

GAZİ

JOURNAL OF ENGINEERING SCIENCES

Comparison of Machine Learning Models For Traffic Volume Estimation at Smart Intesections

Seyitali İlyas^a, Yalçın Albayrak^b

Submitted: 19.11.2023 Revised: 15.12.2023 Accepted: 15.12.2023 doi: 10.30855/gmbd.0705S14

ABSTRACT

Keywords: Smart Intersection, Digital Twin, Big Data, Machine Learning, Traffic Volume Estimation

^a Akdeniz University,
Faculty of Engineering,
Department of Electrical and Electronics
Engineering
07000 - Antalya, Türkiye
Orcid: 0000-0003-2009-3454
e mail: seyitaliant@gmail.com

^b Akdeniz University,
Faculty of Engineering,
Department of Electrical and Electronics
Engineering
07000 - Antalya, Türkiye
Orcid: 0000-0002-5632-0869

^{*}Corresponding author:
seyitaliant@gmail.com

The population growth and mobility of modern cities have made efficient management of the transportation systems increasingly critical. According to this, the digital twin concept has become a powerful tool for understanding, managing, and optimizing complex systems by creating a digital reflection of the physical world. Transportation is one of these areas, moreover, with the help of such smart intersections and technological infrastructure, large amounts of data are made up of estimation models which is one of the important elements of digital twins. This study aims to provide an academic understanding of which model should be chosen to have accurate and more effective results during the application of traffic value prediction in smart intersections. Hourly vehicle count data from the arrival branches of two smart intersections in Antalya were utilized for this. This data is separated as training and test data created based on learning machine models including Linear Regression Polynomial Regression, Support Vector Regression (SVR), and Random Forest Regression. Thus, traffic volume estimation was conducted for each branch at smart intersections. Forecasting models were evaluated using Mean Absolute Error (MAE) and Least Squares (R^2) methods. Accordingly, it was seen that the Random Forest model performed better than the other proposed models.

Akıllı Kavşaklarda Trafik Hacmi Tahmini İçin Makine Öğrenmesi Modellerinin Karşılaştırılması

ÖZ

Modern şehirlerin artan nüfusu ve hareketliliği, ulaşım sistemlerinin verimli yönetimini gittikçe daha kritik hale getirmiştir. Bu bağlamda, dijital ikiz kavramı, fiziksel dünyanın dijital bir yansımını oluşturarak karmaşık sistemleri anlamak, yönetmek ve optimize etmek için güçlü bir araç haline gelmiştir. Ulaşım da bu alanlardan biridir ki akıllı kavşaklar gibi teknolojik altyapılar sayesinde büyük miktarda veri toplanmakta ve bu verilerle dijital ikiz mimarisinin önemli bir unsuru olan tahmin modelleri oluşturulmaktadır. Bu çalışma akıllı kavşaklarda trafik hacmi tahmini uygulamalarında daha etkili ve doğru sonuçlar elde etmek için hangi modelin seçilmesi gerektiğine dair akademik bir anlayış sunmayı amaçlamaktadır. Bunun için Antalya'da bulunan seçilmiş iki akıllı kavşağın her bir geliş kolu için saatlik araç sayım verileri kullanılmıştır. Bu veriler eğitim ve test verisi olarak ayrılmış olup Lineer Regresyon, Polinomal Regresyon, Destek Vektör Regresyonu (SVR) ve Rastgele Orman Regresyonu tabanlı makine öğrenmesi modelleri oluşturulmuştur. Böylece akıllı kavşaklarda her bir kol için trafik hacmi tahmini yapılmıştır. Ortalama Mutlak Hata (MAE) ve En Küçük Kareler (R^2) yöntemleri ile tahmin modellerinin performans karşılaştırması yapılmıştır. Buna göre Rastgele Orman modelinin diğer önerilen modellere göre daha başarılı performans göstermiş olduğu görülmüştür.

Anahtar Kelimeler: Akıllı Kavşak, Dijital İkiz, Büyük Veri, Makine Öğrenmesi, Trafik Hacmi Tahmini

1. Giriş (Introduction)

Dijital ikiz teknolojisi ilk defa 2002 yılında Dr. Michael Grieves tarafından Ulusal Havacılık Uzay İdaresi'nin (NASA) Apollo Programı'nda oluşturulmuştur [1]. Dijital ikiz verilerin fiziksel ve sanal ortam arasında her iki yönde gerçek zamanlı tam entegrasyonu olarak tanımlanmaktadır. Burada amaç gerçekleştirilecek sorunları yaşanmadan tespit ederek maliyet kaybını engelleyecek eylemde bulunmaktır. Dijital ikiz teknolojisinde fiziksel nesnede yapılan bir değişiklik ile otomatik olarak dijital nesnede değişiklik meydana gelmekte olup bu durumun tersi de oluşabilmektedir [2, 3].

Dijital ikiz, dijital gölge, dijital model kavramları birbiriyle ilişkili ancak farklı kavramlardır ve literatürde yapılmış olan çalışmalar tanımlanırken bunların birbiriyle karıştırılabildiği görülmüştür. Dijital model, önceden var olan veya planlanan bir fiziksel nesnenin dijital versiyonu olarak tanımlanmaktadır. Dijital modelde, fiziksel sistem ile dijital model arasında hiçbir otomatik veri alışverişi olmamaktadır. Bu, dijital model oluşturulduktan sonra fiziksel nesnede yapılan bir değişikliğin dijital model üzerinde hiçbir şekilde etkisi olmayacağı anlamına gelmektedir. Dijital gölge kavramı ise fiziksel ve dijital nesne arasında tek yönlü bir akışa sahip olan bir dijital durum temsilidir [4, 5]. Fiziksel nesnedeki bir değişiklik, dijital nesnede bir değişikliğe yol açmaktadır ancak tersi durum gerçekleşmemektedir. Bu tanımlar ışığında dijital ikiz üç unsurdan oluşmaktadır: Nesnelerin İnterneti (IoT) sensörleriyle sağlanan geçmiş veriler, gerçek zamanlı veriler ve makine öğrenmesi algoritmalarıyla oluşturulan modeller kullanılarak elde edilen gelecekteki verilerdir [4, 6, 7].

Ulaşım konusunda dijital ikiz teknolojisi kapsamında büyük bütçeli uygulamalarının yapıldığı ancak uygulama yöntemlerinin detaylı içeriğine az değinildiği görülmüştür [8]. Bu amaçla Literatür incelemesinde dijital ikiz mimarisinin bileşenlerinden biri olan tahmin modellerinin oluşturulması ile ilgili çalışmalar incelenmiştir.

Janković vd. (2021) Sırbistan Cumhuriyeti'ndeki yollarda bulunan otomatik sayaçlarından tarih, saat, yön, araç sayısı özelliklerini içeren verileri toplayarak günlük ve saatlik trafik hacmi tahmin etmişlerdir. Bunun için Doğrusal Regresyon, Çok Katmanlı Algılayıcı, k-En Yakın Komşu, M5P, Rastgele Orman, REPTree algoritmalarıyla oluşturulmuş modeller kullanılmıştır. İlk senaryo da günlük trafik hacmi tahmini yapılmıştır. M5P algoritmasına (kolerasyon katsayısı eğitim için 0.9778, test için 0.9756) dayalı model en iyi performansı göstermiştir. İkinci senaryo için saatlik trafik hacmi tahmin edilmiştir. İkinci senaryo için en iyi sonuçları k-En Yakın Komşu algoritması vermiştir [9]. Bratsas vd. (2020) çalışmalarında, Yunanistan'ın Selanik karayolu ağından elde ettikleri trafik verilerini kullanarak araçların ortalama hızını tahmin etmişlerdir. Bu amaçla, Doğrusal Regresyon, Çok Katmanlı Algılayıcı, Destek Vektör Regresyonu (SVR), Rastgele Orman gibi makine öğrenmesi modellerini kullanmışlardır. Bu modellerin tahmin başarı oranlarını farklı senaryolar altında birbirleriyle karşılaştırmışlardır. Sonuç olarak, Destek Vektör Regresyon modeli belirli koşullar altında daha istikrarlı bir performans sergilerken, Çok Katmanlı Algılayıcı Sinir Ağı (MLP-NN) modeli daha küçük hatalara sahip olmanın yanı sıra değişken koşullarda en iyi sonuçları vermiştir. Bu çerçevede, her iki modelin de farklı senaryolarda farklı avantajlara sahip olduğu gözlemlenmiştir [10]. Ali ve Hamed (2018) tarafından gerçekleştirilen çalışmada Suudi Arabistan'ın Al Ghat şehrindeki yıl, saat, kaza türü, konum, araç sayısı, sürücü vatandaşlık durumu, yaralı, ölü bilgilerini içermekte olan trafik veri setini kullanılarak, kazaya neden olan durumlar tahmin edilmiştir. Apriori algoritması ve Bulanık Kümeleme algoritması arasında yapılan performans karşılaştırmasında, Apriori algoritmasının daha başarılı sonuçlar verdiği gözlemlenmiştir [11]. Navarro vd. (2022) dört kollu sinyalizasyon kavşaktaki trafik akımlarını tahmin etmek için Münih Araştırma Merkezi'nden alınan veri kümesini kullanmışlardır. Tekrarlayan Sinir Ağları (RNN'ler), MLP-NN, Uzun Kısa Süreli Bellek (LSTM), Gradyan Arttırma, Rastgele Orman, Doğrusal Regresyon ve Stokastik Gradyan algoritmaları trafik akımı tahminindeki başarıları açısından karşılaştırılmıştır. Yapılan analizler, tüm makine öğrenmesi ve derin öğrenme algoritmalarının kurulacak akıllı kavşaklarda trafik yönetiminde uygulanabileceğini göstermiştir. Ancak, performans sonuçları incelendiğinde, MLP-NN'nin daha iyi sonuçlar verdiği (R^2 0,93) ve aynı zamanda daha kısa bir eğitim süresinde (18 s) gerçekleştiği belirlenmiştir [12]. Hou vd. (2021) hava değişikliklerinin trafik akışı üzerindeki etkisini dikkate alarak yeni bir tahmin modeli oluşturmuşlardır. Önerilen modelin daha başarılı performans ortaya koyduğunu tespit etmişlerdir [13]. Alajali vd. (2018) bu çalışmada, Avustralya'nın Melbourne merkezi iş bölgesindeki kavşaklarda sensörlerden gelen trafik hacmi veri setiyle, kaza ve yol çalışmalarını içeren olay veri setlerini entegre ederek yeni özelliklerin tahmin modellerindeki hata değerlerine etkisini incelemişlerdir. Tahmin için Regresyon Ağaçları (RT), Gradyan Arttırıcı Karar Ağaçları (GBRT), Rastgele Orman (RF), Aşırı Gradyan Arttırma (XGBoost), SVR modellerini kullanmışlardır. Sadece sensör verileri kullanıldığında SVR, RT, GBRT, RF ve

XGBoost modelleri için Ortalama Kare Hata (MSE) değerlerinin sırasıyla 0.8883, 0.8838, 0.8772, 0.8815, 0.8773 olduğu belirlenmiştir. Ancak olay verileri sensör verileriyle entegre edildiğinde MSE'ler sırasıyla 0.8859, 0.8829, 0.8684, 0.8770 ve 0.8678 olarak elde edilmiştir. Bu sonuçlar, olay verilerinin entegrasyonunun tahmin modellerinin MSE değerini iyileştirdiğini göstermiştir. Özellikle XGBoost için en iyi MSE değeri elde edildiği görülmüştür [14].

Antalya Büyükşehir Belediyesi, TC. Ulaştırma ve Altyapı Bakanlığının Ulusal AUS Strateji Belgesi ve 2020-2023 Eylem Planı doğrultusunda 'İl Trafik Kontrol Merkezlerinin Kurulması' eylem planını tamamlayarak 'Akıllı Kavşaklar Projesi' ile trafik veri akışını sahada kurulumu yapılan 40 adet Akıllı Kavşaktan sağlamaya başlamıştır [15, 16]. Ancak bu Akıllı Kavşaklar dijital ikiz teknolojisi kapsamında değildir çünkü gelecek tahmini yapmamaktadır. Bu çalışmada Antalya ili özelinde ardışık akıllı kavşak verileri kullanılarak dijital ikiz mimarisi için altyapı oluşturulması planlanmaktadır. Bunun için Trafik Kontrol Merkezinde toplanan büyük veri akışıyla makine öğrenmesi tabanlı tahmin modelleri oluşturulmuştur [17]. Bu modellerin trafik hacmi tahminindeki performans başarıları ortaya koyulmuştur.

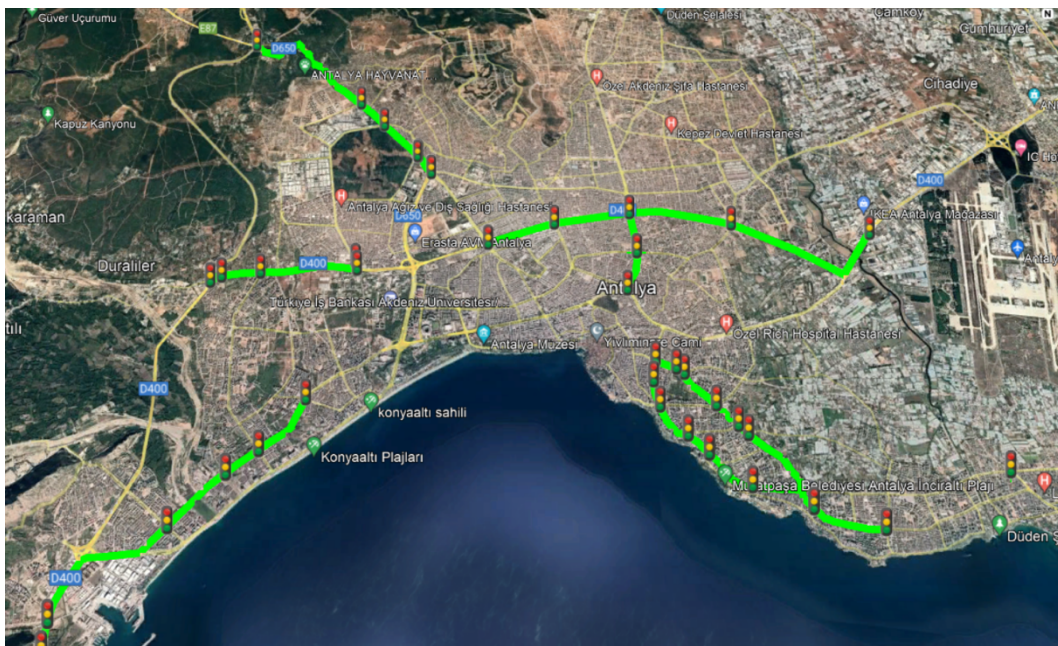
2. Metodoloji ve Yöntemler (Methodology And Methods)

Bu çalışmada Akıllı Kavşaklardan trafik hacmi verileri toplanarak Python yazılımında veriler düzenlenmiştir. Düzenlenen verilerle Python da Lineer Regresyon, Polinomal Regresyon, SVR ve Rastgele Orman Regresyonu makine öğrenme algoritmaları kullanılarak tahmin modelleri oluşturulmuştur. Daha sonra oluşturulan bu modellerin araç tahminindeki başarısını değerlendirmek için MAE ve en R^2 yöntemleri kullanılmıştır.

2.1. Veri kümesi (Dataset)

Makine öğrenmesi modellerinin oluşturulması için veri setine ihtiyaç vardır. Yapılan çalışmada kullanılan veri setleri Atatürk Bulvarı ile Öğretmenler Caddesi kesişiminde bulunan (1 numaralı) ve Atatürk Bulvarı ile Gazi Mustafa Kemal Bulvarı kesişiminde bulunan (2 numaralı) 4 kollu kavşakların her bir kollundan alınmıştır. Python yazılımı kullanılarak kavşaklardaki her bir yön için tarih ve zaman serisini temsil eden bağımsız değişken verilerinin olduğu sütun ve saatlik araç sayımlarını temsil eden yön 1, yön 2, yön 3, yön 4 olarak adlandırılan bağımlı değişken verilerinin olduğu 4 sütundan oluşmaktadır. Veri kümesi içerisinde bulunan bağımlı ve bağımsız değişkenler Tablo 1'de veri tablosu üzerinde gösterilmiştir.

Şekil 1'de Antalya Büyükşehir Belediyesinin Trafik Kontrol Merkezine veri akışını yapan Akıllı Kavşakların Google haritalar üzerinden görüntüsü verilmiştir.



Şekil 1. Akıllı Kavşakların Google Earth Gösterimi
(Representation of Smart Intersections on Google Earth)

Tablo 1’de Atatürk Bulvarı ile Gazi Mustafa Kemal Bulvarı kesişiminde bulunan (2 numaralı) kavşağa ait 10 satırlık örnek veri kümesi gösterilmiştir.

Tablo 1. Atatürk Bulvarı ile Gazi Mustafa Kemal Bulvarı kesişim kavşağına ait 10 Satırlık Örnek Veri Kümesi Tablosu (Sample Dataset Table for the intersection of Atatürk Street and Gazi Mustafa Kemal Street)

Tarih	Zaman	Yön 1 (araç/saat)	Yön 2 (araç/saat)	Yön 3 (araç/saat)	Yön 4 (araç/saat)	Toplam (araç/saat)
14.09.2022	00:00	836	314	629	159	1938
14.09.2022	01:00	532	173	390	110	1205
14.09.2022	02:00	245	116	244	63	668
14.09.2022	03:00	179	72	179	54	484
14.09.2022	04:00	137	58	175	41	411
14.09.2022	05:00	167	62	184	29	442
14.09.2022	06:00	321	116	483	60	980
14.09.2022	07:00	1059	365	1044	182	2650
14.09.2022	08:00	1797	594	1334	417	4142
14.09.2022	09:00	1697	626	1276	437	4036

1 numaralı kavşak verileri ise 2022-09-07 ile 2022-09-21 tarihleri arasındaki 15 günlük sayımdan oluşmaktadır. Veri kümesi 1800 adet bağımlı değişken, 360 adet bağımsız değişken olmak üzere toplamda 2160 veriden oluşmaktadır. 2 numaralı kavşak verileri 2022-09-07 ile 2022-09-14 tarihleri arasındaki 8 günlük sayımdan oluşmaktadır. Veri kümesi 960 adet bağımlı değişken, 192 adet bağımsız değişken olmak üzere toplamda 1152 veriden oluşmaktadır.

Tahmin modellerinin performansını objektif bir şekilde değerlendirmek için daha önce bilinmeyen veri kümesi gerekmektedir [18]. Bu sebeple 1 numaralı kavşak veri kümesi %20’si (3 gün) test verisi, %80’i eğitim verisi olarak, 2 numaralı kavşak veri kümesi %12,5’i test (1 gün) %87,5’i eğitim verisi olarak ayrılarak regresyon modellerinin oluşturulmasında kullanılmıştır.

2.2. Regresyon modellerinin oluşturulması (Creating regression models)

Lineer Regresyon, Polinomal Regresyon, Destek Vektör Regresyon, Rastgele Orman Regresyon algoritmaları bağımsız değişkenin (X) bağımlı bir değişkene (Y) olan istatistiksel ilişkisini açıklamak için kullanılmıştır. Veri kümesi eğitim verisi ile eğitilip tahmin modelleri oluşturulmuş ve test verisi ile de model başarısı gösterilmiştir.

2.2.1. Lineer regresyon modeli (Linear regression model)

Lineer regresyon modeli bir doğru denklemi kullanarak X değerlerinin Y değerlerine nasıl etki ettiğini gösterir. Lineer regresyon modeli, bu değişkenler arasındaki ilişkiyi doğrusal bir eğri ile ifade eder. Lineer regresyon modeli, bu eğriyi temsil eden denklemi kullanarak, bağımsız değişkenlerin Y değerlerini tahmin edebilecektir. Bu model, özellikle doğrusal bir ilişkinin olduğu durumlarda, öngörücü analiz ve tahmin yapmak için kullanılmaktadır. Lineer regresyon, Eşitlik.1’deki denklem ile ifade edilmektedir.

$$Y = a + bX + \varepsilon \quad (1)$$

Burada Y bağımlı değişkeni X bağımsız değişkeni, a kesme terimini, b eğim terimini (X ’teki birim artışın Y ’deki birim artışı ile ilişkisini), ε hata terimini (modelin gerçek değerlerden ne kadar sapabileceğini) ifade etmektedir [19].

2.2.2. Polinomal regresyon modeli (Polynomial regression model)

Polinomal regresyon, bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki doğrusal olmayan ilişkiyi modellemek için kullanılan bir regresyon analiz yöntemidir. Polinomal regresyon, lineer regresyon modellerinden farklı olarak, bir polinom denklemi kullanarak bağımlı ve bağımsız değişkenler arasındaki ilişkiyi ifade etmektedir. Polinomal regresyon, yüksek dereceli polinomlar kullanarak, verilerin

uydurulması ve tahmininde daha esnek bir yaklaşım sağlamaktadır. Özellikle, verilerin doğrusal bir şekilde değişmediği veya bir eğri veya dalgalı desen gösterdiği durumlarda kullanılmaktadır. Polinomal regresyon, Eşitlik.2'deki denklemlerle ifade edilmektedir.

$$Y = b_0 + b_1X + b_2X^2 + b_3X^3 + \dots + b_nX^n + \varepsilon \quad (2)$$

Burada X bağımsız değişkeni, Y bağımlı değişkeni, $b_1, b_2, b_3 \dots b_n$ polinom katsayılarını ε hata terimini ifade etmektedir [20].

2.2.3. Destek vektör regresyonu modeli (Support vector regression model)

SVR modeli, doğrusal olmayan bir hiper düzlemi kullanarak bir bağımlı değişkenin değerlerini bir veya daha fazla bağımsız değişkenle ilişkilendirmektedir. Bu model, çoğu regresyon probleminde kullanılan geleneksel yöntemlerden daha verimli bir şekilde çalışmaktadır. SVR modeli, Eşitlik.3'teki denklemlerle ifade edilmektedir.

$$Y = wX + b \quad (3)$$

Burada, Y bağımlı değişkeni, X bağımsız değişkenleri, w ağırlık vektörünü, b sabit terimi ifade etmektedir [21].

2.2.4. Rastgele orman modeli (Random forest model)

Rastgele Orman makine öğrenmesi modeli hem sınıflandırma hem de regresyon problemlerini çözmek için kullanılan bir öğrenme yöntemidir. Rastgele Orman modeli, birden fazla karar ağacını kullanarak tahminler yapmakta ve bu karar ağaçlarının sonuçlarını bir araya getirerek daha doğru bir tahmin sonuçları elde etmektedir.

Rastgele Orman modeli, aşağıdaki adımlardan oluşmaktadır:

- Eğitim verilerinden rastgele bir alt küme oluşturulur.
- Bu alt küme, rastgele seçilen özellikler üzerinde birçok karar ağacı eğitir.
- Oluşturulan her ağaç, rastgele bir şekilde alt özellikler ve veri noktaları seçerek oluşturulur.
- Her ağacın sonucu, oybirliğiyle veya ağırlıklı ortalamayla birleştirilir.

Rastgele Orman modeli, aşırı öğrenmeyi engellemek ve daha sağlam bir genelleme elde etmek amacıyla kullanılmaktadır. Her bir ağaç, farklı bir alt kümeden öğrenildiği için, model eğitim verilerine aşırı uyum sağlamamaktadır. Bu durum, daha genel bir model elde edilmesini sağlamaktadır.

Rastgele Orman modeli, özellikle yüksek boyutlu veri kümelerinde etkili olmakta olup ağaç sayısı, özellik seçimi ve derinlik gibi parametreleri ayarlamak suretiyle modelin doğruluğu artırılmaktadır [12, 22].

2.3. Model performansının değerlendirilmesi (Evaluation of model performance)

Önerilen modellerin performansı test edilmiştir. Deneysel aşamada oluşturulmuş olan modellerin performansını değerlendirmek için MAE ve R^2 yöntemleri kullanılmıştır.

2.3.1. Ortalama mutlak hata (Mean absolute error)

Ortalama Mutlak Hata, bir tahminin gerçek değerden ortalama olarak ne kadar uzak olduğunu ölçen bir istatistiksel ölçüttür. MAE, her bir tahmin hatasının mutlak değerlerinin ortalamasını hesaplar. Mutlak değer, bir sayının işaretini göz ardı ederek o sayının pozitif değerini alır. Bu nedenle, MAE, bir tahminin gerçek değerden ne kadar farklı olduğunu ölçer, ancak bu farkın yönünü dikkate almaz.

MAE, özellikle regresyon modellerinin performansını ölçmek için kullanılır. Regresyon modeli, bir girdi değişkeniyle bir çıktı değişkeni arasındaki ilişkiyi modelleyen bir matematiksel fonksiyondur. MAE, modelin gerçek değerlerden ne kadar sapma gösterdiğini ölçerek, modelin ne kadar iyi performans gösterdiğini anlamamıza yardımcı olur. MAE Eşitlik.4'teki denklemlerle gösterilmektedir.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \left(\frac{1}{n}\right) \sum_{i=1}^n |e_i| \quad (4)$$

Yukarıdaki denklemlerde n veri noktalarının sayısını, y_i gerçek değeri (gözlem değeri), x_i tahmin edilen değeri (model tarafından üretilen değer), e_i mutlak hatayı ifade etmektedir.

2.3.2. R-kare (R-square)

Önerilen tahmin modellerinin performansını karşılaştırmak için R-Kare (R^2) yöntemi kullanılmıştır. Bu yöntem, bağımlı ve bağımsız değişkenler arasındaki etkinin matematiksel model sonuçlarına dayalı olarak karşılaştırılmasını sağlamaktadır [23]. R^2 , belirli bir regresyon modelinin ne kadar iyi uyduğunu ölçen bir istatistiksel ölçüdür. R^2 Eşitlik.4'teki denklemle hesaplanmaktadır.

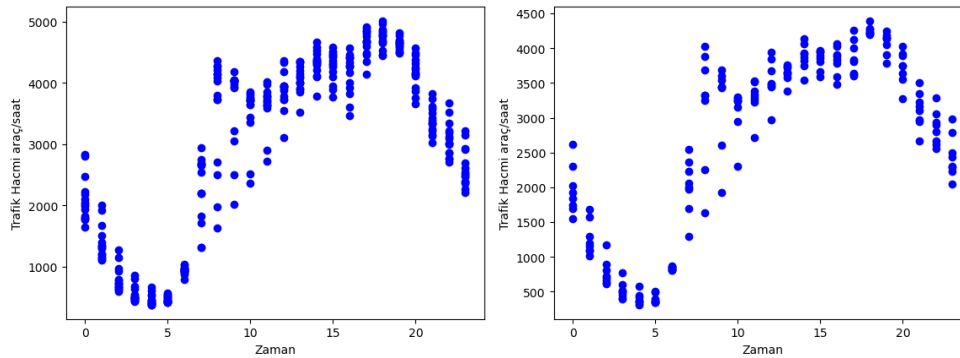
$$R^2 = 1 - \left(\frac{\sum_{i=1}^n (y_i - x_i)}{\sum_{i=1}^n (y_i - x_0)} \right) \quad (5)$$

Yukarıdaki denklemde x_0 ortalama değeri (en kötü yapılabilecek tahmin) ifade etmektedir. R^2 değerinin bire yaklaşması durumunda modelin tahmin performansının yüksek olduğu, sıfıra yaklaşması durumunda modelin başarısız olduğu sonucu çıkarılabilmektedir [24].

3. Deney Sonuçları (Experiment Results)

Veri setinin genel yapısını görmek seçilecek tahmin modelleri açısından önem taşımaktadır. Veri setine göre, örneğin normal bir dağılıma sahip veriler için parametrik modeller (Lineer Regresyon gibi) daha uygun olabilirken, dağılımı bilinmeyen veriler için daha esnek modeller (Random Forest gibi) tercih edilebilmektedir.

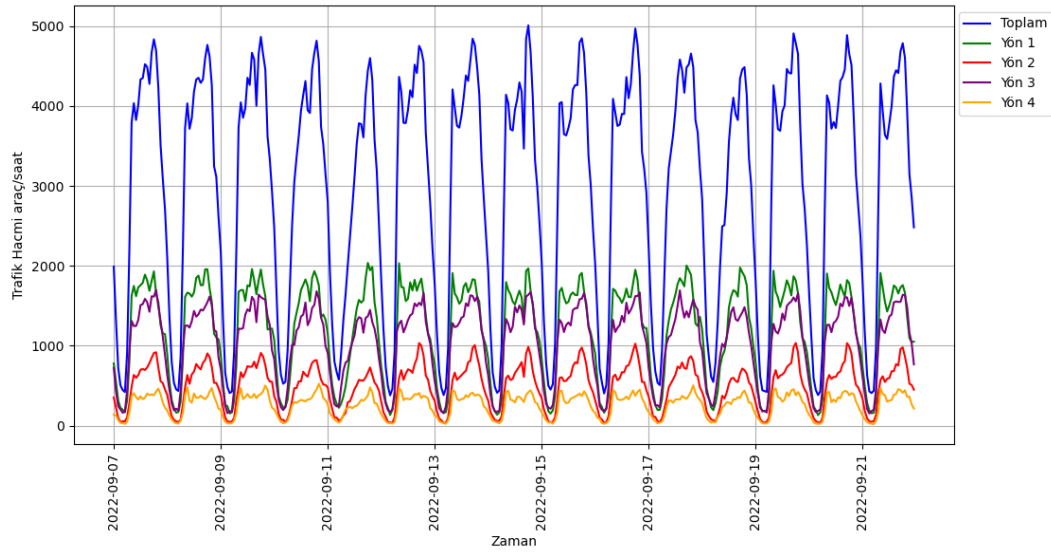
Şekil 2'de Python çıktısı olarak elde edilmiş olan 1 numaralı ve 2 numaralı kavşaklara ait toplam trafik hacim verilerinin dağılımı gösterilmiştir. Burada her iki kavşak içinde sabah saatlerinde trafik yoğunluğunun daha az olduğu 1 numaralı kavşak için akşam 17:00:00 ile 18:00:00 (2022-09-07 ile 2022-09-21 tarihleri) arasındaki saatin pik saat, 2 numaralı kavşak için 18:00:00 ile 19:00:00 (2022-09-07 ile 2022-09-14 tarihleri) arasındaki pik saat olduğu belirlenmiştir.



Şekil 2. 1 Numaralı ve 2 Numaralı Kavşak Trafik Hacmi Verileri
(Traffic Volume Data for Junction No. 1 and No. 2)

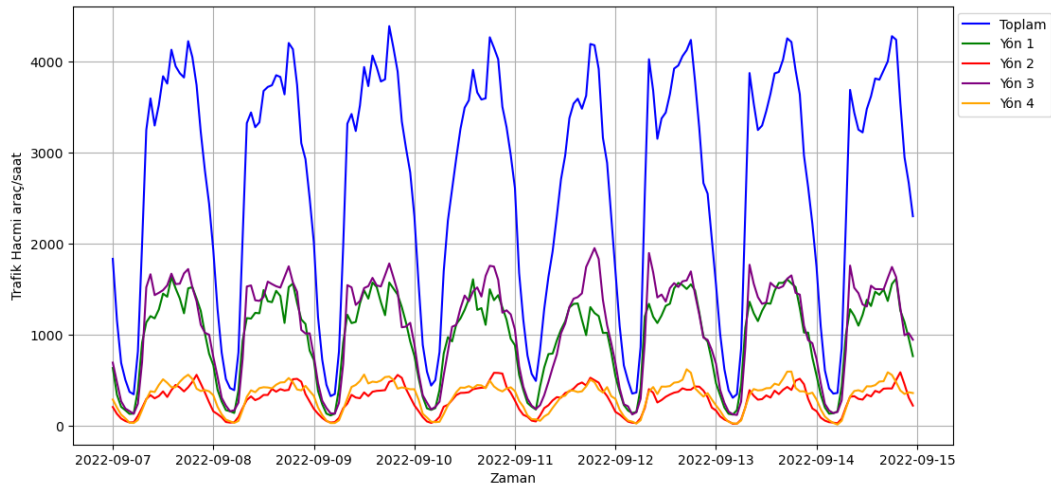
Trafik Akışının Yönetimi için geliş kollarındaki trafik hacminin bilinmesi, kavşakta trafik akışını düzenlemek açısından önemlidir. Özellikle yoğun saatlerde veya özel etkinliklerde, trafik akışını dengelemek ve trafik sıkışıklığını önlemek için bu bilgilere ihtiyaç vardır. Akıllı kavşaklarda sinyal süreleri kavşağa geliş kollarındaki yoğunluklara göre optimize edilmektedir.

Aşağıdaki şekilde 1 numaralı kavşak için 2022-09-07 ile 2022-09-21 tarihleri arası trafik hacim verileri gösterilmiştir. Yön 1 ve Yön 3'ten gelen araç sayılarının ana arter olmaları sebebiyle diğer yönlerden daha fazla olduğu, asıl trafik yoğunluğunun bu kollarında yaşandığı görülmüştür. Trafik hacim verilerine göre sıralama yapıldığında Yön 1, Yön 3, Yön 2, Yön 4 olacak şekilde kavşağa geliş güzergahları sıralanmıştır.



Şekil 3. 1 Numaralı Kavşak Toplam ve Tüm Geliş Kolları Trafik Hacim Verileri
(Traffic Volume Data for Junction No. 1, Total and All Arrival Branches)

Aşağıdaki şekilde 2 numaralı kavşak için 2022-09-07 ile 2022-09-15 tarihleri arası trafik hacim verileri gösterilmiştir. Yön 1 ve Yön 3 güzergahlarından gelen trafik akışının birbirine yakın ve diğer yönlere göre daha fazla olduğu, Yön 2 ve Yön 4 güzergahlarından gelen trafik akışının birbirine yakın olduğu görülmüştür.



Şekil 4. 2 Numaralı Kavşak Toplam ve Tüm Geliş Kolları Trafik Hacim Verileri
(Traffic Volume Data for Junction No. 2, Total and All Arrival Branches)

1 numaralı kavşak ve 2 numaralı kavşak kolları için oluşturulan modellerin analiz sonuçları Tablo 2 ve Tablo 3'te gösterilmiştir. Lineer Regresyon modelin Şekil 2 grafiğinde veri setinin dağılımından anlaşılacağı üzere MAE ve R^2 değerleri açısından en düşük değerleri almıştır. Trafik hacmi tahmininde rastgele orman modellerine dayalı tahminlerin daha başarılı sonuçlar verdiği gözlenmiştir.

Tablo 2. 1 Numaralı Kavşak Kolları Tahmin Modelleri Performans Değerleri

Model	y^1 ' Güzergahı		y^2 ' Güzergahı		y^3 ' Güzergahı		y^4 ' Güzergahı	
	MAE	R^2	MAE	R^2	MAE	R^2	MAE	R^2
Lineer Regresyon	433,4680	0,3340	186,5940	0,4867	323,2860	0,4850	97,2162	0,4217
Polinomal Regresyon	201,5408	0,7888	84,3496	0,8603	124,8740	0,8931	52,8634	0,7519
Destek Vektör Regresyon	162,3194	0,8281	64,6678	0,9010	98,1503	0,9353	44,8088	0,8026
Rasgele Orman Regresyon	118,9882	0,9272	50,4261	0,9505	68,6889	0,9699	34,2801	0,9157

Tablo 3. 2 Numaralı Kavşak Kolları Tahmin Modelleri Performans Değerleri

Model	y ¹ Güzergahı		y ² Güzergahı		y ³ Güzergahı		y ⁴ Güzergahı	
	MAE	R ²	MAE	R ²	MAE	R ²	MAE	R ²
Lineer Regresyon	325,3016	0,4771	71,7533	0,6229	382,3645	0,3748	107,3552	0,5126
Polinomal Regresyon	124,2187	0,8869	56,1171	0,8269	172,7752	0,7928	39,6911	0,9114
Destek Vektör Regresyon	101,5467	0,9208	39,4431	0,9022	138,6960	0,8267	34,7533	0,9219
Rasgele Orman Regresyon	62,0924	0,9767	20,8846	0,9692	75,4847	0,9565	22,8378	0,9718

4. Sonuçlar (Conclusion)

Bu çalışmada trafik hacim verilerinin öngörülmesi kapsamında Dijital İkiz mimarisinin yapısı incelenmiştir. Makine öğrenmesi algoritmaları kullanarak oluşturulan tahmin modelleriyle literatürdeki çalışmalardan farklı olarak Antalya'da seçilen Akıllı Kavşaklarda, trafik hacimleri tahmin edilmiştir. Doğrulama parametrelerine göre tahmin modelleri birbiriyle karşılaştırılmış ve başarı oranı yüksek tahmin modeli dijital ikiz mimarisinde uygulanmak üzere seçilmiştir. Böylece Antalya Büyükşehir Belediyesinin akıllı kavşaklarda işleyen adaptif trafik yönetim sisteminin dijital ikiz mimarisi kapsamında altyapısının oluşturularak iyileştirilmesi amaçlanmıştır.

Trafik ışıklarının ve trafik yönetim şemalarının planlanmasını ve işletilmesini desteklemek için trafik tahminlerinin oluşturulması dijital ikiz için önemli bir unsurdur. Gelecek trafik hacim verilerinin tahmin edilmesi, Akıllı Kavşaklarda oluşabilecek arıza, bakım gibi durumlarda kavşağın adaptif trafik yönetimi sağlayamayacak olması sebebiyle sinyalizasyon kavşağına sabit sinyal süresi verecek operatörlere yol gösterebilecektir. Operatörler gün içerisindeki farklı saatlerdeki trafik tahmin verine göre en uygun sinyal sürelerini çoklu planlar oluşturularak verebilecektir.

Bu amaçla çalışmada oluşturulmuş olan modellerden her iki kavşak ve tüm güzergahlar için Rasgele Orman Regresyon Modelinin MAE ve R² (0,91-0,98 arasında) performans parametreleri açısından diğer modellerden daha başarılı sonuçlar göstermiş olduğu görülmüştür. Tahmin modelleri başarı sırasına göre Rasgele Orman Regresyon Modeli, Destek Vektör Makineleri Modeli, Polinomal Regresyon Modeli, Lineer Regresyon Modeli olmak üzere belirlenmiştir. Böylece trafik hacmi yapılan Akıllı Kavşaklardaki veri türüne doğru göre hangi tahmin modellerinin seçilebileceği tespit edilmiştir.

Gelecekte yapılması hedeflenen çalışmalar:

- Bu çalışmada önerilen tahmin modellerinin benzetim programıyla oluşturulan ulaşım ağında uygulaması yapılarak trafik parametreleri açısından çıktılar elde edilecektir. Böylece literatürde çok rastlanılmayan bir şekilde tahmin modelleri ulaşım parametreleri açısından karşılaştırılacaktır.
- Daha fazla eğitim verisi toplanarak, modelin daha iyi geliştirilecek büyük veri kümesi, modelin daha iyi sonuçlar elde etmesine yardımcı olacaktır.
- Veri türüne göre doğru model seçimi tahmin başarısını artırmaktadır. Zaman serisi verileri için özel modeller (ARIMA, LSTM) oluşturularak daha büyük ve karmaşık veriler üzerinde tahmin başarıları araştırılacaktır.
- Bu çalışmada bağımsız değişken saatlik zaman dilimi olarak seçilmiştir. Gelecek çalışmalarda büyük veriler üzerinden günlerin ve ayların bağımsız değişken olarak eklenip modellerin daha iyi genelleme yapması beklenmektedir.

Teşekkür (Acknowledgment)

Çalışma '1649B032303124 'numaralı '2211-C Öncelikli Alanlara Yönelik Yurt İçi Doktora Programı' ile TÜBİTAK tarafından desteklenmiştir. Kullanılan gerçek trafik verilerini sağladığı için Antalya Büyükşehir Belediyesi, Ulaşım Planlama ve Raylı Sistem Dairesi Başkanlığı'na ve MOSAŞ GRUP'a kurumuna teşekkür ederiz.

Çıkar Çatışması Beyanı (Conflict of Interest Statement)

Yazarlar tarafından herhangi bir çıkar çatışması bildirilmemiştir.

Kaynaklar (References)

- [1] İ. Aynacı, "Dijital ikiz ve sağlık uygulamaları," *İzmir Kâtip Çelebi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 3, no. 1, pp. 70–79, 2020.
- [2] Z. Wang, K. Han and P. Tiwari, "Digital twin-assisted cooperative driving at non-signalized intersections," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 198–209, Jun. 2022, doi: 10.1109/TIV.2021.3100465.
- [3] E. Tomanik, A. J. J. Reyes, V. Tomanik and B. Tormos, "Machine-learning-based digital twins for transient vehicle cycles and their potential for predicting fuel consumption," *Vehicles*, vol. 5, no. 2, pp. 583–604, May 2023, doi: 10.3390/vehicles5020032.
- [4] A. Fuller, Z. Fan, C. Day and C. Barlow, "Digital twin: Enabling technologies, challenges and open research," *IEEE access*, vol. 8, pp. 108952–108971, May 2020, doi: 10.1109/ACCESS.2020.2998358.
- [5] T. Y. Fujii, V. T. Hayashi, R. Arakaki, W. V. Ruggiero, R. Bulla, F. H. Hayashi and K. A. Khalil, "A digital twin architecture model applied with MLOps techniques to improve short-term energy consumption prediction," *Machines*, vol. 10, no. 1, Jan. 2022, doi: 10.3390/machines10010023.
- [6] H. Xu, A. Berres, S. B. Yoginath, H. Sorensen, P. J. Nugent, J. Severino, S. A. Tennille, A. Moore, W. Jones and J. Sany, "Smart mobility in the cloud: enabling real-time situational awareness and cyber-physical control through a digital twin for traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, pp. 3145–3156, March 2023, doi: 10.1109/TITS.2022.3226746.
- [7] J. Lai, Z. Chen, J. Zhu, W. Ma, L. Gan, S. Xie and G. Li, "Deep learning based traffic prediction method for digital twin network," *Cognitive Computation*, vol. 15, no. 5, pp. 1748–1766, September 2023, doi: 10.1007/s12559-023-10136-5
- [8] "Real-time transportation management" aimsun.com, [Online]. Available: <https://www.aimsun.com/real-time-transportation-management/>. [Accessed: Mar. 27, 2023.]
- [9] S. Janković, A. Uzelac, S. Zdravković, D. Mladenović, S. Mladenović and I. Andrijanić, "Traffic volumes prediction using big data analytics methods," *International Journal for Traffic and Transport Engineering*, vol. 2, no. 0, pp. 184–198, 2021, doi: 10.7708/ijtte2021.11(2).01.
- [10] C. Bratsas, K. Koupidis, J. M. Salanova, K. Giannakopoulos, A. Kaloudis and G. Aifadopolou, "A comparison of machine learning methods for the prediction of traffic speed in urban places," *Sustainability (Switzerland)*, vol. 12, no. 1, Jan. 2020, doi: 10.3390/SU12010142.
- [11] F. M. N. Ali and A. A. M. Hamed, "Usage apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents," *Journal of Information and Telecommunication*, vol. 2, no. 3, pp. 231–245, Jul. 2018, doi: 10.1080/24751839.2018.1448205.
- [12] A. N. Espinoza, O. R. L. Bonilla, E. E. G. Guerrero, E. T. Cuatle, D. L. Mancilla, C. H. Mejia and E. I. Gonzalez, "Traffic flow prediction for smart traffic lights using machine learning algorithms," *Technologies (Basel)*, vol. 10, no. 1, p. 5, January 2022. <https://doi.org/10.3390/technologies10010005>
- [13] Y. Hou, Z. Deng and H. Cui, "Short-term traffic flow prediction with weather conditions: based on deep learning algorithms and data fusion," *Complexity*, vol. 2021, doi: 10.1155/2021/6662959.
- [14] W. Alajali, W. Zhou and S. Wen, "Traffic flow prediction for road intersection safety," in *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 812–820. October 2018, Guangzhou, China. Available: IEEE Xplore, <https://ieeexplore.ieee.org/document/8560129> [Accessed: 20 Sept. 2023].
- [15] "Strateji ve Eylem Planları - Haberleşme Genel Müdürlüğü," uab.gov.tr, Available: <https://hgm.uab.gov.tr/strateji-eylem-planlari>. [Accessed: Sept. 15, 2023].
- [16] "MAtchUP Antalya - Antalya Büyükşehir Belediyesi," matchupantalya.org, [Online]. Available: <http://www.matchupantalya.org/> [Accessed: Nov. 1, 2023].
- [17] A. Kemeç and H. Gül, "Antalya Büyükşehir Belediyesi örneğinde akıllı kent uygulamaları," *Kamu Yönetimi Ve Politikaları Dergisi*, 2(3), pp. 355–382.
- [18] M. Akın and Ş. Sağıroğlu, "Büyük veri perspektifinden trafik tahmini traffic prediction based on big data perspective," in *2019 1st International Informatics and Software Engineering Conference (UBMYK), 2019, Ankara, Turkey*, Available: IEEE Xplore <https://ieeexplore.ieee.org/document/8965545>. [Accessed: 02 Dec. 2023].
- [19] A. C. Müller and S. Guido, *Introduction to machine learning with Python*, O'Reilly Media, 2022.
- [20] J. Brownlee, *Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras*, Machine Learning

Mastery, 2016.

[21] A. Kowalczyk, *Support vector machines succinctly*,” Synfusion Inc, 2017.

[22] A. Boukerche and J. Wang, “Machine learning-based traffic prediction models for intelligent transportation systems,” *Computer Networks*, vol. 181, Nov. 2020. doi: 10.1016/j.comnet.2020.107530.

[23] “R Kare ve Düzeltmiş R Kare - Veri bilimi okulu,” veribilimiokulu.com, Available: <https://www.veribilimiokulu.com/r-kare-ve-duzeltilmis-r-kare/>, [Accessed: February. 16, 2023].

[24] “Regresyon analizi,” wikipedia.org, Available: https://tr.wikipedia.org/wiki/Regresyon_analizi, [Accessed: May. 26, 2023].

* This paper was presented at the 5th International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAIAME 2023) and the abstract was published as an e-book.

This is an open access article under the CC-BY license

