



Gerçek Hayat Verileriyle Makine Öğrenmesi Algoritmalarına Dayalı Otobüs Durak Süresi Tahmini

Bus Stop Time Prediction Based on Machine Learning Algorithms with Real-Life Data

Kevser Şahinbaş

İstanbul Medipol Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, ksahinbas@medipol.edu.tr
ORCID: <https://orcid.org/0000-0002-8076-3678>

MAKALE BİLGİLERİ

Makale Geçmişi:

Geliş 23 Mayıs 2022
Revizyon 29 Temmuz 2022
Kabul 01 Ağustos 2022
Online 30 Eylül 2022

Anahtar Kelimeler:

tahmin, otobüs varış süresi, toplu taşıma, veri yönetimi, YSA, DVR

ÖZ

Toplu taşıma sistemleri, gelişmekte olan ülkelerde ve nüfus yoğunluğunun yüksek olduğu bölgelerde büyük bir önem arz etmektedir. Yüksek popülasyona sahip şehirlerde kent içi aktif ulaşım süreçlerinin ve buna yönelik ihtiyaçların giderek yoğunlaştığı gözlemlenmektedir. Bu gereksinimden doğan araç sayısı fazlalığı ve yoğun trafik, büyük bir zaman dilimini kapsayarak günlük yaşamımızın önemli bir parçası haline gelmiştir. Bu sebeple ulaşım sistemleri yönetimi, toplu taşımacılık planlaması, planlamaların sürekli revize halinde olması ve kontrolü, kalabalık kentlerdeki günlük hayat akışında en önemli ihtiyaçlardan biridir. Bu çalışma, karayolu toplu taşımada kilit nokta olan otobüs verilerine dayanmaktadır. Çalışmanın amacı, İstanbul'da belirli bir hatta yapılan seferlerin yolculuk süre verilerinin analizi, duraklar arası sürenin ve durağa varış saati verilerinin analiz edilmesi ve gelecek günlere yönelik tahmin yapılmasıdır. Çalışma sırasında analiz edilen 522B hattı gidiş yönü verilerin tamamı gerçek verilerdir. Bu güzergaha ait veri seti 2021 yılının Temmuz ve Ağustos ayları bazında incelenmiştir. Makine öğrenmesi algoritmalarından Yapay Sinir Ağları (YSA) ve Destek Vektör Regresyon (DVR) yöntemlerinin, çeşitli trafik koşulları altında tahminler gerçekleştirirken oldukça rekabetçi olduğu ortaya çıkmaktadır. Karşılaştırmalı çalışmalar, YSA'nın daha doğru tahmin sonuçları sağladığını ve bir duraktan diğer durağa geçme süresi dağılımındaki belirsizlikleri daha etkin bir şekilde tahmin etme eğiliminde olduğunu göstermektedir.

ARTICLE INFO

Article history:

Received 23 May 2022
Received in revised form 29 July 2022
Accepted 01 August 2022
Available online 30 September 2022

Keywords:

prediction, arrival time, public transport, data management, ANN, SVR

Doi: 10.24012/dumf.1120379

* Sorumlu Yazar

ABSTRACT

Public transport systems are of great importance in developing countries and regions with high population density. In cities with high populations, it is observed that active urban transportation processes and the needs for this are increasingly intensified. The excess number of vehicles and heavy traffic arising from this need have become an important part of our daily life, covering a considerable time period. So much so that the moments when transportation comes to a standstill have become a routine in our lives. For this reason, transportation systems management, public transportation planning, constantly revising and controlling the plans are one of the most important needs in the flow of daily life in crowded cities. This study is based on bus data, which is the key point in road public transport. Studies on the prediction of stop arrival and journey times for future strategic planning in transportation management for vehicles are very important. The aim of the study is to analyze the travel time data of the trips made on a certain line in Istanbul, to analyze the time between stops and arrival time data and to make predictions for the future days. All of the 522B line going direction data analyzed during the study are real data. The data set for this route was examined on the basis of July and August 2021. Results from examines indicate that that Artificial Neural Networks (ANN) and Support Vector Regression (SVR) methods from machine learning algorithms are quite competitive when performing predictions under various traffic conditions. Comparative studies show that ANN provides more accurate estimation results and tends to predict uncertainties in the distribution of transition time from one stop to another more effectively.

Giriş

Nüfusun yoğun olduğu büyükşehirlerde doğru otobüs varış bilgilerinin sağlanması, yolcuların otobüs duraklarında bekleme sürelerini azaltmak için hayati önem taşımaktadır. Öğrencilerin hafta içi okul yolculuğu, vatandaşların iş yerlerine ulaşım yolculuğu, tatil başlangıç ve bitiş dönemlerinde oluşan yolculuk yoğunluğu, bozuk yol problemleri, kaza vb. faktörler ve mevsimsel şartların her biri bu ulaşım akışında çeşitli tıkanıklıklara ve olumsuz durumlara sebebiyet vermektedir. Bu rutindeki olumsuzlukların yanı sıra, sürekli artan nüfusun yolculuk taleplerinin ve istedikleri ulaşım hizmetlerinin de arttığı ve yoğunlaştığı gözlemlenmektedir. Fakat tüm bu faktörler günlük yaşantımızın bir parçası olmakla birlikte, zaruri olarak yaşanmaktadır. Tüm bu hayat akışı sebebiyle oluşan trafik kaçınılmazdır. Fakat özel araçlarla ulaşım dışında toplu taşıma sistemlerindeki stratejik planlamalar ve sefer kontrolleri daha planlı ve halkı mağdur etmeyecek bir yolculuk süreci sağlamaktadır. Bu sebeple kent içi toplu taşıma kuruluşlarının varlığı, bu kuruluşlarda yapılan sefer planlama ve yönetim stratejileri büyük bir önem arz etmektedir.

Bir ülkenin büyükşehirlerindeki şehir içi ulaşım sistemlerinin düzensiz ve plansız olması o şehre maliyet, zaman yönetimi ve mutsuz bir halk olarak olumsuz bir şekilde yansımaktadır. Böyle bir durumda yaşanan ulaşım tıkanıklıkları ve zaman kaybı, sıvı yakıtların gereksiz kullanımına sebebiyet vererek çevre kirliliğine de yol açmaktadır [1].

Türkiye’de tüm kentlerde toplu taşımacılık mevcut yerel yönetimlerin sorumluluğu olarak görülmektedir. Bu yüzden her şehirde belediyeler kentte ulaşım hizmeti vermek için, metro, otobüs, tramvay ve tren gibi vasıtaların güzergâh ve seferlerini organize eden, ulaşım akışını sağlayan çeşitli kuruluşları oluşturmuşlardır [2]. Bu kuruluşlardan biri de İstanbul Büyükşehir Belediyesine bağlı olan, İstanbul’da kent içi ulaşımın nabzını tutan İstanbul Elektrik Tramvay ve Tünel İşletmeleri, herkesçe bilinen adıyla İETT’dir. Bu çalışmada İETT kuruluşundan sağlanmış tek bir hat ve güzergaha ait 2 aylık sefer verileriyle analiz ve tahmin yapılmıştır. Toplu taşımada en sık kullanılan araçlardan biri olan otobüsle yapılan seferler incelenmiştir.

Otobüs yolculukları, çok sayıda seferle, kent içinde fazla sayıda güzergahla, güzergaha bağlı olarak belirlenmiş durak sayılarıyla kilit bir konumdadır. Seferlerin toplam yolculuk süresi dışında vatandaşı en fazla ilgilendiren husus ise, duraklar arası mesafede geçirilen süredir. İnsanlar toplu taşımayı kullanmak için buldukları lokasyona aracın geliş süresini bilerek yolculuk planı yapmak istemektedirler. Metro, tren, tramvay gibi araçların geliş süresi dakika cinsinden net bir şekilde gösterilmektedir. Fakat otobüslerin geliş süreleri kesin bir sonuç verememektedir. Çünkü yolun durumu, trafik yoğunluğu, aracın sefere planlanan vakitte başlayıp başlamadığı, araçtan doğru veri gelip gelmediği, hava şartları gibi birçok faktör anlık olarak çeşitli ulaşım uygulamalarında belirtilen süreler yansımayaabilmektedir. Bu sebeple insanlar yolculuk esnasında mağdur olabilmektedir.

Makine öğrenimi yöntemlerinden Yapay Sinir Ağları (YSA) ve Destek Vektör Regresyonu (DVR) otobüs varış ve geçiş işleminin doğrusal olmayan bozucu faktörlerini simüle etmek için kullanılır.

Otobüs varış süresi tahminiyle ilgili geniş bir araştırma grubu makine öğrenimi modelleri, Yapay Sinir Ağları [3]-[7], Rastgele Ormanlar [8], Destek Vektör Makinesi (DVM) [9]-[11]; regresyona dayalı yöntemler [12]- [13] ve hibrit modeller [14]-[15] olarak kategorize edilebilir.

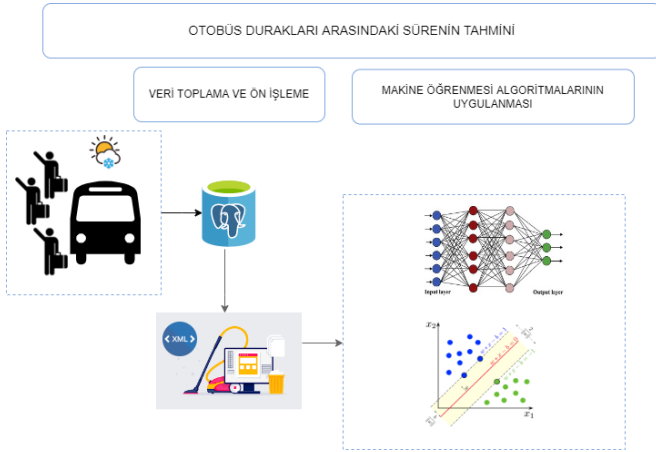
Pang ve diğerleri [16] tahmin doğruluğunu geliştirmek için uzun bir kısa süreli bellek bloğu ile önerilen tekrarlayan sinir ağını kullanarak farklı konumlarda otobüs varış zamanının uzun vadeli bağımlılıklarını yakalayan bir model önermektedir. Lin ve diğerleri [17] hiyerarşik sinir ağını kullanarak otobüs varış zamanını tahmin eden bir tahmin modeli önermiştir. Jeong ve diğerleri [18] otobüs varış zamanı tahmininde doğrusal regresyon modeli, parametrik olmayan regresyon modeli ve YSA modelinin performanslarını karşılaştırmıştır. YSA'nın tahmin doğruluğundaki üstünlüğü, gerçek dünyadaki veri yolu işletim verileriyle doğrulamaktadır. Yu ve diğerleri [19] otobüs varış zamanını tahmin etmek için DVR uygulamış ve uygulanabilirliğini kanıtlamıştır. DVR'nin diğer uygun modellerle geliştirildiği uygulama ile kanıtlanmış, belirli koşullar altında daha doğru tahminler elde edilebilmiştir [20]. Benzer şekilde, DVR'ye dayalı bazı yeni yöntemler önerilmiş ve kanıtlanmıştır [21]-[22].

Bu çalışmada amaç geçmişe dayalı gerçek verilerle, binen yolcu, durma süresi, bir duraktan diğer durağa varış süresi ve hava şartları gibi durumların da dahil olduğu bir zaman dilimiyle makine öğrenmesi algoritmalarına dayalı tutarlı bir tahmin çalışması yapmaktır. Bu çalışmanın diğer çalışmalarından farkı durma süresinin ve meteoroloji değerlerinin diğer çalışmalarda bulunmamasıdır. Bu tahmin çalışması doğrultusunda araçların duraklara ulaştığı sürenin en doğru şekilde sunulması için model gelişimi hedeflenmektedir. Veri kümesi analiz edilerek Python programlama diliyle algoritma geliştirilmiştir. Çalışmada makine öğrenmesi tahmin algoritmalarından Yapay Sinir Ağları ve Destek Vektör Regresyon yöntemleri kullanılmıştır.

Çalışmanın birinci bölümünü giriş kısmı oluşturmaktadır. İkinci bölümü ise veri setini tanıma, veri ön işleme, çalışmada kullanılan algoritmaların tanıtımı yer almaktadır. Üçüncü bölümde ise geliştirilen algoritmanın tahmin sonucunun değerlendirmesi ve sonuç kısımları yer almaktadır.

Yöntem

Bu çalışmada otobüs hattına ait; yolcu, durma süresi, duraklar arasındaki süre, hava durumu verileri elde edilmiş, ön işlemeden geçmiş ve daha sonra YSA ve DVR algoritmalarıyla duraklar arası tahmin yapılmıştır. Bu algoritmaların model performansını test etmek için Python yazılımı kullanılmıştır. Şekil 1’de çalışmanın kurgusu gösterilmektedir.



Şekil 1. Otobüs durakları arasındaki süre tahmini için çalışma kurgusu.

Bu çalışmada yapay sinir ağının çok katmanlı ileri beslemeli (multi layer feed forward) yapay sinir ağı mimarisi ve ağı öğrenme algoritması için de Geri Yayılım (Backpropagation) algoritması kullanılmıştır. 8 sürekli sayılardan oluşan girdi, yine sürekli sayılardan oluşan çıktı verisi ve gizli katmanlardan oluşan çok katmanlı algılayıcı olarak tanımlanmıştır. Binen Yolcu, DurmaSuresi, Durak, Sıcaklık, Dew_Point, NemOrani, RuzgarHizi, Basinc 8 girdi setini oluşturmaktadır. Çıktı seti ise varisSuresi verilerinden oluşmaktadır. Gizli katmandaki nöronların sayısı 8 ila 15 arasında olabilmektedir. Çıkış katmanında optimal sayıda nöron bulmak için deneme-yanılma yöntemini gizli katmandaki çeşitli sayılarla (100 ila 150 arasında değişen) nöronlar kullanılmıştır. Aktivasyon fonksiyonu olarak çıktı katmanı için mean_squared_error ve gizli katmanlar için selu fonksiyonu kullanılmıştır. Maksimum yineleme sayısı deneme yapılarak 5000 olarak belirlenmiştir. Diğer tüm parametreler Python yazılımında varsayılan değerler olarak atanmıştır. Tablo 1'de YSA algoritmasına ait parametreler gösterilmektedir.

Tablo 1. YSA Parametreleri.

YSA Parametreleri	Değer
Gizli Katman	100-150
Çekirdek fonksiyon	Selu
Yineleme Sayısı (epoch)	5000

Veriseti

Bu çalışmada kullanılan veriler, Mecidiyeköy-Yenidoğan bölgesine hizmet veren bir şehir içi güzergâh olan 552B No'lu otobüs güzergâhından toplanan otobüs gidiş verileri ve 30 Temmuz 2021-30 Ağustos 2021 tarihleri arasında 12.000'e yakın veriden ve 600'e yakın otobüs yolculuğundan oluşmaktadır.

Çalışılan veriler 2 aylık bir döneme ait olup, tek bir hat ve güzergâha aittir. Hat kodu isimli kolonda hattın ismi olan '522B' verisi görülmektedir. Güzergâh verisi ise hat kodunu ve gidiş dönüş bilgisini vermektedir. Çalışmada kullanılan veri yalnızca gidiş güzergâhı verisidir.

Güzergâh kolonunda görülen '522B_G_D0' bilgisindeki 'G' harfi gidiş yönü olduğu bilgisini vermektedir. Gidiş ve dönüş güzergâhlarındaki durak sayıları farklı olduğu için, tahminde devamlılık sağlamak adına tek bir güzergâh seçilmiştir.

Kapı No isimli veri, araçların üzerinde yazan tekil bir değeri göstermektedir. Her aracın kapı numarası farklıdır.

Kayıt günü verisi, seferin kaydının düzenlendiği tarihi vermektedir. Görev id verisi ise, her bir seferin kodunu ayrı bir şekilde vermektedir. Bir seferde, tek yönde 37 adet durak bulunmaktadır. 1. duraktan başlayıp 37. durağa gelerek yolculuğu bitirmek tek bir seferi belirtmektedir. Sıra kolonu 1'den 37'ye kadar olan durak sıralarını vermektedir. Durak kodu ise her durağa özel id bilgisini vermektedir. Durak kodları ve sıraları değişmemektedir.

Durma süresi verisi, durağa gelip durulduğunda beklenen süreyi saniye cinsinden yansıtmaktadır. Binen yolcu bilgisi ise her durakta binen yolcu sayısını ifade etmektedir. Geçiş zamanı verisi aracın ilgili duraktan geçtiği tarih ve saati vermektedir. Bazı seferler gece yarısından sonraya uzayabilir. Bu sebeple kayıt günü ile geçiş zamanındaki tarih bilgisi her zaman aynı olmayabilir. Etldate ise sisteme veri aktarımının yapıldığı tarihi belirtmektedir.

Verisetinin hazırlanması

Geçiş zamanı verilerinden faydalanarak duraklar arası süreyi hesaplamak için yeni bir kolon ekleme ihtiyacı doğmuştur. Bunu yapmak için duraktan durağa her bir geçiş zamanı yeni bir kolonla yan yana getirilmiştir. Her bir geçiş değerinden bir önceki zaman saniye bazında çıkarılarak 'Sonraki Fark' isimli yeni bir kolona yazılmıştır. Bu kolon türetilirken Sql dilinde lead, lag ve datediff fonksiyonları kullanılmıştır. Veride geline son durumda, sıra 1 de görünen sonraki fark değeri 2. durağa giderken geçen sürenin saniye bazında gösterimi olarak açıklanabilmektedir. Son durağa gelindiğinde 'null' değer geldiği için bu çalışmada sonuncu sıralardaki 'null' değerler çıkarılmıştır. Bir seferin tamamlanma süresi için yeni bir kolon daha oluşturulmuştur. 2. durağa giderken, 1. ve 2. durak süreleri toplanarak aynı işlemler diğer 37 durak için de kümülatif bir şekilde toplanarak gidilmiştir. Her bir sefer için kümülatif toplam tekrar edilmiştir. Bu sayede bir seferin tamamlanma süresi bulunmuştur. Bu işlemler sonunda veri setinin 9 nitelikten oluştuğu görülmektedir. Kalan veri seti analiz edildiğinde, durak sıralarındaki eksiklik saptanmıştır. Belli seferlerde art arda 3-4 adet duraktan veya 1. ve 37. duraklarda araçtan veri gelmediği anlaşılmıştır. Bu eksiklikler bölgedeki bağlantıdan, araç cihazlarının bozulmasından veya sinyal bozukluğundan kaynaklanabilmektedir. Veri setinin ön işleme aşamasının son etabında, eksik verisi olan seferler temizlenmiştir. Tüm verilerin eksiksiz bir şekilde geldiği görevlerin verileriyle çalışmaya devam edilmiştir. Veriler normalize edilmiştir.

Tahmin için kullanılan nitelikler Tablo 1'de gösterilmektedir. Hava durumuna ait veriler ise Meteoroloji Genel Müdürlüğünden alınmıştır.

Tablo 2. Tahmin için kullanılan nitelikler.

Nitelik Adı	Veri Tipi	Açıklama
DurmaSuresi	Nümerik	Otobüsün duraklardaki durma süresi
Durak	Nümerik	1'den 37'ye kadar olan durak sayısı
Temperature	Nümerik	Hava Sıcaklığı
Dew_Point	Nümerik	Çiğ Noktası
Humidity	Nümerik	Nem Oranı

3) Doğrusal olmayan problemler için, destek vektörü regresyonu, düşük boyutlu uzaydaki doğrusal olmayan problemleri istatistiksel öğrenme ve katı optimizasyon teorilerine dayalı olarak yüksek boyutlu Hilbert uzayına eşlemek için çekirdek fonksiyon teknolojisini kullanır ve eğitim hızı yalnızca sayı ile ilgilidir. “Aşırı öğrenme” problemlerinin üstesinden gelmekte ve iyi bir genelleme elde etmek için eğitim için yalnızca az sayıda örnek kullanabilmektedir [32].

Model Değerlendirme Kriteri

Sınıflandırma algoritmalarının sonuçlarını doğrulamak ve değerlendirmek için bu çalışmada Ortalama karesel hata (RMSE), Ortalama Kare Hatası (Mean Square Error-MSE) ve ortalama mutlak hata (MAE) ve belirleme katsayısı (R²) kullanılmıştır [33]. Denklemler aşağıdaki gibi tanımlanmaktadır:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2 \quad (4)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y(i) - \hat{y}(i)| \quad (5)$$

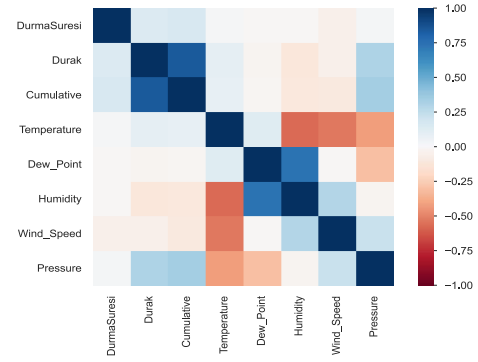
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2} \quad (6)$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}(i) - y(i))^2}{\sum_i (y_{mean}(i) - y(i))^2} \quad (7)$$

burada $y(i)$ gerçek değerdir, $\hat{y}(i)$ $y(i)$ 'nin tahmin edilen değeridir, $y_{mean}(i)$ $y(i)$ 'nin ortalama değeridir ve N örnek sayısıdır. MAE, mutlak hataların ortalamasını hesaplayarak duraklar arası tahmin hatalarını gösterir; RMSE, büyük sapmalarla örnekleme noktalarının katkılarını yükselterek bir tahmin modelinin kararlılığını gösterir; ve R^2 , durak varış süresinin değerlerinin ölçülen değerlere uygunluğunu gösterir. MAE ve RMSE'nin daha küçük değerlere sahip olması göstergelerin daha güçlü olduğunu göstermektedir.. R^2 aralığı [0, 1]'dir ve değeri 1'e ne kadar yakınsa, model ölçülen değerlere o kadar iyi uymaktadır.

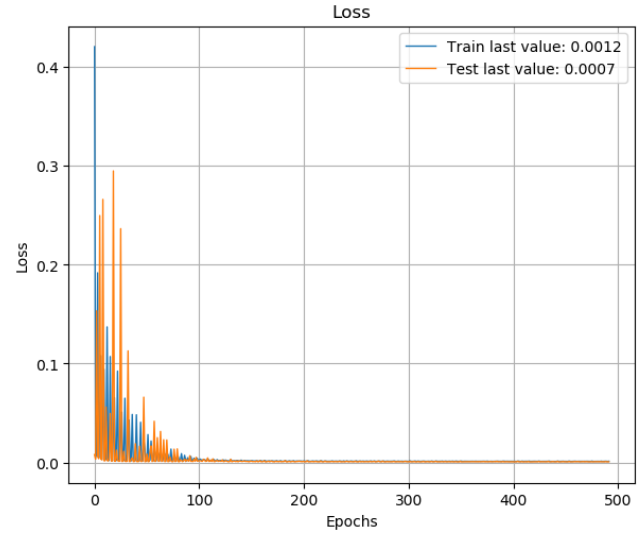
Analiz Sonuçları

Çalışmanın bu bölümünde Yapay Sinir Ağları ve Destek Vektör Regresyon sınıflandırma algoritmalarının otobüs sefer verileri üzerinde model önerisinde bulunmuş ve analiz sonuçları ayrıntılı bir şekilde gösterilmiştir. YSA ve DVR algoritmaları sınıflandırma modellerinde sıklıkla tercih edilmektedir.



Şekil 4. Veriler arasında korelasyon.

Şekil 4'te Pearson's korelasyon katsayıları görülmektedir. Korelasyon analizinin amacı, veri setindeki bir veya daha fazla bağımsız değişken arasındaki ilişkinin yönünü ve şiddetini göstermektir [34]. Durak değişkeni ile varma süresi değişkeni arasında yüksek ve pozitif bir ilişki olduğu görülmektedir. Durak sayısı ilerlerken varma süresinin de pozitif yönde arttığı gözlemlenmektedir. Durak değişkeni ile Basınç orta düzeyde ve çığ noktası arasında ise çok zayıf pozitif yönde ilişki gözlemlenmektedir. Durma süresi ile durak ve varma süresi arasında zayıf düzeyde pozitif yönde bir ilişki olduğu görülürken diğer değişkenler arasında çok zayıf ama pozitif bir ilişki olduğu görülmektedir. Çığ noktası ve nem oranı arasında yüksek ve pozitif yönde, sıcaklık ve nem oranı arasında yüksek düzeyde negatif yönde ilişki görülmektedir.



Şekil 5. YSA Tahmin Performansı

Tablo 3. YSA ve DVR Algoritmalarının RMSE, MSE, MAE ve R² Performans Değerleri

Model	RMSE	MSE	MAE	R ²
YSA	0.0223	0.00049	0.01396	0.853
DVR_rbf	0.06142	0.00377	0.0557	0.767
DVR_linear	0.042408	0.00179	0.0320	0.784
DVR_polynomial	0.050734	0.00257	0.0423	0.755

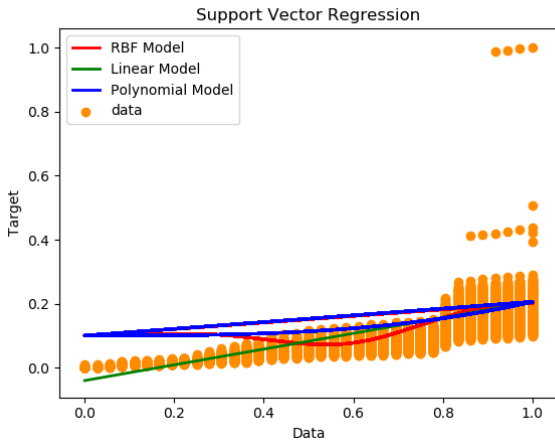
Bu çalışma, YSA ve DVR tabanlı modellerin performanslarını RMSE, MSE ve MAE ve R^2 açılarından sunmaktadır. YSA modelin performansı hesaplanmış ve sonuçlar Tablo 5'te gösterilmiştir. YSA performans değerlerine göre RMSE (0.0223), MSE (0.00049), MAE (0.01396)) skorları elde edilmiştir. Düşük skorlar değerlendirme için daha iyi olmaktadır. Bunun yanında R^2 (0.853) skoru elde edilmiştir, 1'e yakın olması modelin başarılı olduğunu göstermektedir.

Önerilen tahmin modeli duraklar arası geçişi tahmin etmektedir. Şekil 5 gözlemlenen değerleri ve tahmin değerlerini göstermektedir.

Gözlem verileri mavi çizgi ile gösterilmekte ve tahmin değerleri turuncu çizgi ile gösterilmektedir ve dolu alan tahmin ve gözlem arasındaki farkı gösterir. Modelleme performansları, mavi ve turuncu çizgiler arasındaki alan tarafından hesaplanır. Daha iyi performanslar alt alandadır. Tablo 3'teki sonuçlara göre, YSA modeli, duraklar arası süre tahmini için bariz bir avantaja sahiptir. YSA ile analiz edilen verilerin RMSE, MSE ve MAE ve R^2 değerleri sırasıyla 0.0223, 0.00049, 0.01396 ve 0.853'tür.

Veriseti DVR sınıfından kernel çekirdeği (rbf), doğrusal (linear), polinomial (poly) için türetilen nesnelere kullanılarak tahmin modelleri oluşturulmuştur. Şekil 6'da her bir çekirdeğe ait tahmin sonuçları görülmektedir.

Bu sonuçlara göre duraklar arası geçiş sürelerini tahmin etmede YSA algoritmasının çok başarılı olduğu gözlemlenmektedir.



Şekil 6. DVR çekirdekleri Tahmin Performansı

DVR modelinin performans değerleri Tablo 4'te gösterilmektedir. DVR performans değerlerine göre RMSE_rbf (0.06142), MSE_rbf (0.00377) ve MAE_rbf (0.0557), RMSE_lienar (0.042408), MSE_linear (0.00179) ve MAE_linear (0.0320) ve RMSE_polynomial (0.050734), MSE_polynomial (0.00257) ve MAE_polynomial (0.0423) skorları elde edilmiştir.

Şekil 6 üç modelin sonuçlarını göstermektedir. Veriler turuncu noktalarla, polynomial model mavi çizgi ile, linear model yeşil çizgi ile ve rbf model kırmızı çizgi ile gösterilmektedir.

Bu sonuçlara göre duraklar arası geçiş sürelerini tahmin etmede DVR algoritmasının başarılı olduğu gözlemlenmektedir.

Tablo 2'deki sonuçlara göre linear çekirdeği rbf ve polynomial çekirdeklerinden daha üstün performans göstermektedir. Performans hata oranı 0.00179 olarak oldukça düşüktür. Çok iyi performans sergilediği sonuçlarla kanıtlanmaktadır.

Tablo 2'den elde edilen sonuçlara göre en iyi sonucu veren algoritma YSA algoritması olduğu görülmektedir. Ardından DVR_linear gelmektedir. Her iki algoritma oldukça yüksek başarı performansı göstermektedir. Otobüs seferlerinde duraklar arasındaki süreyi hesaplamak için YSA ve DVR algoritmaları oldukça başarılı performans göstermektedir.

Sonuç ve Değerlendirme

Araçlara dair ulaşım yönetiminde yapılan, geleceğe yönelik stratejik planlamalar için durak varış ve yolculuk sürelerinin tahmin edilmesine yönelik çalışmalar oldukça önemlidir.

Nüfus yoğunluğunun git gide yoğunlaştığı, iş ve okul gereksinimlerinin fazlalaştığı ve dolayısıyla günlük hayattaki yolculuk akışı sebebiyle artan yolculuk hizmeti taleplerinin karşılanması büyükşehirlerdeki en büyük problemlerden biri olmaktadır. Bu süreçte yaşanan problemlerin ve tıkanıklıkların çözüme kavuşturulmaması, günlük hayatı durma noktasına getirerek ve ulaşım hizmeti talep eden halkı mutsuz ederek daha büyük olumsuzluklara neden olmaktadır. Trafik problemi ve düzgün ulaşım planlamalarını yapılmayıp, yolda geçen gereksiz bekleme süresi sebebiyle fazladan akaryakıt kullanımına neden olarak maliyet hususunda ve çevre kirliliği anlamında çok daha büyük problemlere neden olabilmektedir. Bu sebeple şehirlerdeki ulaşım hizmetleri titizlikle planlanmalı, daha kapsamlı çalışmalar ve iyileştirmeler yapılmalıdır. Ulaşım planlama stratejilerinde teknolojinin yetkin kullanımı oldukça önemlidir. Ulaşım planlamalarında kullanılan yöntemler, günümüz teknolojisinin metotlarından oldukça uzak konumdadır. Bu alanda yapay zekâ ve makine öğrenmesi çalışmaları yapılarak iyileştirmeler yapılmalıdır. Trafik akışına direkt veya dolaylı olarak etki eden unsurlar üzerinde titizlikle çalışılmalıdır. Araç sayıları, yol durumu, hava durumu, mevsimler koşulları ve kaza durumları üzerinde çalışılmalıdır. Her alanda makine öğrenmesi algoritmalarına dayalı tahminleme çalışmaları yapılmalıdır.

522B numaralı otobüs güzergahının otobüs geçiş tarihine göre ampirik bir çalışma yapılmıştır. YSA ve DVR yöntemlerinin, çeşitli trafik koşulları altında tahminler gerçekleştirirken oldukça rekabetçi olduğu ortaya çıkmıştır. Karşılaştırmalı çalışmalar, YSA'nın daha doğru tahmin sonuçları sağladığını ve bir duraktan diğer durağa geçme süresi dağılımındaki belirsizlikleri daha etkin bir şekilde tahmin etme eğiliminde olduğunu göstermektedir.

Bu çalışmada duraklar arasındaki süre hesaplatılarak bu süre üzerinden farklı modellerle eğitim yapılmıştır. Modeller arası kıyaslama yapılarak, her modelde grafiğin seyrinin aynı olduğu gözlemlenmiştir. Tüm modellerde hata değeri 0'a çok yakındır ve çok başarılı sonuçlar vermiştir. Gelecekteki çalışmalar için süre kapsamında akışın daha farklı seyrettiği güzergahlarla yeniden çalışmalar yapılmalı, kaza, yol durumu da çalışmaya dahil edilerek sistemlerin birbiriyle

entegre çalışması ile daha iyi sonuçlar elde edilerek çıktıları yolculuk planlama stratejilerinde kullanılmalıdır.

Etik kurul onayı ve çıkar çatışması beyanı

“Hazırlanan makalede etik kurul izni alınmasına gerek yoktur”

“Hazırlanan makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır”

Yazar Katkıları

Tüm araştırma ve yazma adımları ilgili yazara aittir.

Teşekkür

Çalışmada kullanılmak üzere veri paylaşımı sağlayan İstanbul Büyükşehir Belediyesine desteğinden dolayı teşekkür ederim.

Kaynakça

- [1] G. Temelcan, G. Sistem optimal bulanık trafik atama probleminin optimizasyonu, 2020.
- [2] B. T. Palamutçuoğlu, Şehir içi toplu taşıma planlamasında akıllı kart veri madenciliği ile sefer sıklığı optimizasyonu, 2020.
- [3] I.J. Chien, Y. Ding, Wei, C. “Dynamic bus arrival time prediction with artificial neural networks”, *J. Transp. Eng.*, 128 (5), 429–438, 2002.
- [4] R. Jeong, R. Rilett, “Bus arrival time prediction using artificial neural network model”. *In Proceedings. The 7th international IEEE conference on intelligent transportation systems (IEEE Cat. No. 04TH8749)* (pp. 988-993). IEEE, 2004.
- [5] H. Liu, K. Zhang, R., He, J., Li, “A neural network model for travel time prediction”, *in: Proc IEEE Conf Intell Comput Intell Syst (ICIS)*, Vol. 1, pp. 752–756, 2009.
- [6] M. Chen, X., Liu, J. Xia, S.I. Chien, A dynamic bus-arrival time prediction model based on APC data, *Comput.-Aided Civ. Inf.*, 19 (5), 364–376, 2004.
- [7] Y. Lin, X. Yang, N. Zou & L. Jia. “Real-time bus arrival time prediction: case study for Jinan, China”. *Journal of Transportation Engineering*, 139(11), 1133-1140, 2013.
- [8] B. Yu, H. Wang, W. Shan, B. Yao., Prediction of bus travel time using random forests based on near neighbors, *Comput.-Aided Civ. Inf.*, 33 (4) (2018) 333–350, 2018.
- [9] B. Yu, Z. Yang, J. Lin. “Bus arrival time prediction using support vector machines”, *J. Intell. Transp. S* 10 (4), 151–158, 2006.
- [10] D. S. Huang. *The Study of Data Mining Methods for Gene Expression Profiles*, 2009.
- [11] C. Bai, Z. Peng, Q. Lu, J. Sun. “Dynamic bus travel time prediction models on road with multiple bus routes”, *Comput. Intell. Neurosci.* (432389), 2015.
- [12] J. Patnaik, S. Chien, Bladikas, A. “Estimation of bus arrival times using APC data”, *J. Public Transp*, 7 (1), 1–20, 2004.
- [13] Y. Zhou, L. Yao, Y. Chen, Y. Gong, J. Lai. “Bus arrival time calculation model based on smart card data”, *J. Comput.* 74 (8), 1634–1638, 2017.
- [14] B Yu, Z. Z., Yang, K. Chen, , B. Yu,” Hybrid model for prediction of bus arrival times at next station”, *J. Adv. Transp.* 44 (3), 193–204, 2010.
- [15] S Zhong, J. Hu, S. Ke, X. Wang , J. Zhao, B. Yao, (2015). “A hybrid model based on support vector machine for bus travel-time prediction”, *Promet - Traffic – Traffico*, 27 (4), 291–300, 2015.
- [16] J. Pang, J. Huang, Y. Du, H. Yu, Q. Huang, & B. Yin, .”Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network”. *IEEE Transactions on Intelligent Transportation Systems*, 20(9), 3283-3293, 2018.
- [17] Y. Lin, X. Yang, N. Zou, N. ve L. Jia. “Real-time bus arrival time prediction: Case study for Jinan, China,” *J. Transp. Eng.*, 139 (11), 1133–1140, 2013.
- [18] R. Jeong, L.R. Rilett, L.R. “Bus arrival time prediction using artificial neural network model”, *in: Proc IEEE 7th Conf Intell Transp Syst (ITSC)*, pp. 988–993, 2004.
- [19] H Yu, D. Chen, Z. Wu, X. Ma, & Y. Wang, “Headway-based bus bunching prediction using transit smart card data”. *Transportation Research Part C: Emerging Technologies*, 72, 45-59, 2016.
- [20] J. Chai, C. Wu, C. Zhao, H. L. Chi, X. Wang, B. W. K. Ling, & K. L. Teo, “Reference tag supported RFID tracking using robust support vector regression and Kalman filter”. *Advanced Engineering Informatics*, 32, 1-10, 2017.
- [21] B. Yu., W.H.K. Lam, L.T. Mei. “Bus arrival time prediction at bus stop with multiple routes”, *Transp. Res. C* 19 (6), 1157–1170, 2011.
- [22] X. Liu, J. Jin, W. Wu, & F. Herz. “A novel support vector machine ensemble model for estimation of free lime content in cement clinkers”. *ISA transactions*, 99, 479-487, 2020.
- [23] D. J. Sargent, “Comparison of artificial neural networks with other statistical approaches-results from medical data sets”, *Cancer*, 91(8), 1636-1642, 2001.
- [24] A. Abraham, (2004). “Meta-Learning Evolutionary Artificial Neural Networks”, *Neurocomputing Journal*, 56, 36-37.
- [25] Z. A. Vassilis, M. P. Dimitris ve G. A Vassilis, V. A. Athanasios. “Solar radiation estimation methods using ANN and empirical models”, *Computers and Electronics in Agriculture*, 160, 160–167, 2019.
- [26] C. Fyfe, C. (2000). *Artificial neural networks and information theory*, 2000.
- [27] C. Cortes, V. Vapnik. “Support-vector networks”, *Machine Learning*, 20, 273-97, 1995.
- [28] A. J. Smola, “A tutorial on support vector regression”. *Statistics and Computing*, 1998.

- [29] D.S. Huang & W. B. Zhao. "Determining the centers of radial basis probabilistic neural networks by recursive orthogonal least square algorithms". *Applied Mathematics and Computation*, 162(1), 461-473, 2005.
- [30] Z. Liu, Y. Xu, G. Duan, C. Qiu., & J. Tan. (2021). "Accurate on-line support vector regression incorporated with compensated prior knowledge". *Neural Computing and Applications*, 33(15), 9005-9023, 2021.
- [31] X. Tang., Z. Ma, Q. Hu & W. Tang. "A real-time arrhythmia heartbeats classification algorithm using parallel delta modulations and rotated linear-kernel support vector machines". *IEEE Transactions on Biomedical Engineering*, 67(4), 978-986, 2019.
- [32] C. S. Lo, & C. M. Wang, C. M. "Support vector machine for breast MR image classification". *Computers & Mathematics with Applications*, 64(5), 1153-1162.
- [33] C. J. Willmott. "Some comments on the evaluation of model performance". *Bulletin of the American Meteorological Society*, 63(11), 1309-1313, 1982.
- [34] R. Taylor. "Interpretation of the correlation coefficient: a basic review". *Journal of diagnostic medical sonography*, 6(1), 35-39, 1990.