



Türkiye'deki il yolları için Yıllık Ortalama Günlük Trafik (YOGT) tahmininde makine öğrenmesi algoritmalarının karşılaştırmalı analizi

A comparative analysis of machine learning algorithms for estimating Annual Average Daily Traffic (AADT) on provincial roads in Turkey

Serkan Biçici^{1,*} 

¹Artvin Çoruh Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, 08100, Artvin, Türkiye

Öz

Yıllık Ortalama Günlük Trafik (YOGT), bir yoldan bir yıl boyunca geçen ortalama araç sayısını ifade etmektedir. Türkiye'de YOGT verileri Karayolları Genel Müdürlüğü tarafından üç farklı idari yol sınıfı için toplanmaktadır. Devlet ve otoyollar için YOGT verisi her yıl düzenli olarak elde edilmekte, ancak il yolları için bu veri genellikle üç veya dört yılda bir toplanmaktadır. Bu çalışmada, il yolları için YOGT tahmini amacıyla yedi farklı makine öğrenmesi algoritması kullanılmış ve performansları karşılaştırılmıştır. Yolu karakteristik özellikleri, diğer ulaşım sistemleriyle ilişkisini ve yol çevresinin demografik/sosyo-ekonomik özelliklerini yansıtan değişkenler kullanılmıştır. Rastgele orman ve destek vektör regresyonu algoritmaları en başarılı sonuçları vermiştir. Yolu karakteristik özellikleri ile demografik/sosyo-ekonomik faktörleri temsil eden değişkenlerin YOGT tahminleri üzerindeki etkisinin, diğer ulaşım sistemleri merkezleriyle olan ilişkileri yansıtan değişkenlere kıyasla çok daha belirleyici olduğu bulunmuştur.

Anahtar kelimeler: YOGT, İl yolları, Makine öğrenmesi, Rastgele orman, Destek vektör regresyonu

1 Giriş

Yıllık Ortalama Günlük Trafik (YOGT), belirli bir noktadan bir yıl boyunca geçen ortalama araç sayısını ifade eder [1, 2]. YOGT değerlerinin toplanması ve hesaplanması hem önemli ölçüde zaman ve kaynak gerektirir hem de masraflıdır [3]. Bu yüzden, her ülke YOGT ile ilgili verilerini toplarken çeşitli standartlar benimseyebilir ve farklı teknolojiler kullanabilir [2, 3]. Türkiye'de YOGT ile ilgili veriler, Ulaştırma ve Altyapı Bakanlığı'na bağlı Karayolları Genel Müdürlüğü (KGM) tarafından toplanmaktadır. Bu süreçte iki farklı teknoloji kullanılmaktadır. İlki, yol altına yerleştirilen sabit manyetik döngülü cihazlarla, yıl boyunca kesintisiz şekilde trafik verileri toplanmaktadır. 2023 yılı itibarıyla Türkiye genelinde yol ağının 527 noktasında (yaklaşık yol ağının %13'ü) bu cihazlar yer almaktadır. Diğer teknoloji ise, yolun üzerine yerleştirilen taşınabilir hava basınçlı hortumlu cihazlardır. Yıl boyunca her mevsimde belirli zaman aralıklarında bu cihazlar kullanılarak trafik verileri elde

Abstract

Annual Average Daily Traffic (AADT) is the average number of vehicles passing on a road over the course of a year. In Turkey, AADT data is collected by the General Directorate of Highways for three different road classes. For state roads and highways, AADT data is collected regularly every year, but for provincial roads, this data is usually collected every three or four years. In this study, seven different machine learning algorithms were used to estimate AADT for provincial roads and their performances were compared. Variables reflecting the characteristics of the road, its relationship with other transportation systems and the demographic/socio-economic characteristics of the road environment were used. Random forest and support vector regression algorithms produced the most successful results. Variables representing road characteristics and demographic/socio-economic factors were found to have a much more significant impact on AADT estimates than variables reflecting the relationship with other transportation systems.

Keywords: AADT, Provincial roads, Machine learning, Random forest, support vector regression

edilmektedir. Her iki teknolojiyi kullanarak toplanan veriler, YOGT hesaplamalarında kullanılmaktadır.

Türkiye'de karayolları, idari sınıflarına göre dört gruba ayrılmaktadır: otoyollar, devlet yolları, il yolları ve köy yolları. Manyetik döngülü cihazlar genellikle otoyollar ve devlet yollarında kullanılmakta olup, KGM bu yol sınıfları için YOGT değerlerini her yıl düzenli bir şekilde hesaplayarak çevrimiçi olarak sunmaktadır [4]. Öte yandan, il yollarına ilişkin YOGT değerleri üç ila dört yılda bir toplanıp hesaplanmaktadır. Ayrıca, bu yollarda yıl boyunca kesintisiz şekilde veri toplayan manyetik döngülü cihazlar bulunmamaktadır. Dolayısıyla, il yollarındaki YOGT değerleri yalnızca yılın belirli dönemlerinde sayım yapan taşınabilir hava basınçlı hortumlu cihazlar aracılığıyla elde edilmektedir. Bu nedenle, il yolları için YOGT ile ilgili veri toplamak daha masraflı ve zaman alıcıdır. Ayrıca, YOGT değerini hesaplamak bazı varsayımlar kullanılmaktadır.

Bu çalışmada, Türkiye'deki il yolları için YOGT değerlerinin tahmin edilmesi amacıyla çeşitli makine öğrenmesi (MÖ) algoritmaları kullanılmıştır. Köy yolları ise

* Sorumlu yazar / Corresponding author, e-posta / e-mail: serkanbicici@artvin.edu.tr (S. Biçici)

Geliş / Received: 19.06.2025 Kabul / Accepted: 29.09.2025 Yayınlanma / Published: 15.10.2025

doi: 10.28948/ngumuh.1723436

bu çalışmaya dahil edilmemiştir; zira bu yollar, köyler arası bağlantıyı sağlayan, düşük trafik hacmine sahip ve gerekli geometrik standartları karşılamayan yollar olarak değerlendirilmektedir.

YOGT değerlerinin olabildiğince doğru bir şekilde tahmin edilmesi, yol tasarımı, bakım-onarım planlaması, yatırım önceliklendirmesi, kaza analizi ve çevresel etki değerlendirmeleri gibi birçok alanda daha verimli ve etkili kararlar alınmasına olanak sağlar [2, 3]. Bu sayede kamu kaynakları daha etkin kullanılırken, güvenli ve sürdürülebilir bir ulaşım sistemi oluşturulabilir.

Bu makalenin devamı aşağıdaki şekilde düzenlenmiştir. Bir sonraki bölümde YOGT değerlerini tahmin etmek için kullanılan yöntemlerin literatür taraması özetlenmektedir. Ardından, bu çalışmada toplanan çeşitli veriler ve kullanılan yöntem açıklanmaktadır. Daha sonra bulgular tartışılmakta ve çalışma sonuç bölümü ile sona ermektedir.

2 Literatür taraması

Yollara ait YOGT değerlerinin tahminine yönelik çalışmalar yaklaşık 30 yıllık geçmişe sahiptir. İlk olarak çeşitli çalışmalarda basit çoklu regresyon analizi (ÇRA) yöntemi kullanıldı [5-7]. Bu yöntem, bağımlı bir değişken ile birden fazla bağımsız değişken arasındaki ilişkileri modellemek için kullanılan bir istatistiksel yöntemdir. ÇRA çok sık kullanıldığı gibi coğrafi ağırlıklı regresyon [8, 9], bayes doğrusal regresyon [10], lojistik regresyon [11] ve kantil regresyon [12] gibi farklı yaklaşımlarla regresyon modelleri YOGT değerleri tahminlerinde uygulandı.

Son zamanlarda, bazı çalışmalar daha kapsamlı değişkenler ve MÖ algoritmaları gibi gelişmiş teknikler kullanarak YOGT değerlerini daha doğru bir şekilde tahmin etmektedir. Örneğin, Sfyridis ve Agnolucci [2] Galler ve İngiltere'deki tüm yolların YOGT değerlerini tahmin etmek için iki farklı MÖ algoritması kullandı. Bu algoritmalar destek vektör regresyonu (SVR) ve rastgele orman (RF) algoritmalarıdır. Çalışmada, bu modellerin performansları değerlendirmek için ortalama mutlak yüzde hatası (MAPE) ve kök ortalama karesel hata (RMSE) ölçütleri kullandı. SVR algoritması, %277'ye kadar MAPE ve 10.410'a kadar RMSE ölçüleri üretti. Benzer şekilde, RF algoritması, %288'e kadar MAPE ve 2.609'a kadar RMSE ölçüleri üretti. Bu iki algoritmanın ÇRA yöntemine göre daha iyi bir performans gösterdiği tespit edildi. Zhang ve Chen [3] iki farklı MÖ algoritması (sinir ağları ve RF) kullandı ve performansları ÇRA yöntemiyle karşılaştırdı. RF algoritmasının 0,92 R-kare ile en iyi performansı gösterdiği bulundu. Sun ve Das [13] Amerika'nın Louisiana eyaletinde bulunan yolların YOGT değerlerini tahmin etmek için bir MÖ algoritması olan SVR kullandı. Bu algoritma ile elde edilen YOGT tahminleri, diğer modellerle elde edilen tahminlere göre daha iyi sonuçlar verdi. Ayrıca, bu çalışmada SVR algoritmasının tahminleri YOGT değeri 1.500'den yüksek olan yollarda daha düşük tahmin etme eğiliminde olduğu gözlemlendi. Dünya genelinde farklı bölgelerde çeşitli yöntemlerle YOGT tahminine ilişkin çalışmalar yapılmasına rağmen, Türkiye bağlamında bu konuyu ele alan kapsamlı bir çalışmaya rastlanmamıştır.

RF ve SVR algoritmaları YOGT tahminlerinde çok sık kullanılan iki farklı MÖ algoritmasıdır. Ayrıca, k-en yakın komşu (KNN), aşırı gradyan artırma (XgbLinear ve XgbTree), genelleştirilmiş doğrusal model (GLM) ve gradyan boosting makinesi (GBM) gibi MÖ algoritmaları; trafik sıkışıklığı, trafik kazaları, seyahat süresi, toplu taşımadaki yolcu akışı, gerçek zamanlı trafik akışı gibi ulaştırma ile ilgili konuların tahmininde kullanıldı [14-18]. Dolayısıyla, bu çalışmada, Türkiye'deki il yollarının 2023 yılına ait YOGT değerlerini tahmin etmek için RF, SVR, KNN, XgbLinear, XgbTree, GLM ve GBM olmak üzere yedi farklı MÖ algoritması kullanılmıştır ve performansları değerlendirilmiştir.

3 Materyal ve metod

Bu çalışmada iki farklı türde veri elde edilmiştir. Bu veriler, her il yolu için YOGT değerleri ile bu değerlere karşılık gelen yol ve çevresel özellikleri temsil eden bağımsız değişkenlerden oluşmaktadır.

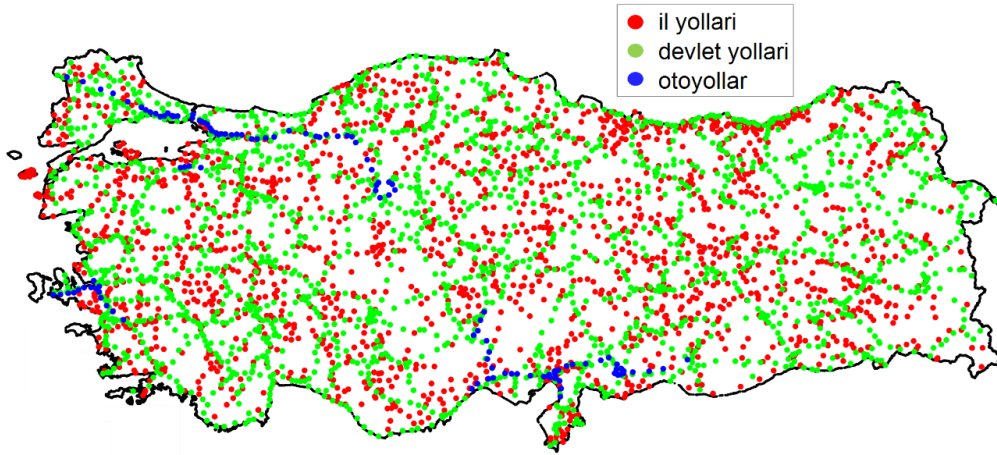
3.1 YOGT değerleri

KGM, her yıl mayıs ayında bir önceki yılın yollarına ait YOGT değerlerini kapsamlı bir rapor halinde KGM'nin resmi sitesinde çevrimiçi olarak sunmaktadır [4]. YOGT değerleri; otoyollar, devlet yolları ve il yolları olmak üzere üç farklı yol sınıfı için hesaplanmaktadır. 2023 yılında Türkiye'de 1.889 il yolu, 1.968 devlet yolu ve 118 otoyol bulunmaktadır. Şekil 1, bu üç yol sınıfına ait yolların lokasyonlarını Türkiye haritası üzerinde göstermektedir. KGM, otoyollar ve devlet yolları için her sene YOGT değerlerini düzenli olarak hesaplamaktadır. Fakat il yollarına ait YOGT ile ilgili veriler her üç veya dört yılda bir toplanmaktadır. Bu nedenle bu çalışmada il yolları için YOGT değerleri tahmin edilmiştir.

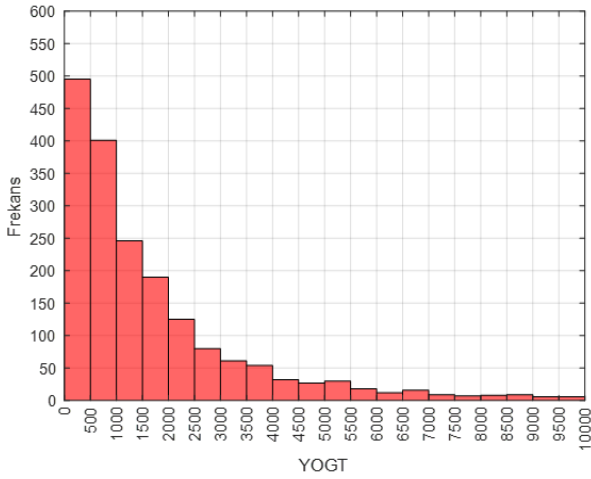
Şekil 2'de, 2023 yılına ait il yollarının YOGT değerleri için oluşturulan histogram grafiği sunulmaktadır. Toplam 1.889 il yolunun ortalama YOGT değeri 2.219,7 araç/gün olarak hesaplanmıştır. Farklı YOGT değer aralıklarındaki il yollarının Türkiye üzerindeki mekansal dağılımını incelemek amacıyla, YOGT değeri en düşük 500 il yolunun konumları Şekil 3(a)'da, en yüksek 500 il yolunun konumları ise Şekil 3(b)'de gösterilmektedir. Şekil 3(a)'ya bakıldığında, en düşük 500 YOGT değerine sahip il yollarının doğu ve iç kesimlerde yoğunlaştığı gözlemlenmiştir. Buna karşılık, Türkiye'nin batı ve kıyı bölgelerinde yer alan il yollarının YOGT değerlerinin diğer bölgelere kıyasla daha yüksek olduğu görülmüştür (Şekil 3(b)).

3.2 Bağımsız değişkenler

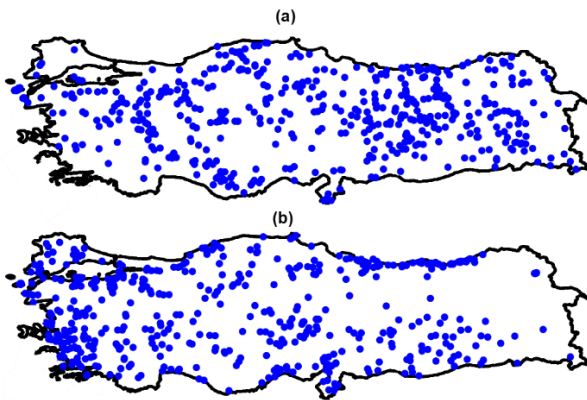
Bu çalışmada YOGT değerleri ile birlikte yol ile ilgili çeşitli bağımsız değişkenler de kullanılmıştır. Bu bağımsız değişkenler iki farklı kaynaktan elde edilmiştir. Bunlar, Türkiye Ulusal Coğrafi Bilgi Sistemi (TUCBS) ve Türkiye İstatistik Kurumu (TÜİK) web sitesidir.



Şekil 1. 2023 yılına ait otoyollar, devlet yolları ve il yollarının lokasyonları.



Şekil 2. 2023 yılına ait il yollarının YOGT değerlerine göre histogram grafiği.



Şekil 3. 2023 yılına ait YOGT değerlerine göre sınıflandırılmış il yollarının lokasyonları.

TUCBS, kamu kurum ve kuruluşlarına ait bilgilerin harita tabanlı bir ortamda görüntülenmesi ve paylaşılması amacıyla kullanılan bir platformdur [19]. Bu platform kamuya açık olmadığından, veriler resmi bir talep yoluyla temin edilmiştir. Bu platform kullanılarak, Ulaştırma ve Altyapı Bakanlığı bünyesindeki KGM'den, il yolları ile ilgili

lokasyon, uzunluk, kavşak tipi, otokorkuluk tipi, refüj genişliği ve cinsi, sinyalizasyon sistemi, kaplama genişliği ve cinsi, şerit adedi, platform tipi ve banket genişliği ve cinsi bilgileri elde edilmiştir. Bu bilgiler yolun karakteristik özelliklerini ifade eden bağımsız değişkenlerdir.

Uzunluk, refüj genişliği, kaplama genişliği ve banket genişliği gibi değişkenler sayısal (SAY) verilerdir. Diğer yandan, kavşak tipi, otokorkuluk tipi, refüj cinsi, sinyalizasyon sistemi, kaplama cinsi, şerit adedi, platform tipi ve banket cinsi kategorik (KAT) veri türüne sahiptir. Kategorik veriler, farklı sınıfları temsil eden kategorik değişkenlerden oluşmaktadır. Örneğin, il yollarında iki farklı kavşak tipi bulunmakta; benzer şekilde üç farklı otokorkuluk tipi, beş farklı refüj cinsi, iki farklı kaplama cinsi ve iki farklı banket cinsi yer almaktadır. Bu tür değişkenler, modellerde sayısal değer yerine kategorik değer olarak yer alır. Tablo 1'de, bu çalışmada kullanılan 12 adet yol karakteristiğini temsil eden bağımsız değişkenler, kısaltmaları, açıklamaları ve veri türleri ile birlikte listelenmektedir. Ayrıca, açıklamalar bölümünde kategorik bağımsız değişkenlerin aldığı kategorik değerler detaylı şekilde açıklanmaktadır.

Bu çalışmada, yol karakteristik özelliklerini ifade eden değişkenlere ek olarak, il yollarının diğer ulaşım sistem merkezlerine olan yakınlıklarının da YOGT değerleri üzerindeki etkisi incelenmiştir. Bu nedenle, TUCBS platformu kullanılarak Devlet Hava Meydanları İşletmesi Genel Müdürlüğü'nden havaalanlarının tür ve lokasyon bilgileri, Denizcilik Genel Müdürlüğü'nden limanların tür ve lokasyon bilgileri ile Demiryolları Genel Müdürlüğü'nden demiryolu istasyonlarının tür ve lokasyon bilgileri temin edilmiştir. Bu bilgiler doğrultusunda, il yollarının diğer ulaşım sistemleri merkezlerine olan yakınlıklarının YOGT değerleri üzerindeki etkisi analiz edilmiştir. Bu amacı gerçekleştirmek için çeşitli kategorik değişkenler tanımlanmıştır. Her bir il yolunun, en yakın ulaşım sistemi merkezine olan mesafesinin belirli bir eşik değerinin altında ya da üstünde olmasına göre kategorik veriler oluşturulmuştur. Tanımlanan 6 adet kategorik bağımsız değişkenin açıklamaları Tablo 1'de sunulmaktadır.

Yol çevresine ait demografik/sosyo-ekonomik veriler, TÜİK web sitesinden temin edilmiştir [20]. TÜİK, Türkiye

genelinde adalet, sağlık, sanayi ve nüfus gibi çeşitli alanlara ilişkin kapsamlı istatistiksel verileri neredeyse her yıl güncelleyerek çevrimiçi olarak erişime sunmaktadır. Bu çalışma kapsamında, 2023 yılına ait il ve/veya ilçe düzeyinde nüfus yoğunluğu, hane halkı bilgileri, araç kayıt istatistikleri ve gelir düzeylerine ilişkin veriler elde edilmiştir.

Hane halkı ve gelir düzeyine ilişkin bilgiler, sayısal veri olarak değerlendirilmiştir. Öte yandan, nüfus yoğunluğu ve kayıtlı araç sayısı değişkenleri için kategorik veriler üretilmiştir. Her bir il yolunun, belirli bir nüfus yoğunluğuna veya kayıtlı araç sayısına sahip yerleşim yerlerine olan mesafesinin belirli bir eşik değerin altında ya da üstünde olması durumuna göre bu kategorik veriler oluşturulmuştur. Bu dönüşüm, aynı şehirde yer almasına rağmen farklı coğrafi konumlara sahip yolları ayırt edebilmek amacıyla yapılmıştır. Tanımlanan 8 adet demografik/sosyo-ekonomik değişkenin açıklamaları Tablo 1’de sunulmaktadır.

Bu çalışmada kullanılan eşik ve sınır değerleri, yapılan ön analizler sonucunda belirlenmiştir. Nüfus ve araç sayısına ilişkin sınır değerlerin model sonuçları üzerinde sınırlı etkisi olduğu gözlemlenmiştir. Buna karşılık, eşik değerlerin daha belirgin etkisi olduğu tespit edilmiştir. Bu nedenle birden fazla eşik değeri kullanılmıştır. Kullanılan tüm sınır ve eşik değerleri, ön değerlendirme sonuçlarına dayalı olarak seçilmiştir.

Sonuç olarak, 12 tane bağımsız değişken yolun karakteristik özelliklerini ifade etmekte, 6 tane bağımsız değişken yol ile diğer ulaşım sistemleri arasındaki ilişkiyi ifade etmekte ve 8 tane bağımsız değişken ise demografik/sosyo-ekonomik özellikleri ifade etmektedir. Yani, bu çalışmada toplam 26 tane bağımsız değişken kullanılmıştır. Bu 26 tane bağımsız değişken farklı kombinasyonlar ile kullanılarak daha iyi YOGT değerleri tahminleri yapılması amaçlanmıştır.

Tablo 1. Yol ve yol çevresi ile ilgili 26 tane bağımsız değişkenler, kısaltmaları, açıklamaları ve veri türleri.

	Bağımsız Değişkenler	Kısaltma	Açıklama	Tür
Yolun Karakteristik Özellikleri	Uzunluk	uz	Kilometre cinsinden yolun uzunluğu	SAY
	Kavşak tipi	kav	Eğer yolda kavşak bulunmuyorsa 0, eğer yolda 1.tip kavşak bulunuyorsa 1, eğer yolda 4.tip kavşak bulunuyorsa 2	KAT
	Otokorkuluk tipi	oto	Eğer yolda otokorkuluk bulunmuyorsa 0, eğer yolda 1.tip otokorkuluk bulunuyorsa 1, eğer yolda 2.tip otokorkuluk bulunuyorsa 2	KAT
	Refüj genişliği	ref_g	Metre cinsinden yoldaki ortalama refüj genişliği	SAY
	Refüj cinsi	ref_c	Eğer yolda refüj bulunmuyorsa 0, eğer yolda 1.tip refüj bulunuyorsa 1, eğer yolda 2.tip refüj bulunuyorsa 2, eğer yolda 3.tip refüj bulunuyorsa 3, eğer yolda 4.tip refüj bulunuyorsa 4, eğer yolda 5.tip refüj bulunuyorsa 5	KAT
	Lamba tipi	lam	Eğer yolda lamba bulunmuyorsa 0, eğer yolda 1.tip lamba bulunuyorsa 1, eğer yolda 3.tip lamba bulunuyorsa 2, eğer yolda 4.tip lamba bulunuyorsa 3	KAT
	Kaplama genişliği	kap_g	Metre cinsinden yoldaki ortalama kaplama genişliği	SAY
	Kaplama cinsi	kap_c	Eğer yolda 1.tip kaplama bulunuyorsa 1, eğer yolda 3.tip kaplama bulunuyorsa 2	KAT
	Şerit adedi	şerit	Yol üzerinde her iki yöndeki toplam şerit sayısı	KAT
	Platform tipi	pla	Eğer yol platformu bölünmemişse 0, eğer yol platformu bölünmüşse 1	KAT
	Banket genişliği	ban_g	Metre cinsinden yoldaki ortalama banket genişliği	SAY
	Banket cinsi	ban_c	Eğer yolda banket bulunmuyorsa 0, eğer yolda 1.tip banket bulunuyorsa 1, eğer yolda 3.tip banket bulunuyorsa 2	KAT
Diğer Ulaşım Sistemleri Özellikleri	Havaalanı lokasyonları	h20	Eğer 20 km içinde sivil havaalanı varsa 1, yoksa 0	KAT
		h50	Eğer 50 km içinde sivil havaalanı varsa 1, yoksa 0	KAT
	Liman lokasyonları	l20	Eğer 20 km içinde liman varsa 1, yoksa 0	KAT
		l50	Eğer 50 km içinde liman varsa 1, yoksa 0	KAT
	Demiryolu lokasyonları	d20	Eğer 20 km içinde demiryolu istasyonu varsa 1, yoksa 0	KAT
		d50	Eğer 50 km içinde demiryolu istasyonu varsa 1, yoksa 0	KAT
Demografik/Sosyo-Ekonomik Özellikleri	Hane halkı	hh	Yola en yakın il/ilçedeki ortalama hane sayısı	SAY
	Gelir	tl	Yola en yakın il/ilçenin TL cinsinden ortalama geliri	SAY
	Toplam nüfus	n20	Eğer 20 km içinde toplam 500.000 nüfuslu bir il/ilçe varsa 1, yoksa 0	KAT
		n50	Eğer 50 km içinde toplam 500.000 nüfuslu bir il/ilçe varsa 1, yoksa 0	KAT
		n100	Eğer 100 km içinde toplam 500.000 nüfuslu bir il/ilçe varsa 1, yoksa 0	KAT
	Araç kayıt istatistiği	m20	Eğer 20 km içinde toplam 100.000 kayıtlı araç olan bir il varsa 1, yoksa 0	KAT
		m50	Eğer 50 km içinde toplam 100.000 kayıtlı araç olan bir il varsa 1, yoksa 0	KAT
		m100	Eğer 100 km içinde toplam 100.000 kayıtlı araç olan bir il varsa 1, yoksa 0	KAT

3.3 Yöntem

Makine öğrenmesi (MÖ) algoritmaları, son yıllarda farklı alanlarda sınıflandırma, tahmin ve öngörü için en sık kullanılan yöntemlerden bazıları haline gelmiştir [21-24]. MÖ algoritmalarında, denetimli [25] ve denimsiz [26] makine öğrenmesi olarak iki farklı yaklaşım bulunmaktadır. Bu çalışmada yedi tane denetimli MÖ algoritması kullanılmıştır. Bunlar rastgele orman (RF), destek vektör regresyonu (SVR), k-en yakın komşu (KNN), aşırı gradyan artırma (XgbLinear ve XgbTree), genelleştirilmiş doğrusal model (GLM) ve gradyan boosting makinesi (GBM) algoritmalarıdır. Denetimli öğrenme tekniğinde, model üretmek için eğitim verisi içindeki girdiler ve çıktılar tanımlanır. Farklı yaklaşımlar/yöntemler kullanılarak girdiler ve çıktılar arasındaki ilişki modellenir. Daha sonra, eğitilen model tahmin yapmak için kullanılır.

Bu çalışmada kullanılan yedi MÖ algoritması hem sınıflandırma hem de regresyon problemleri için kullanılan yöntemlerdir. RF algoritması, ilk olarak Ho [27] tarafından önerildi, ardından Breiman [28] tarafından “torbalama” (bagging) yöntemi de entegre edilerek daha da geliştirildi. Genel olarak bu yöntemde, verinin alt kümeleri kullanılarak birden fazla karar ağacı ile model üretilir ve bu ağaçların çıktılarının ortalaması alınarak tahmin yapılır. Bu yöntem, yüksek doğruluk sağlarken aşırı öğrenmeye karşı dayanıklıdır.

SVR algoritması, veri setinin yapısına bağlı olarak doğrusal ve doğrusal olmayan modeller şeklinde ikiye ayrılır [29]. Eğer veriler doğrusal bir şekilde ayrılabilirse, bu model doğrusal SVR olarak adlandırılır. Ancak veriler doğrusal olarak ayrıştırılamıyorsa, doğrusal olmayan SVR tercih edilir. Bu çalışmada doğrusal SVR kullanılmıştır. Bu algoritma, tahmin edilen değerlerin belirli bir tolerans bandı içinde kalmasını hedefler ve bu bandın dışındaki hataları minimize etmeye çalışır. SVR, özellikle küçük veri setlerinde ve yüksek boyutlu problemlerde başarılı sonuçlar vermektedir.

KNN algoritması, belirli bir veriye en yakın k veri noktasını bularak ve bu komşuların değerlerine dayalı tahminler yapar [30]. Parametrik olmayan bir yöntem olması nedeniyle veri dağılımı hakkında ön kabul gerektirmez.

Aşırı gradyan artırma (XgbLinear ve XgbTree) algoritması sıralı bir şekilde karar ağaçları oluşturur ve her yeni karar ağacı için önceki karar ağaçlarında yapılan hataları düzelteren “boosting” adı verilen bir teknik kullanır [31]. Yani, her adımda önceki modelin hatalarını minimize eden yeni bir ağaç öğrenir ve bu şekilde model hatası kademeli olarak azaltılır. XgbTree, karar ağacı tabanlı iken; XgbLinear, lineer regresyon tabanlı bir varyanttır.

GLM, doğrusal regresyonun farklı bir versiyonu olarak bilinir ve bağımlı değişkenin normal dağılım yerine başka bir dağılıma (örneğin Binom, Poisson vb.) sahip olmasına olanak veren bir MÖ algoritmasıdır [32]. Bu esneklik sayesinde, farklı türdeki veri yapılarına uygulanabilir.

GBM ise, modelleri sırayla oluşturan ve her yeni modelin öncekiler tarafından yapılan hataları düzeltmeye çalıştığı bir topluluk öğrenme tekniğidir [16]. GBM, zengin parametre

ayarları sayesinde esnek ve güçlü bir tahmin aracıdır. Bu yedi MÖ algoritmalarının detayları, literatürde bulunan çeşitli çalışmalarda daha kapsamlı biçimde açıklanmıştır [33-37].

Bu çalışmada, *R* programındaki *caret* paketi kullanılarak bu yedi MÖ algoritması uygulanmıştır [38]. *Caret* paketi, bu çalışmada kullanılan yedi MÖ algoritması da dahil olmak üzere çok sayıda MÖ algoritması içermektedir. Bu algoritmalar kara kutu modeller olup, iç işleyişlerinin yorumlanması zordur. Ayrıca, bu MÖ algoritmaları için birçok parametre gerekmektedir. Bu çalışmada *R* yazılımında varsayılan parametreler kullanılmıştır ve hiperparametre optimizasyonu uygulanmamıştır. Hiperparametre optimizasyonu, modelde önceden belirlenen parametrelerin en uygun kombinasyonunun sistematik olarak aranma sürecidir. Bazı literatür çalışmaları varsayılan parametreler ile optimize edilen parametreler arasında kıyaslama yapmaktadır [39-41].

Bu MÖ algoritmalarıyla, farklı değişkenler kullanılarak modeller üretebilmektedir. Bu çalışmada kullanılan 26 bağımsız değişken farklı kombinasyonlarda kullanılarak dokuz farklı veri seti elde edilmiştir. Tablo 2’de hangi veri setinde hangi bağımsız değişkenlerin kullanıldığı işaretlenmektedir.

Veri seti 1 (VS1), veri seti 2 (VS2) ve veri seti 3 (VS3), yalnızca yolun karakteristik özelliklerini temsil eden bağımsız değişkenleri içermektedir. Örneğin, VS1’de kavşak tipi, otokorkuluk tipi, refüj cinsi, lamba tipi, kaplama cinsi, şerit sayısı, platform tipi ve banket cinsi gibi kategorik değişkenler yer almaktadır. VS2 ise uzunluk, refüj genişliği, kaplama genişliği ve banket genişliği gibi sayısal değişkenleri içermektedir. VS3’te hem VS1’deki kategorik hem de VS2’deki sayısal değişkenler birlikte kullanılmaktadır.

Veri seti 4 (VS4) ve veri seti 5 (VS5), yolun karakteristik özelliklerine ek olarak yolların diğer ulaşım sistemleri merkezleriyle olan ilişkisini yansıtan bağımsız değişkenleri de içermektedir. VS4’te, VS1’de yer alan değişkenlere ilave olarak bu ilişkiyi tanımlayan 6 kategorik değişken eklenmektedir. Benzer şekilde, VS5’te, VS3’teki tüm değişkenlere ek olarak aynı 6 kategorik değişken kullanılmaktadır.

Veri seti 6 (VS6) ve veri seti 7 (VS7) ise yolun karakteristik özelliklerinin yanı sıra demografik/sosyo-ekonomik faktörleri yansıtan bağımsız değişkenleri de kapsamaktadır. VS6, VS1’in değişkenlerine ek olarak 8 adet demografik/sosyo-ekonomik (hem kategorik hem de sayısal) değişken içermektedir. Benzer şekilde, VS7, VS3’ün değişkenlerine ek olarak 8 adet demografik/sosyo-ekonomik (hem kategorik hem de sayısal) değişken içermektedir.

Veri seti 8 (VS8) yalnızca yolların diğer ulaşım sistemleriyle olan ilişkisini yansıtan değişkenleri ve demografik/sosyo-ekonomik değişkenleri içermektedir. Son olarak, veri seti 9 (VS9) bu üç ana değişken grubunun tamamını kapsamaktadır.

Tablo 2. 26 tane bağımsız değişkenler ile üretilen dokuz farklı veri seti.

	Bağımsız Değişkenler	Veri Seti 1 (VS1)	Veri Seti 2 (VS2)	Veri Seti 3 (VS3)	Veri Seti 4 (VS4)	Veri Seti 5 (VS5)	Veri Seti 6 (VS6)	Veri Seti 7 (VS7)	Veri Seti 8 (VS8)	Veri Seti 9 (VS9)
Yolun Karakteristik Özellikleri	uz		✓	✓		✓		✓		
	kav	✓		✓	✓	✓	✓	✓		✓
	oto	✓		✓	✓	✓	✓	✓		✓
	ref_g		✓	✓		✓		✓		
	ref_c	✓		✓	✓	✓	✓	✓		✓
	lam	✓		✓	✓	✓	✓	✓		✓
	kap_g		✓	✓		✓		✓		
	kap_c	✓		✓	✓	✓	✓	✓		✓
	şerit	✓		✓	✓	✓	✓	✓		✓
	pla	✓		✓	✓	✓	✓	✓		✓
	ban_g		✓	✓		✓		✓		
	ban_c	✓		✓	✓	✓	✓	✓		✓
Diğer Ulaşım Sistemleri Özellikleri	h20				✓	✓			✓	✓
	h50				✓	✓			✓	✓
	l20				✓	✓			✓	✓
	l50				✓	✓			✓	✓
	d20				✓	✓			✓	✓
	d50				✓	✓			✓	✓
Demografik/Sosyo-Ekonomik Özellikleri	hh						✓	✓	✓	✓
	tl						✓	✓	✓	✓
	n20						✓	✓	✓	✓
	n50						✓	✓	✓	✓
	n100						✓	✓	✓	✓
	a20						✓	✓	✓	✓
	a50						✓	✓	✓	✓
	a100						✓	✓	✓	✓

Oluşturulan veri setleri, üç ana değişken grubunun farklı kombinasyonlarından oluşturulmuştur. Yapılan ön analizler sonucunda tahmin performansı çok düşük olan bazı veri setleri bu çalışmaya dahil edilmemiştir.

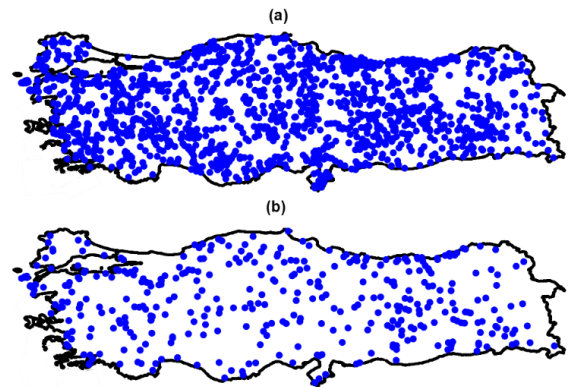
Bu çalışmada, dokuz farklı veri seti kullanılarak yedi MÖ algoritmasının performansları değerlendirilmiştir. Bunun için, 1.889 il yolu *eğitim* ve *test* olarak iki farklı parçaya bölünmüştür. *Eğitim* verisi, MÖ algoritmasını eğitmek ve geliştirmek için kullanılmıştır. Daha sonra, eğitilen model *test* verisi üzerinde kullanılarak YOGT değerleri tahmin edilmiştir. Bu değerler gözlenen YOGT değerleri ile karşılaştırılarak MÖ algoritmasının performansı değerlendirilmiştir.

İki farklı koşulda *eğitim* ve *test* verisi hazırlanmıştır. İlk olarak verinin %80'i *eğitim* verisi, %20'si ise *test* verisi olarak kullanılmıştır. Şekil 4, *eğitim* ve *test* veri seti için rastgele seçilen il yollarının lokasyonlarını örnek olarak göstermektedir. Çalışmanın devamında bu şekilde üretilen veri seti Veri_20 olarak adlandırılmıştır.

İkinci olarak verinin %50'si *eğitim* ve %50'si *test* verisi olarak seçilmiştir. Bu şekilde oluşturulan veri seti, çalışmanın devamında Veri_50 olarak adlandırılmıştır. Bu iki farklı *eğitim* ve *test* verisi seçimi kullanılarak, MÖ algoritmalarının YOGT tahminlerindeki etkisi incelenmiştir.

Eğitim ve *test* verisi bu iki koşulda rastgele olarak seçilmiştir. Rastgele seçimden kaynaklanan belirsizliği azaltmak için bu seçim 100 defa tekrarlanmıştır. Yani,

Veri_20 için %80'i *eğitim* verisi, %20'si ise *test* verisi seçimi 100 defa tekrarlanmış ve her bir tekrarda MÖ algoritmalarının performansları değerlendirilmiştir. Benzer işlemler Veri_50 için de yapılmıştır. Böylece *eğitim* ve *test* verisi seçerken oluşacak rastgelelik azaltılmıştır.



Şekil 4. %80 *eğitim* ve %20 *test* verisi olarak seçilen il yollarının lokasyonları (a) *Eğitim* ve (b) *Test*.

4 Bulgular ve tartışma

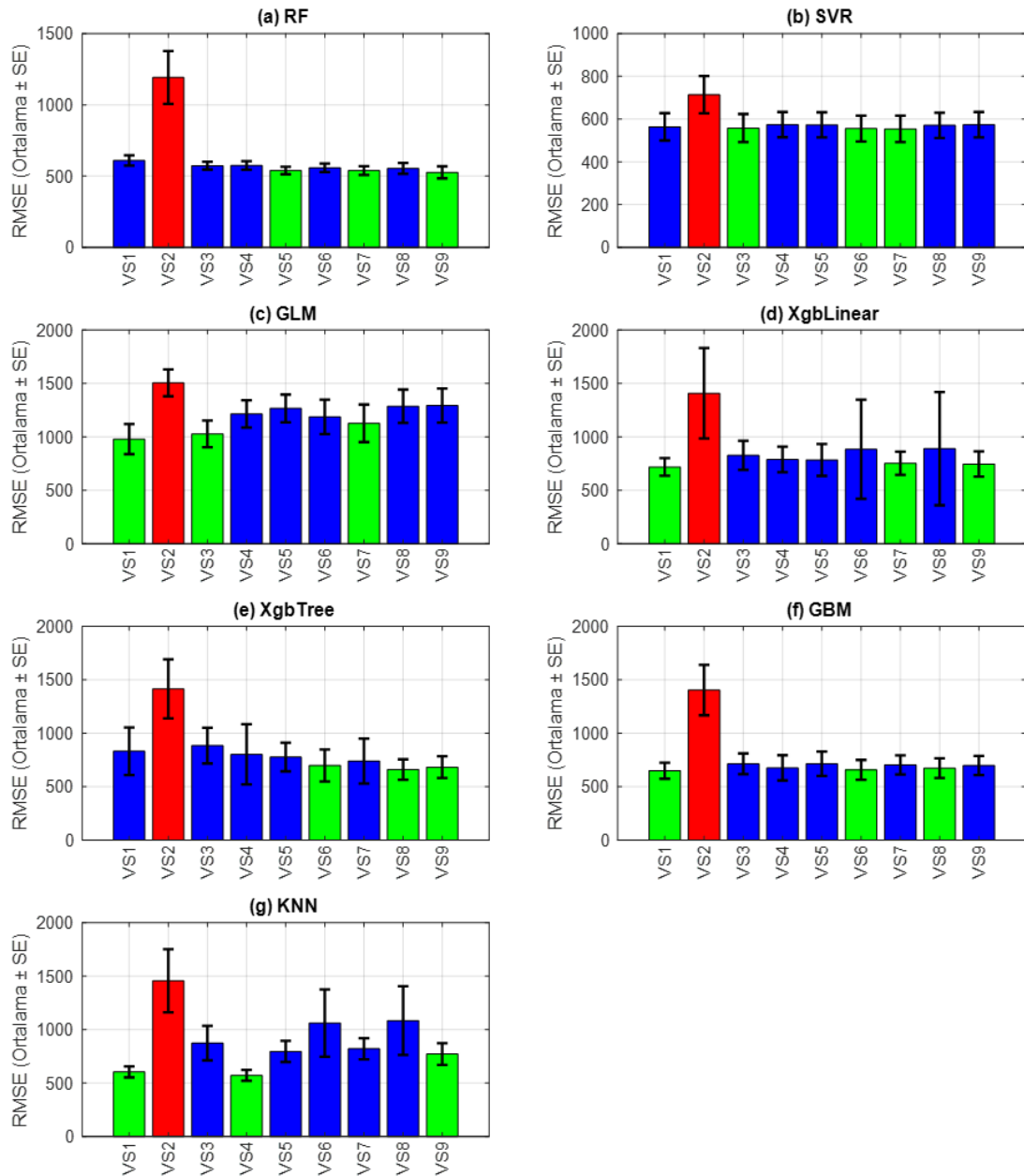
Bu çalışmada, farklı MÖ algoritmalarının çeşitli veri setleri (VS) altındaki performanslarını karşılaştırmak için iki farklı ölçüt kullanılmıştır. İlk ölçüt, kök ortalama karesel hata (RMSE) ölçütüdür. RMSE, Denklemler (1)'de gösterildiği

gibi gözlenen değerler ile tahmin edilen değerler arasındaki karesel farkların ortalamasının kareköküdür.

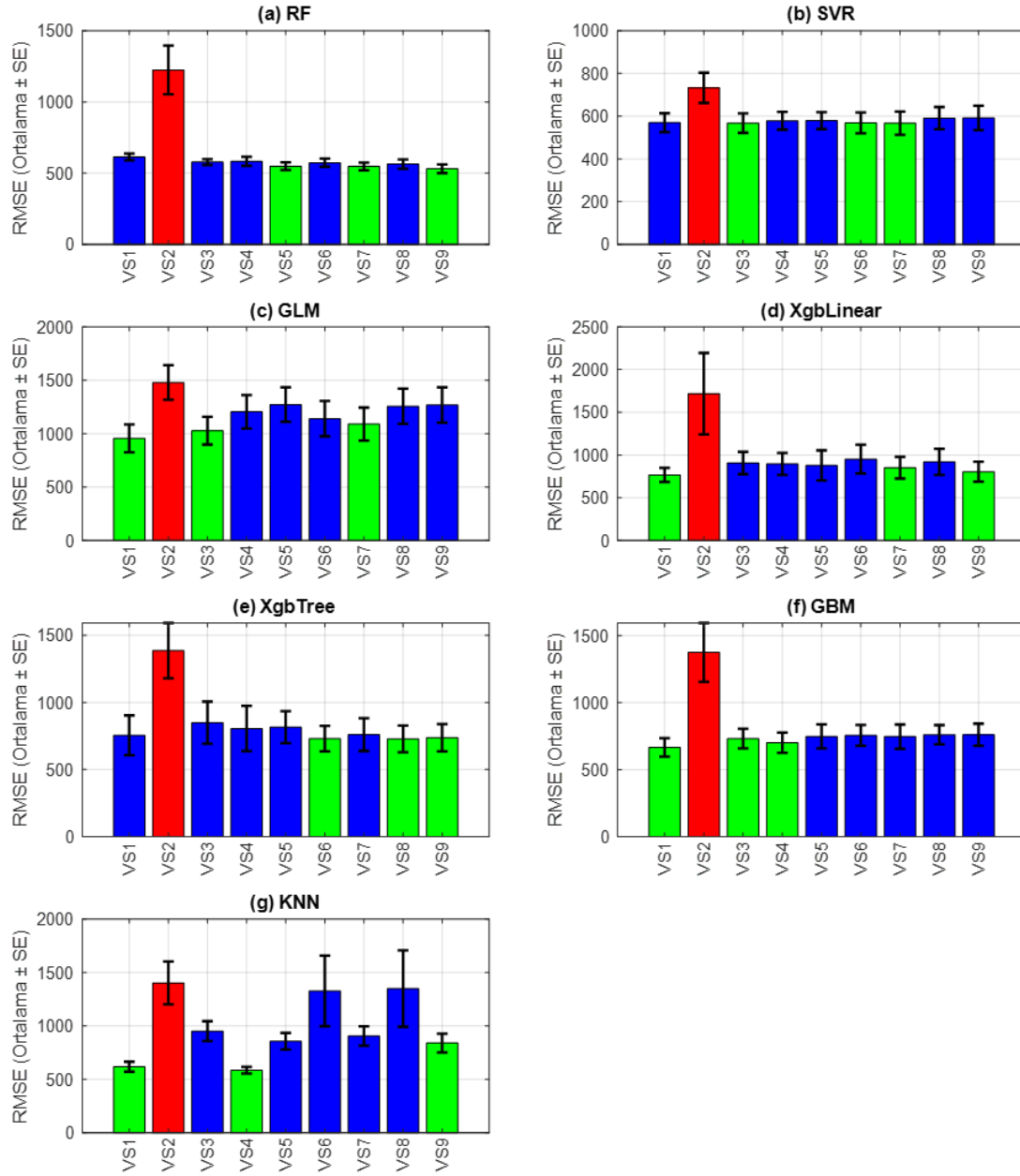
$$RMSE = \sqrt{\frac{\sum (x_{tahmin} - x_{gözlem})^2}{N}} \quad (1)$$

Bu denklemde, x_{tahmin} MÖ algoritması kullanılarak tahmin edilen YOGT değerini ifade eder. Benzer şekilde, $x_{gözlem}$ gerçekte gözlenen YOGT değerini ifade eder. N ise tahmin edilen yolların sayısıdır. Bu ölçütün daha küçük bir değere sahip olması daha iyi YOGT tahmin performansına işaret eder.

Daha önce anlatıldığı gibi *eğitim* ve *test* verisi seçimi 100 defa tekrarlanmıştır. Yani her bir koşul için 100 tane RMSE ölçütü elde edilmiştir. Bu RMSE ölçütlerini sunmak için çubuk grafiği kullanılmıştır. Şekil 5'te, yedi farklı MÖ algoritmasının, dokuz farklı veri seti (VS) kullanılarak elde edilen RMSE değerlerini çubuk grafiği şeklinde sunulmaktadır. Bu sonuçlar Veri_20 kullanılarak elde edilmiştir. Yani, her bir tekrarda verinin %80'i *eğitim*, %20'si ise *test* verisi olarak ayrılmıştır. Benzer şekilde, Şekil 6 ise Veri_50 kullanılarak elde edilen RMSE değerlerinin çubuk grafiğini göstermektedir. Ayrıca, her iki şekilde yer alan çubuk grafikleri YOGT tahmin performanslarına göre renklendirilmiştir. Kırmızı renk en kötü sonucu, yeşil renk ise en iyi üç sonucu temsil etmektedir.



Şekil 5. Dokuz farklı veri seti (VS) için farklı MÖ algoritmalarının Veri_20 ile elde edilen RMSE değerleri.



Şekil 6. Dokuz farklı veri seti (VS) için farklı MÖ algoritmalarının Veri_50 ile elde edilen RMSE değerleri.

Her iki şekilde de VS2 kullanıldığında tüm MÖ algoritmalarında en kötü sonuçlar açıkça gözlemlenmiştir. Şekil 5 ve Şekil 6'da gösterilen tüm MÖ algoritmalarında, VS2 kırmızı ile gösterilmektedir. VS2, yalnızca yolun karakteristik özelliklerini temsil eden uzunluk, refüj genişliği, kaplama genişliği ve banket genişliği gibi sayısal değişkenleri içermektedir. Buna karşılık, en iyi üç sonucu veren veri setleri kullanılan MÖ algoritmasına göre değişiklik göstermiştir. Örneğin, Şekil 5(a)'da gösterildiği gibi, RF algoritması kullanıldığında VS5, VS7 ve VS9 en iyi üç sonucu vermişken, bu durum Şekil 5(b)'de gösterilen SVR algoritmasında ise VS3, VS6 ve VS7 olarak gözlemlenmiştir. Benzer şekilde, kullanılan MÖ algoritmasına bağlı olarak en iyi üç veri seti farklılık göstermektedir.

Diğer ulaşım sistemleri merkezleriyle olan ilişkileri yansıtan bağımsız değişkenlerin YOGT tahminleri üzerindeki etkisini incelemek için VS1 ile VS4 ve VS3 ile VS5 karşılaştırılmıştır. Tablo 2'de özetlendiği gibi, VS1 ve VS3 sadece yolun karakteristik özelliklerini temsil eden sayısal ve/veya kategorik değişkenleri içerirken, VS4 ve VS5 buna ek olarak yolun diğer ulaşım sistemleri merkezleriyle olan ilişkisini yansıtan bağımsız değişkenleri de içermektedir. Hem Şekil 5 hem de Şekil 6'da, tüm MÖ algoritmalarında VS4 ve VS5'in sırasıyla VS1 ve VS3'e göre daha kötü veya benzer sonuçlar verdiği gözlemlenmiştir. Dolayısıyla, diğer ulaşım sistemleri merkezleriyle olan ilişkileri yansıtan değişkenlerin YOGT tahminleri üzerinde olumlu bir etkisinin bulunmadığı görülmüştür.

Benzer şekilde, demografik/sosyo-ekonomik faktörleri yansıtan bağımsız değişkenlerin etkisini incelemek için VS1 ile VS6 ve VS3 ile VS7 karşılaştırılmıştır. Tablo 2’de özetlendiği gibi, VS1 ve VS3 sadece yolun karakteristik özelliklerini temsil eden sayısal ve/veya kategorik değişkenleri içermektedir. Buna karşın VS6 ve VS7, karakteristik özelliklere ek olarak demografik/sosyo-ekonomik faktörleri temsil eden değişkenleri de içermektedir. Şekil 5 ve Şekil 6’da görüldüğü üzere, bazı MÖ algoritmalarında VS6 ve VS7 daha başarılı olurken, bazı algoritmalarda VS1 ve VS3 daha iyi performans göstermiştir. Yani, demografik/sosyo-ekonomik faktörleri yansıtan değişkenlerin YOGT tahminleri üzerindeki etkisinin, diğer ulaşım sistemleri merkezleriyle olan ilişkileri yansıtan değişkenlerin etkisinden daha fazla katkı sağladığı görülmüştür.

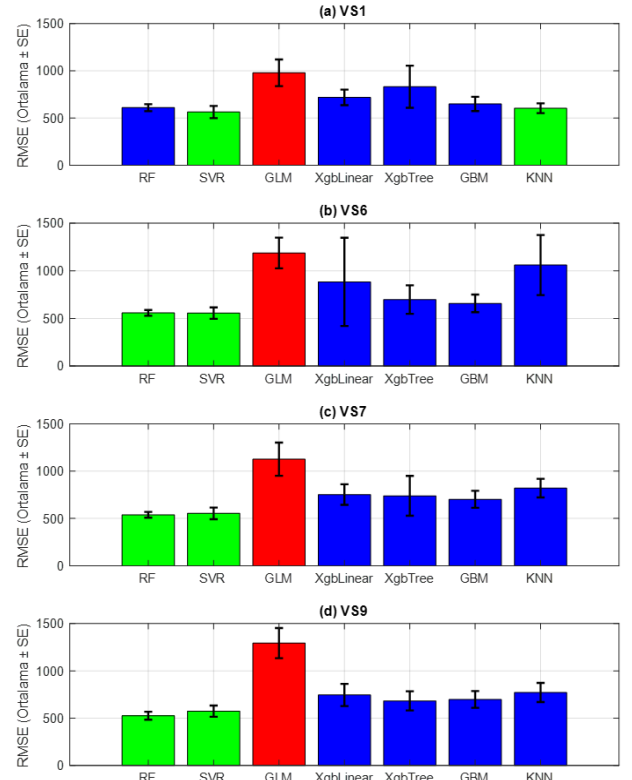
Demografik/sosyo-ekonomik faktörler ile yolun diğer ulaşım sistemleri merkezleriyle olan ilişkilerini temsil eden değişkenleri birlikte içeren VS8 incelendiğinde, yalnızca XgbTree algoritmasında (Şekil 5(e) ve Şekil 6(e)) ve GBM algoritmasında (Şekil 5(f)) iyi sonuçlar elde edildiği görülmüştür. Yani, veri setinde yolun karakteristik özelliklerini temsil eden değişkenler bulunmadığında MÖ algoritmalarının YOGT tahmin performansı genel olarak düşmüştür. Son olarak, tüm değişkenleri içeren VS9 incelendiğinde, birçok MÖ algoritmasında iyi YOGT tahmin sonuçları elde edilmiştir.

Bu bulgular doğrultusunda, sadece yolun karakteristik özelliklerini yansıtan kategorik değişkenleri içeren VS1, hem yolun karakteristik özelliklerini yansıtan değişkenler hem de demografik/sosyo-ekonomik değişkenleri içeren VS6 ile VS7 ve son olarak tüm üç farklı türdeki değişkenleri içeren VS9 veri setleri üzerinde MÖ algoritmasının performansları değerlendirilmiştir.

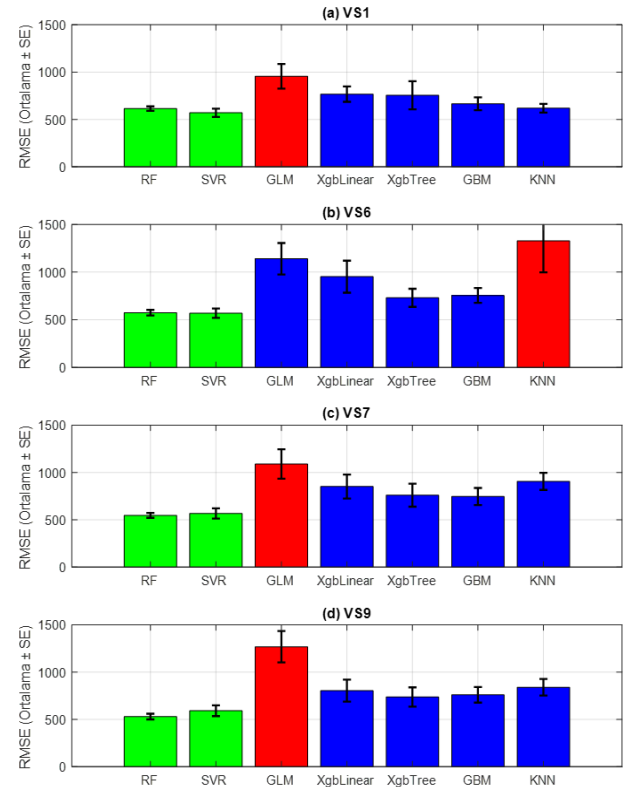
Şekil 7 ve Şekil 8’de, yedi farklı MÖ algoritmasının dört seçili veri seti (VS1, VS6, VS7 ve VS9) altında elde edilen RMSE değerlerini çubuk grafiği halinde sunulmaktadır. Benzer şekilde, Şekil 7’deki sonuçlar Veri_20 kullanılarak, Şekil 8’deki sonuçlar ise Veri_50 kullanılarak elde edilmiştir. Ayrıca bu şekillerde, Şekil 5 ve Şekil 6’da olduğu gibi performanslara göre renklendirme yapılmıştır.

Bu iki şekilde, yedi farklı MÖ algoritması açık bir şekilde karşılaştırılabilmektedir. Örneğin, GLM algoritması neredeyse tüm durumlarda en zayıf performansı göstermiştir. Sadece Şekil 8(b)’de görüldüğü gibi, VS6 ve Veri_50 kullanıldığı durumda KNN algoritması en kötü sonucu vermiştir. Buna karşılık, RF ve SVR algoritmaları çoğu durumda en iyi sonuçları veren iki MÖ algoritması olmuştur. Ancak, Şekil 7(a)’da görüldüğü gibi, VS1 ile Veri_20 kullanıldığında en iyi iki sonucu SVR ve KNN algoritmaları sağlamıştır.

Bu çalışmada, RMSE ölçütüne ek olarak, ortalama mutlak yüzde hata (MAPE) ölçütü de kullanılmıştır. MAPE, Denklem (2)’de gösterildiği gibi tahmin edilen değerlerin gerçek değerlere göre ortalama yüzde sapmasını ifade eden alternatif bir hata ölçütüdür.



Şekil 7. Yedi farklı MÖ algoritmasının dört seçili veri seti altında Veri_20 ile elde edilen RMSE değerleri.



Şekil 8. Yedi farklı MÖ algoritmasının dört seçili veri seti altında Veri_50 ile elde edilen RMSE değerleri.

Tablo 3. Yedi farklı MÖ algoritması ve dokuz farklı veri seti (VS) için Veri_20 ile elde edilen ortalama MAPE değerleri.

	VS1	VS2	VS3	VS4	VS5	VS6	VS7	VS8	VS9	Genel Ortalama
RF	0.80	1.24	0.78	0.82	0.80	0.88	0.69	0.80	0.72	0.83
SVR	0.63	1.15	0.69	0.63	0.69	0.64	0.62	0.70	0.65	0.71
GLM	1.50	2.42	1.22	1.70	1.84	1.43	1.31	1.69	1.69	1.64
XgbLinear	0.83	1.40	0.83	0.83	0.82	0.80	0.80	0.85	0.77	0.88
XgbTree	0.82	1.71	0.90	0.88	0.87	0.80	0.86	0.84	0.74	0.93
GBM	1.04	1.10	1.11	1.03	1.11	0.98	1.02	0.89	0.89	1.01
KNN	0.84	1.56	1.17	0.80	1.18	0.90	1.11	1.16	1.00	1.08
Genel Ortalama	0.92	1.51	0.95	0.95	1.04	0.91	0.91	0.99	0.92	

Tablo 4. Yedi farklı MÖ algoritması ve dokuz farklı veri seti (VS) için Veri_50 ile elde edilen ortalama MAPE değerleri.

	VS1	VS2	VS3	VS4	VS5	VS6	VS7	VS8	VS9	Genel Ortalama
RF	0.76	1.25	0.72	0.79	0.73	0.74	0.68	0.72	0.71	0.78
SVR	0.70	1.16	0.68	0.64	0.67	0.63	0.62	0.66	0.69	0.71
GLM	0.95	2.43	1.24	1.67	1.84	0.98	0.98	1.67	1.11	1.43
XgbLinear	0.92	1.49	0.97	0.85	0.87	0.86	0.86	0.86	0.82	0.94
XgbTree	0.84	1.74	0.92	0.92	0.91	0.80	0.86	0.78	0.83	0.95
GBM	1.04	1.89	1.09	1.06	0.99	0.91	0.97	0.91	0.94	1.08
KNN	0.85	1.57	1.25	0.85	1.16	0.94	1.20	1.48	0.99	1.14
Genel Ortalama	0.86	1.64	0.98	0.96	1.02	0.83	0.88	1.00	0.87	

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_{tahmin} - x_{gözlem}}{x_{gözlem}} \right| \quad (2)$$

x_{tahmin} , $x_{gözlem}$ ve N değişkenleri yukarıda açıklanmıştır. RMSE'de olduğu gibi, MAPE değerinin düşük olması, modelin YOGT tahmin performansının yüksek olduğunu göstermektedir.

Tablo 3, yedi farklı MÖ algoritması kullanılarak, dokuz farklı veri setiyle elde edilen ortalama MAPE değerlerini göstermektedir. Bu sonuçlar, Veri_20 kullanılarak elde edilmiştir. Benzer şekilde, Tablo 4'te ise Veri_50'ye dayalı ortalama MAPE değerleri sunulmaktadır.

Tablo 3 ve Tablo 4'te farklı veri setleri altında YOGT tahmin performansları incelendiğinde, VS2'nin ortalama MAPE değerlerinin diğer veri setlerine göre daha yüksek olduğu görülmüştür. Bu da VS2'nin en zayıf YOGT tahmin performansı sergilediğini göstermiştir. Buna karşın, en iyi üç YOGT tahmin performansını gösteren veri setleri kullanılan MÖ algoritmasına göre değişiklik göstermektedir. Ancak, genel ortalama değerlere bakıldığında, hem Tablo 3'te hem de Tablo 4'te VS1, VS6, VS7 ve VS9 en başarılı YOGT performanslarını vermiştir.

Tablo 3 ve Tablo 4'te farklı MÖ algoritmalarının YOGT tahmin performansları karşılaştırıldığında, GLM algoritmasının ortalama MAPE değerlerinin diğer MÖ algoritmalarına göre yüksek olduğu gözlemlenmiştir. Bu durum, GLM algoritmasının en düşük YOGT tahmin performansı sergilediğini göstermiştir. Öte yandan, en düşük ortalama MAPE değerleri RF ve SVR algoritmalarıyla elde edilmiştir. Yani, RF ve SVR algoritmaları en başarılı sonuçları vermiştir. Sonuç olarak, ortalama MAPE değerlerine dayanan bu bulgular, önceki RMSE tabanlı değerlendirmelerle tutarlılık göstermiştir.

Ayrıca, MÖ algoritmaları işlem süreleri açısından da değerlendirilmiştir. MÖ algoritmalarının çalışması için gerekli olan işlem süresi kullanılan bilgisayarın donanım özelliklerine bağlıdır. Tablo 5, her bir MÖ algoritması için ortalama işlem sürelerini saniye cinsinden göstermektedir. Bu süreler, 3.60 GHz AMD Ryzen 5 5500 işlemcili ve 32 GB RAM'e sahip Windows tabanlı bir masaüstü bilgisayar (Donanım 1) ile 1.8 GHz Intel Core i5 işlemcili MacBook Air dizüstü bilgisayarda (Donanım 2) elde edilmiştir.

Tablo 5. İki farklı bilgisayar donanımında MÖ algoritmaların çalışması için gerekli olan saniye cinsinden işlem süreleri.

	Donanım 1		Donanım 2	
	Veri_20	Veri_50	Veri_20	Veri_50
RF	19.39	11.63	25.16	23.74
SVR	6.31	6.54	19.18	15.22
GLM	3.35	3.28	4.75	5.02
XgbLinear	39.72	36.22	61.35	43.96
XgbTree	25.97	24.15	46.40	36.26
GBM	4.71	4.47	7.99	6.82
KNN	3.29	3.24	5.18	4.65

GLM, GBM ve KNN algoritmaları, diğer MÖ algoritmalarına kıyasla daha kısa işlem süresiyle çalışmıştır. Öte yandan, XgbLinear ve XgbTree algoritmaları en uzun işlem süresine sahip algoritmalar olmuştur. RF ve SVR algoritmalarının işlem süreleri ise bu iki grup arasında yer almıştır.

Şekil 5-8 ve Tablo 3-5 boyunca sunulan MÖ algoritmalarının farklı koşullardaki karşılaştırmaları, algoritmaların varsayılan parametreleri kullanılarak gerçekleştirilmiştir. Ancak, hiperparametre optimizasyonunun yapılması durumunda algoritmaların

performanslarında değişiklikler ortaya çıkabileceği düşünülmektedir.

Sonuç olarak, hem RMSE ve MAPE ölçütleri hem de işlem süresini dikkate alındığında, RF ve SVR algoritmaları diğer MÖ algoritmalarına göre daha iyi performans verdiği gözlemlenmiştir. Ayrıca, yolun karakteristik özelliklerini ifade eden bağımsız değişkenler ile yol çevresine ait demografik/sosyo-ekonomik değişkenler birlikte kullanıldığı zaman YOGT tahminlerinin performansı artmıştır.

Bu çalışmada ayrıca, yüksek YOGT değerlerine sahip il yollarının veya düşük YOGT değerlerine sahip il yollarının MÖ algoritması kullanılarak elde edilen tahminlerdeki performansları da ayrıntılı incelenmiştir. RF algoritması kullanıldığı zaman, YOGT değeri 1.000 araç/gün'den az olan yolların %94'ünde abartılı tahmin yapıldığı gözlemlenmiştir. Yani, YOGT değeri gözlenen YOGT değerine göre daha yüksek tahmin edilmiştir. SVR algoritması kullanıldığı zaman, YOGT değeri 3.000 araç/gün'den fazla olan yolların %98'inde tahminlerin altında yapıldığı gözlemlenmiştir. Yani, YOGT değeri gözlenen YOGT değerine göre daha küçük tahmin edilmiştir.

Son olarak bu çalışmadaki MÖ algoritmasının performansları benzer analizler yapan ve RMSE ile MAPE ölçülerini kullanan üç farklı çalışma ile karşılaştırılmıştır. **Tablo 6**, bu çalışmada RF ve SVR algoritmaları kullanıldığında elde edilen ortalama RMSE ve MAPE değerleri ile diğer üç çalışmadan raporlanan RMSE ve MAPE değerini göstermektedir.

Tablo 6. Bu çalışmada bulunan RMSE ve MAPE ölçütlerinin literatürdeki benzer çalışmalar ile karşılaştırılması.

Çalışma	Çalışma Alanı	Yöntem	RMSE	MAPE
Bu çalışma	Türkiye	RF	539	0.68
Bu çalışma	Türkiye	SVR	554	0.62
[2]	İngiltere	RF	2119	0.03-2.88
[2]	İngiltere	SVR	2336	0.02-2.76
[12]	ABD	ÇRA	229	-
[42]	ABD	ÇRA	993	0.26-0.35

Bu üç çalışmadan biri, bu çalışmada olduğu gibi RF ve SVR algoritmalarını kullanmış; diğerleri ise literatürde anlatılan ÇRA gibi yöntemleri kullanmışlardır. Ayrıca, bu üç çalışma bu çalışmadan farklı olarak daha çeşitli bağımsız değişkenler kullanmıştır. Örneğin, Sfyridis ve Agnolucci [2], bu çalışmaya kıyasla daha ayrıntılı demografik/sosyo-ekonomik değişkenlerin (örneğin, çalışan nüfus vb.) yanı sıra, yolun topoğrafik özelliklerini (kentsel ya da kırsal bölgede bulunması gibi) de dikkate almışlardır. Benzer şekilde, Wang vd. [12] çalışmasında, topoğrafik özelliklere ek olarak sağa/sola dönüş şeritlerinin varlığı, refüj tipi, yol kenarına park edilebilirlik durumu ve kaldırım varlığı gibi fiziksel yol özellikleri de modele dahil edilmiştir. Pulugurtha ve Kusam [40] ise yol özellikleri, demografik/sosyo-ekonomik bilgiler ve arazi kullanımı verilerini içeren, 40'tan fazla değişkenden oluşan oldukça kapsamlı bir veri seti kullanmıştır.

Tablo 6'da görüldüğü üzere, bu çalışmada bulunan RMSE ölçütü Wang vd. [12] tarafından yapılan çalışmada bulunan RMSE ölçütünden büyük olduğu gözlemlenmiştir. Diğer iki çalışmada ise RMSE ölçütü dikkate alındığında YOGT tahminleri bu çalışmaya göre daha kötü olduğu bulunmuştur. Benzer şekilde, bu çalışmada bulunan MAPE ölçütü Pulugurtha ve Kusam [40] tarafından yapılan çalışmada bulunan MAPE ölçütünden daha büyük olduğu gözlemlenmiştir. Bu durum, söz konusu çalışmalarda kullanılan veri setinin daha homojen bir yapıya sahip olması ve model performansını daha güçlü biçimde açıklayabilecek bağımsız değişkenlerin bulunmasıyla açıklanabilir.

5 Sonuçlar

YOGT belirli bir noktadan bir yıl boyunca geçen ortalama araç sayısını ifade etmekte olup, ulaşımla ilgili birçok durumda önemli bir girdi olarak kullanılmaktadır. Her ülkenin YOGT verilerini toplama standartları ve teknolojileri farklıdır. Türkiye'de YOGT verileri, yol altına yerleştirilen sabit manyetik döngülü cihazlar ve yol üzerine kurulan taşınabilir hava basınçlı hortumlu cihazlar kullanılarak KGM tarafından toplanmaktadır.

Bu çalışmada, Türkiye'deki il yolları için YOGT değerlerinin tahmin edilmesi amacıyla yedi farklı MÖ algoritması kullanılmıştır. Bu algoritmalar; rastgele orman (RF), destek vektör regresyonu (SVR), k-en yakın komşu (KNN), aşırı gradyan artırma (XgbLinear ve XgbTree), genelleştirilmiş doğrusal model (GLM) ve gradyan boosting makinesi (GBM) olarak sıralanmaktadır. MÖ algoritmalarında kullanılmak üzere yolun karakteristik özellikleri, diğer ulaşım sistemleriyle olan ilişkisini yansıtan değişkenler ve yol çevresinin demografik/sosyo-ekonomik özellikleri olmak üzere üç farklı grupta toplam 26 kapsamlı değişken toplanmıştır. Bu veriler farklı kombinasyonlarla değerlendirilerek, MÖ algoritmalarının performansları analiz edilmiştir.

Bu çalışmada elde edilen bulgular, YOGT tahmininde kullanılan değişken türlerinin ve MÖ algoritmalarının model üzerindeki etkisini ortaya koymuştur. Sadece yolun fiziksel özelliklerini temsil eden sayısal değişkenler kullanıldığında tahmin doğruluğu düşük kalmıştır. Diğer ulaşım merkezleriyle olan ilişkileri yansıtan değişkenlerin modele eklenmesi ise tahmin performansında belirgin bir iyileşme sağlamamıştır. Buna karşın, demografik/sosyo-ekonomik değişkenlerin dahil edilmesi diğer ulaşım merkezleriyle olan ilişkileri yansıtan değişkenlerin modele eklenmesine göre daha etkili olduğu görülmüştür. Modellerin karşılaştırılmasında RF ve SVR algoritmaları en başarılı sonuçları verirken, GLM algoritması en düşük tahmin performansını göstermiştir. Ayrıca RF algoritması, düşük hacimli (1.000 araç/gün'den az) yolların büyük çoğunluğunda (yaklaşık %94'ü) tahminlerin gerçek değerinde kaldığını gözlemlenmiştir. SVR algoritması ise yüksek hacimli (3.000 araç/gün'den fazla) yolların %98'inde gerçek değerinde tahmin üretmiştir.

Bu bulgular doğrultusunda, RF ve SVR algoritmalarının, yolun karakteristik özellikleri ile yol çevresine ait demografik/sosyo-ekonomik değişkenlerin birlikte kullanıldığı durumlarda YOGT tahminlerinde daha başarılı

olduğu sonucuna varılmıştır. Böylece, YOGT değerlerinin daha doğru bir şekilde tahmin edilmesi sayesinde, yol tasarımı, bakım-onarım planlaması, kaza analizi ve çevresel etki değerlendirmeleri gibi pek çok alanda daha etkin kararlar alınmasına katkı sağlaması beklenmektedir.

Bu çalışmanın bazı metodolojik sınırlamaları bulunmaktadır ve bu sınırlamalara bağlı olarak gelecekteki araştırmalar için çeşitli öneriler sunulabilir. İlk olarak, *eğitim* ve *test* verileri rastgele seçilmiş olsa da, coğrafi yakınlık nedeniyle benzer yol karakteristiklerinin her iki veri setinde de yer alması, model performansını artırmış olabilir. Bu nedenle, gelecekte mekânsal çapraz doğrulama gibi coğrafi temelli doğrulama yöntemlerinin kullanılması, MÖ algoritmalarının bölgesel genellenebilirliğini daha doğru bir şekilde değerlendirmek açısından önemlidir. Ayrıca, modelin farklı trafik koşullarında da test edilebilmesi için trafik hacmine dayalı doğrulama yönteminin uygulanması, çalışmanın geliştirilmesi için önemli bir gelecek adımı olarak görülmektedir.

Ayrıca bu çalışma MÖ algoritmaları varsayılan parametreler ile değerlendirilmiştir. Hiperparametre optimizasyonu yapılmaması, algoritmaların potansiyel performanslarının tam olarak yansıtılamamasına neden olabileceği için, bu durum çalışmanın önemli bir metodolojik sınırlılığı olarak ele alınmıştır.

Teşekkür

Bu çalışma, Artvin Çoruh Üniversitesi Bilimsel Araştırma Projeleri Koordinatörlüğü (AÇÜBAP) tarafından desteklenmiştir (Proje No: 2024.F40.02.02).

Ayrıca, bu çalışma büyük ölçüde, hem internet üzerinden hem de özel talepler yoluyla erişilen devlet belgelerine dayanmaktadır. Bu belgelerin temin edilmesindeki değerli yardımları için KGM ve TUCBS personeline en içten teşekkürlerimi sunarım.

Çıkar çatışması

Yazarlar çıkar çatışması olmadığını beyan etmektedir.

Benzerlik oranı (iThenticate): %6

Kaynaklar

- [1] R. P. Roess, E. S. Prassas and W. R. McShane, Traffic engineering. Pearson/Prentice Hall, 2004.
- [2] A. Sfyridis and P. Agnolucci, Annual average daily traffic estimation in England and Wales: An application of clustering and regression modelling. Journal of Transport Geography, 83, 102658, 2020. <https://doi.org/10.1016/j.jtrangeo.2020.102658>.
- [3] X. Zhang and M. Chen, Enhancing statewide annual average daily traffic estimation with ubiquitous probe vehicle data. Transportation Research Record, 2674(9), 649-660, 2020. <https://doi.org/10.1177/0361198120931100>.
- [4] Ulaştırma ve Altyapı Bakanlığı - Karayolu Genel Müdürlüğü, Trafik ve Ulaşım Bilgileri. <https://www.kgm.gov.tr/Sayfalar/KGM/SiteTr/Trafik/TrafikveUlasimBilgileri.aspx>, Accessed 01 May 2025.
- [5] D. Mohamad, K. C. Sinha, T. Kuczek and C. F. Scholer, Annual average daily traffic prediction model for

- county roads. Transportation research record, 1617(1), 69-77, 1998. <https://doi.org/10.3141/1617-10>.
- [6] Q. Xia, F. Zhao, Z. Chen, L. D. Shen and D. Ospina, Estimation of annual average daily traffic for nonstate roads in a Florida county. Transportation Research Record, 1660(1), 32-40, 1999. <https://doi.org/10.3141/1660-05>.
- [7] L. Pun, P. Zhao and X. Liu, A multiple regression approach for traffic flow estimation. IEEE access, 7, 35998-36009, 2019. <https://doi.org/10.1109/ACCESS.2019.2904645>.
- [8] F. Zhao and N. Park, Using geographically weighted regression models to estimate annual average daily traffic. Transportation research record, 1879(1), 99-107, 2004. <https://doi.org/10.3141/1879-12>.
- [9] B. Selby and K. M. Kockelman, Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression. Journal of Transport Geography, 29, 24-32, 2013. <https://doi.org/10.1016/j.jtrangeo.2012.12.009>.
- [10] M. Doustmohammadi and M. Anderson, A Bayesian regression model for estimating average daily traffic volumes for low volume roadways. International Journal of Statistics and Probability, 8(1), 143-149, 2019. <https://doi.org/10.5539/ijsp.v8n1p143>.
- [11] D. Apronti, K. Ksaibati, K. Gerow and J. J. Hepner, Estimating traffic volume on Wyoming low volume roads using linear and logistic regression methods. Journal of traffic and transportation engineering (English edition), 3(6), 493-506, 2016. <https://doi.org/10.1016/j.jtte.2016.02.004>.
- [12] K. Wang, R. DeVine, N. Huynh, W. Jin, G. Comert and M. Chowdhury, Comparison of models with and without roadway features to estimate annual average daily traffic at non-coverage locations. International Journal of Transportation Science and Technology, 15, 244-259, 2024. <https://doi.org/10.1016/j.ijtst.2023.10.001>.
- [13] X. Sun and S. Das, Developing a method for estimating AADT on all Louisiana roads. Louisiana Transportation Research Center, Louisiana, USA, Technical Report FHWA/LA.14/548, July 2015.
- [14] F. Harrou, A. Zeroual and Y. Sun, Traffic congestion monitoring using an improved kNN strategy. Measurement, 156, 107534, 2020. <https://doi.org/10.1016/j.measurement.2020.107534>.
- [15] D. Xu, Y. Wang, P. Peng, S. Beilun, Z. Deng and H. Guo, Real-time road traffic state prediction based on kernel-KNN. Transportmetrica A: Transport Science, 16(1), 104-118, 2020. <https://doi.org/10.1080/23249935.2018.1491073>.
- [16] Y. Zhang and A. Haghani, A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies, 58, 308-324, 2015. <https://doi.org/10.1016/j.trc.2015.02.019>.
- [17] Y. Han, T. Peng, C. Wang, Z. Zhang and G. Chen, A hybrid GLM model for predicting citywide spatio-temporal metro passenger flow. ISPRS International

- Journal of Geo-Information, 10(4), 222, 2021. <https://doi.org/10.3390/ijgi10040222>.
- [18] B. K. Koo, J. W. Baek and K. Y. Chung, Weight feedback-based harmonic MDG-ensemble model for prediction of traffic accident severity. Applied Sciences, 11(11), 5072, 2021. <https://doi.org/10.3390/app11115072>.
- [19] Türkiye Ulusal Coğrafi Bilgi Sistemleri, Ulusal Coğrafi Bilgi Platformu. <https://tucbs.gov.tr>, Accessed 01 May 2025.
- [20] Türkiye İstatistik Kurumu (TÜİK). <https://tuik.gov.tr>, Accessed 01 May 2025.
- [21] J. Li, J. Boonaert, A. Doniec and G. Locenguez, Multi-models machine learning methods for traffic flow estimation from floating car data. Transportation Research Part C: Emerging Technologies, 132, 103389, 2021. <https://doi.org/10.1016/j.trc.2021.103389>.
- [22] İ. Bütün, B. Kaplan ve K. Adem, Rseslibknın makine öğrenmesi yöntemi kullanılarak Parkinson Hastalığının Tanısı. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 9(2), 715-721, 2020. <https://doi.org/10.28948/ngumuh.655720>.
- [23] S. Zeybek, Derin öğrenme ve makine öğrenmesi yöntemleri ile sosyal medya verilerinden suç tespiti. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 14(1), 175-182, 2025. <https://doi.org/10.28948/ngumuh.1551734>.
- [24] Y. Dokuz, A. Bozdağ ve B. Gökçek, Hava kalitesi parametrelerinin tahmini ve mekansal dağılımı için makine öğrenmesi yöntemlerinin kullanılması. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 9(1), 37-47, 2020. <https://doi.org/10.28948/ngumuh.654092>.
- [25] Y. M. Kızılkaya ve A. Oğuzlar, Bazı denetimli öğrenme algoritmalarının R programlama dili ile kıyaslanması. Karadeniz Uluslararası Bilimsel Dergi, 37, 90-98, 2018. <https://doi.org/10.17498/kdeniz.405746>.
- [26] T. Jiang, J. L. Gradus and A. J. Roselline, Supervised machine learning: a brief primer. Behavior therapy, 51(5), 675-687, 2020. <https://doi.org/10.1016/j.beth.2020.05.002>.
- [27] T. K. Ho, Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition IEEE, pp. 278-282, Montreal, Canada, 1995
- [28] L. Breiman, Random forests. Machine learning, 45, 5-32, 2001. <https://doi.org/10.1023/A:1010933404324>.
- [29] D. Basak, S. Pal and D. C. Patranabis, Support vector regression. Neural Information Processing-Letters and Reviews, 11(10), 203-224, 2007.
- [30] X. Luo, D. Li, Y. Yang and S. Zhang, Spatiotemporal traffic flow prediction with KNN and LSTM. Journal of Advanced Transportation, 2019(1), 4145353, 2019. <https://doi.org/10.1155/2019/4145353>.
- [31] I. L. Cherif and A. Kortebi, On using extreme gradient boosting (XGBoost) machine learning algorithm for home network traffic classification. In: 2019 Wireless Days (WD), pp. 1-6, Manchester, UK, 2019. <https://doi.org/10.1109/WD.2019.8734193>.
- [32] A. J. Dobson and A. G. Barnett, Generalized linear models. Routledge, New York, 2019.
- [33] B. Mahesh, Machine learning algorithms-a review. International Journal of Science and Research, 9(1), 381-386, 2020. <https://doi.org/10.21275/ART20203995>.
- [34] A. Singh, N. Thakur and A. Sharma, A review of supervised machine learning algorithms. In: 2016 3rd international conference on computing for sustainable global development (INDIACom), pp. 1310-1315, New Delhi, India, 2016.
- [35] G. Bonaccorso, Machine Learning Algorithms: Popular algorithms for data science and machine learning. Packt Publishing Ltd, 2018.
- [36] G. James, D. Witten, T. Hastie, and R. Tibshirani, An introduction to statistical learning: with applications in R. New York: springer, 2013.
- [37] J. H. Friedman, Greedy function approximation: a gradient boosting machine. Annals of statistics, 29(5), 1189-1232, 2001.
- [38] K. Max, W. Jed, W. Steve, W. Andre, K. Chris, E. Allan, C. Tony, M. Zachary and et.al. Package 'caret'. The R Journal, 223.7: 48, 2020.
- [39] P. Probst, A. L. Boulesteix, and B. Bischl, Tunability: Importance of hyperparameters of machine learning algorithms. Journal of Machine Learning Research, 20(53), 1-32, 2019.
- [40] A. Bagnall and G. C. Cawley, On the use of default parameter settings in the empirical evaluation of classification algorithms. arXiv preprint arXiv:1703.06777, 2017.
- [41] R. G. Mantovani, A. L. Rossi, J. Vanschoren, B. Bischl and A. C. Carvalho, To tune or not to tune: Recommending when to adjust SVM hyper-parameters via meta-learning. In 2015 International joint conference on neural networks (IJCNN), pp. 1-8, Killarney, Ireland, 2015.
- [42] S.S. Pulugurtha and P.R. Kusam, Modeling annual average daily traffic with integrated spatial data from multiple network buffer bandwidths. Transportation research record, 2291(1), 53-60, 2012. <https://doi.org/10.3141/2291-07>.

