Amerikan İnşaat Mühendisleri Birliği üyelerinden Davis ve diğ. [8] tarafından yazılan bir makalede ise zaman serileri ile YSA yöntemleri karşılaştırılmıştır. Çalışma sonuçları YSA yöntemi olan K-En Yakın komşular (KNN) yönteminin Box and Jenkins yöntemine göre daha iyi sonuçlar ürettiği anlaşılmıştır. 1994 yılında Amerika'da yapılan bir araştırmada ise, YSA ve K-En Yakın Komşular modelleri arasında bir kıyaslama yapılmıştır. Smith ve diğ. [9] yaptıkları araştırmada Telegraph Road lokasyonu verileri için KNN modelinin daha iyi sonuçlar verdiği, Woodrow Wilson köprüsü lokasyonu verileri için ise YSA

modelinin daha iyi sonuçlar verdiği göstermiştir. YSA modelinin büyük hata oranlarının çok daha fazla olduğu görülmektedir.

Hinsbergen ve diğ. [10] tarafından 2007 yılında yapılan çalışmada Naïve, Parametrik ve Parametrik olmayan modeller üzerinde çalışılmıştır. Çalışma sonuçlarında Bütünleşik Otoregresif Hareketli Ortalama (ARIMA) gibi parametrik modellerin bazı durumlarda iyi sonuçlar verirken bazı durumlarda sonuçların beklenenin altında olduğu görülmüştür. Aynı çalışmada Sinir Ağlarının bu problem özelinde en çok kullanılan görece de iyi sonuçlar veren model olduğu sonucuna varılmıştır.

Dong ve diğ. [11] tarafından yapılan ve devam eden bir araştırmada ise Destek Vektör Makinesi modeli ile eXtreme Gradient Boosting (XGBoost) modeli arasında bir karşılaştırma yapılmıştır. Devam eden çalışmanın 2018 yılındaki sonuçlarına göre XGBoost modelinin SVM modeline göre daha hızlı çalıştığı ve daha doğru sonuçlar ürettiği anlaşılmıştır.

Yiğit ve Haldenbilen [12] tarafından yapılan çalışmada ise Yapay Sinir Ağları (YSA) modelinin ARIMA modelinden daha başarılı olduğu gözlemlenmiştir. Çalışmada haftanın 3 gününün belli saatlerinde alınan veriler kullanılmış olup modellerin farklı parametrelerde farklı sonuçlar verebileceğinden bahsedilmiştir.

Sun ve diğ. [13] tarafından yapılan araştırmada Çin'de bulunan bir otoyoldan toplanan 80 günlük veri seti kullanılmıştır. Çalışmada Evrişimli Sinir Ağları (CNN), Uzun Kısa Süreli Bellek (LSTM), Mevsimsel Otoregresif Entegre Hareketli Ortalama (SARIMA) ve XGBoost modelleri test edilmiştir. SARIMA hariç XGBoost'un geleneksel tahmin yöntemlerinden daha iyi performans gösterdiği sonucuna varılmıştır. Fakat çalışma gösteriyor ki XGBoost modeli SARIMA modelinden çok daha hızlı çalışmaktadır. Tüm bu sebeplerle XGBoost modelinin çalışmada kullanılan modellerin arasında en iyi seçenek olduğu kabul edilmiştir.

Wang ve diğ. [14] tarafından geliştirilen Time-Variant Graph Convolutional Network (TVGCN) üç farklı gerçek zamanlı veri seti üzerinde test edildiğinde benzer modellere göre çok daha iyi sonuçlar vermiştir. Graph Theory kullanılan model Support Vector Regression (SVR), LSTM gibi modellerin yanı sıra, Diffusion Convolutional Recurrent Neural Network (DCRNN), Attention Based Spatial–Temporal Graph Convolutional Network (ASTGCN) gibi yapay sinir ağları modellerine oranla da

oldukça düşük hata oranlarına sahip tahminler yapmıştır. Test sonuçlarında SVR: %19, LSTM: %17, DCRNN: %14, ASTGCN: % 16, TVGCN: %12 ortalama mutlak hata yüzdesinde hatalar yapmıştır.

2. BİLİMSEL ARKAPLAN

2.1 Kısa Vadeli Trafik Tahmini

Kısa vadeli trafik tahmini 30 yılı aşkın süredir ulaşırma araştırmalarının birçok alanında kullanılmıştır ve kullanılmaya devam edilmektedir. Sürücülere doğru bilgilerin verilmesi, sinyalizasyon optimizasyonu, yol fizibilite çalışmaları vb. alanlarda oldukça sık kullanılmaktadır. Veri analizi ve hesaplama gücünün de artmasıyla popülerliği her geçen gün artmaktadır. En geniş tanımı ile kısa vadeli trafik tahmini, verilerin analizi ve çeşitli modeller (Makine Öğrenmesi, Derin öğrenme, Zaman serileri vb.) kullanılarak işlenmesi sonucunda birkaç saniye ile birkaç saat arasında değişen bir periyotta oluşacak trafiğin tahmin edilmesidir.

2.2 Makine Öğrenmesi

Makine Öğrenmesi (MÖ), genel olarak optimizasyon veya doğru tahminler yapmak için deneyimi kullanan 1950'lerde ortaya çıkışmış hesaplama yöntemleri olarak tanımlanabilir. Deneyim ise toplanan ve analiz edilmek için dijitalize edilmiş, öğrenecek bilgisayar için geçmiş veriyi ifade eder. Bu verilerin kalitesi, MÖ modellerinin başarısı için oldukça önemlidir [15]. Makine öğrenmesi öğrenme biçimine göre üç başlıkta toplanabilir. Bunlar;

- Gözetimli Öğrenme (Supervised Learning)
- Gözetimsiz Öğrenme (Unsupervised Learning)
- Pekiştirmeli Öğrenme (Reinforcement learning)

Makine Öğrenmesi ile çözülebilecek problem türleri;

- Metin ve Belge Sınıflandırması
- Doğal Dil İşleme
- Bilgisayarlı Görü

- Hesaplamlı Biyoloji
- Kısa Vadeli Trafik Tahmini vb.

2.2.1 Doğrusal Regresyon

Doğrusal regresyon , tahmin edilmek istenilen değişken(bağımlı değişken) ile onu etkileyen değişkenler(bağımsız değişkenler) arasındaki ilişkinin ve bağımsız değişkenlerin bağımlı değişken üzerindeki etkilerinin belirlenmesinde kullanılır. Basit doğrusal regresyon analizi değişkenler arasındaki ilişkinin yapısı ve derecesi ile ilgilenmektedir. Denklem 2.1'de matematiksel modeli görülmektedir.

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.1)$$

Burada;

Y : Bağımlı değişken

X : Bağımsız değişken

β_0 : Doğrunun y ekseni kestiği noktası

β_1 : Doğrunun eğimi

ε : Hata oranıdır.

Doğrusal regresyon model karmaşıklığı gørece az olsa da doğrusal veri setlerinde oldukça başarılı ve hızlı sonuçlar üretebilmektedir. Bunun yanında aykırı değerler ve aşırı öğrenme modeli kötü etkileyebilir.

2.2.2 Lojistik Regresyon

Lojistik regresyon, bağımlı ve bağımsız değişkenler arasında doğrusal bir ilişki varsayımdan sınıflandırma işlemi yapmaya yarayan bir regresyon yöntemidir. Doğrusal bir ilişki varsayımlı yerine açıklayıcı değişkenlerin logitleri ile yanıt arasında doğrusal ilişki olduğunu varsayar. Parametreleri tahmin etmek için sıradan en küçük kareler yerine maksimum olasılık tahmini kullanır ve bu nedenle büyük örneklem yaklaşımlarına dayanır. Lojistik Regresyon analizinin merkezinde, bir olayın log oranını tahmin eden görev bulunur. Matematiksel olarak, lojistik regresyon aşağıdaki gibi tanımlanmış çoklu doğrusal regresyon fonksiyonunu tahmin eder:

$$\text{logit}(p) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k \quad (2.2)$$

Denklem 2.2' de görüleceği üzere p ; karakteristik özelliğinin var olma olasılığıdır.

$$Possibility = (p)/(1 - p) \quad (2.3)$$

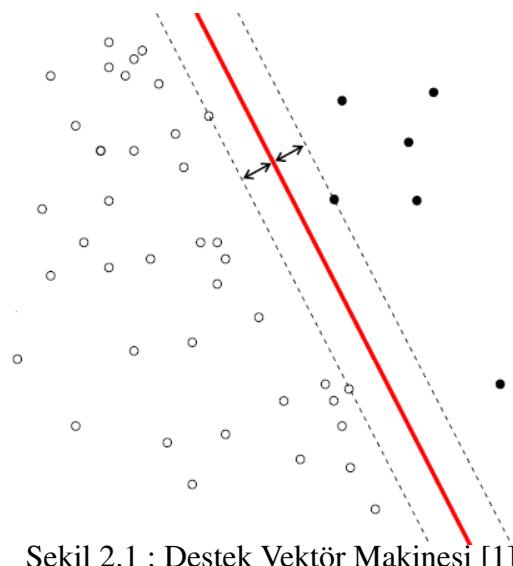
Denklem 2.3' de görüleceği üzere; olasılık, karakteristik özelliğin var olma olasılığının, karakteristik özelliğin var olmama olasılığına bölünmesi ile elde edilir.

$$\text{logit}(p) = \ln \frac{p}{1 - p} \quad (2.4)$$

Karekök hataların toplamını en aza indirgeyen parametreleri seçmek yerine (doğrusal regresyon gibi), lojistik regresyonda tahmin, örnek değerlerin gözlem olasılığını en yükseğe çıkarıyan parametreleri seçer.

2.2.3 Destek Vektör Makinesi

Destek vektör makineleri özellikle sınıflandırma problemlerinde küçük ve orta büyüklükteki verileri sınıflandırmak için kullanılan denetimli makine öğrenmesi algoritmasıdır. Temelde, veri düzlemindeki farklı sınıflara ait noktaları sınıflandırmak için doğrular oluşturur ve bu doğruların farklı sınıflardaki noktalara olan uzaklıklarının maksimum olmasını amaçlar.



Şekil 2.1 : Destek Vektör Makinesi [1]

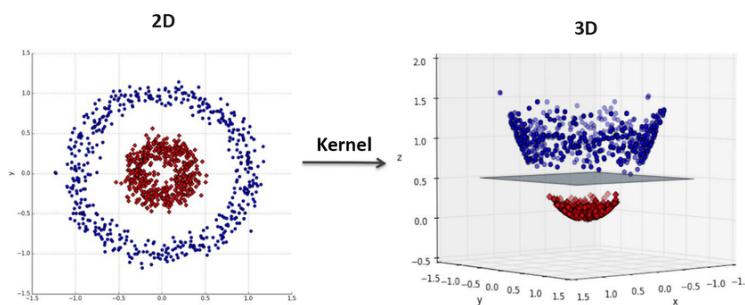
Şekil 2.1'de iki sınıfı ayıran bir doğru görülmektedir. Doğru ile sınıflar arasında kalan alana marjin adı verilir ve bu alan ne kadar büyükse sınıflandırma o kadar başarılı olmuş olur.

$$\hat{y} = \begin{cases} 0 & \text{if } \mathbf{w}^T \cdot \mathbf{x} + b < 0, \\ 1 & \text{if } \mathbf{w}^T \cdot \mathbf{x} + b \geq 0 \end{cases}$$

Şekil 2.2 : SVM Doğru Denklemi

Şekil 2.2'de görüldüğü gibi w ; ağırlık vektörü, x ; girdi vektörü, b ; sapmadır. Yeni bir değer için çıkan sonuç 0'dan küçükse, siyah noktalara daha yakın olacaktır. Eğer, çıkan sonuç 0'a eşit veya büyükse, bu durumda beyaz noktalara daha yakın olacaktır. Her veri seti iki boyutlu uzayda sınıflandırılamayabilir. Böyle durumlarda boyut artırmak işlem yükünü artıracacağı için bunun yerine farklı çekirdek (kernel)'ler kullanılır. Bunlardan bazıları;

- Polynomial Kernel
- Gaussian RBF (Radial Basis Function) Kernel
- Sigmoid Kernel vb.



Şekil 2.3 : Polinomal Kernel [2]

Şekil 2.3'de Polinomal Kernel kullanıldıktan sonraki veri uzayı görülmektedir.

2.2.3.1 Radial Basis Function Kernel

RBF kerneller gauss dağılımına benzerliği ve en fazla genelleştirilmiş kernel olması sebebiyle en çok kullanılanlardan birisidir. Temelde iki noktanın birbirine ne kadar benzediğini veya aralarındaki uzaklıği hesaplar. Matematiksel olarak şu şekilde gösterili;

$$K(X_1, X_2) = \exp\left(-\frac{\|X_1 - X_2\|^2}{2\sigma^2}\right) \quad (2.5)$$

Denklemde görülen;

σ : varyans değeri

$||X_1 - X_2||^2$: iki nokta arasındaki uzaklıktır.

2.2.3.2 Polynomial Kernel

Polynomial kernel benzerlik kestirimi için yalnızca verilen girdilerin özelliklerine değil bunların kombinasyonlarına da bakar. Regresyon analizinde bu tür kombinasyonlar etkileşim özellikleri olarak bilinir. Matematiksel olarak şu şekilde ifade edilir;

$$K(x, y) = (x^T y + c)^d \quad (2.6)$$

Denklemde;

x, y : Öznitelik vektörleri

c : Hiperparametre

d : Denklem derecesi

2.2.4 Extreme Gradient Boosting

Extreme Gradient Boosting (xgboost), karar ağacı temelli ve eğim artırmalı bir makine öğrenmesi algoritmasıdır. Son yıllarda popülerliği artan model, oldukça iyi sonuçlar üretirken aynı zamanda da karar ağaçlarını paralel bir şekilde oluşturduğu için benzer modellerden daha hızlı çalışmaktadır.

2.2.4.1 Karar Ağacı

Karar Ağacı, kök, karar düğümleri ve yaprak düğümlerden oluşur. Ağaç yapısı şeklinde model oluşturan bir sınıflandırma yöntemidir. Temelde algoritma, etiketli eğitim verilerini ve tümevarım yöntemini kullanarak karar yapıları oluşturur. Daha sonra bu yapıları kullanarak verinin hangi sınıfı ait olduğuna karar verir. Modeli en çok etkileyen parametreler ise;

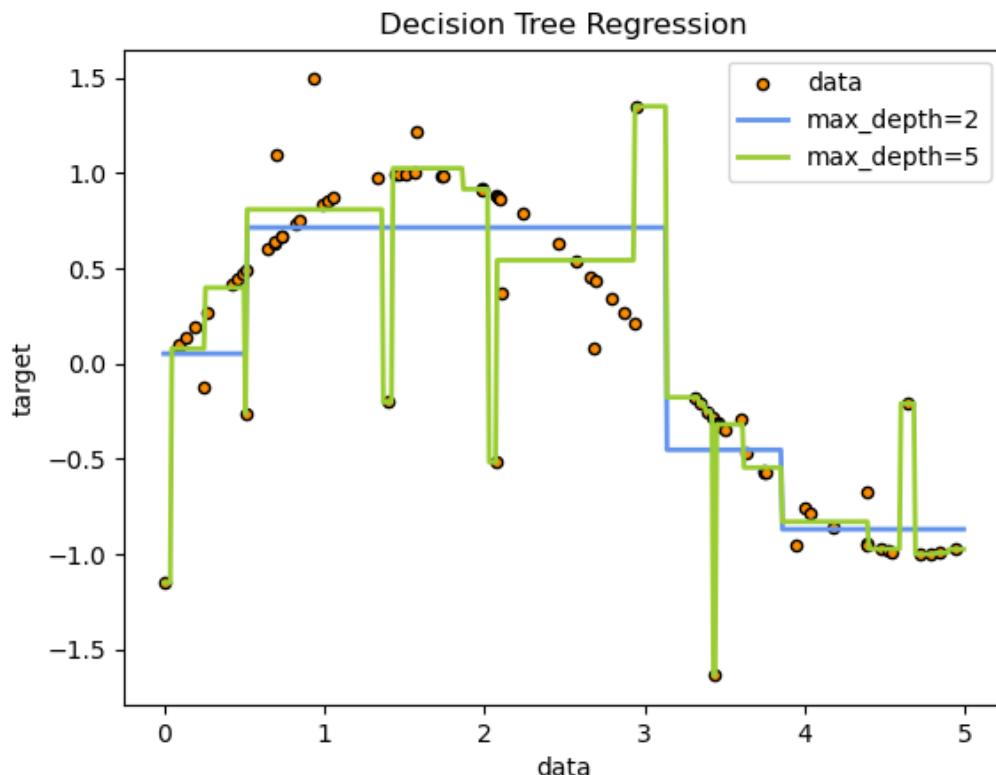
Alt örnek(Subsample) : Eğitim verisinin ne kadar oranla eğitimde kullanılacağı

Örnek kolon(Colsample) : Eğitim verisindeki kolonların ne kadarının eğitimde kullanılacağı

Maksimum Derinlik : Karar ağacının derinliği

Minimum çocuk ağırlığı: Bir düğümdeki çocukların alabileceği en düşük ağırlıktır.

Şekil 2.4'te iki farklı karar ağacı ile yapılan regresyon sonucu görülmektedir. Bu



Şekil 2.4 : Karar ağacı ile regresyon [3]

sonuçlardan maksimum derinliği daha fazla olan model aykırı değerleri daha iyi yakalayabilirken bu durum model karmaşıklığını artttırduğu gibi aşırı öğrenmeye de yol açabilmektedir.

2.2.4.2 Gradyan Artırma

Gradyan artırma, karar ağaçları gibi zayıf tahmin modelleri topluluğu şeklinde bir tahmin modeli üreten, regresyon , sınıflandırma ve diğer görevler için kullanılan bir makine öğrenimi teknigidir. Diğer artırma yöntemleri gibi, gradyan artırma da zayıf öğrencileri yinelemeli bir şekilde tek bir güçlü öğrencide birleştirir. Matematiksel model şu şekilde açıklanabilir;

$$y = \frac{1}{n} \sum_i (\hat{y}_i - y_i)^2 \quad (2.7)$$

Burada;

\hat{y}_i : Tahmin edilen değer

y_i : Gerçek değer

n : Toplam örnek sayısıdır.

Çoğu gözetimli makine öğrenmesi algoritmasında olduğu gibi burada da amaç giriş değerlerinden çıkış değerlerini tahmin edebilmektir. Her adımda model kendinden önceki hatayı tahmin eder geliştirir ve bir sonraki adıma iletir.

3. METODOLOJİ

Bu bölümde tez çalışmasında kullanılan veri seti, metodlar, teknolojiler hakkında detaylı bilgi verilmektedir.

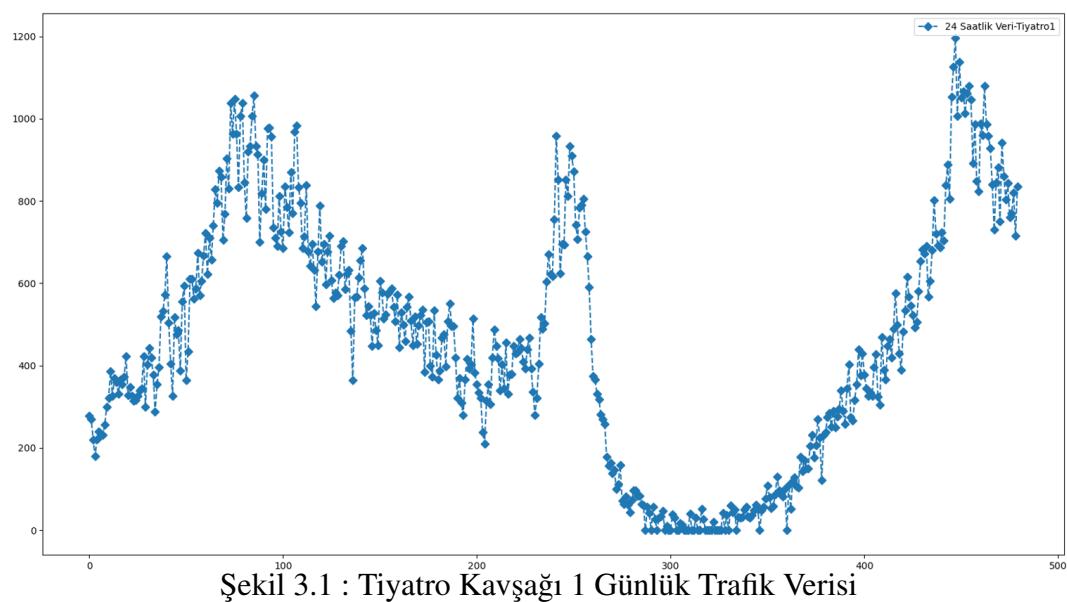
3.1 Veri Seti

Çalışmada, Denizli Büyükşehir Belediyesi bünyesinde bulunan Ulaşım Daire Başkanlığına ait trafik verileri kullanılmıştır. Daire başlanlığının kullanmış olduğu akıllı ulaşım sistemine bağlı 165 kavşak bulunmaktadır. Çalışmada 165 kavşak arasından yüksek yoğunluğu olan ve stratejik konumu sebebiyle Tiyatro Kavşağı ve Halley Kavşağı seçilmiştir. Veri kümesi 05/09/2021 ile 30/09/2021 tarihleri arasındaki 27 günden oluşmaktadır ve verinin oluşturulduğu tarih, saatlik araç yoğunluğu ve sinyalizasyon bilgilerini içermektedir.

Tablo 3.1 : Veri kümesi

| Kavşak adı | Yön sayısı | Veri sayısı |
|------------|------------|-------------|
| Tiyatro | 4 | 40000 |
| Halley | 4 | 40000 |

Şekil 3.1'de Tiyatro kavşağı lise geliş yönüne ait bir günlük trafik verisi görülmektedir.



3.2 Ön İşleme

Veri setinin Ulaşım Daire Başkanlığından alındığı haliyle oldukça temiz olmasına karşın işlenebilmesi veya bir model eğitilmesi mümkün değildir. Bu sebeple üzerinde çalışılmadan önce bazı işlemler yapılması gerekti.

3.2.1 Eksik değerlerin tamamlanması

Veri setinin içerisinde az sayıda eksik değer mevcuttu. Model eğitimini kötü yönde etkilememesi adına bu değerler bulundukları saat aralıklarının ortalaması alınarak dolduruldu. Böylece aykırı değer oluşumunun önüne geçilmiş oldu.

3.2.2 Kavşak ve yön özelliklerinin eklenmesi

Veri setinde her bir kavşağa ait her bir yön farklı veri setleri içerisinde tutulmaktadır. Model eğitiminde tüm verileri bir arada kullanmak istediğim için verileri birleştirmem ve hangi verinin hangi kavşağa ve yöné ait olduğunu bilgisini eklemem gerekti.

3.2.3 Tarih kolonunun parçalanması

Verilerin oluşturulma tarihi tek bir kolonda AA/GG SA/DK/SN şeklindeydi ve bu haliyle üzerinde işlem yapılamazdı. Bu sebeple bu bilgiyi ay, gün, saat ve dakika olmak üzere dört farklı kolona böldüm. Böylece verinin son hali 3.2'de görülmektedir.

Tablo 3.2 : Ön işleme sonrası veri kümesi

| Veri Etiketi | Yön | Ay | Gün | Saat | Dakika | Değer |
|--------------|-----|-------|-------|-------|--------|-------|
| 2 | 8 | 80000 | 80000 | 80000 | 80000 | 80000 |

3.3 Model seçimi

Literatürde, kısa vadeli trafik tahmini problemi için kullanılan çok çeşitli algoritmalar bulunmaktadır. Bunlardan büyük çoğunluğu derin öğrenme ve yapay sinir ağlarıdır. Hem halihazırda literatürde çok fazla bulundukları için hem de Denizli'de kullanılan AUS'nin çevirim süresi oldukça kısa olduğu için bu karmaşık modelleri kullanmak hesaplama güçlüklerine yol açabilecektir. Bu sebepler ile literatürde başarılarını kanıtlamış olan Destek Vektör Makinesi ve Gradyan Artırma algoritmalarının kullanılmasına karar verilmiştir.

4. BAŞARI

Makine öğrenmesi modellerinin başarılarını ölçmek için farklı yollar bulunmaktadır. Sınıflandırma bir sınıfı ait etiketi tahmin etmeyi amaçlarken regresyon ise sayısal bir değeri tahmin etmeyi amaçlar. Bu sebeple sınıflandırma modellerinde doğruluk, doğru sınıflandırılmış verilerin tüm verilere bölümü ile bulunabilirken regresyonda belirli hata metrikleri kullanmak gerekmektedir.

4.1 Hata Ölçütleri

4.1.1 Ortalama Hata Karesi

Ortalama hata karesi(MSE) regresyon problemlerinde çokça kullanılan bir hata metriğidir. MSE, bir veri kümesindeki tahmin edilen ve gerçek değerler arasındaki farkın karesinin veri sayısına bölümü ile hesaplanır.

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

Denklem 4.1'de;

MSE : Ortalama Hata Karesi

y_i : Gerçek değer

\hat{y}_i : Tahmin edilen değerdir.

Gerçek değer ile tahmin edilen değer arasındaki farkın karesi alınması sebebiyle hata değeri her zaman pozitif çıkmaktadır. Ayrıca bu yöntem hata değerini şişirme eğilimindedir. Aynı zamanda büyük hataların metrik üzerinde etkisi artırılmak istendiğinde oldukça kullanışlıdır.

4.1.2 Karekök Ortalama Hata

Karekök ortalama hata(RMSE) MSE'nin bir uzantısıdır. Aralarındaki fark ise hatanın karesi alındıktan sonra karekökünün alınmasıdır. Böylece hatalar şişirilmeden olduğu gibi kullanılır. RMSE metriğinde de MSE'de olduğu gibi hata her zaman pozitiftir.

Matematiksel formülasyonu 4.2'de verilmiştir.

$$RMSE = \sqrt{\left(\frac{1}{N} \sum_1^N (y_i - \hat{y}_i)^2\right)}, \quad RMSE = \sqrt{MSE} \quad (4.2)$$

4.1.3 Ortalama Mutlak Hata

Ortalama mutlak hata(MAE), RMSE gibi hata puanının birimleri tahmin edilen hedef değerin birimleriyle eşleşir. RMSE'den farklı olarak MAE'deki değişimler doğrusaldır ve bu sebeple sezgisel olarak sınıflandırılır.

$$MAE = \frac{\sum_1^N |y_i - \hat{y}_i|}{N} \quad (4.3)$$

5. SONUÇLAR VE DEĞERLENDİRME

5.1 Uygulama

Çalışmada, Denizli Büyükşehir Belediyesi trafik verileri kullanılarak kısa vadeli trafik tahmin modeli geliştirilmiştir. Veri setini model eğitimine hazır hale getirebilmek için eksik verilerin tamamlanması, verilerin normalizasyonu, yeni özniteliklerin eklenmesi gibi ön işlemler yapılmıştır. Tahmin modelinden elde edilen trafik verileri faz süresi hesaplama modeli olan Webster (İngiliz) modeli kullanılarak yeni değerler hesaplanmıştır. Son işlem olarak hesaplanan faz süreleri trafik akış simülasyonu olan PTV Vissim'de simüle edilmiştir.

5.2 Sonuçlar

Çalışmanın sonuçları iki başlık altında inceleneciktir. Bunlar;

- Makine Öğrenimi Modeli Sonuçları
- Hesaplanan Faz Değerlerinin Simülasyon Sonuçları

5.2.1 Makine Öğrenimi Modeli Sonuçları

5.2.1.1 Destek Vektör Makinesi

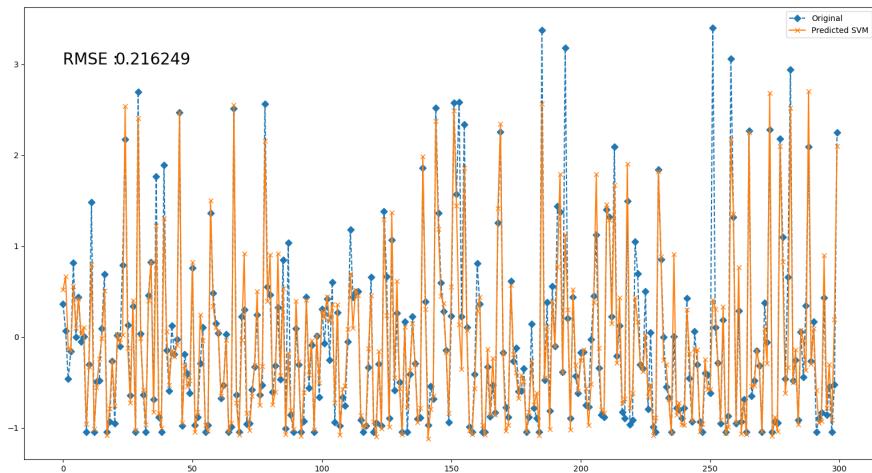
Destek vektör makinesi modeli, eğitim ve tahmin sürelerinin uzunluğu ve test veri seti üzerindeki doğruluğu ile beklenenden daha düşük bir performans göstermiştir. Tablo

Tablo 5.1 : SVM Sonuçları

| Model | Eğitim Süresi(saat) | Tahmin Süresi(saniye) | Doğruluk(%) |
|-------|---------------------|-----------------------|-------------|
| SVM | ~ 7.5 | ~ 79.2 | ~ 82.6 |

5.1'de modele ait sonuçlar görülmektedir.

Test veri setine ait tahmin değerleri ise şekil 5.1'de görülmektedir.



Şekil 5.1 : SVM Tahmin Sonuçları

5.2.1.2 Extreme Gradient Boosting

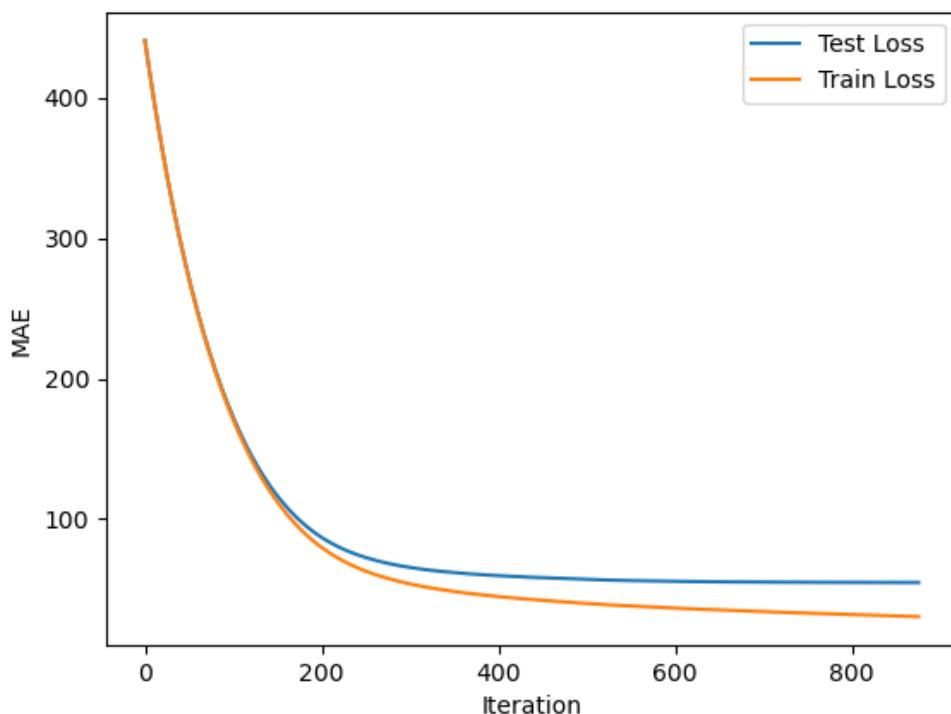
Xgboost gerek eğitim ve tahmin süreleri ile gerek tahmin doğruluğu ile oldukça iyi sonuçlar üretmiştir. Çevrim süresinin kısıtlı olduğu akıllı ulaşım sistemlerinde kullanılmaya uygun görülmüştür.

Tablo 5.2 : Xgboost Sonuçları

| Model | Eğitim Süresi(saat) | Tahmin Süresi(saniye) | Doğruluk(%) |
|---------|---------------------|-----------------------|-------------|
| Xgboost | ~ 0.2 | ~ 0.003 | ~ 92 |

Modele ait sonuçlar tablo 5.2'de görülmektedir.

Şekil 5.2'de de modele ait eğitim ve test kayıpları görülmektedir.



Şekil 5.2 : Xgboost Eğitim ve Test Loss Değerleri

5.2.2 Hesaplanan Faz Değerlerinin Simülasyon Sonuçları

Çalışmada üç farklı hesaplama yönteminden Webster yöntemi seçilmiştir. Bunun sebebi, tüm Denizli'de faz sürelerinin bu yöntem ile hesaplanmasıdır. Trafiğin hızlı arttığı veya hızlı azaldığı anlarda, AUS'nın sinyalizasyon verisi ile trafik tahmin modelinden elde edilen yoğunluğun Webster yöntemi ile hesaplanması sonucunda elde edilen sinyalizasyon verisi simülasyonda test edilmiştir.

Simülasyon sonuçları tablo 5.11 ve tablo 5.12'de görülmektedir.

| Yön | Yeşil Süresi(sn) | Gecikme(sn) | Ort Hız(km/s) | Doluluk(%) |
|-----|------------------|-------------|---------------|------------|
| 1 | 18 | 33.70 | 41.50 | 0.32 |
| 2 | 28 | 26.45 | 45.20 | 2.14 |
| 3 | 17 | 41.60 | 45.67 | 0.06 |
| 4 | 31 | 28.36 | 46 | 4.87 |

Tablo 5.3 : AUS Verisi Tiyatro Kavşağı Trafik Simülasyonu Sonuçları

| Yön | Yeşil Süresi(sn) | Gecikme(sn) | Ort Hız(km/s) | Doluluk(%) |
|-----|------------------|-------------|---------------|------------|
| 1 | 19 | 33.65 | 43 | 0.25 |
| 2 | 22 | 31.85 | 45.3 | 1.78 |
| 3 | 12 | 26.30 | 40.75 | 0.13 |
| 4 | 41 | 19.88 | 46.6 | 4.63 |

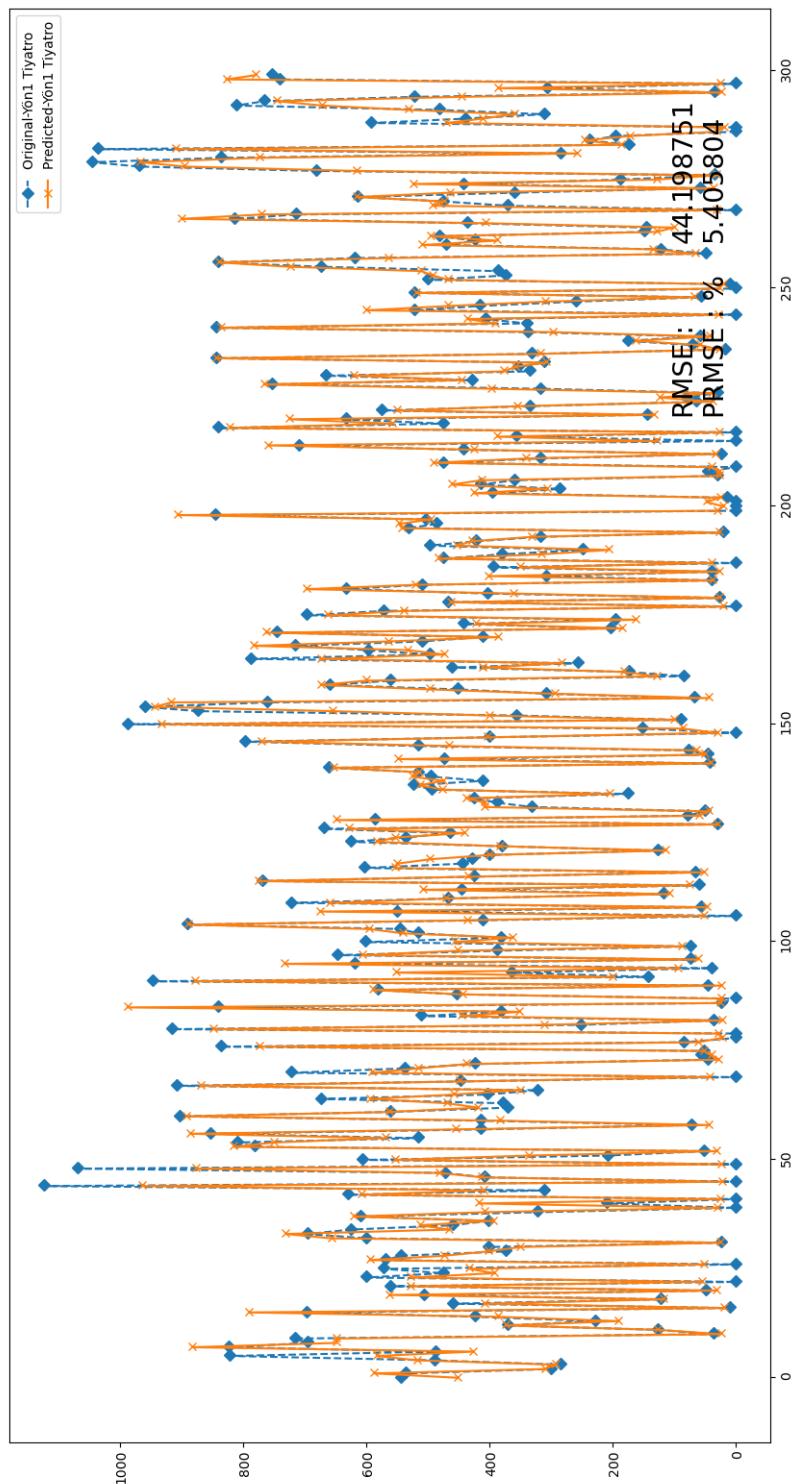
Tablo 5.4 : Tahmin Modeli Verisi Tiyatro Kavşağı Trafik Simülasyonu Sonuçları

5.3 Genel Değerlendirme

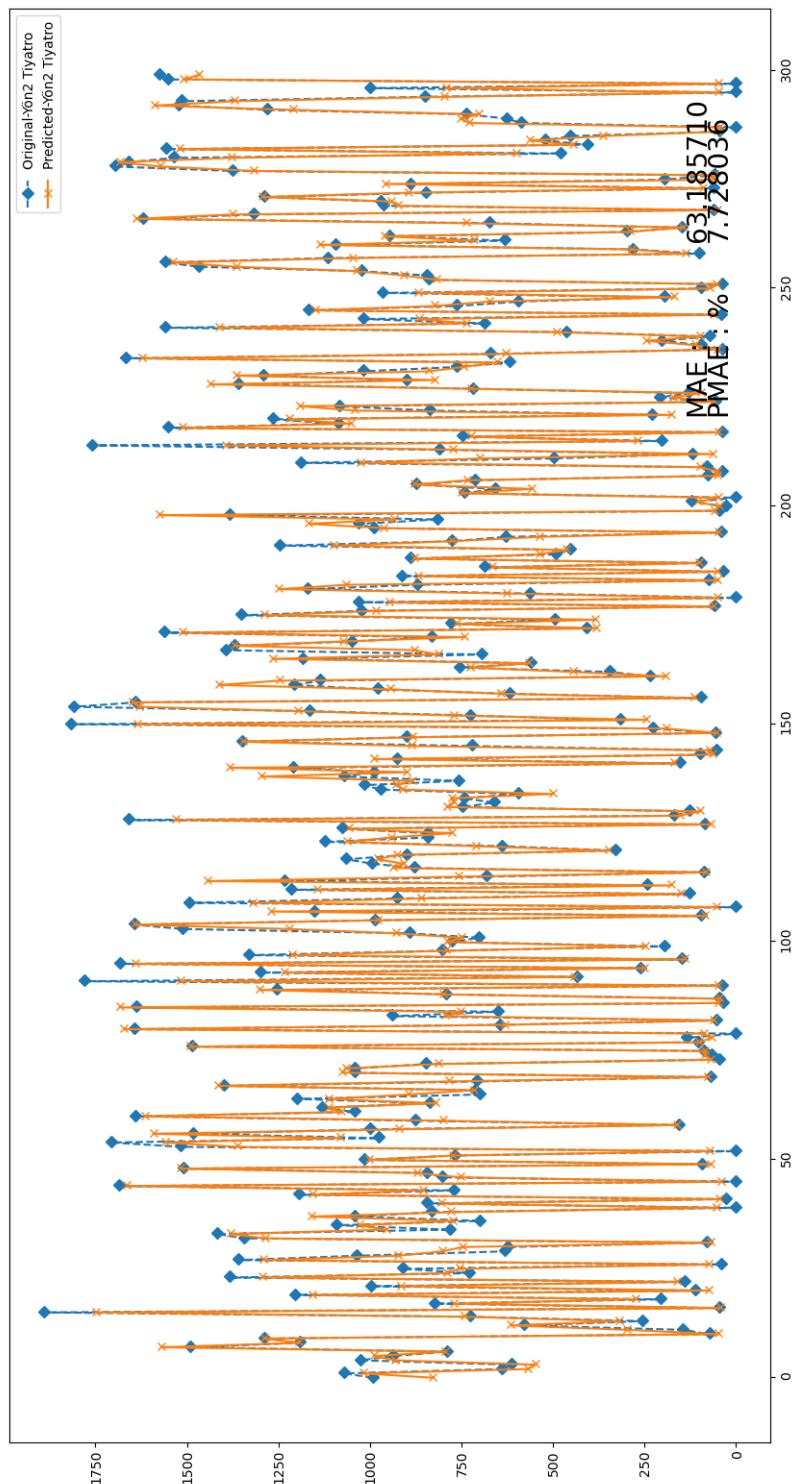
Simülasyon sonuçları hızlı artan veya azalan anlarda trafik tahmin modelinin kuyruklama gecikmesinde $\sim 10\%$ ile $\sim 15\%$, kavşak yoğunlığında $\sim 3\%$ ile $\sim 8\%$ azalma olduğunu göstermiştir. Bu iyileştirmeler değişimin az olduğu saatlerde $\sim 2\%$ ile $\sim 5\%$ arasında kalmıştır. Sisteme bağlı 164 kavşak olduğu göz önüne alınırsa kayda değer sonuçlar elde edilmiştir.

5.4 Tartışma

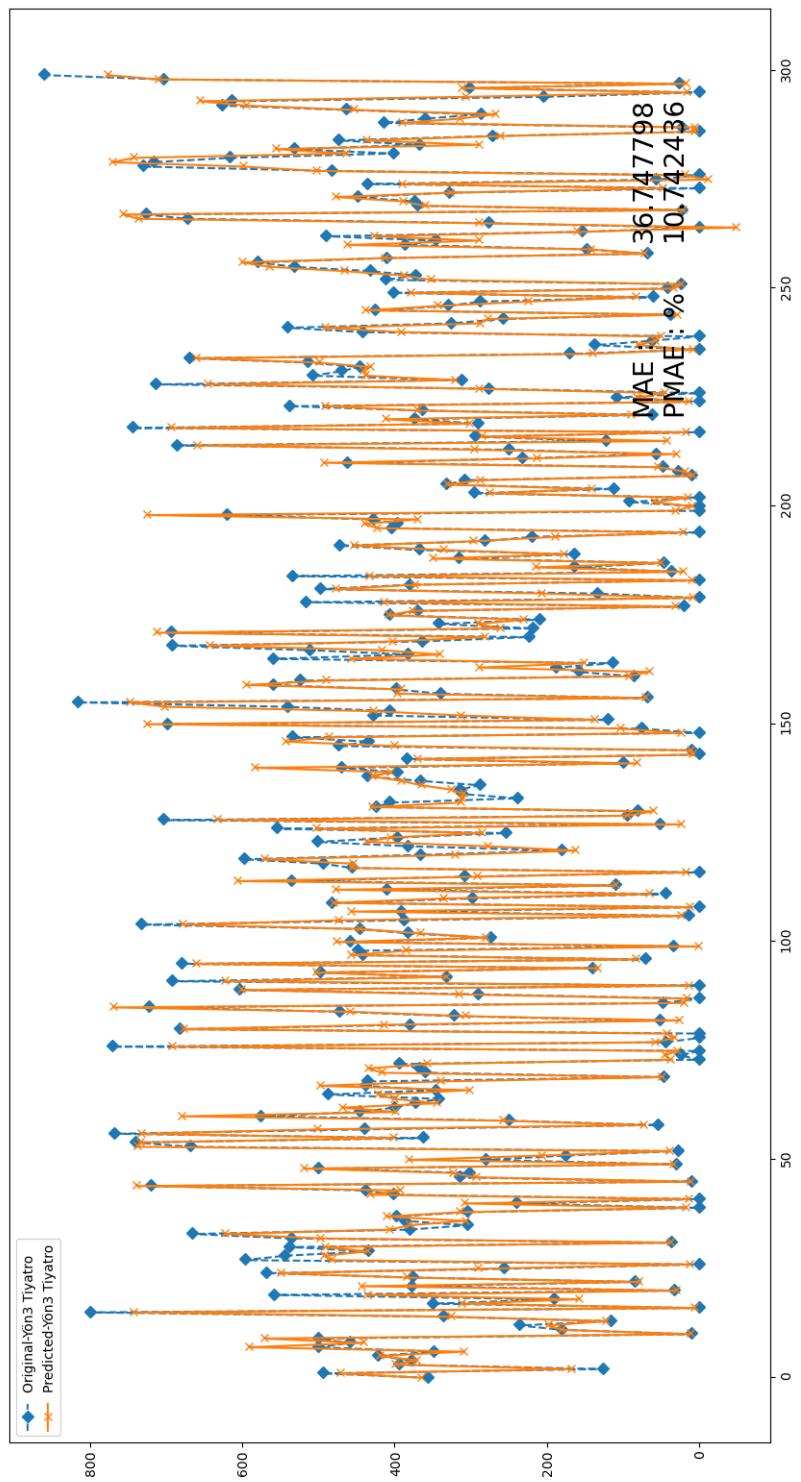
Çalışma sonucunda, kısa vadeli trafik tahmininin hızlı artan veya hızlı azalan anlarda akıllı ulaşım sistemlerine çeviklik kazandırdığı görülmüştür. Testler ve simülasyonlar tek bir kavşakta yapılmış olup, trafiğin büyük bir sistem olduğu göz önüne alındığında yöntemin etkilerini tam olarak yansıtamamıştır. Model belirli bir süre kullanılıp veriler alındıkça etkileri daha net anlaşılabilecektir.



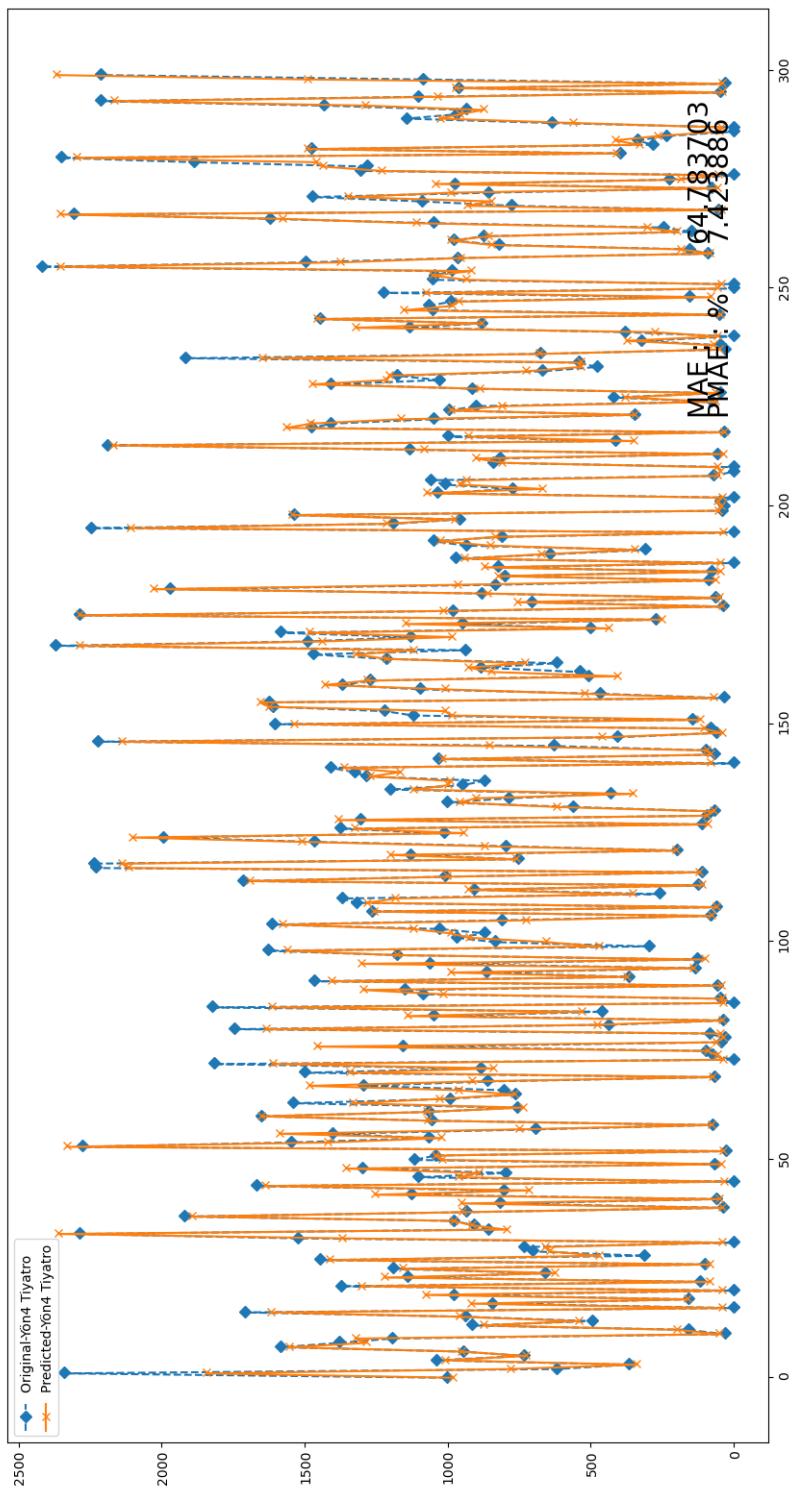
Şekil 5.3 : Xgboost Tiyatro Yön 1 Tahmin Değerleri



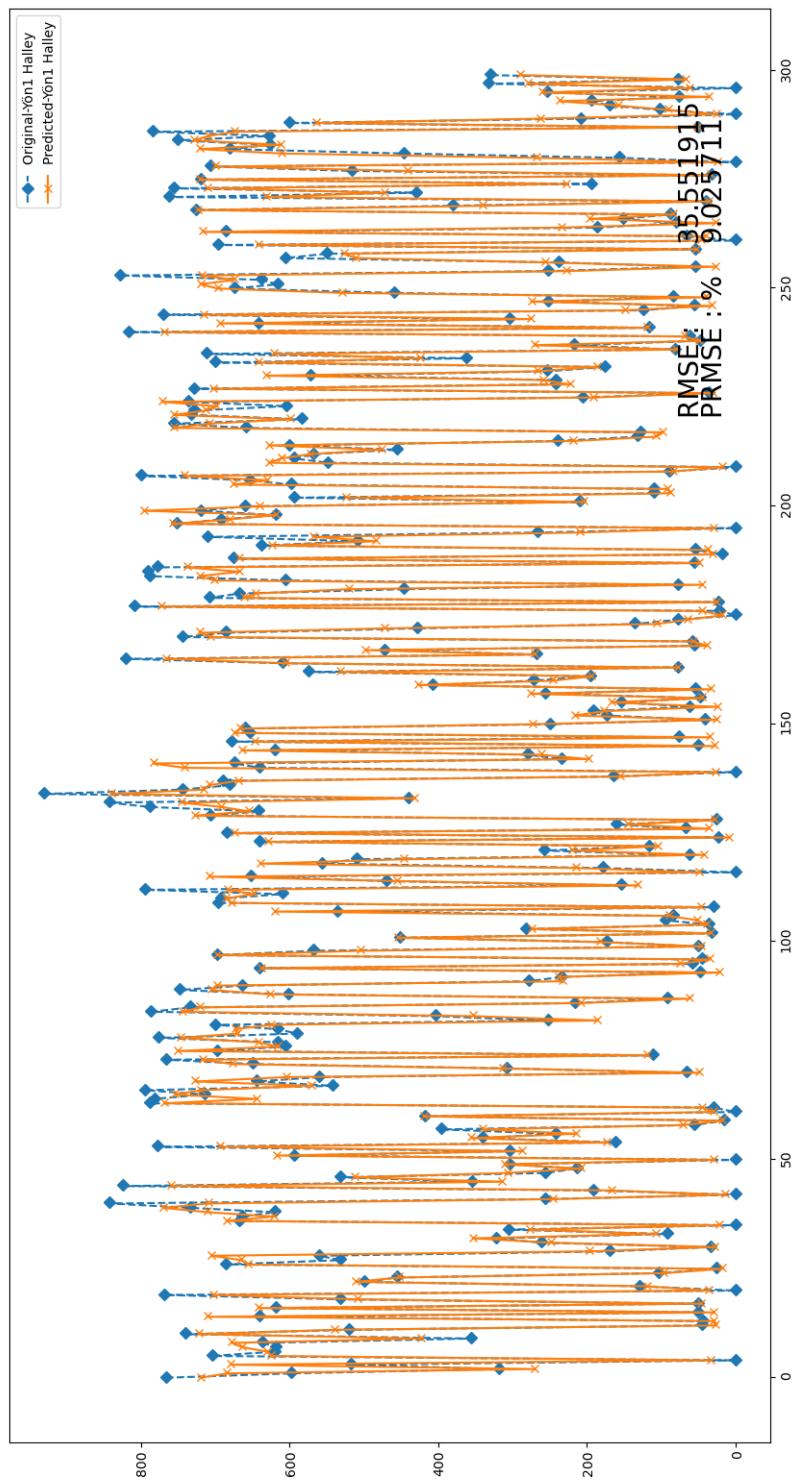
Şekil 5.4 : Xgboost Tiyatro Yön 2 Tahmin Değerleri



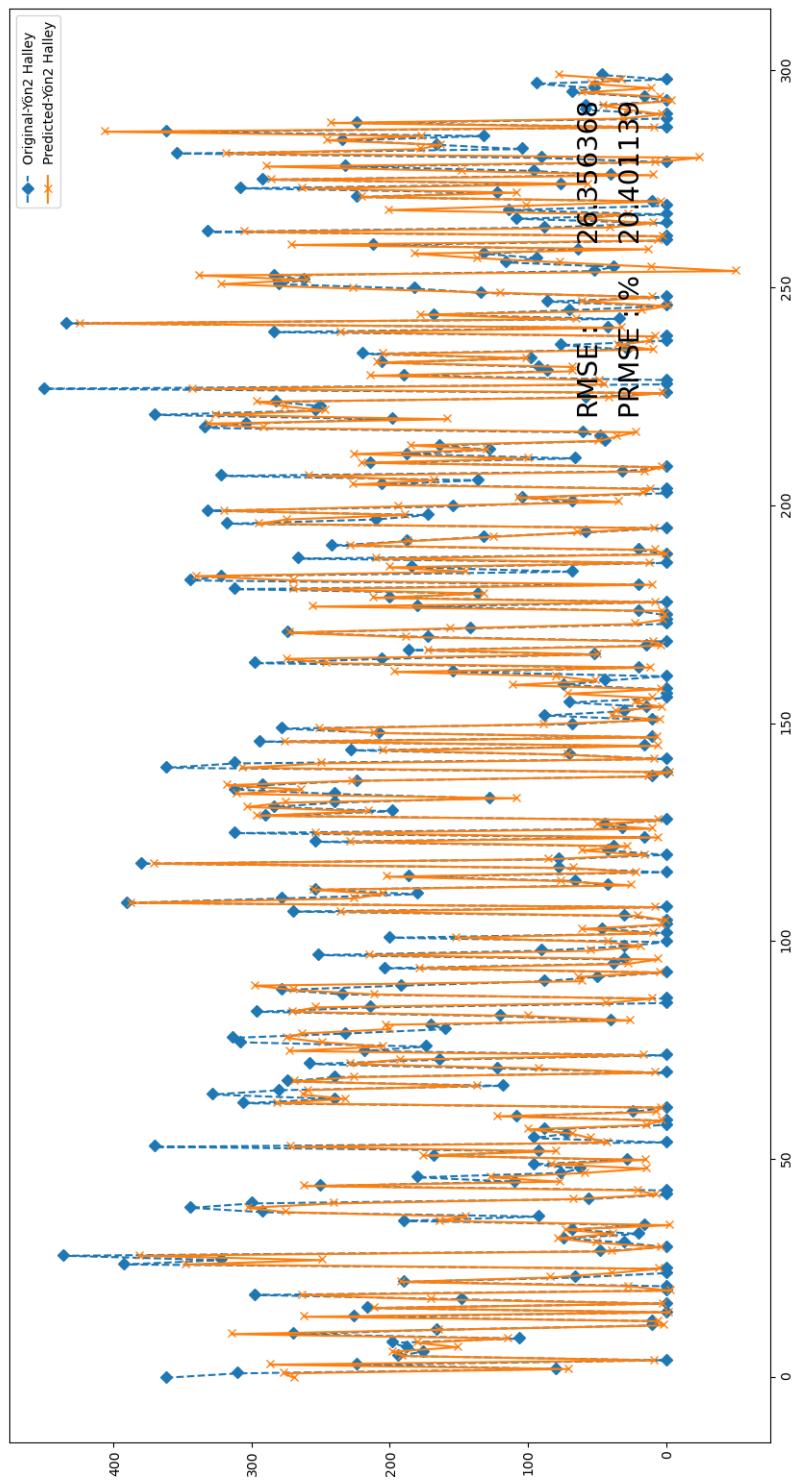
Şekil 5.5 : Xgboost Tiyatro Yön 3 Tahmin Değerleri



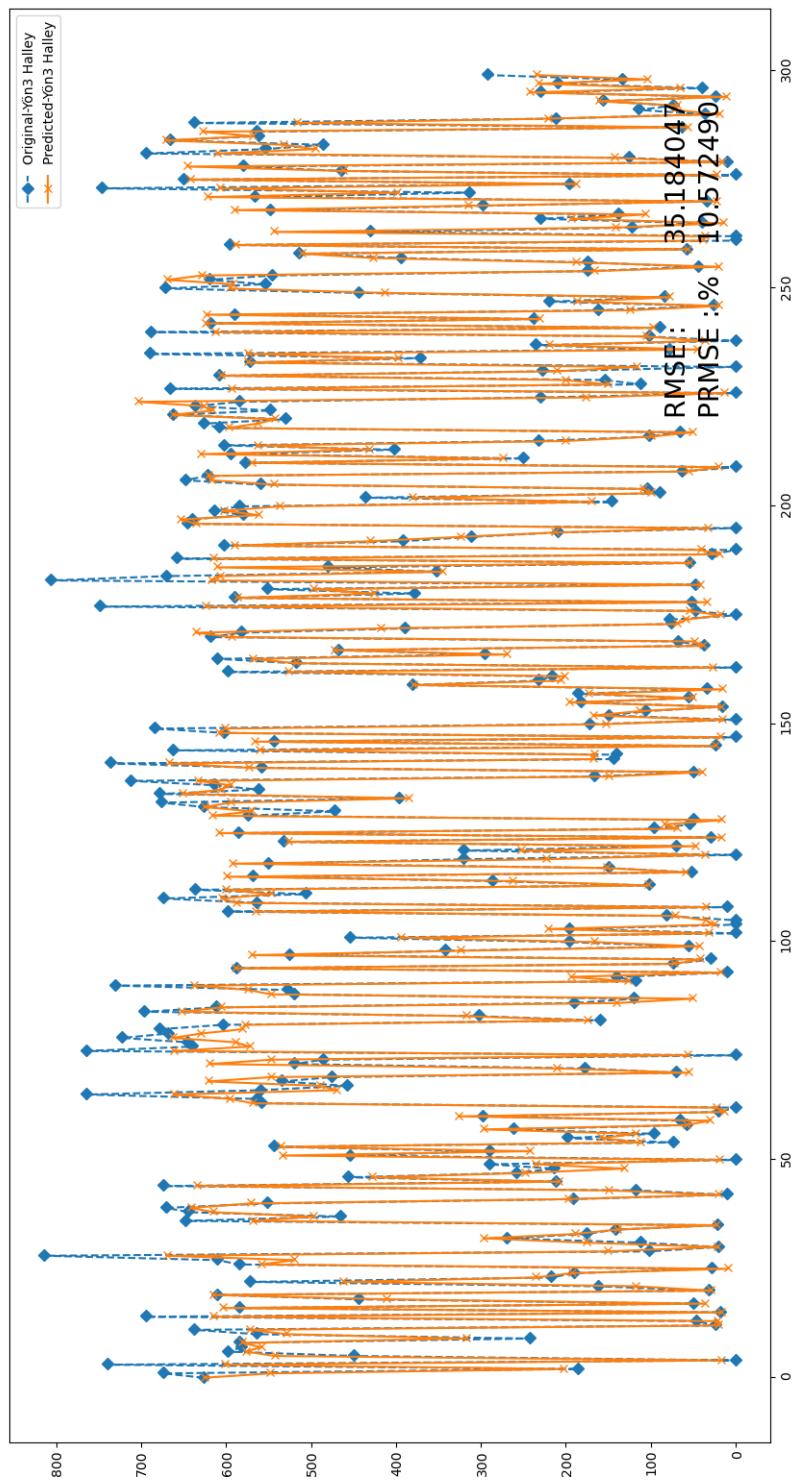
Şekil 5.6 : Xgboost Tiyatro Yön 4 Tahmin Değerleri



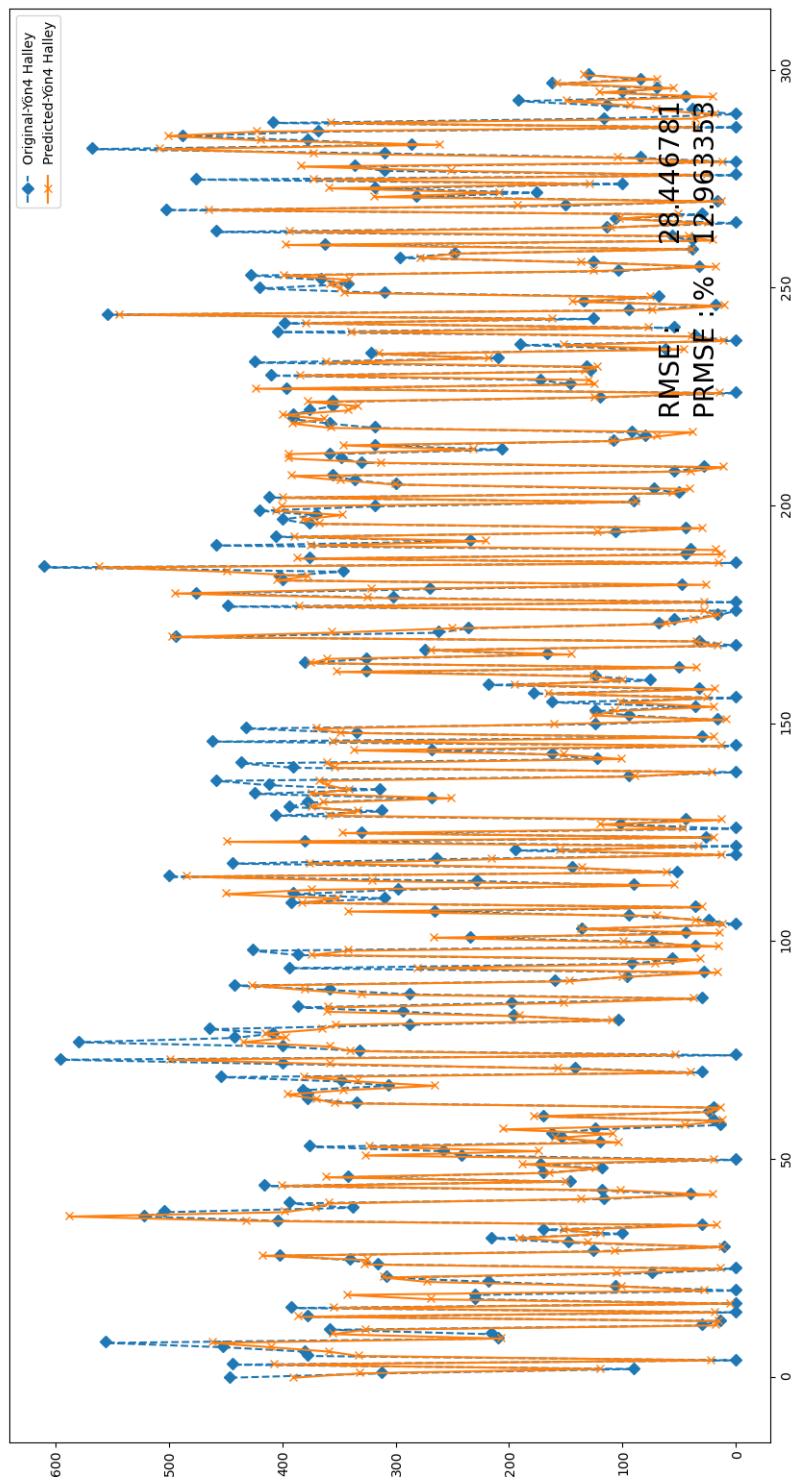
Şekil 5.7 : Xgboost Halley Yön 1 Tahmin Değerleri



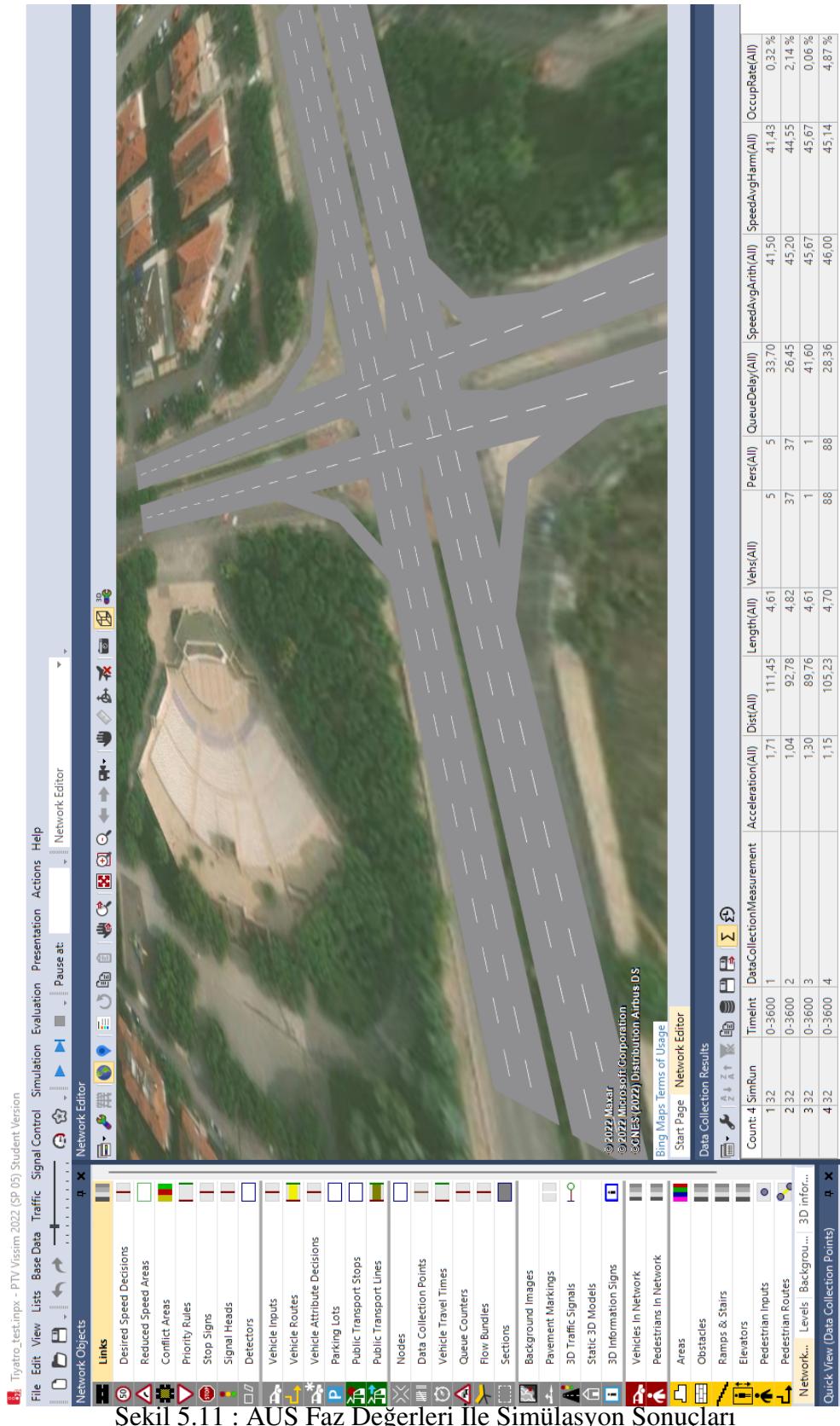
Şekil 5.8 : Xgboost Halley Yön 2 Tahmin Değerleri



Şekil 5.9 : Xgboost Halley Yön 3 Tahmin Değerleri



Şekil 5.10 : Xgboost Halley Yön 4 Tahmin Değerleri



Şekil 5.11 : AUS Faz Değerleri İle Simülasyon Sonuçları



Şekil 5.12 : Tahmin Modeli Faz Değerleri İle Simülasyon Sonuçları

Kaynakça

- [1] **ŞİRİN, E.**, (2019), Support Vector Machine (SVM) İle Sınıflandırma: Python örnek uygulaması, <https://www.veribilimiokulu.com/support-vector-machine-svm-ile-siniflandirma-python-ornek-u>
- [2] **AKCA, M.F.**, (2020), Nedir Bu Destek Vektör Makineleri? (Makine öğrenmesi serisi-2), <https://medium.com/deep-learning-turkiye/nedir-bu-destek-vekt%C3%B6r-makineleri-makine-%C3%B6%C4%9Frenmesi-serisi-2-94e576e4223e>.
- [3] 1.10. decision trees, <https://scikit-learn.org/stable/modules/tree.html>.
- [4] **Placek, M.**, (2021), Car production: Number of cars produced worldwide, <https://www.statista.com/statistics/262747/worldwide-automobile-production-since-2000/>.
- [5] **Zhang, K. ve Batterman, S.** (2013). Air pollution and health risks due to vehicle traffic, *The Science of the total environment*, 450-451, 307–316.
- [6] **Makino, H., Tamada, K., Sakai, K. ve Kamijo, S.** (2018). Solutions for urban traffic issues by ITS technologies, *IATSS research*, 42(2), 49–60.
- [7] **Nihan, N.L. ve Holmesland, K.O.** (1980). Use of the box and Jenkins time series technique in traffic forecasting, *Transportation*, 9(2), 125–143, <https://doi.org/10.1007/BF00167127>.
- [8] **Davis, G.A. ve Nihan, N.L.** (1991). Nonparametric regression and short-term freeway traffic forecasting, *Journal of Transportation Engineering*, 117(2), 178–188.
- [9] **Smith, B. ve Demetsky, M.** (1994). Short-term traffic flow prediction models-a comparison of neural network and nonparametric regression approaches, *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, cilt 2, s.1706–1709 vol.2.
- [10] **Hinsbergen, C., Lint, J. ve Sanders, F.** (2007). Short Term Traffic Prediction Models, *14th World Congress on Intelligent Transport Systems, ITS 2007*, 7.
- [11] **Dong, X., Lei, T., Jin, S. ve Hou, Z.** (2018). Short-term traffic flow prediction based on XGBoost, *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*, IEEE, s.854–859.

- [12] **Ravza Nur YİĞİT, S.H.** (2021). Kısa Zamanlı Trafik Tahmini ile Devre Süresi Optimizasyonu ve Gecikme Analizi, *Teknik Dergi*.
- [13] **Sun, B., Sun, T. ve Jiao, P.** (2021). Spatio-Temporal Segmented Traffic Flow Prediction with ANPRS Data Based on Improved XGBoost, *Journal of Advanced Transportation*, 2021.
- [14] **Wang, Y., Fang, S., Zhang, C., Xiang, S. ve Pan, C.** (2021). TVGCN: Time-Variant Graph Convolutional Network for Traffic Forecasting, *Neurocomputing*.
- [15] **Mohri, M., Rostamizadeh, A. ve Talwalkar, A.** (2018). *Foundations of machine learning*, MIT press.

ÖZGEÇMİŞ

Ad Soyad: Osman Doğukan URKAN

Doğum Tarihi ve Yeri: 06.01.1995 - Uşak

E-Posta: osman.urkan@hotmail.com

ÖĞRENİM DURUMU:

- **Lise:** 2013, Uşak Lisesi
- **Lisans:** Pamukkale Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği

MESLEKİ DENEYİMLER VE ÖDÜLLER:

- Staj I: Leonardo Turkey Havacılık Savunma ve Güvenlik Sistemleri A.Ş, 2020