



Farklı regresyon modelleriyle kestirilen zenit troposferik gecikmelerin değerlendirilmesi

Evaluation of zenith tropospheric delays estimated by different regression models

Ali Utku Akar^{1,*} , Cevat İnal² 

^{1,2} Konya Teknik Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Harita Mühendisliği Bölümü, 42250, Konya, Türkiye

Öz

Yapay zekâ ve makine öğrenimi alanındaki gelişmeler, GNSS uygulamalarındaki sorunlara alternatif çözümler sunmakta veya mevcut çözümlerin verimliliğini artırmaya imkân sağlamaktadır. GNSS’de karşılaşılan birçok hata vardır ve bu hatalar kullanıcılar için problem oluşturmaktadır. Troposferik gecikme bunlardan birisidir. GNSS teknolojisine ilginin artmasıyla beraber öğrenme algoritmalarının atmosfer/troposfer çalışmalarında kullanımı, troposferik gecikme kestirimi için yeni modellerin oluşturulmasını önemli hale getirmiştir. Bu çalışmada, Radyal Tabanlı Destek Vektör Regresyonu (RTF-DVR), Ridge ve Elastik-Net regresyon modelleriyle zenit troposferik gecikmenin (ZTD) kestirilmesi amaçlanmış, makine öğrenimi esasına göre eğitilmiş yeni regresyon modellerinin ZTD kestiriminde alternatif olarak tercih edilebilirliği araştırılmıştır. Bunun için farklı yöntemlerden elde edilen sonuç ZTD modelleri karşılaştırılarak analiz edilmiştir. Analiz sonuçlarından RTF-DVR modelinin daha iyi sonuçları verdiği, bunu Elastik-Net ve Ridge modellerinin takip ettiği tespit edilmiştir. Sonrasında yeni modellerden elde edilen ZTD değerleri, Canadian Spatial Reference System–Precise Point Positioning (CSRS-PPP) ZTD değerleriyle karşılaştırılıp modellerin performansı değerlendirilmiştir. Değerlendirme sonuçlarına göre, CSRS-PPP’ye en uyumlu modelin RTF-DVR olduğu sonucuna varılmıştır

Anahtar kelimeler: ZTD kestirimi, Makine öğrenimi, RTF-DVR, Ridge ve elastik-net regresyonu, CSRS-PPP

1 Giriş

Uydularla Konum Belirleme Sistemleri (GNSS: Global Navigation Satellite Systems) içerisinde önemli bir yeri olan sinyaller, atmosferde yayılırken iyonosferdeki elektron içeriğinden ve nötr troposferdeki hava kütlelerinden etkilenmektedir [1]. GNSS sinyallerinde, iyonosferik etkinin giderilebilmesi için tekli, ikili ve üçlü faz farkları kullanılır veya uygulamada çift frekanslı GNSS alıcıları tercih edilir [2, 3]. Bu durum iyonosferik etkiyi elimine ederken troposfer etkisini ortadan kaldırmaz. GNSS sinyallerinin troposferdeki kütleler ile etkileşimi sonucu doğrusal hareket etmemesi ve buna bağlı olarak sinyallerin alıcıya beklenen sürede ulaşmaması troposferik gecikme olarak tanımlanmaktadır. İzdüşüm fonksiyonları yardımıyla zenit yönündeki

Abstract

Advances in artificial intelligence and machine learning provide alternative solutions to the problems in GNSS applications or enable to increase the efficiency of existing solutions. There are many errors encountered in GNSS and these errors pose a problem to users. The tropospheric delay effect is one of them. The increasing interest in GNSS technology as well as the use of learning algorithms in atmosphere/troposphere studies have made it important to create new models for tropospheric delay estimation. In this study, it was aimed to estimate the zenith tropospheric delay (ZTD) by using Radial Basis Function-Support Vector Regression (RBF-SVR), Ridge and Elastic-Net regression models. The new regression models trained according to machine learning was investigated its preferability as an alternative in ZTD estimation. For this reason, the results obtained from different methods were analyzed by comparing the ZTD models. From the analysis results, it was determined that the RBF-SVR model gave the best results, followed by the Elastic-Net and Ridge models. Afterward, the ZTD values obtained from the new models were compared with the Canadian Spatial Reference System–Precise Point Positioning (CSRS-PPP) ZTD values and the performance of the models was evaluated. According to the evaluation results, it was concluded that RBF-SVR is the most compatible model to CSRS-PPP.

Keywords: ZTD Estimation, Machine learning, RBF-SVR, Ridge and elastic-net regression, CSRS-PPP

indirgenmiş durumu ise zenit troposferik gecikmesi (ZTD: Zenith Tropospheric Delay) şeklinde ifade edilmektedir. Gecikme etkisi, hidrostatik ve ıslak bileşenden oluşur. Genellikle hidrostatik bileşen deneysel (ampirik) modellerle doğrudan hesaplanabilmektedir [4]. Ancak su buharının dağılımı düzensiz olduğu için ıslak bileşene ait kesin bir model söz konusu değildir. Islak gecikmenin belirlenmesinde, öncül olarak kullanılan hidrostatik bileşen değerlerinin doğru tespiti, ıslak bileşenin gerçeğe yakın kestirilebilmesinde önemli bir unsurdur böylelikle bileşenlerin her ikisinin de iyi kestirilmesinde fayda vardır [5]. Söz konusu bileşenler, kısıtlayıcı faktörler olduğu için konum belirleme veya meteoroloji alanında daha yüksek

* Sorumlu yazar / Corresponding author, e-posta / e-mail: auakar@ktun.edu.tr (A. U. Akar)

Geliş / Received: 15.03.2022 Kabul / Accepted: 07.04.2022 Yayımlanma / Published: 18.07.2022

doi: 10.28948/ngumuh.1088375

hassasiyetin sağlanması adına GNSS gözlemlerine dâhil edilmelidir [6].

Troposfer tabakası, atmosferin yeryüzüne en yakın bölgesinde olup, kalınlığı ekvatoran kutuplara doğru gidildikçe azalmaktadır. Bu farklılığı göstermesinin temel nedeni, ekvatoral bölgelerde hem sıcaklığa bağlı olarak ısınan havanın yükselmesi hem de dış kuvvet etkilerinin kutuplara nazaran daha büyük olmasıdır [7]. Ayrıca kalınlık farkları mevsimlere göre çok fazla değişiklik gösterebilmektedir. Bu açıdan troposfer tabakasına bağlı gecikme etkisi; bölgeden bölgeye, yüksekliğe, hava değişimine ve meteorolojik olaylara göre değişmektedir. Mevsimsel ve konumsal değişkenlik, GNSS konumlandırma ve navigasyon uygulamalarında önemli bir hata kaynağı olarak kabul edilir [8].

Günümüze kadar, troposferik hatanın giderilmesi için farklı kuruluşlar/kişiler tarafından çeşitli troposfer modelleri geliştirilmiştir [9, 10]. Üstelik modeller; atmosfer fiziği değişimi, hava ve iklim çalışmaları, ZTD ve yoğunlaşabilir su buharı (PWV: Precipitable Water Vapor) gibi birtakım atmosferik ürünlerin elde edilmesinde bir araç olarak kullanılmaya başlamıştır. Geliştirilen modeller, birbirinden farklı olsa bile her birinin kendine özgü kısıtlamaları bulunmaktadır. Örneğin; Hopfield [11], Saastamoinen [12] ve Black [13] gibi yüzey meteorolojik modelleri, yüzeyde ölçülen sıcaklık, basınç ve bağıl nem parametrelerini kullanarak gecikme etkisini belirler. Ölçülemeyen yüzey parametreleri ise matematiksel model üzerinden belirlendiği için ZTD tahmininde kullanılan meteorolojik parametreler hava koşullarından bağımsız olabilir. Dolayısıyla gerçek zamanlı uygulamalarda anlık meteorolojik verileri elde etmek zordur ve bu da pratik çalışmaları kısıtlar [14]. Küresel Basınç ve Sıcaklık/Küresel İzdüşüm Fonksiyonu (GPT: Global Pressure and Temperature/GMF: Global Mapping Function) serileri ve UNB3m (University of New Brunswick) gibi diğer ampirik modeller, belirli periyotlarda ortalama enlem-boylam değerleri ile belirlenen meteorolojik gridleri kullanır [15]. Bu modeller uzun vadeli ortalamalara dayandığı için küresel ısınma sonucu ani hava değişimlerine karşı duyarsız hale gelebilmektedir [16]. Diğer taraftan yaygın internet ağı, donanım arızaları, sinyal bozulmaları, elektrik kesintisi gibi sorunlar istasyondaki veri akışlarını bozarak model hatalarında artışa sebep olur. Veri akışındaki tutarsızlık nedeniyle veri arşivi ve ZTD ürünlerinin kalitesi bozulur. Mevcut troposfer ürünlerini böyle kullanmaya ek olarak, istasyon bazında elde edilen meteorolojik verilerle oluşturulmuş makine öğrenimi modeli, alternatif bir yaklaşım olarak görülebilir. Öğrenme modelleriyle veri arşivindeki hataların/eksikliklerin ortaya çıkartılıp giderilmesi de söz konusu olabilmektedir [17].

GNSS teknolojisine ilginin artmasıyla birlikte öğrenme algoritmalarının troposfer çalışmalarında kullanılmasının önemi ortaya çıkmıştır [18]. Çalışmalarda ilk olarak, atmosferdeki sıcaklık, basınç, kütsel gazlar ve su buharı basıncı gibi karmaşık parametrelerin kestirimi için sinir ağlarına dayalı modellerin oluşturulması üzerine ağırlık verilmiştir [7, 19]. Parametre kestirimleri dışında Łoś vd. [20], Polonya için rastgele orman sınıflama yaklaşımını kullanarak 0-2 saatlik fırtınaları tahmin etmede entegreli su

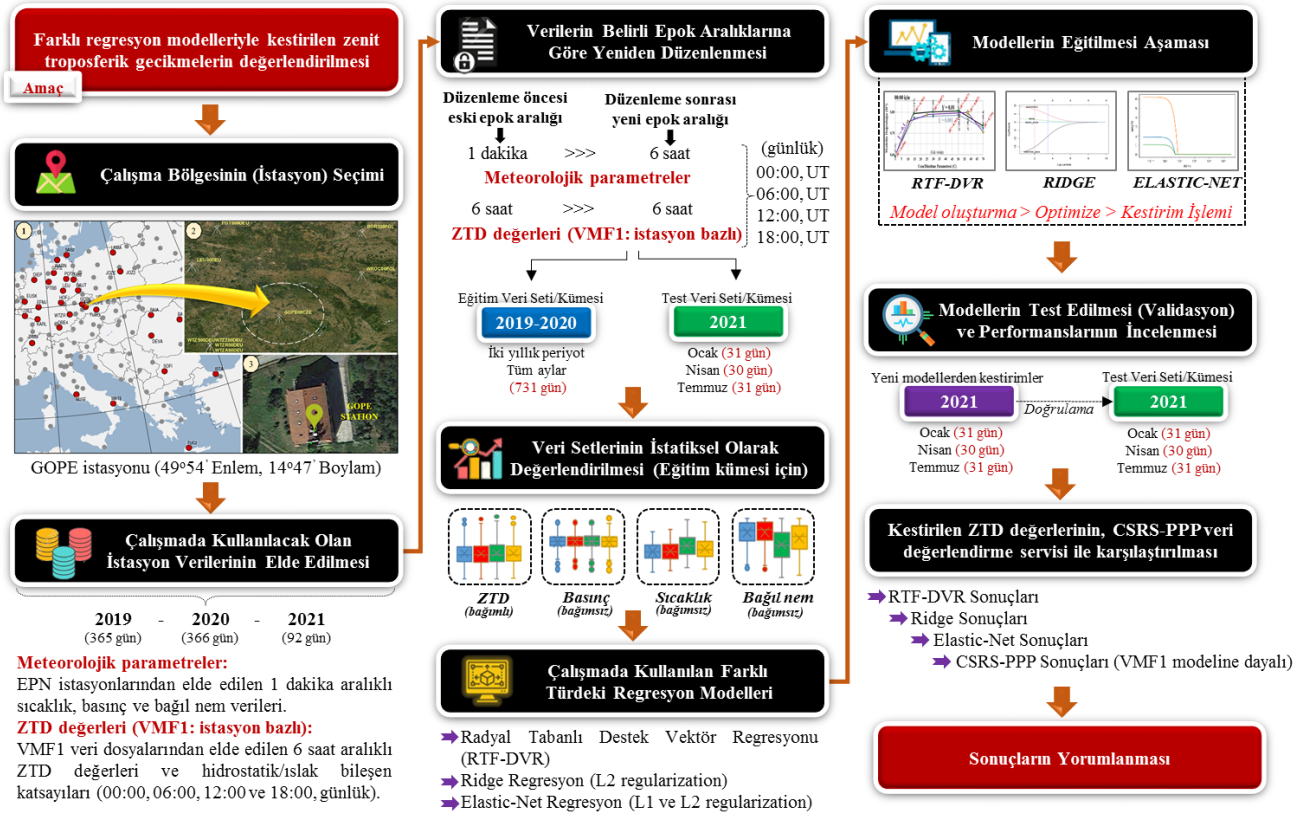
buharı ve tomografi tabanlı düşey profillerin sinerjisini kullanmak için model sunmuştur. Manandhar vd. [21] tarafından, su buharı basıncının düzensizliği üzerine çalışmalar gerçekleştirilmiş olup PWV ve meteorolojik sensörler kullanılarak saatlik yağış tahmini yapılmıştır. Selbesoğlu [22], GNSS istasyonları ve meteorolojik verilerin entegre edilmesi ile Yapay Sinir Ağı (YSA) tabanlı troposferik ıslak gecikmenin enterpolasyonu için doğruluk araştırması gerçekleştirmiştir. Deng vd. [23], uzun mesafeli az sayıdaki referans istasyonu altında troposferik gecikme enterpolasyonunun düşük hassasiyeti problemini çözmek için Destek Vektör Regresyonu'na (DVR) dayalı yöntem önermiştir. Çeşitli yöntemlerin troposfer modellemesinde kullanımı, troposferik gecikme kestirimi için yeni modellerin oluşturulmasını tetiklemiştir [24].

Bu çalışmada, makine öğrenimine dayalı Radyal Tabanlı Destek Vektör Regresyonu (RTF-DVR), Ridge ve Elastik-Net regresyon modellerinin ZTD kestirimi için alternatif yöntemler olarak tercih edilebilirliği araştırılmıştır. Modellemeler için ilk olarak, çalışma bölgesi (istasyon) seçimi gerçekleştirilmiş ardından çalışmada kullanılacak olan meteorolojik ve istasyon bazlı VMF1 (Vienna Mapping Function) ZTD verileri temin edilmiştir. Veriler; 2019-2020 olmak üzere iki yıllık dönem ile 2021 yılındaki Ocak, Nisan ve Temmuz'u içeren üç aylık dönemi kapsamaktadır. 2019 ve 2020 yılı verileri modellerin eğitilmesi, 2021 yılındaki üç aya ait veriler ise test/doğrulama amacıyla kullanılmıştır. Eğitilen modeller üzerinden ZTD'ler kestirilerek farklı regresyon modeli türleri analiz edilmiştir. Son olarak, yeni modellerden elde edilen ZTD değerleri, internet tabanlı GNSS değerlendirme servislerinden Canadian Spatial Reference System–Precise Point Positioning (CSRS-PPP) ZTD değerleriyle karşılaştırılıp yeni modellerin performansı değerlendirilmiştir.

2 Materyal ve metod

Küresel ısınma, iklim değişikliği ve atmosfer fiziğindeki değişimler sonucu atmosferik parametrelerde ani hareketlilik veya beklenmeyen dönemlerde beklenmedik hava koşulları gözlemlenmektedir. GNSS sinyalleri bu durumdan doğrudan etkilenir ve troposferik gecikmenin tahmini zorlaşır. Dolayısıyla mevcut durumdaki yüzey meteorolojik ve ampirik troposfer modelleri gecikme etkisini tam olarak yansıtamamaktadır. Ayrıca referans istasyonlarındaki elektrik kesintisi, donanım arızaları ve sinyal bozulmaları gibi çeşitli faktörler, troposferik gecikme dosyalarının arşivlenmesini engeller. Bu türden sorunlara alternatif çözümler sunma açısından makine öğrenimi çerçevesindeki farklı regresyon modelleri kullanılmış ve modellerin ZTD kestirimindeki performansının değerlendirilmesi amaçlanmıştır. Çalışma kapsamında hazırlanan iş akış süreci Şekil 1'de özetlenmiştir.

Uygulamadaki süreci takiben, ilk olarak istasyon seçimi gerçekleştirilmiş ve istasyon verileri düzenlenerek modellemeler için hazır hale getirilmiştir. Veri setlerinin istatistiksel olarak değerlendirilmesi yapıldıktan sonra farklı regresyon modelleri kurulmuş ayrıca test edilmiştir. En son, yeni modellerden kestirilen ZTD değerleri, CSRS-PPP'ye ait ZTD değerleriyle karşılaştırılmıştır.



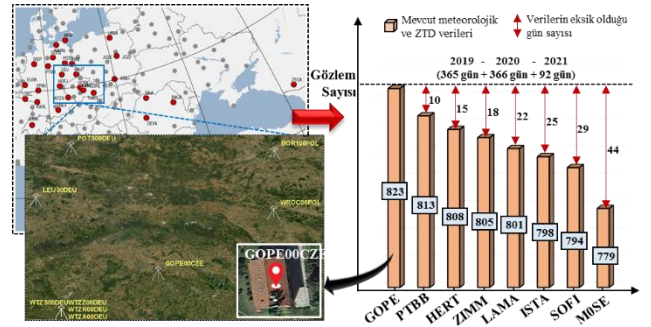
Şekil 1. Çalışma kapsamında takip edilen iş akış süreci

2.1 Çalışma bölgesinin (istasyon) seçimi

ZTD kestirimi yapabilmek için girdi verilerine yani meteorolojik parametrelere ihtiyaç vardır. Bu parametreler, referans istasyonu üzerindeki meteorolojik sensörlerden üretilmektedir. Ancak dünya genelinde, GNSS ve meteorolojik sensörlerin birlikte çalıştığı çok az sayıda referans istasyonu (46 istasyon) mevcuttur. Bu çalışmada, EUREF Sabit GNSS Ağı (EPN: EUREF Permanent GNSS Network) içerisinde farklı bölgelerde bulunan GNSS-meteorolojik sensörlü istasyonlar araştırılmış ve bu yönde istasyon seçimi gerçekleştirilmiştir. İstasyon seçiminde, regresyon modellerinden yüksek performans elde edilmesi açısından yeterli veri sayısının sağlanması, eksiksiz, düzenli ve doğru atmosferik parametrelerin modelde kullanılması gibi unsurlara dikkat edilmiştir. 2019-2020 yılları arasındaki tüm günler ile 2021'deki üç farklı ayı (Ocak-31, Nisan-30 ve Temmuz-31) kapsayan 92 güne ait veriler incelenmiş ve ardından eksik gözlem sayıları bulunmuştur. Burada 2019 ve 2020 yılı verileri regresyon modellerinde eğitim veri seti açısından, 2021 yılındaki ocak, nisan ve temmuz ayları ise test veri seti açısından önemlidir. Dolayısıyla, tüm incelemeler sonucunda, farklı regresyon teknikleriyle ZTD kestirimi yapılabilmesi için Çek Cumhuriyeti bünyesindeki GOPE (GOPE00CZE) istasyonu kullanılmıştır (Şekil 2).

GOPE istasyonu, Avrupa bölgesinde yer almakta olup yaklaşık olarak 49°54'(N) enlemi ve 14°47'(E) boylamında konumlanmaktadır. Bu istasyon, Avrupa'yı kapsayan EPN ile Uluslararası GNSS Servisi (IGS: International GNSS

Service) olmak üzere her iki farklı GNSS ağına da aktif rol oynamaktadır.



Şekil 2. Çalışma kapsamındaki istasyonun seçilmesi

Ek olarak EPN'de güvenilir istasyon olması, anlık meteorolojik verileri toplayacak sensörün bulunması ve üretilen ham verilerin VMF/GPT Serileri gibi veri servislerinde de kabul görmesi GOPE istasyonunun tercih edilmesinde etken olmuştur.

2.2 Verilerin elde edilmesi ve düzenlenmesi

Çalışmada kullanılan veriler; yerinde ölçülen meteorolojik parametreler ve VMF1 istasyon bazındaki ZTD değerleridir. Meteorolojik parametreler, EPN istasyonlarına ait ham veri ve analiz sonuçlarının 1 dakika aralıklarla yayınlandığı <https://www.epncb.oma.be/> adresinden temin

edilmiştir. ZTD değerleri ise, VMF veri sunucusundan elde edilen analiz sonuçlarının 6 saat aralıklarla (00:00, 06:00, 12:00 ve 18:00, günlük) daha sonra yayımlandığı <https://vmf.geo.tuwien.ac.at/> adresinden alınmıştır. VMF1 modeline dayalı ZTD değerlerinin kullanılmasının iki nedeni vardır:

1) VMF1 modelinin, ZTD değerini kestirmede en yüksek doğruluğa sahip olması, güvenilir model olması ve yerinde gözlemlerle atmosferik parametreleri iyi belirlemesi,

2) Regresyon modellerinden kestirilen ZTD değerlerini karşılaştırmak amacıyla kullanılan CSRS-PPP'nin troposferik model olarak VMF1'i kullanması ve sonuç dosyasının bu formatta hazırlanmasıdır.

2019-2020 yılları ile 2021 yılındaki üç ayı içeren veri dosyaları ilgili yerlerden elde edildikten sonra verilerin hazırlanması aşamasına geçilmiştir. Farklı kuruluşlar/veri servisleri, istasyon verilerini farklı epok aralıklarında paylaştıkları için veri setleri çalışmanın içeriğine göre oluşturulmalıdır. Bu bağlamda dosyaların belirli formata getirilmesi için Python v3.8 programlama dili kullanılmıştır. CSRS-PPP'den 30 saniye aralıklarla elde edilen gecikme değerleri, düzenleme sonucunda 6 saat aralıklı çıktı dosyasına dönüştürülmüştür. Benzer işlem meteorolojik parametreler için de uygulanmıştır (1 dakikadan 6 saatliğe). Sonuç olarak tüm kullanılan veriler dört farklı zaman dilimine (00, 06, 12, 18) göre düzenlenmiştir.

2.3 Veri setlerinin istatistiksel değerlendirilmesi

Kullanıcının kararlarını bilimsel forma dönüştürme, modellerin en etkin bir şekilde çalışmasına yardımcı olma, veri yapısını kolay tanıma ve yorumlama açısından verilerin istatistiksel olarak değerlendirilmesi büyük önem taşımaktadır. Özellikle belirsizlik durumlarında, çeşitli değerlendirmeler sonucu çözüm üretilerek veride nereye bakılacağı, hangi noktaya odaklanılacağı ve yapısının nasıl olacağı konularında karar vericiye fikirler sunulur. Bu bakımdan çalışma kapsamında modellerin eğitilmesi sırasında kullanılacak olan 2019 ve 2020 verileri değerlendirilmiştir. Basınç, sıcaklık, bağıl nem ve ZTD değerleri üzerinden kutu grafikleri (boxplot) oluşturulmuştur (Şekil 3).

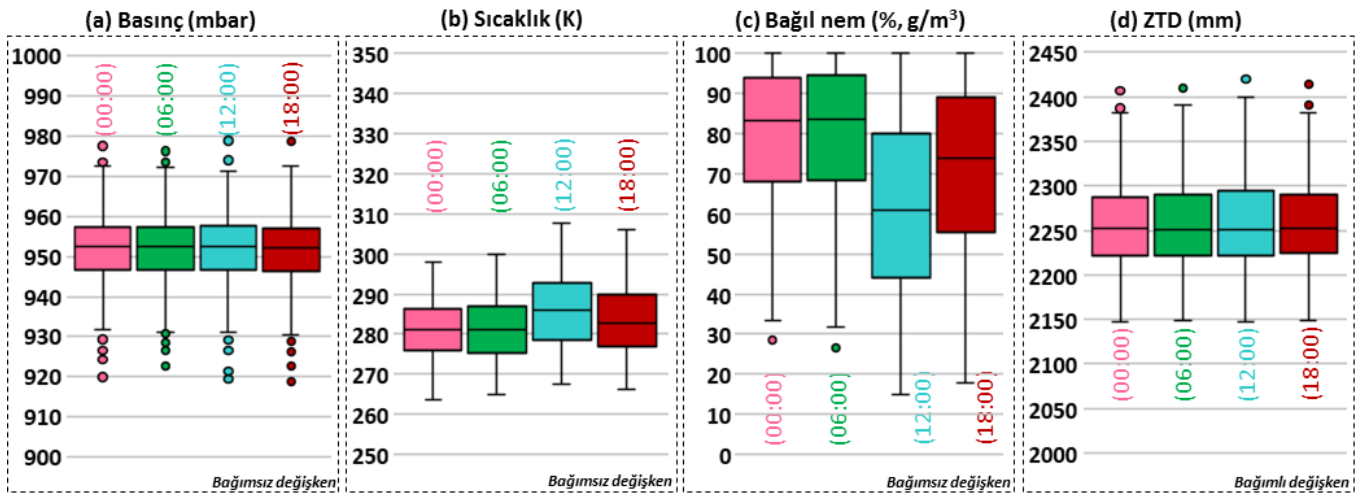
Şekil 3'deki değişkenlere bakıldığında, uç sınırların dışında kalan bazı noktalar bulunmaktadır. Bu noktalar gözlem yapılan gündeki verinin genel dağılıma uymadığını yani aykırı olduğunu ifade etmektedir. Normal koşullarda aykırı olan verilerin veri setinden çıkartılması gerekir, ancak burada farklı yol izlenmiştir. Aykırı olarak belirlenen veri/değişkenin diğer değişkenlerle olan ilişkisi incelenmiş gözlem dosyalarındaki genel dağılımına bakılmıştır. Eğer aykırılık tüm değişkenlerde aynı güne ait ise o günün verileri modellerin eğitilmesinde tercih edilmemiştir. Bu bağlamda iki yıllık periyot içerisinde 6 günün (2019: 23 ve 30 Temmuz, 10 ve 21 Ağustos, 23 Ekim; 2020: 4 Eylül) verileri aykırı olarak tespit edilmiş ve modelin başarısını düşürmemesi adına bu veriler eğitim kümesinde kullanılmamıştır. Sadece sonuç grafikleri oluşturma ve yorumlama aşamasında ilgili günün verileri dâhil edilmiştir.

2.4 Radyal tabanlı destek vektör regresyonu (RTF-DVR)

Destek vektör regresyonu (DVR), destek vektör makinelerinin (DVM) bir formu olup regresyon problemlerinin çözümünde kullanılmaktadır. DVR'de çözülmesi gereken ana konu, gözlem yapılan eğitim veri setindeki hatayı en az yapmak yerine genelleştirilmiş hata sınırını minimuma indirmektir [25]. Regresyonda kullanılan destek vektörlerinin genel ifadesi Denklem (1)'de verilmiştir.

$$(X_1, Y_1) \dots (X_n, Y_n); X \in R^d; Y \in R; f(X) = w \cdot X_i + b \quad (1)$$

Burada, X_n : d-boyutlu uzayda giriş vektörünü, Y_n : giriş vektörüne karşılık gelen çıkış vektörünü, w : ağırlık vektörünü, b : sapmayı ifade etmektedir. DVR'nin matematiği, e-duyarsız (ϵ) kayıp fonksiyonu olarak adlandırılan bir kayıp fonksiyonu tanımlamaya dayanmaktadır ve buna bağlı olarak gerçek değerlerin belirli bir mesafesi içinde olan hatalar görmezden gelinir. Amaç, her bir X_i değerini kullanarak, önceden kestirilen e-duyarsız (ϵ)



Şekil 3. Basınç (a), sıcaklık (b), bağıl nem (c) ve ZTD (d) değerlerine ait kutu grafikleri

değerinden daha büyük olmayan bir mesafede Y_i değerini tespit edebilen $f(X)$ fonksiyonunu bulmaktır. Yöntemin genel fonksiyonu Denklem (2)'de gösterilmektedir.

$$f(X) = |Y_i - (w \cdot X_i + b)| \leq \varepsilon \quad (2)$$

Burada algoritma, ε 'den küçük olduğunda hata olmadığını, ancak ε 'den büyük olduğunda ise bir hataya sahip olduğunu bulur. Destek vektörleri, yüksek boyutlu öznelik uzayında ε fonksiyonunu kullanarak lineer regresyon gerçekleştirir ve eş zamanlı olarak $\|w\|^2$ 'yi en aza indirerek model karmaşıklığını ortadan kaldırır. Ancak lineerlik her zaman oluşmaz. Dolayısıyla, doğrusal olmayan uygulamalarda bir çekirdek fonksiyonu kullanılarak girdi verilerinin yüksek boyutlu uzaya dönüştürülmesiyle sonuç bulunabilir [26, 27].

DVR yönteminde model kestirim performansını üç kontrol mekanizması etkiler. Bu mekanizmalar; 1) Çekirdek tipi, 2) Çekirdek fonksiyonu ve 3) Düzeltme (C) parametresidir. Çekirdek tipi olarak ifade edilen şey modelin lineer, polinom ve radyal tabanlı fonksiyon (RTF) çekirdeklerinden en uygun olanının seçimidir. Çekirdek fonksiyonu ve düzeltme parametreleri, seçilen çekirdek türünün özellikleridir. Doğrusal, polinom gibi çekirdeklerin çözemediği karmaşık problemlerde ise RTF tercih edilir [28]. Radyal tabanlı fonksiyonun temel yapısı Denklem (3),

$$e^{-\gamma \|X_i - Y_i\|^2} \quad (3)$$

ile gösterilmektedir. e , doğal logaritmik tabanı ifade ederken, γ parametresi kullanıcı tarafından modelin optimizasyonu esnasında belirlenmektedir.

2.5 Ridge regresyonu

1970'de Hoerl ve Kennard tarafından önerilen yöntemin amacı, hataların kareleri toplamını en aza indiren katsayıları düzeltme uygulayarak bulmaktır [29]. En Küçük Kareler (EKK) yönteminden farklı olarak, parametre kestirimlerinin varyanslarını küçültmek ve daha güvenilir kestirimler elde etmek için korelasyon matrisine λ_2 parametresi eklenir [30]. Diğer bir deyişle, EKK yönteminden elde edilen katsayılar yanlış değildir. Ancak varyans önemli ölçüde yüksektir ve katsayılar gerçek değerlerden uzaktır. Ridge regresyonunda ise, kestirimlere bir miktar yanlışlık eklenir ve bu da standart sapmanın azalmasına neden olur. Bu nedenle daha güvenilir kestirimler elde edilebilir. L_2 normu ya da Ridge regresyon olarak bilinen yöntemin matematiği Denklem (4)'te verilmiştir [31].

$$\beta_{Ridge} = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 \quad (4)$$

(hataların kareleri toplamı) (düzeltme terimi)

Düzeltme terimi içerisinde yer alan λ_2 parametresi $\beta_1, \beta_2, \dots, \beta_j$ katsayılarını sıfıra doğru daraltmakta olup düzeltme teriminin regresyon katsayılarını kontrol etmeyi sağlar. λ_2 parametresi sadece $\beta_1, \beta_2, \dots, \beta_j$ katsayıları ile etkileşim

halindedir β_0 'ı etkilememektedir [32]. Ridge regresyonu için düzeltme terimindeki λ_2 'nin 0'a eşit olması, parametrenin hiçbir etkisinin olmadığı ve eşitliğin EKK yöntemine dönüştüğü anlamına gelmektedir. λ_2 'nin sonsuza doğru ilerlemesi ise, parametrenin büyüdüğünü ve regresyon katsayılarının neredeyse sıfıra eşit olacağı anlamını taşımaktadır. EKK yöntemi katsayı kestirimleri için sadece tek bir sonuç verirken, Ridge regresyon λ_2 'nin her farklı değeri için farklı katsayı kestirimleri sunmaktadır. Dolayısıyla katsayı kestirimleri sürekli değiştiğinden, Ridge regresyonda en iyi λ_2 değerinin seçimi önemlidir. Modelin elde edilmesinde kullanılacak olan en iyi λ_2 değerinin belirlenmesi için çapraz doğrulama (cross-validation) yöntemi tercih edilmektedir [33]. Ridge regresyon yönteminin bazı avantajları şunlardır:

* Tüm değişkenler ile modeli kurar ve modeldeki ilişkisiz değişkenleri modelden çıkarmak yerine katsayılarını sıfıra yaklaştırır.

* Aşırı öğrenmeye (overfitting) karşı dirençli bir yapıdadır.

* Değişken sayısı fazlaysa EKK'ya göre daha iyi sonuçlar sunar ve çoklu doğrusallık problemi için etkilidir.

* Yanlıdır ancak varyansı düşüktür.

2.6 Elastik-Net regresyonu

Elastik-Net regresyonu, Ridge ve Lasso regresyon yöntemlerinden yola çıkılarak 2005 yılında Zou ve Hastie tarafından geliştirilmiştir [34]. Ridge regresyonunda olduğu gibi aynı işlem adımı takip edilerek düzeltme işlemi gerçekleştirilir. Her bir katsayının, hatalarının kareleri toplamındaki rolüne göre λ_2 parametresi, β katsayılarına düzeltme uygular. Değişken seçimi ise Lasso regresyon yöntemindeki gibi yapılır. Anlamsız değişkenlerin katsayıları sıfır yapılır ve otomatik değişken seçimi gerçekleşmiş olur [35]. L_1 ve L_2 normları olarak tanımlanan Elastik-Net regresyonunun matematiği Denklem (5),

$$\beta_{els} = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1 \sum_{j=1}^p |\beta_j| \quad (5)$$

(hataların kareleri toplamı) (düzeltme terimleri)

ile gösterilmektedir. λ_1, λ_2 parametrelerinin seçimi, k-katlı çapraz doğrulama yöntemiyle yapılır. λ_1 ve λ_2 için, belirli değerleri içeren bir küme seçilir ve her bir değer test hataları hesaplanır. En düşük hata değerini veren $\lambda_{1,2}$ parametreleri seçilerek elastik net regresyon modeli oluşturulur [36].

2.7 Model performans değerlendirme ölçütleri

Farklı regresyon yöntemlerini ve yöntemlerden elde edilen kestirim değerleri ile CSRS-PPP'nin yayınladığı gecikme değerlerini karşılaştırmak için R^2 değeri (Belirleme katsayısı), Karesel Ortalama Hata (KOH), Ortalama Mutlak Hata (OMH) ve Ortalama Mutlak Yüzde Hata (OMYH) sonuçları performans analizlerinde kullanılmıştır. R^2 değeri, kestirilen regresyon doğrusunun gözlemler kümesinin eğilimini ne derecede temsil ettiğini gösteren bir katsayı olduğu için çalışmada önem arz etmektedir. Model

performans değerlendirme ölçütleri, Denklem (6), Denklem (7), Denklem (8) ve Denklem (9)'da sunulmaktadır.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{\sum_{i=1}^n (\bar{y}_i - \hat{y})^2} \quad (6)$$

$$KOH = \pm \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (7)$$

$$OMH = \frac{1}{n} \sum_{i=1}^n l y_i - \bar{y}_i l \quad (8)$$

$$OMYH = \frac{\sum_{i=1}^n l y_i - \bar{y}_i l}{n} * 100 \quad (9)$$

y_i : gözlenen/gerçek değerlerini, \bar{y}_i : model kestirim değerlerini, \hat{y} : ortalama değerleri ve i : 1,2, ..., n 'i temsil etmektedir.

3 Bulgular ve tartışma

3.1 Modellerin eğitim sonuçları

Çalışma kapsamında GOPE referans istasyonunun iki yıllık (2019-2020) verileri düzenlenerek RTF-DVR, Ridge ve Elastik-Net regresyon modellerinin eğitilmesi amaçlanmıştır. Modellemeler için kullanılan eğitim veri setleri, meteorolojik veriler ve istasyon bazında VMF1 modelinden elde edilen ZTD değerleri içermektedir. Modellerden yüksek performansa sahip kestirim sonuçları elde etmek için en uygun parametrelerle modelin eğitilmesi gerekir. Dolayısıyla her bir modele ait en uygun parametrelerin belirlenmesi amacıyla optimizasyon işlemi uygulanmıştır. Eğitime, optimizasyon ve kestirim işlemleri numPy, scikit-learn, matplotlib, pandas, seaborn, statsmodels ve sciPy kütüphaneleri aracılığıyla Python v3.8

programlama dilinde gerçekleştirilmiştir. Modellerin eğitim ve test doğrulukları için parametrelerin seçiminde 10-katlı çapraz doğrulama ve grid arama (grid-search) yöntemlerinden yararlanılmıştır. Farklı modellerin optimizasyon öncesi başlangıç parametreleri, optimizasyon sonrası en uygun parametreleri ve bu parametrelere bağlı R^2 ile KOH değerleri Tablo 1'de sunulmuştur. Optimizasyon yöntemleri sonucu parametre değişikliklerine dayalı model doğrulukları değerlendirilmiş ve regresyon modelleri için en uygun parametreler seçilmiştir.

Tablo 1'de, RTF-DVR modelinde kullanıcı tarafından belirlenmesi gereken parametrelerin, C ve γ parametreleri olduğu görülmektedir. RTF için iki parametre olan C ve γ 'nın model sonucuna etkisi kesinlikle söz konusudur. Ancak kullanılan hangi parametre çiftinin modelde en iyi sonucu verdiği bilinmemektedir. Optimize edilmeden önce parametreler, başlangıç değeri olarak C=1 ve γ =1 şeklinde modelde tanımlanmaktadır. Optimizasyon yöntemleri ile bu parametreler, oluşturulacak olan gecikme kestirim modeli için tespit edilmiştir. İlk aşamada C parametresi; [1, 1000] eşik aralıklarında değeri 10 artırılarak denenmiş ve en uygun eşik değeri "C=160" olarak tespit edilmiştir. C'nin 160 olduğu durumda model, yüksek R^2 ve düşük KOH değerlerine sahip olmuştur. İkinci aşamada ise ilk aşamada belirlenen parametrelerin yakınında daha küçük bir aralıkta ikinci bir arama yapılarak daha iyi parametre değerlerine ulaşılmaya çalışılmıştır. Ancak bu kısımda sadece C parametresi değil γ parametresi de hesaba katılmıştır. C; [150, 170] eşik aralıklarında değeri 1 artırılarak denenmiştir. Diğer taraftan γ parametresi; $[10^{-3}, 10^{-1}]$ eşik aralığında değeri 0.01 artırılarak sonuçlardaki farklılıklar araştırılmıştır. C ve γ ikili parametresi tüm değerler için irdelenmiş olup değerlendirmeye alınmıştır. Değerlendirme sonucuna göre, C'nin 165 ve γ parametresinin 0.02 olduğu durumda model en yüksek R^2 ve en düşük KOH değerlerine ulaşmıştır. Yani, RTF-DVR modelinin "C=165" ve " γ =0.02" parametre değerleriyle eğitilmesi sonucuna varılmıştır.

Tablo 1. Modellerin optimizasyon öncesi ve sonrası parametre değerleri

Model	RTF-DVR (Radyal tabanlı fonksiyon)	
Parametre	Optimizasyon öncesi: (C = 1, γ = 1)	Optimizasyon sonrası: (C = 165, γ = 0.02)
Saat	Tüm saatler için (Ort.)	Tüm saatler için (Ort.)
R ² değeri	0.831	0.871
KOH (mm)	± 20.2	± 17.1
Model	Ridge Regresyonu (L ₂ normu)	
Parametre	Optimizasyon öncesi: λ_2 (alpha) = [0, ∞]	Optimizasyon sonrası: λ_2 (alpha) = 0.005
Saat	Tüm saatler için (Ort.)	Tüm saatler için (Ort.)
R ² değeri	0.768	0.793
KOH (mm)	± 24.6	± 22.7
Model	Elastik-Net Regresyon (L ₁ ve L ₂ normları)	
Parametre	Optimizasyon öncesi: λ_1 = [0, ∞]; λ_2 = [0, ∞]	Optimizasyon sonrası: λ_1 = 0.520; λ_2 = 0.524
Saat	Tüm saatler için (Ort.)	Tüm saatler için (Ort.)
R ² değeri	0.770	0.794
KOH (mm)	± 24.4	± 22.5

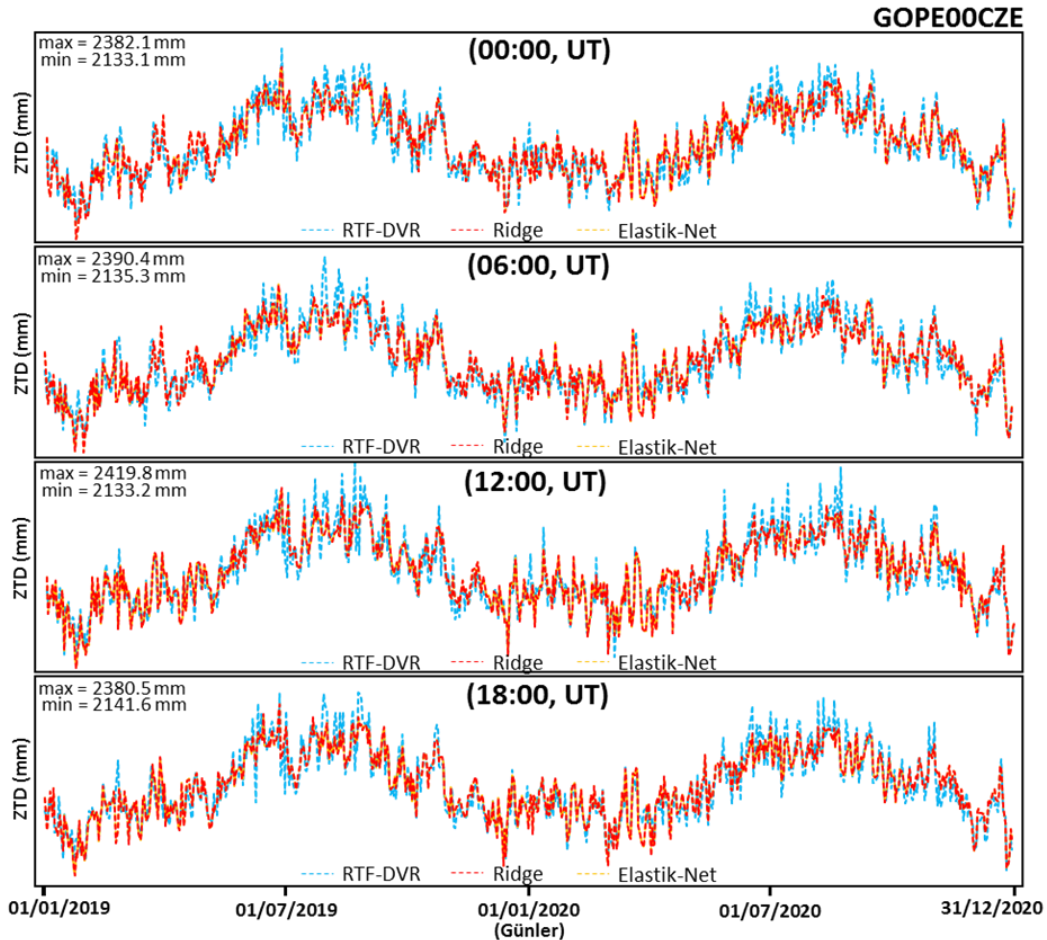
Ridge ve Elastik-Net regresyon modelinde ise belirlenmesi gereken parametreler ve işlem adımları, birbirine benzerlik göstermektedir. Bu modellerde düzeltme teriminin olduğu kısımdaki parametreler (λ_1 ve λ_2) kullanıcı tarafından belirlenir. Burada λ_1 ve λ_2 'in 0'a yaklaşması, parametrenin hiçbir etkisinin olmadığı ve eşitliğin EKK yöntemine dönüştüğünü, sonsuza doğru ilerlemesi ise, parametrenin büyüdüğünü ve regresyon katsayılarının neredeyse sifira eşit olacağını ifade etmektedir.

Optimize edilmeden önce bu parametreler, başlangıç değeri olarak $\lambda_1=0.00$ ve $\lambda_2=0.00$ şeklinde modelde tanımlanmakta ve sonsuza (∞) doğru gidebilmektedir. Optimizasyon işlemleri ile bu parametreler, oluşturulacak olan gecikme kestirim modeli için tespit edilmiştir. Optimum $\lambda_{1,2}$ parametrelerine ulaşmak için $[0, \infty]$ eşik aralıklarında değerler 0.001 artırılarak sonuçlar incelenmiştir. Çalışmadaki veri setinin düşük $\lambda_{1,2}$ parametrelerinde daha iyi sonuçlar verdiği görülmüş ve bu bağlamda $[0, 1]$ eşik aralığında optimum parametrelerin araştırılması düşünülmüştür. Ridge regresyonu için λ_2 parametresinin 0.005 olduğu durumda, en yüksek R^2 ve en düşük KOH değerlerine ulaşılmıştır. Bu nedenle " $\lambda_2(\alpha)=0.005$ " parametresi Ridge modelinin eğitilmesinde kullanılmıştır. Elastik-Net regresyonunun eğitilmesi için " $\lambda_1=0.520$ " ve " $\lambda_2=0.524$ " en uygun parametreler olarak bulunmuştur.

Parametreler aracılığıyla eğitilen modellerin, 2019-2020 yıllarındaki sonuç ZTD değerleri (eğitim için) arasındaki uyum değerlendirilmiştir (Şekil 4).

Buna göre Şekil 4'de, Ridge ve Elastik-Net regresyon modellerinin birbirine çok yakın sonuçlar ürettiği, RTF-DVR modelinin ise bu iki modele göre biraz daha farklı sonuçlar elde ettiği gözlemlenmiştir. Ridge ve Elastik-Net ile RTF-DVR modelleri arasındaki farklılıkların özellikle, yaz dönemine karşılık gelen temmuz ve ağustos aylarında arttığı tespit edilmiştir. Bu farklılıkların nedeni, sıcaklık ve nemin yükselmesiyle, su buharı basıncı ve gecikme etkisinin büyümesidir. Buna bağlı olarak yaz aylarında elde edilen gecikme değerlerinin tüm yıl içerisinde elde edilen değerlere nazaran normal dağılım sergilememesi sonucu modellerin matematiğine göre sonuçlar farklılaşmaktadır. Ridge ile Elastik-Net modelinin arka planındaki matematiksel denklemin benzerlik göstermesinden dolayı büyük ZTD farklılıkları oluşmamıştır. Yaz aylarının dışındaki dönemlerde ise farklılıklar saptanmamakla birlikte modeller arasındaki en az değişimler kış aylarında gerçekleşmiştir. Şekil içerisinde yer alan min ve max değerleri çeşitli saat dilimlerinde gözlemlenen en düşük ve en büyük ZTD değerlerini ifade etmektedir.

Modellere ait dört farklı zaman dilimi için hesaplanan R^2 , KOH, OMH ve OMYH değerleri Tablo 2'de sunulmuştur.



Şekil 4. Eğitilen modellerdeki sonuç ZTD değerlerinin birbiriyle uyumu

Tablo 2. Modellerdeki ilgili zaman dilimleri için elde edilen sonuçlar

MODEL:	RTF-DVR				Ridge Regresyonu				Elastik-Net Regresyon			
	00:00	06:00	12:00	18:00	00:00	06:00	12:00	18:00	00:00	06:00	12:00	18:00
Saat:	00:00	06:00	12:00	18:00	00:00	06:00	12:00	18:00	00:00	06:00	12:00	18:00
R ² değeri	0.855	0.845	0.901	0.883	0.791	0.773	0.811	0.798	0.792	0.775	0.812	0.799
KOH (mm)	± 18.2	± 18.9	± 15.2	± 16.3	± 23.0	± 23.8	± 21.9	± 22.3	± 22.8	± 23.6	± 21.6	± 22.0
OMH (mm)	10.7	11.5	8.6	9.4	18.0	18.8	16.9	17.1	17.9	18.6	16.9	17.0
OMYH	0.47	0.51	0.38	0.41	0.79	0.82	0.74	0.75	0.79	0.81	0.74	0.72

Tablo 2'ye göre, tüm modeller için en iyi sonuç değerlerin 12:00'da, en kötü sonuçların 06:00 saat diliminde elde edildiği sonucuna varılmıştır. Ridge ve Elastik-Net yönteminde tüm saatlerdeki değerler birbirine yakınlık göstermektedir. Ancak bu yöntemlerin eğitime başarısı RTF-DVR yöntemine göre düşüktür. Tablo 1 ve Tablo 2'de de görüldüğü üzere RTF-DVR yönteminin eğitimdeki model başarısı daha yüksek olarak tespit edilmiştir.

3.2 Yeni ZTD değerlerinin kestirilmesi ve test edilmesi

Modellerin eğitimi aşaması tamamlandıktan sonra, sonuç modeller üzerinden ZTD kestirimleri gerçekleştirilmiştir. Sonuç modellerin doğruluğu ve performansı, 2021 yılındaki ocak, nisan ve temmuz aylarına ait 92 günün verisi kullanılarak test edilmiştir. Uygulamada bu ayların seçilme nedeni, troposferik gecikme etkisinin bu dönemlerde sırasıyla en az, ortalama seviyede ve en fazla olarak görülmesidir. Çalışmada ilk olarak modellerden kestirilen ZTD değerleri, ilgili yerden temin edilen ZTD değerleriyle test edilmiştir. Test etmek için temin edilen veri seti, VMF veri sunucusundan istasyon bazında sonradan yayınlanan

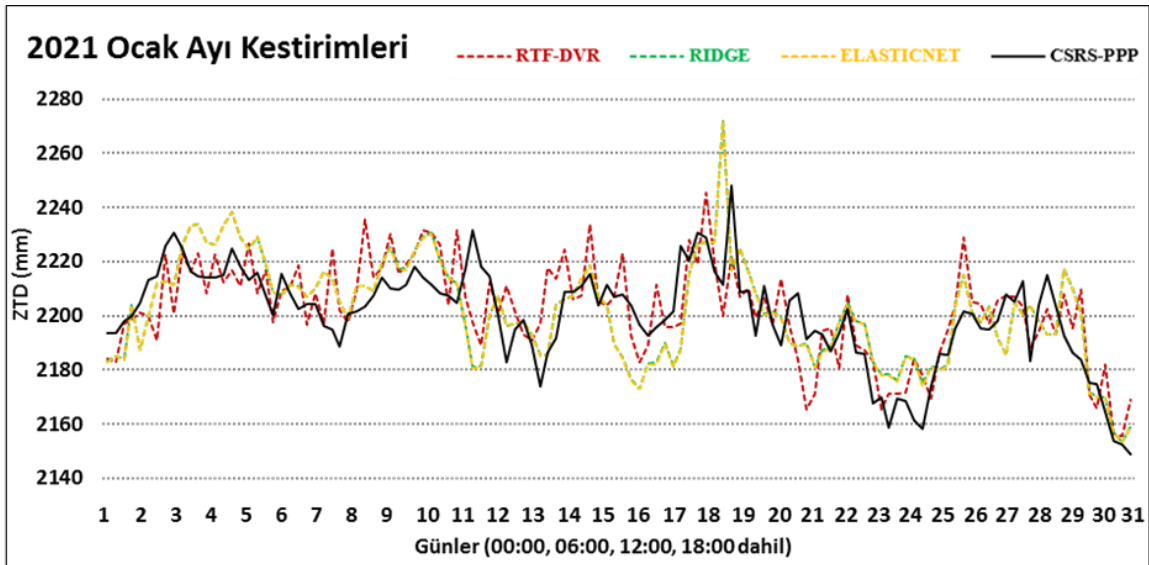
gecikme değerleridir. Model performanslarının değerlendirilmesi için KOH ölçütü kullanılmıştır. Analiz sonuçlarına göre; RTF-DVR ile VMF veri sunucusundan temin edilen veriler arasındaki KOH; ocak, nisan ve temmuz ayları için sırasıyla; ± 13.3 mm, ± 16.4 mm ve ± 24.8 mm olarak hesaplanmıştır. Benzer şekilde Ridge ve Elastik-Net ile VMF veri sunucusuna ait veriler arasındaki KOH'lar sırasıyla, ± 13.9 mm, ± 16.8 mm ve ± 27.2 mm bulunmuştur. Tüm model sonuçları ele alındığında, yaz günlerinin model başarılarını düşürdüğü diğer kış ve ilkbahar günlerinin başarıyı çok fazla etkilemediği sonucuna varılmıştır.

3.3 Kestirilen ZTD değerlerinin CSRS-PPP ile karşılaştırılması

Çalışmanın son aşamasında, yeni modellerden kestirilen ZTD değerleri, CSRS-PPP'ye ait ZTD değerleriyle karşılaştırılmıştır. Farklı yöntemlere göre üç farklı ay için kestirilen ZTD'lere ait istatistiksel sonuçlar Tablo 3'de verilmiş ve Şekil 5, 6 ve 7'de grafik olarak sunulmuştur. Grafiklerde günler içerisindeki saat dilimleri (00:00, 06:00, 12:00 ve 18:00) ardışık olarak devam etmektedir.

Tablo 3. Farklı yöntemlerden elde edilen ZTD ile CSRS-PPP ZTD'nin karşılaştırılması

Servis	Aylar (2021)	RTF-DVR			Ridge Regresyonu			Elastik-Net Regresyon		
		KOH (mm)	OMH (mm)	OMYH	KOH (mm)	OMH (mm)	OMYH	KOH (mm)	OMH (mm)	OMYH
CSRS-PPP	Ocak	± 13.8	10.6	0.48	± 15.1	11.7	0.53	± 15.0	11.6	0.52
	Nisan	± 14.4	11.5	0.52	± 15.8	11.7	0.53	± 15.0	11.6	0.52
	Temmuz	± 25.4	20.4	0.88	± 27.3	22.0	0.94	± 27.2	21.9	0.93



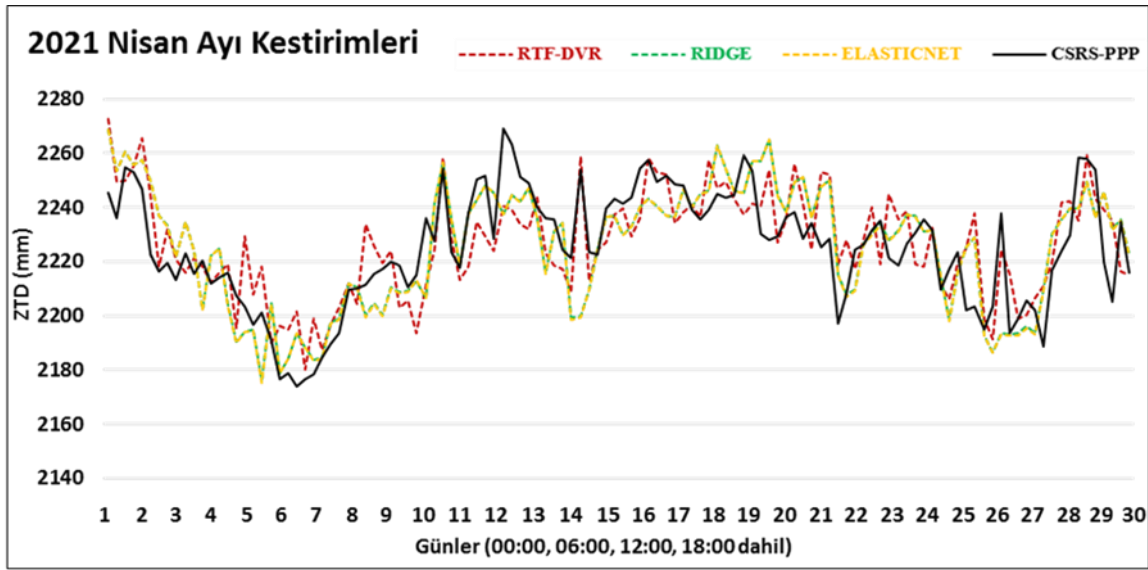
Şekil 5. 2021 Ocak ayındaki ZTD değerlerinin karşılaştırılması

RTF-DVR modelinin ocak ve nisan aylarında, daha düşük karesel ortalama hata değeriyle (sırasıyla ± 13.8 ve ± 14.4 mm) CSRS-PPP'ye yakın sonuçlar elde ettiği gözlemlenmiştir. Bu aylarda Ridge ile Elastik-Net modelleri RTF-DVR'ye kıyasla daha düşük başarılar elde etse de sonuçlar birbirine yakındır ve CSRS-PPP ile aralarındaki hata farkları düşüktür. Ridge ve Elastik-Net modellerinden elde edilen değerlerin birbirine çok yakın olduğu görülmektedir (Şekil 5 ve Şekil 6).

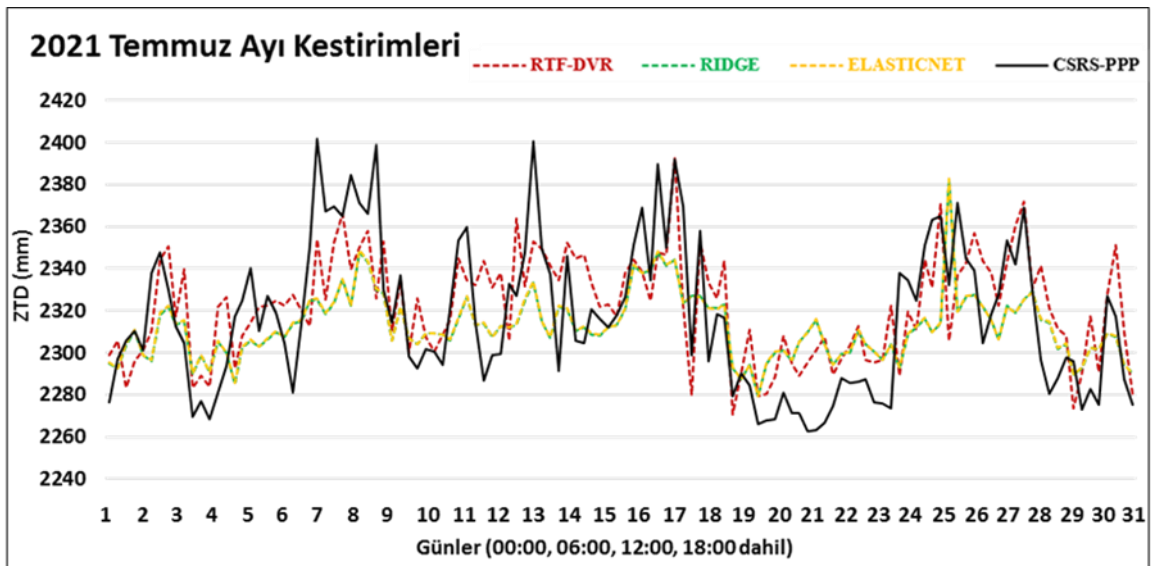
Genel olarak, alternatif yaklaşım şeklinde değerlendirilebilecek RTF-DVR modeli, CSRS-PPP gibi uluslararası kullanımı olan servislere göre ± 14.1 mm KOH

(ocak ve nisan ayı için) ile sonuçlar üretmektedir. Ridge ve Elastik-Net modelleri ise sırasıyla, ± 15.4 ve ± 15.0 mm KOH ile sonuçları elde etmektedir.

Diğer taraftan, tüm kestirim modelleri CSRS-PPP'ye göre temmuz ayında daha çok sapmalara sebebiyet vermiştir. Bunun nedeni modelin eğitilmesi aşamasında da ifade edildiği gibi yaz mevsiminde sıcaklık ve bağıl neme bağlı düzensiz hareketlerdir. Yeni modellerdeki KOH değerlerinin CSRS-PPP'ye göre yaz günlerinde ± 25 mm ile ± 27 mm aralığında değiştiği gözlemlenmektedir. Şekil 7'de CSRS-PPP ile modeller arasındaki sapmaların en çok 7-9 ve 19-23 arasındaki günlerde ortaya çıktığı söylenebilir.



Şekil 6. 2021 Nisan ayındaki ZTD değerlerinin karşılaştırılması



Şekil 7. 2021 Temmuz ayındaki ZTD değerlerinin karşılaştırılması

4 Sonuçlar ve öneriler

Son zamanlarda makine öğrenme algoritmaları; GNSS meteorolojisi, konum belirleme, navigasyon, hassas tarım ve atmosferik araştırmalar olmak üzere pek çok alanda kullanılmaktadır. Bu algoritmalar, GNSS açısından çok önemli katkılar sunmaktadır. GNSS uygulamalarında karşılaşılan birçok temel hata kaynağı vardır ve bu hatalar GNSS kullanıcıları için bir problem olarak ortaya çıkmaktadır. Troposferik gecikme etkisi bunlardan biridir. Özellikle uydu bazı jeodezik ölçmelerdeki gecikme etkilerinin belirlenebilmesi amacıyla mevcut kestirim modelleri kullanılmaktadır. Ayrıca bu modeller halen günümüzde geliştirilmektedir ya da alternatif yeni çözümler aranmaktadır.

Bu çalışmada, ZTD kestirimi için RTF-DVR, Ridge ve Elastik-Net'in kullanılması yeni alternatif yöntemler olarak sunulmuştur. Modelden kestirimler farklı regresyon tiplerine dayalı olarak gerçekleştirilmiştir. Modellemeler için GOPE istasyonunda 2019-2020-2021 dönemlerine ait basınç, sıcaklık ve bağıl nemden oluşan meteorolojik parametreler ve VMF1 ZTD değerleri temin edilmiştir. 2019-2020 yılına ait veri kümesi modellerin eğitilmesinde, geri kalan 2021 yılındaki üç ayı kapsayan günler ise testte kullanılmıştır. Model performansları; R^2 , KOH, OMH ve OMYH ölçütleri ile değerlendirilmiştir.

Modellerin eğitilmesi sonucu, 2019-2020 yıllarındaki tüm günler için en iyi sonuçları RTF-DVR modeli sağlamış ve KOH değerleri ± 2 cm'nin altında kalmıştır. Diğer iki regresyon modeli ise, biraz daha düşük başarıyla birbirine çok yakın sonuçlar elde etmiştir. Eğitilen modeller kullanılarak, 2021 yılı için farklı meteorolojik şartlara sahip ocak, nisan ve temmuz ayında ZTD kestirimleri yapılmıştır. Kestirim sonuçlarına göre, RTF-DVR'deki KOH değerleri; ocak, nisan ve temmuz ayları için sırasıyla; ± 13.3 mm, ± 16.4 mm ve ± 24.8 mm olarak bulunmuştur. Ridge ve Elastik-Net'te ise bu değerler sırasıyla, ± 13.9 mm, ± 16.8 mm ve ± 27.2 mm hesaplanmıştır. Benzer sonuçlar CSRS-PPP karşılaştırmasında da görülmüştür. RTF-DVR modeli CSRS-PPP ile en uyumlu sonuçları vermiştir. Tüm analiz sonuçları birlikte değerlendirildiğinde, yaz günlerinin model başarılarını düşürdüğü diğer kış ve ilkbahar günlerinin başarıyı çok fazla etkilemediği sonucuna varılmıştır. Özet olarak, model başarısı açısından sıralama yapılacak olursa, en yüksek başarıyı RTF-DVR modeli yakalamış ve bunu Elastik-Net ile Ridge modelleri takip etmiştir.

Çalışma sonuçlarından, günlük gecikmeyi tahmin etmek için farklı regresyon modellerinin alternatif olabileceği ve mevcut ZTD ürünlerinde eksik/hatalı verileri iyileştirme aracı olarak tercih edilebileceği görülmektedir. ZTD kestirimlerinde daha iyi sonuçlar elde edilmek istenirse, modellerin uzun vadeli ve farklı arazi koşullarını içeren verilerle kurulması önerilebilir.

Teşekkür

Bu çalışma, Ali Utku AKAR'ın "GNSS gözlemlerindeki troposferik gecikmenin destek vektör makineleri algoritması ile kestirimi" isimli yüksek lisans tezinden üretilmiştir.

Çıkar çatışması

Yazarlar çıkar çatışması olmadığını beyan etmektedir.

Benzerlik oranı (iThenticate): %7

Kaynaklar

- [1] V. B. Mendes, Modeling the neutral-atmospheric propagation delay in radiometric space techniques. UNB geodesy and geomatics engineering technical report 199(10), 1999.
- [2] A. Angrisano, S. Gaglione, C. Gioia, M. Massaro and U. Robustelli, Assessment of NeQuick ionospheric model for Galileo single-frequency users. Acta Geophysica, 61(6), 1457-1476, 2013. <https://doi.org/10.2478/s11600-013-0116-2>.
- [3] M. Kahveci ve F. Yıldız, Uydularla Konum Belirleme Sistemleri (GPS/GNSS): Teori-Uygulama, 8. Basım, Nobel Yayıncılık, Ankara, 2017.
- [4] P. Tregoning and T. A. Herring, Impact of a priori zenith hydrostatic delay errors on GPS estimates of station heights and zenith total delays. Geophysical Research Letters, 33(23), 2006. <https://doi.org/10.1029/2006GL027706>.
- [5] R. Dach, S. Lutz, P. Walser and P. Fridez, Bernese GNSS Software Version 5.2, User manual. Astronomical Institute, University of Bern, Bern Open Publishing, 2015. DOI: 10.7892/boris.72297, ISBN: 978-3-906813-05-9.
- [6] K. Wilgan, F. Hurter, A. Geiger, W. Rohm and J. Bosy, Tropospheric refractivity and zenith path delays from least-squares collocation of meteorological and GNSS data. Journal of Geodesy, 91(2), 117-134, 2017. <https://doi.org/10.1007/s00190-016-0942-5>.
- [7] P. Benevides, J. Catalao and G. Nico, Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors. Remote Sensing, 11(8), 966, 2019. <https://doi.org/10.3390/rs11080966>.
- [8] S. Li, T. Xu, N. Jiang, H. Yang, S. Wang and Z. Zhang, Regional zenith tropospheric delay modeling based on least squares support vector machine using GNSS and ERA5 data. Remote Sensing, 13(5), 1004, 2021. <https://doi.org/10.3390/rs13051004>.
- [9] J. Böhm, A. Niell, P. Tregoning and H. Schuh, Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data. Geophysical research letters, 33(7), 2006. <https://doi.org/10.1029/2005GL025546>.
- [10] J. Boisits, D. Landskron and J. Böhm, VMF3o: the Vienna Mapping Functions for optical frequencies. Journal of Geodesy, 94(6), 1-11, 2020. <https://doi.org/10.1007/s00190-020-01385-5>.
- [11] H. S. Hopfield, Two-quartic tropospheric refractivity profile for correcting satellite data. Journal of Geophysical research, 74(18), 4487-4499, 1969. <https://doi.org/10.1029/JC074i018p04487>.
- [12] J. Saastamoinen, Contributions to the theory of atmospheric refraction. Bulletin Géodésique (1946-

- 1975), 105(1), 279-298, 1972. <https://doi.org/10.1007/BF02521844>.
- [13] H. D. Black, An easily implemented algorithm for the tropospheric range correction. *Journal of Geophysical Research: Solid Earth*, 83(B4), 1825-1828, 1978. <https://doi.org/10.1029/JB083iB04p01825>.
- [14] J. Liu, X. Chen, J. Sun and Q. Liu, An analysis of GPT2/GPT2w+Saastamoinen models for estimating zenith tropospheric delay over Asian area. *Advances in Space Research*, 59(3), 824-832, 2017. <https://doi.org/10.1016/j.asr.2016.09.019>.
- [15] R. Leandro, M. Santos and R. Langley, UNB neutral atmosphere models: development and performance. In *Proceedings of the 2006 national technical meeting of the institute of navigation*, pp. 564-573, 2006.
- [16] S. Serrano-Vincenti, T. Condom, L. Campozano, J. Guamán and M. Villacís, An empirical model for rainfall maximums conditioned to tropospheric water vapor over the Eastern Pacific Ocean. *Frontiers in Earth Science*, 8, 198, 2020. <https://doi.org/10.3389/feart.2020.00198>.
- [17] S. Li, T. Xu, Y. Xu, N. Jiang and L. Bastos, Forecasting GNSS Zenith Troposphere Delay by Improving GPT3 Model with Machine Learning in Antarctica. *Atmosphere*, 13(1), 78, 2022. <https://doi.org/10.3390/atmos13010078>.
- [18] T. Hadas, F. N. Teferle, K. Kazmierski, P. Hordyniec and J. Bosy, Optimum stochastic modeling for GNSS tropospheric delay estimation in real-time. *GPS solutions*, 21(3), 1069-1081, 2017. <https://doi.org/10.1007/s10291-016-0595-0>.
- [19] M. Ding, A neural network model for predicting weighted mean temperature. *Journal of Geodesy*, 92(10), 1187-1198, 2018. <https://doi.org/10.1007/s00190-018-1114-6>.
- [20] M. Łoś, K. Smolak, G. Guerova and W. Rohm, GNSS-based machine learning storm nowcasting. *Remote Sensing*, 12(16), 2536, 2020. <https://doi.org/10.3390/rs12162536>.
- [21] S. Manandhar, Y. H. Lee, Y. S. Meng, F. Yuan and J. T. Ong, GPS-derived PWV for rainfall nowcasting in tropical region. *IEEE transactions on geoscience and remote sensing*, 56(8), 4835-4844, 2018. <https://doi.org/10.1109/TGRS.2018.2839899>.
- [22] M. O. Selbesoğlu, *Global Navigasyon Uydu Sistemleri (GNSS) Gözlemlerinden Elde Edilen Islak Troposfer Gecikmesinin Yapay Sinir Ağları ile Modellenmesi*, Doktora Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Türkiye, 2017.
- [23] J. Deng, M. Xu, X. Yu and A. Zhang, Interpolation estimation method of tropospheric delay for long baseline network RTK based on support vector machine. In *IOP Conference Series: Earth and Environmental Science*, Vol. 192, No. 1, p. 012069. IOP Publishing, 2018. DOI: [10.1088/1755-1315/192/1/012069](https://doi.org/10.1088/1755-1315/192/1/012069).
- [24] L. Miotti, E. Shehaj, A. Geiger, S. D'Aronco, J. D. Wegner, G. Moeller and M. Rothacher, Tropospheric delays derived from ground meteorological parameters: comparison between machine learning and empirical model approaches. In *2020 European Navigation Conference (ENC)*. pp. 1-10, IEEE, 2020.
- [25] M. Palaniswami and A. Shilton, Adaptive support vector machines for regression. In *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02. Vol. 2*, pp. 1043-1049, IEEE, 2002.
- [26] V. Kecman, *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. MIT press, 2001.
- [27] A. Mathur and G. M. Foody, Multiclass and binary SVM classification: Implications for training and classification users. *IEEE Geoscience and remote sensing letters*, 5(2), 241-245, 2008. DOI: [10.1109/LGRS.2008.915597](https://doi.org/10.1109/LGRS.2008.915597).
- [28] R. Debnath and H. Takahashi, Kernel selection for the support vector machine. *IEICE transactions on information and systems*, 87(12), 2903-2904, 2004.
- [29] A. E. Hoerl and R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67, 1970.
- [30] R. R. Wilcox, Multicollinearity and ridge regression: results on type I errors, power and heteroscedasticity. *Journal of Applied Statistics*, 46(5), 946-957, 2019. <https://doi.org/10.1080/02664763.2018.1526891>.
- [31] G. James, D. Witten, T. Hastie and R. Tibshirani, *Resampling methods*. In *An introduction to statistical learning* (pp. 175-201). Springer, New York, NY, 2013. https://doi.org/10.1007/978-1-4614-7138-7_5.
- [32] J. Al-Jararha, New approaches for choosing the ridge parameters. *Hacettepe Journal of Mathematics and Statistics*, 47(6), 1625-1633, 2016.
- [33] J. Friedman, T. Hastie and R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1), 1, 2010.
- [34] H. Zou and T. Hastie, Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2), 301-320, 2005. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.
- [35] R. J. Tibshirani, The lasso problem and uniqueness. *Electronic Journal of statistics*, 7, 1456-1490, 2013. DOI: [10.1214/13-EJS815](https://doi.org/10.1214/13-EJS815).
- [36] S. Aslan ve T. Yıldız, Makine Öğrenmesinde Rastgele Oran ve Sıralı Küme Örneklemesi Yöntemlerinin Doğrusal Regresyon Modellerine Etkisi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 24(70), 29-36, 2022. <https://doi.org/10.21205/deufmd.2022247004>.

