

AKÜ FEMÜBİD 21 (2021) 035603 (668-680)

AKU J. Sci. Eng. 21 (2021) 035603 (668-680)

DOI: 10.35414/akufemubid.856192

Araştırma Makalesi / Research Article

Düşük Plastisiteli Killi Zeminlerin Kıvam Özelliklerinin Tahmininde Rastgele Orman Yöntemi

Zülal AKBAY ARAMA¹, Seyidcem KARAKAŞ¹, Said Enes NURAY¹, Oğuzhan ALTEN², Muhammed Selahaddin AKIN³, Hazal Berrak GENÇDAL⁴

¹ İstanbul Üniversitesi-Cerrahpaşa, Mühendislik Fakültesi, İnşaat Mühendisliği Bölümü, İstanbul.

² GEODATA Engineering S. p. A., İstanbul.

³ Kartal Belediyesi, Kartal, İstanbul.

⁴ İstanbul Kültür Üniversitesi, Mühendislik Fakültesi, İstanbul.

Sorumlu yazar e-posta: zakbay@iuc.edu.tr
seyidcem.karakas@ogr.iu.edu.tr
enesnuray1@hotmail.com
altenogu@gmail.com
msehahaddinakin@gmail.com
h.gencdal@iku.edu.tr

ORCID ID: <http://orcid.org/0000-0001-8185-7329>
ORCID ID: <http://orcid.org/0000-0001-6250-4501>
ORCID ID: <http://orcid.org/0000-0002-2333-2687>
ORCID ID: <http://orcid.org/0000-0002-7798-7005>
ORCID ID: <http://orcid.org/0000-0003-1980-1812>
ORCID ID: <http://orcid.org/0000-0002-1041-5643>

Geliş Tarihi: 07.01.2021

Kabul Tarihi: 26.06.2021

Öz

Anahtar kelimeler

Düşük plastisiteli killer;
Rastgele orman yöntemi;
Karar ağaçları;
Plastisite indisi;
Likit limit

Bu çalışma kapsamında, düşük plastisiteli killi zeminlerin kıvam özelliklerinin tahmininde Rastgele Orman yönteminin uygulanabilirliği değerlendirilmiştir. Bu amaçla, Birleştirilmiş Zemin Sınıflandırma Sistemi'ne göre düşük plastisiteli kil sınıfında yer alan 665 adet zeminin likit, plastik limit ve su muhtevası deney sonuçları derlenerek bir veri seti oluşturulmuştur. Python yazılımı kullanılarak yapılan tek ve çok değişkenli regresyon analizlerinde likit limit, derinlik, su muhtevası ve doğal birim hacim ağırlığı değerleri bireysel veya gruplar halinde girdi parametreleri olarak değerlendirilmiş ve plastisite indisi tahminindeki etkinlikleri araştırılmıştır. Aynı zamanda, laboratuvar verilerinin tutarsızlık durumları veya belirli bir eğilim izlememesi sebebi ile ortaya çıkan tahmin oranı azalmasına dikkat çekilerek, bu oranlarının Rastgele Orman yöntemi ile nasıl yükseltilebileceği konusu da incelenmiştir. Sonuçlar, Rastgele Orman yönteminin düşük plastisiteli kil zeminlerin plastisite indisi değerinin tahmininde kullanılabilir nitelikte olduğunu göstermektedir.

The Random Forest Method to Predict the Consistency Characteristics of Low Plastic Clayey Soils

Abstract

Keywords

Low plastic clays;
Random forest method;
Decision trees;
Plasticity index;
Liquid limit

Within the scope of this study, the applicability of the Random Forest Method in the prediction process of the consistency properties of low plastic clayey soils was evaluated. For this purpose, a data set was created by compiling the liquid limit, plastic limit, and water content test results of 665 soils. These soils were classified as low plastic clayey soils according to the Unified Soil Classification System. Univariate and multivariate regression analyzes were performed using Python software. The liquid limit, depth, water content, and natural unit weight values were evaluated individually or in groups as input parameters and their effectiveness in plasticity index estimation was investigated. At the same time, by drawing attention to the decrease in the estimation rate caused by the inconsistency of the laboratory data or not following a certain trend, the issue of how these rates can be increased by the Random Forest method was also examined. Consequently, it is shown that the Random Forest method can be used for the estimation of the consistency properties of low plastic clayey soils.

1. Giriş

Geoteknik projelendirmenin asli hedefi olan zemin-yapı etkileşim sürecinin öngörülmesi standart mühendislik uygulamaları kapsamının ötesinde, oldukça zor bir iştir. Tasarımın başarısı, gerçek davranışın tahminindeki başarı seviyesine büyük ölçüde bağlı olmakla birlikte gerek zeminin çok değişken ve heterojen yapısı gerekse yapı sistemlerinin karmaşıklığı bu tahmin ve hesapların, ekonomik kıstasları da gözeterek, yeterli hassasiyette yapılmasını güçleştirmektedir. Bu bağlamda, incelenen proje alanlarındaki zemin profillerinde yer alan tabakaların davranış özelliklerinin laboratuvar ve arazi deneyleri ile elde edilmesi en geçerli yoldur (İyisan 1993). Ancak, söz konusu deneylerin uygulanması, süre-bütçe optimizasyonunun elde edilmesi açısından çoğu zaman olanaklı değildir. Geoteknik parametrelerin elde edilme hızının artırılması ve proje giderlerinin azaltılması amaçlanarak kabul edilebilir parametre tahmin süreçlerinin ortaya konulması, güncel literatür kaynaklarının araştırma konularından birini oluşturmaktadır. Ancak bu durumda, tahmin süreçleri ile belirlenen parametrelerin tasarımlarda doğrudan kullanımının olası etkileri kanıtlanabilir nitelikte olmalıdır (Nuray vd. 2021). Günümüzde kullanılan gelişmiş bilgisayar teknolojileri yeni ve özgün malzeme parametre tahmin ve davranış modellerin ortaya çıkarılmasını ve istatistikî yöntemler ile bütünleştirilerek belirli metrikler ile kontrol edilmelerini sağlamıştır. Bu bakış açısı ile son yıllarda farklı alanlarda uygulanan (Pal 2003, Gislason et al. 2004, Waske et al. 2007, Akar vd. 2010) öğrenme tabanlı bir regresyon yöntemi olan Rastgele Orman yönteminin (RO), analizlerde yüksek doğruluk derecesi ve hızlı bir çözüm sağladığı vurgulanmaktadır. RO, çok sayıda değişken ve sınıfı, karmaşık modellemelere ve parametrelere ihtiyaç duymadan tahmin edebildiği için diğer yöntemlere kıyasla olumlu bir alternatif oluşturmaktadır (Watts and Lawrence 2008). RO yönteminin geoteknik mühendisliği disipliniinde uygulanabilirliği ise hali hazırda çok yeni bir araştırma konusudur. Pham vd. (2020) çalışmalarında, RO ve parçacık sürüsü optimizasyonunun yeni bir hibrit yumuşak hesaplama modelini geliştirilmiş, kil içeriği ile su

muhtevasını girdi ve zeminin drenajsız kesme dayanımını çıktı parametresi olarak kullanmıştır. Pham vd. (2020), serbest basınç dayanımını tahmin etmek için RO algoritmasının uygulama olasılığını araştırmışlardır. Bu amaçla, Vietnam'daki Long Phu enerji santrali projesi kapsamında gerçekleştirilen laboratuvar deneylerinden elde edilen toplam 118 numune ve bunların deneylerini kullanmışlardır. Modelleme için kullandıkları veriler, girdi değişkenleri olarak kil içeriği, nem içeriği, özgül ağırlık, boşluk oranı, likit limiti ve plastik limiti içerirken, çıktı parametresi serbest basınç dayanımı olarak belirlenmiştir. Ly ve Pham (2021), destek vektör makinası, Gauss regresyon süreci ve RO şeklinde üç adet makine öğrenmesi yöntemi kullanarak Vietnam'daki Da Nang-Quang Ngai otoyolunda yapılan deneylerden faydalanarak oluşturdukları 145 verilik bir set kullanmışlardır. Analizlerinde girdi parametreleri olarak kil yüzdesi, su muhtevası, likit limit, plastik limit, özgül ağırlık ve boşluk oranı, çıktı parametresi olarak kohezyon değerini kullanmışlardır. Çalışma sonuçlarını determinasyon katsayısı (R), ortalama mutlak hata (MAE) ve karesel ortalama hata (RMSE) metriklerini hesaplayarak kontrol etmişlerdir. Nuray vd. (2021) yüksek plastisiteli kil zeminlere ait literatürden toplanarak elde edilen 387 adet kıvam limiti ve 59 adet kompaksiyon-kıvam limiti deney çiftinin kullanılması ile oluşturulan bir veri seti kullanılarak plastisite indisinin ve kompaksiyon parametrelerinin RO yöntemi ile tahmin olasılığını araştırmışlardır. Zhang vd. (2021), Bayesian optimizasyonuna dayalı olarak yeni veri güdümlü aşırı gradyan artırma (XGBoost) ve RO yöntemlerini kullanarak, yumuşak hassas killi zeminlerin geoteknik özellikleri ile drenajsız kayma mukavemetleri arasındaki ilişkiyi incelemişlerdir. Ön konsolidasyon basıncı, düşey efektif gerilme, kıvam limitleri ve doğal su muhtevası değerleri ile drenajsız kayma mukavemeti arasındaki ilişkiyi tahmin etmek için genel bir yaklaşım geliştirmişlerdir. Tüm bu çalışmaların sonuçları göz önüne alındığında RO yönteminin; hızlı sonuç veren, yüksek doğruluk sağlayan, sayısal tahminlerde iyi performans sergileyen ve öğrenmede avantaj sağlayan bir yöntem olduğu söylenebilmektedir. Bu bağlamda, bu makale kapsamında, RO yöntemi kullanılarak

geoteknik parametre tahmin süreci modellemesi yapılmıştır. Bu amaçla, düşük plastisiteli kil içerikli zeminlerin likit limit (w_L), plastik limit (w_P), doğal su muhtevası (w_n), doğal birim hacim ağırlık (γ_n), kuru birim hacim ağırlık (γ_k) değerlerinin kullanılması ile oluşturulan bir veri seti kullanılmıştır. Söz konusu veri seti kullanılarak oluşturulan tek ve çok değişkenli regresyon analizleri ile plastisite indisi (I_p) değerinin hesaplanmasında etkili olan faktörler incelenmiş ve plastisite indisi değerinin RO 'nun kullanımı ile doğrudan tahmin edilmesi olasılığı araştırılmıştır. I_p değeri matematiksel olarak w_L değeri ile w_P değerinin farkı ile hesaplanmakta olup malzeme davranışı açısından ise zeminin plastik davrandığı su muhtevası aralığı olarak tanımlanabilmektedir. Plastik davranış ise ince daneli zeminlerin sabit hacimde ve kırılma olmaksızın kalıcı deformasyona uğrama niteliğinin ölçüsüdür (Çinicioğlu vd. 2002). I_p değerine göre zeminler ilk olarak Atterberg tarafından sınıflandırılmış olup, bu sınıflandırma I_p 'nin zeminin türü, plastisite derecesi ve kohezivite derecesi arasındaki ilişki hakkında bilgi vermektedir (Prakash and Jain 2002). Aynı zamanda, I_p Birleştirilmiş Zemin Sınıflandırma Sistemi'nde (USCS) ince dane içerikli zeminlerin, sınıflarının belirlenmesinde w_L 'nin yanı sıra kullanılan ikinci parametredir. Dolayısı ile geoteknik mühendisliği yapılarının tasarımı sürecinde uygulanan arazi karakterizasyonu çalışmalarında, likit limit deneyinin uygulanması kaçınılmazdır. Aynı zamanda, I_p değerinin doğru bir şekilde belirlenmesi özellikle ince dane içerikli zeminlerde malzeme davranışının değerlendirilebilmesi açısından büyük önem taşımaktadır. I_p değerinin, uygulanan likit ve plastik limit deneyleri ile belirlenmesi için faydalanan standartlardan biri de ASTM D4318'dir. Bu makale kapsamında derlenen veri setinin elde edilmesi amacı ile de uygulanan tüm deneylerde bu standart kullanılmıştır. Ancak, ASTM D4318'de belirtildiği üzere bu standardın ürettiği sonucun kalitesi, onu uygulayan personelin yeterliliğine ve kullanılan ekipman ve tesislerin uygunluğuna bağlı olmaktadır. Likit limit deneyinin uygulanmasında Casagrande kabı kullanılmakta olup, kap-kam mekanizması ile değişen su içeriklerinde uygulanan darbe sayısı kaydedilerek w_L değeri hesaplanabilir. Plastik limit

deneyinin uygulanmasında ise elipsoidal boyutlu bir zemin kütesinin cam bir plaka üzerinde elle tekrarlı yuvarlanmasıyla deney gerçekleştirilir (Das 2010). Whyte (1982)'in plastik limit değerinin belirlenmesine dair yapmış olduğu çalışmalarda, zeminlerde tam doygunluk ve sıkışmazlık koşullarının geçerliliği kabul edilse dahi, zeminin akma gerilmesinin bazı deney uygulama koşullarının fonksiyonu olacağını ifade etmiştir. Bu koşullar; deney sırasında hazırlanmakta olan zemin numunesine uygulanan basınç; el-numune arasındaki temas düzleminin geometrisi; el-numune-cam düzlemi arasında sürtünme ve operatörün numuneyi yuvarlama hızı şeklindedir. Bu bağlamda, söz konusu faktörlerin etkisini minimize etmek ve deney uygulama süresini kısaltmak amacı ile plastik limit deneyi uygulanmadan I_p değerini, sadece w_L kullanımı ile veya farklı basit fiziksel deneylerin uygulanması ile tahmin etme süreci günümüz araştırmalarında sıklıkla rastlanan bir konu durumundadır (Naveena et al., 2018; Akbay Arama vd. 2021). Shimobe ve Spagnoli (2020), literatürden elde ettikleri 121 farklı ince daneli zeminin düşen koni ve Casagrande yöntemi ile belirlenen kıvam limit sonuçlarını kullanarak, doğal su muhtevası, koni penetrasyon indeksi ve plastisite indisi arasındaki ilişkiyi araştırmışlardır. Ermias ve Vishal (2020), ileri beslemeli üç katmanlı sinir ağı ile yaptıkları çok değişkenli regresyon analizlerinde aktivite, kıvam limitleri ve ince dane yüzdesi değerlerini kullanarak ince daneli zeminlerin şişme yüzdesini tahmin etmişlerdir. Kayabası (2020), ince daneli zeminlerin mühendislik özellikleri arasındaki istatistiksel ilişkileri inceleyebilmek için Türkiye/Ankara şehrinden elde ettiği 440 adet ince daneli zemin kullanmıştır. Elde edilen zemin türlerinin şişme potansiyeli, şişme basıncı ve geoteknik özellikleri arasındaki ilişkiyi belirlemiş, ayrıca şişme basıncını plastisite indeksi değerinden tahmin edebilmek için tekli regresyon analizleri yapmıştır. Ahmed vd. (2020), yüksek plastisiteli killi zeminlere uçucu kül ilavesinin geoteknik özellikler üzerindeki etkilerini incelemiştir. Bu amaçla literatürden elde ettiği çeşitli laboratuvar test sonuçlarını kullanmıştır. Killi zeminlerin kıvam limitlerini, serbest basınç dayanımını ve sıkıştırma özelliklerini tahmin etmek için en uygun istatistiksel

modellerin geliştirilmesine odaklanmış, elde ettiği amprik ifadeleri literatürdeki diğer çalışmalarla karşılaştırmıştır. Söz konusu çalışmalarda kıvam limitleri ile ilgili tahminleme süreçleri yürütülmüştür ancak analizler dâhilinde RO yönteminin kullanımı söz konusu değildir. Bu nedenlerle, bu makalenin içeriğinde RO'nun I_p değerinin tahmininde uygulanabilirliği, farklı girdi parametreleri kullanılarak kurgulanan tek ve çok değişkenli analizler ile değerlendirilmiştir. Tek değişkenli analizlerde girdi parametresi w_L , çok değişkenli analizlerde girdi parametreleri ise iki set halindedir (1.set: w_L, w_n, γ_d ; 2.set: $w_L, w_n, \gamma_k, \gamma_d$).

2. Malzeme ve Yöntem

Çalışma kapsamında analiz yöntemi olarak kullanılan RO, karar ağacı tabanlı bir regresyon aracıdır (Breiman 2001). RO yöntemi için matematiksel formülasyonun ayrıntılı bir açıklaması Breiman (2001a) ve Liaw ve Wiener (2002)'de bulunabilmektedir. RO, tüm değişkenler arasından en iyi dalı kullanarak her bir düğümü dallara ayırmak yerine, her bir düğümde rastgele olarak seçilen değişkenler arasından en iyisini kullanıp her bir düğümü dallara ayırır (Akar ve Güngör 2012). Her bir veri seti orijinal veri setinden yer değiştirmeli olarak üretilir ve rastgele özellik kullanılarak ağaçlar geliştirilir (Archer 2008, Breiman 2001). Hızlı bir şekilde ve istenilen sayıda regresyon ağacı geliştiren ve çalıştıran RO, bir bağımlı değişken için tahmin yapabilmek adına, girdi verisini arka planda çalışan ormandaki her karar ağacına yerleştirir. Her bir veri seti orijinal veri setinden yeniden örnekleme kullanılarak üretilir. Sonuç olarak RO regresyonu, birden fazla karar ağacına sahip olur ve daha kararlı bir tahmin elde etmek için onları birleştirir (Ouedraogo et al. 2019). Algoritmanın herhangi bir denklem veya iki boyuttan fazla düşünüldüğünde bir düzlem oluşturamaması bu yöntemin literatürde belirtilen ispatlanan en belirgin eksikliğidir. Herhangi bir denklem üretilmediği için bilgisayar haricinde kullanımı mümkün kılınmazken, bu regresyon herhangi bir bilgisayar programına entegre edildiğinde başarılı sonuçlar vererek kullanılabilir (Segal, 2003). RO regresyonu algoritması, verilerin tamamını kullanmak yerine her analizde veri seti içinden seçtiği rastgele veri

gruplarını kullanır. Rastgele alınan veriler sebebi ile algoritma, her analizde, birbirileri ile çok yakın fakat değişik determinasyon katsayıları (R^2) hesaplar. Program, tanımladığı rastgele kesitlerin her biri için karar ağaçları inşa eder ve regresyona bir bağımsız değişken değeri girildiğinde karşılığı olan bağımlı veriyi tahmin edebilmek için bu karar ağaçlarının her birinde girdiyi ait olduğu bölüme yerleştirerek oluşturduğu karar ağaçlarındaki sonuçların ortalamasını alır. Bu regresyon türünde başlangıç olarak kullanılacak olan karar ağacı sayısı genellikle veri sayısının karekökü ile yarısı arasındaki sayılar aralığında seçilip, analizler gerçekleştirilebilir. Analiz sonuçlarının alınacağı ağaç sayısı, belirleme katsayısı (R^2)'nin maksimum olduğu karar ağacı sayısı olarak seçilir. Bununla birlikte, diğer regresyon teknikleriyle karşılaştırılabilecek değişken önem ölçülerinin (metriklerin) hesaplanmasına da olanak sağlamaktadır (Grömping 2009). Bu çalışma kapsamında RO regresyon yöntemi Phyton programı ile kullanılarak, oluşturulan veri seti analizleri gerçekleştirilmiştir. Python, yorumlayıcıları kullanan ve çok geniş bir uygulama alanında tercih edilen üst düzey, dinamik, nesne yönelimli, genel amaçlı bir programlama dilidir. Kolay anlaşılırdır ve kullanımı kolay olacak şekilde tasarlanmıştır. Yeni başlayanlar için uygun bir dil olmasıyla popülerlik kazanmış ve bugüne kadar en popüler giriş dili olarak kullanılan Java'nın yerini almıştır. Dinamik olarak yazılan bir dil olması sebebi ile Python çok esnektir. Bu dil, yapısal ve nesne yönelimli dâhil olmak üzere farklı programlama tekniklerini destekleyebilir (Srinath, 2017; Yogatama and Tirta, 2021). Phyton programında RO regresyon algoritmasının uygulanması ile analiz edilen veriler, algoritmaya girmeden önce ön hazırlık işlemlerinden geçirilebilmektedir. Bu işlemler, aykırı verilerin temizlenmesi, eksik verilerin doldurulması, eğitim ve test için verilerin ayrılması gibi amaçlar doğrultusunda sürdürülmektedir. Bu çalışmanın içeriğinde yapılan tahminlemelerde amaç gerçek veriye yakınlık olduğu için eksik veriler yerine herhangi bir tahminleme (doldurma) işlemi yapılmamıştır. Eğitim ve test verilerinin ayrılma işlemleri bir makine öğrenmesi kütüphanesi olan Sklearn kütüphanesindeki `train_test_split` modülü ile yapılmıştır. Bu modül girdi ve çıktı verilerinin

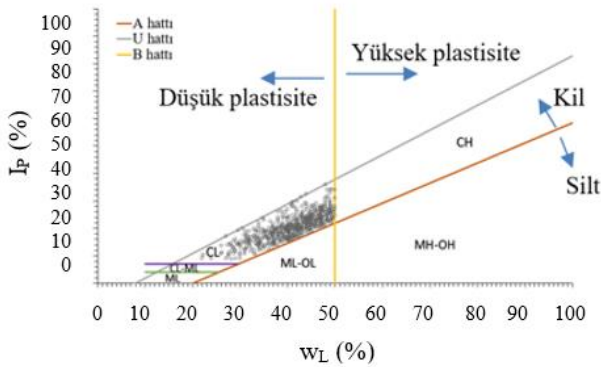
tamamını değerlendirmekte ve istenilen bir oranda verileri eğitim ve test için bölmektedir. Modül parametrelerinden olan shuffle parametresi (shuffle=True olarak kullanılır) ile veriler karıştırılarak bölme işlemi gerçekleştirilmektedir. Ayrıca belli bir yerden bölme sağlayan random_state parametresi (random_state=x (x herhangi bir sayı) olarak kullanılır) kullanılmamıştır çünkü belli bir yerden ayrılan veriler için iyi bir temsil yapıp yapılamadığı belli olamamıştır. O yüzden test sonucu en uygun model ve buna bağlı olarak bölünmüş veri seti esas alınarak analizler uygulanmıştır. Analizlerin sonuçları, uluslararası metrikler olarak kullanılan R (korelasyon katsayısı), R^2 (determinasyon katsayısı), MAE (ortalama mutlak hata), MSE (en küçük ortalama kare hatası), RMSE (karesel ortalama hata), MAPE (ortalama mutlak hata yüzdesi) ve SSE (kare hataları toplamı) değerleri kullanılarak yorumlanmıştır. Kullanılan korelasyonlar ilişkinin güvenilirliğini ölçmektedirler (verinin buna uygunluğunun iyiliği). Diğer bir deyişle, bir değişkenin diğerini ne kadar iyi tahmin edebileceğinin bir ölçüsü niteliğindedirler (verinin bağlamı göz önüne alındığında) ve bir ilişkinin kesinliğini belirlerler. Bu bağlamda, korelasyon katsayısı, R, bir sayının diğerindeki değişikliklerden ne kadar etkilenmesinin beklenebileceğinin matematiksel bir ölçüsüdür. Doğrusal R bazen geliştiricisi Karl Pearson'ın onuruna Pearson ürün moment korelasyon katsayısı olarak adlandırılır. İki değişken X ve Y arasındaki doğrusal ilişkinin derecesinin kesin sayısal değerini verir. Sıfır R, iki sayının ilişkili olmadığı anlamına gelir. Sıfır olmayan R, sayıların ilişkili olduğu anlamına gelir, ancak katsayı 1 veya -1 olmadığı sürece başka etkiler vardır ve iki sayı arasındaki ilişki sabit değildir. R sıfıra ne kadar yakınsa belirsizlik o kadar büyük olur ve düşük korelasyon katsayıları, ilişkinin yararlı olacak kadar kesin olmadığı anlamına gelir. $R=1$, iki sayının mükemmel bir şekilde ilişkili olduğu anlamına gelir: biri büyürse diğeri de büyür ve birindeki değişiklik diğerindeki değişikliğin katıdır. $R= -1$, sayıların mükemmel şekilde ters orantılı olduğu anlamına gelir. Birindeki büyüme, diğerindeki büyümenin negatif bir katıdır (Kandpal and Kandpal, 2014). Aynı zamanda R değerinin karesi olarak da adlandırılan determinasyon katsayısı, R^2 , bir regresyon

uyumunun iyiliği olarak yorumlanır. R^2 ne kadar yüksek olursa, bağımlı değişkenin bağımsız değişken tarafından açıklandığı varyans o kadar iyi olur. R^2 , bir regresyonun kullanılabilirliğinin genel ölçüsüdür ve açıklanan varyasyonun toplam varyasyona oranıdır. R^2 , bir değişkenin varyansının (dalgalanmanın) diğer değişkenden tahmin edilebilirlik oranını verdiği için yararlıdır ve $0 < R^2 < 1$ olacak şekildedir. Sonuç olarak R^2 , en uygun çizgiye en yakın olan verilerin yüzdesini temsil eder. Davis (1971) ve Lewis (1982), R^2 değerine bağlı olarak korelasyon gücünün alt ve üst sınırlarını tanımlamışlardır. R^2 değerinin 0.01-0.09 aralığında korelasyon gücünün çok zayıf, R^2 değerinin 0.10-0.29 aralığında korelasyon gücünün zayıf, R^2 değerinin 0.3-0.49 aralığında korelasyon gücünün orta, R^2 değerinin 0.50-0.69 aralığında korelasyon gücünün güçlü, R^2 değerinin 0.70-1.00 aralığında korelasyon gücünün çok güçlü olarak tanımlanabileceğini belirtmişlerdir. En küçük ortalama kare hatası, MSE, bir regresyon eğrisinin bir diziye ne ölçüde yakın olduğunu göstermektedir. MSE, bir makine öğrenmesi modelinin performansını ölçmekte olan bir metriktir ve daima pozitif ve MSE değeri düşük veya sıfıra yakın olan tahminleyicilerin daha iyi bir performans sergilediği söylenebilmektedir (Veri bilimcisi, 2021). Ortalama mutlak hata, MAE, yönlerini dikkate almadan bir dizi tahmindeki hataların ortalama büyüklüğünü ölçer. Tüm bireysel farklılıkların eşit ağırlığa sahip olduğu tahmin ve gerçek gözlem arasındaki mutlak farkların deney numunesi üzerindeki ortalamasıdır. MAE, 0'dan ∞ 'a kadar değişebilen niteliktedir ve düşük veya negatif yönelimli değerlerde modeller daha iyi performans gösterirler. Karesel ortalama hata, RMSE, hatanın ortalama büyüklüğünü de ölçen ikinci dereceden bir puanlama kuralı olarak tanımlanabilir ve tahmin ile gerçek gözlem arasındaki kare farklarının ortalamasının karekökü olarak hesaplanabilir (Medium, 2021; Veri bilimcisi, 2021). RMSE, 0'dan ∞ 'a kadar değişebilir niteliktedir ve düşük veya negatif yönelimli değerlerin daha iyi performans gösterdiği söylenebilmektedir. $RMSE=0$ modelde hiç hata olmadığı anlamına gelmektedir. Ortalama mutlak hata yüzdesi, MAPE, tahminlerin mutlak yüzdeler hata oranının ortalamasıdır. Hata, gerçek veya gözlemlenen değer eksi tahmin edilen değer olarak tanımlanır. Yüzde hataları, MAPE'yi

hesaplamak için işarete bakılmaksızın toplanır. Ayrıca mutlak yüzde hataları kullanıldığından pozitif ve negatif hataların birbirini yok etmesi sorunu da ortadan kalkar. MAPE ne kadar küçükse tahmin o kadar iyidir (Swamidass, 2000). Kare hataları toplamı, SSE, gözlemlenen değer ile tahmin edilen değer arasındaki farktır. Veri ile doğrusal regresyon gibi bir tahmin modeli arasındaki tutarsızlığın ölçüsüdür. Hata ne kadar küçük olursa, regresyonun tahmin gücü o kadar iyi olur.

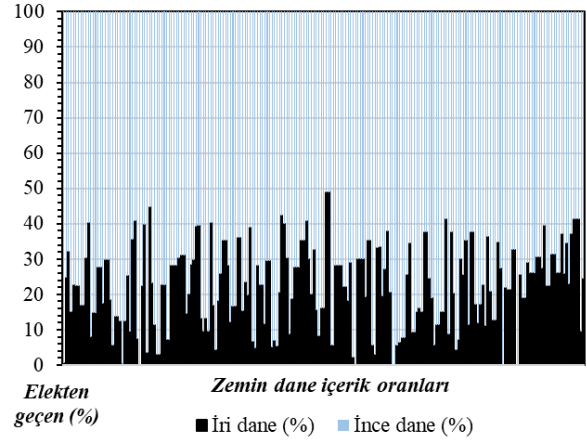
3. Verilerin Sınıflandırılması

Bu çalışma kapsamında, USCS'ye göre düşük ve orta plastisiteli inorganik killer, kumlu, çakıllı ve siltli killer olarak ifade edilen CL grubu zeminler için ASTM D4318'e göre uygulanan kıvam limiti, ASTM D422'ye göre uygulanan elek analizleri, TS 1900-1'e göre uygulanan doğal birim hacim ağırlık deneyleri dikkate alınmıştır (Şekil 1). Bu bağlamda, ağırlıklı olarak İstanbul İli Avrupa yakasının güneybatısında uygulanan zemin etüt raporlarının sayısallaştırılması ile derlenen 665 adet kıvam limitleri-elek analizi; 272 adet kıvam limitleri-doğal su muhtevası-doğal birim hacim ağırlık-elek analizi, 116 adet ise kıvam limitleri-doğal su muhtevası-doğal birim hacim ağırlık-kuru birim hacim ağırlık ve elek analizi deneylerini içeren 3 farklı set veri düzenlenmiştir.



Şekil 1. Makale kapsamında kullanılan zeminlerin Casagrande plastisite kartında gösterimi. Her bir sette uygulanan deney türüne göre setlerdeki veri sayısı değişmektedir. Söz konusu verilerde, zeminlere ait likit limit değerleri 50'den küçük olması sınıflandırmada en temel ayırıcı özelliktir. Ayrıca plastisite grafiğine yerleştirilen verilerin A hattının üstü ve U hattının altında tanımlanan aralıkta kalması, zeminlerin içeriğinde, davranışta etkili olacak seviyede kil içeriği olduğunu

göstermektedir. Bu durum, değerlendirmeye alınan zeminlerin ince ve iri dane oranlarını belirten Şekil 2'de sunulmuştur.



Şekil 2. Makale kapsamında kullanılan zeminlerin dane içerik oranları

Çalışma kapsamında kullanılan veriler, birinci yazar adına düzenlenmiş özel izin belgesi ile İstanbul Valiliği Yatırım İzleme ve Koordinasyon Başkanlığı'ndan temin edilmiştir. İlgili veriler, kopyalama-sayısallaştırma-tasnif etme gibi sıralı adımlar ile bilgisayar ortamına aktarılmıştır. İlgili deneyler, İstanbul'da bulunan kamu binalarının temellerinin veya iksa sistemlerinin veya temel-yapı güçlendirme sistemlerinin veya dayanma yapılarının tasarımında kullanılan deneylerdir. Aynı zamanda, söz konusu deneylerin uygulanışı birinci yazar tarafından kontrol edilebilmiştir. Bu sebepler ile deneylerin hangi standartlar dâhilinde ve çevresel koşullar altında uygulandığı yazarlar tarafından bilinmektedir. Ancak bu durum, deneylerin uygulanmasında farklı operatörlerin yer aldığı gerçeğini değiştirememektedir. Dolayısı ile Whyte (1982)'in öngördüğü, malzeme plastik davranışını etkileyen faktörlerin ortadan kaldırılabilmesi amacı göz önünde tutularak I_p değerinin RO yöntemi kullanılarak tahmin edilmesi sürecine başvurulmuştur.

4. Analiz ve Değerlendirmeler

Makalede öngörülen analizler, tek değişkenli ve çok değişkenli olmak üzere alt gruplara ayrılmıştır. Tek değişkenli analizlerde I_p değerinin, sadece w_L 'nin fonksiyonu olduğu varsayılarak tahminleme süreci yürütülmüştür. Başka bir deyişle, girdi parametresi olarak sadece w_L , çıktı parametresi olarak I_p göz

önüne alınmıştır. Analizlerde kullanılan setteki deney grubu sayısı 665 (665 adet likit limit, 665 adet plastik limit deneyi uygulanmıştır) adettir. Tek değişkenli analizler kapsamında RO algoritmasının uygulanmasının yanı sıra ayrıca doğrusal, logaritmik, polinom, üssel regresyon uygulamaları da yapılmıştır. Ancak, makale kapsamına sadece en tatminkâr metrik değerlerini veren doğrusal regresyon uygulamaları dâhil edilmiştir. Çok değişkenli analizlerde ise iki farklı veri seti kullanılmış ve RO analizleri her iki set için ayrı ayrı uygulanmıştır. Söz konusu veri setlerinden birincisinde toplam deney grubu sayısı 272 adet olup analizlerde girdi parametresi olarak w_L , w_n , γ_n , γ_k çıktı parametresi olarak I_p değerlendirilmiştir. İkinci veri setindeki toplam deney grubu sayısı 116 adet olup analizlerde girdi parametresi olarak w_L , w_n , γ_n , γ_k çıktı parametresi olarak I_p değerlendirilmiştir.

4.1 Tek değişkenli analizler

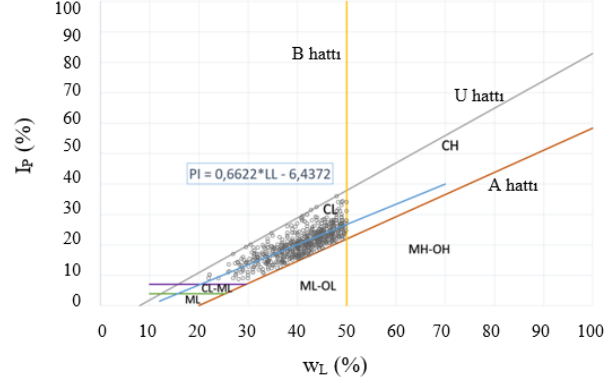
4.1.1 Doğrusal Regresyon uygulaması

Doğrusal regresyon analizlerinde w_L değerleri bağımsız değişken, I_p değerleri ise bağımlı değişken olarak tanımlanmıştır. Analizlerde kullanılan w_L değerlerinin üst sınırı %50 alt sınırı ise %22 olarak belirlenmiştir. Regresyon sonucunda I_p ile w_L arasında elde edilen ilişki Eşitlik (1)'de verilmiştir (Şekil 3).

$$I_p = 0.6622W_L - 6.4372 \quad (1)$$

Veri dağılımları göz önüne alındığında; verilerin belirli bölgede kümelenmesine rağmen parametrelerin arasında doğrusal bir ilişki sağlanabildiği görülmektedir. Bu ilişkiyi ifade eden en uygun denklemin R değeri 0.804, R^2 değeri 0.647, SSE değeri 5371, MSE değeri 8.07, RMSE değeri 2.84, MAPE değeri 11.24 ve MAE değeri 2.21 olarak hesaplanmıştır. Elde edilen R^2 değerinin Davis (1971) ve Lewis (1982) önerisine göre değerlendirilmesi durumunda elde edilen eşitliğin ($R^2= 0.50-0.69$ aralığında) güçlü bir korelasyon olduğu söylenebilmektedir. Ancak diğer metrik değerler de göz önüne alındığında (R ve R^2 dışında), veriler arasında doğrusal bir ilişki gözlemlenmesine rağmen, tatminkâr sonuçlar elde edilemediği görülmektedir. Söz konusu metriklerde hedef,

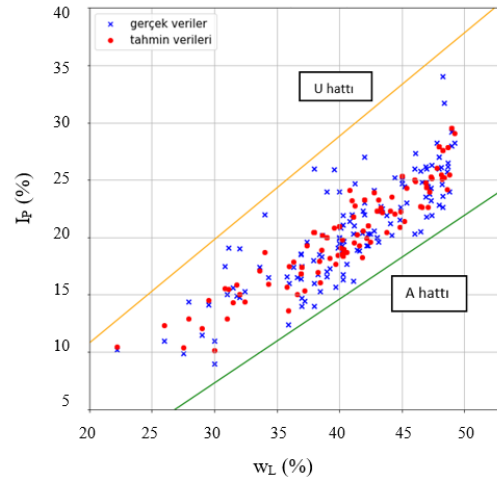
düşük değerler hesaplanması ile daha iyi performans elde edilmesidir. Ancak özellikle SSE değerinin çok büyük olması korelasyon tahmin gücünün düşük olduğu kanısını ortaya çıkarmaktadır.



Şekil 3. Doğrusal regresyon uygulanması ile elde edilen I_p - w_L ilişkisi

4.1.2 RO Regresyonu uygulaması (w_L : Bağımsız değişken, I_p : Bağımlı değişken)

RO regresyon analizlerinde de w_L değerleri bağımsız değişken, I_p değerleri ise bağımlı değişken olarak tanımlanmıştır. Analizlerde kullanılan w_L değerlerinin üst sınırı %49.2 alt sınırı ise %22.2 olarak belirlenmiştir. Analizler için oluşturulan toplamda 665 deney grubundan oluşan veri setinin %80 oranındaki kısmı, RO algoritmasını eğitmek için, %20 oranındaki kısmı ise algoritmayı test etmek için kullanılmıştır. Şekil 4'te test aşamasında kullanılan verilerin tahmin değerleri ile gerçek değerleri karşılaştırmalı bir şekilde sunulmuştur.



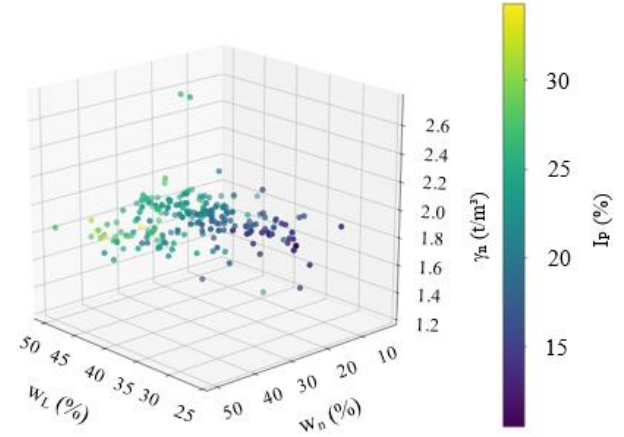
Şekil 4. RO regresyon uygulanması ile elde edilen I_p - w_L ilişkisi

Analizlerin sonucunda, R değeri 0.865, R^2 değeri 0.749, SSE değeri 709, MSE değeri 5.33, RMSE değeri 2.3, MAPE değeri 8.83 ve MAE değeri 1.78 olarak elde edilmiştir. Elde edilen R^2 değerinin Davis (1971) ve Lewis (1982) önerisine göre değerlendirilmesi durumunda elde edilen eşitliğin ($R^2= 0.70-1.0$ aralığında) çok güçlü bir korelasyon olduğu söylenebilmektedir. Ancak bu ölçüt RO algoritması uygulamalarının doğrusal regresyon analizi uygulamalarına göre performansının değerlendirilmesi için yeterli değildir. Analiz sonucunda hesaplanan diğer metriklerin doğrusal regresyon sonucunda elde edilen metrikler ile karşılaştırılması, uygulanabilirliğin denetlenmesi açısından tatminkâr sonuçlar verebilmektedir. Bu bağlamda, RO analizleri ile elde edilen metrik değerlerinde, doğrusal regresyon analizlerinde elde edilen metrik değerlerine oranla belirgin bir düşme gerçekleştiği aşikârdır. Bu durum RO yönteminin uygulanmasının doğrusal regresyon analizlerine göre daha iyi performans sonuçları vereceğinin göstergesidir. Aynı zamanda doğrusal regresyon analizlerinde RO regresyon analizlerine göre daha fazla sayıda veri ile eğitim yapılmış olması ve doğası gereği test sonuçları eğitimin yapıldığı veri ile alınmış olmasına rağmen, RO regresyonunda daha iyi performans elde edilmesi dikkat çekicidir. Ayrıca, SSE değeri doğrusal regresyonda 665 veri, RO regresyonunda ise 133 veri test edilerek hesap edildiği için, bu değer iki regresyonun karşılaştırılmasında kullanılması uygun değildir.

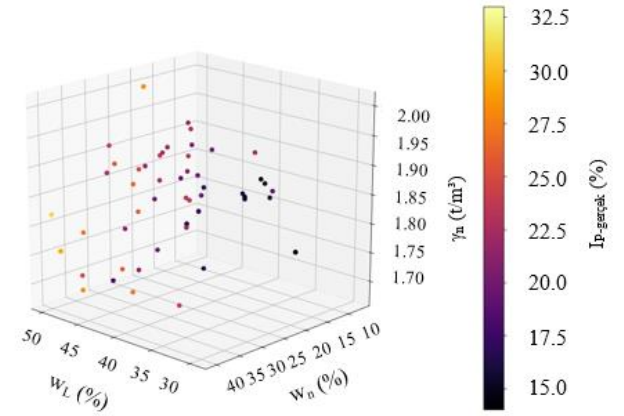
4.3 Çok değişkenli analizler

4.3.1 RO Regresyonu uygulaması (1. Veri seti)

RO regresyon analizlerinde de w_L , w_n , γ_n değerleri bağımsız değişken, I_p değerleri ise bağımlı değişken olarak tanımlanmıştır. Analizlerde kullanılan w_L değerlerinin üst sınırı %49.2 alt sınırı ise %27.6 olarak belirlenmiştir. Analizler için oluşturulan toplamda 272 deney grubundan oluşan veri setinin %80 oranındaki kısmı, RO algoritmasını eğitmek için, %20 oranındaki kısmı ise algoritmayı test etmek için kullanılmıştır. Şekil 5'te, analizlerde kullanılan 1. veri setine ait parametreler ve değerleri sunulmuştur. Şekil 6'da ise 1. veri seti kullanılarak uygulanan çok değişkenli RO regresyonu analizinde kullanılan test kümesi ve veri dağılımları sunulmuştur.

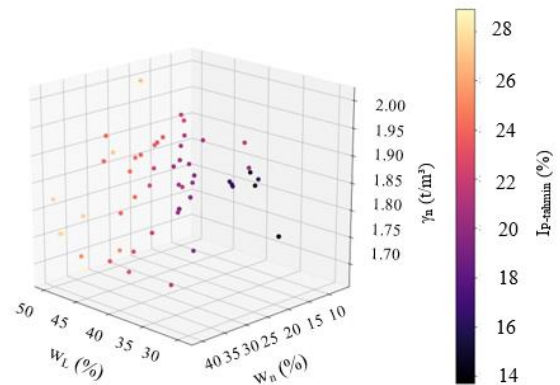


Şekil 5. Çok değişkenli RO regresyonu uygulamasında 1. veri seti



Şekil 6. Çok değişkenli RO regresyonu uygulamasında 1. veri seti: test verileri

Söz konusu test verileri ile uygulanan analiz sonucunda elde edilen tahmin değerleri ise Şekil 7'de verilmiştir.



Şekil 7. Çok değişkenli RO regresyonu uygulamasında 1. veri seti: tahmin verileri

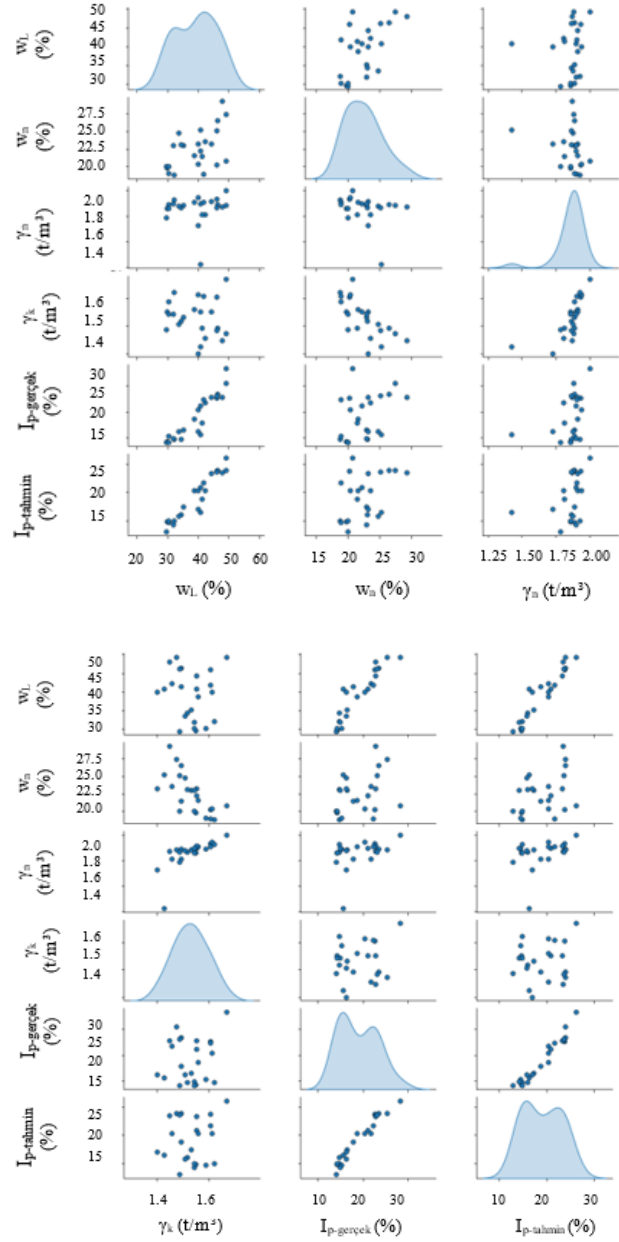
Çok değişkenli RO regresyonu analizinin 1. veri seti için uygulanması sonucunda R değeri 0.964, R^2 değeri 0.93, SSE değeri 93.15, MSE değeri 1.7, RMSE

değeri 1.3 MAPE değeri 4.96 ve MAE değeri 0.95 olarak hesaplanmıştır. Elde edilen R^2 değerinin Davis (1971) ve Lewis (1982) önerisine göre değerlendirilmesi durumunda elde edilen eşitliğin ($R^2= 0.70-1.0$ aralığında) çok güçlü bir korelasyon olduğu söylenebilmektedir. Yapılan analiz sonuçlarına göre elde edilen metrik değerler, RO algoritmasının I_p tahmininde başarılı bir performans sunduğunu göstermektedir. 1. setteki verilerin dağılımı değerlendirildiğinde, dağınık bir yapıya sahip olduğu söylenebilmektedir. RO yöntemi uygulamasının iyi bir performans sunması, kullanılan veri setinin bu şekildeki dağılımı ile ilişkilendirilebilir niteliktedir. Çünkü dağınık veri setleri ile herhangi bir eğri veya düzlem oluşturulması mümkün olmadığından, RO regresyonu dağınık veri setlerinde daha iyi çalışmakta ve çok daha iyi doğruluk değerleri vermektedir. Bu durum, literatürdeki diğer çoğu yöntemin uygulanması açısından tercih edilmeyen bir durum olmasına rağmen, RO algoritmasının avantajlı ve dikkat çeken bir özelliğini oluşturmaktadır. Tahmin performansındaki artışın diğer bir sebebi olarak, analiz girdi parametrelerinin değişmesi öne sürülebilmektedir. Çok değişkenli analizler kapsamında, değişkenlere yeni eklenen γ_n parametresinin veri setindeki tahmin performansını olumlu etkilediği söylenebilmektedir. RO regresyonunun bir diğer özelliği olan fazla parametre ile daha iyi çalışabilme özelliği de bu durumu destekler niteliktedir.

4.3.2 RO Regresyonu uygulaması (2. Veri seti)

RO regresyon analizlerinde w_L , w_n , γ_n , γ_k değerleri bağımsız değişken, I_p değerleri ise bağımlı değişken olarak tanımlanmıştır. Analizlerde kullanılan w_L değerlerinin üst sınırı %49.2 alt sınırı ise %29.5 olarak belirlenmiştir. Analizler için oluşturulan toplamda 116 deney grubundan oluşan veri setinin %80 oranındaki kısmı, RO algoritmasını eğitmek için, %20 oranındaki kısmı ise algoritmayı test etmek için kullanılmıştır. 2. veri seti için uygulanan çok değişkenli RO regresyonu sonucunda elde edilen I_p tahmin değerleri ve diğer tüm parametreler Şekil 8'deki determinasyon matrisi ile görselleştirilmiştir. Determinasyon matrisinin köşegenindeki dağılımlar, değişkenlerin ağırlığının hangi değerlerde ve ne

kadar olduğunu görebilmek amacı ile eklenmiştir. Ayrıca, bu matristen her bir değişkenin kendi aralarındaki ilişkisi de gözlemlenebilmektedir.



Şekil 8. Çok değişkenli RO regresyonunun 2. Veri setindeki uygulamasında çıktı ile girdi değişkenleri arasındaki korelasyon analizi

2. veri seti için uygulanan çok değişkenli RO regresyonunda R değeri 0.969, R^2 değeri 0.94, SSE değeri 21.51, MSE değeri 0.89, RMSE değeri 0.94 MAPE değeri 4.32 ve MAE değeri 0.81 olarak hesaplanmıştır. Elde edilen bu metrik ölçüm değerleri, analizlere eklenen γ_k parametresinin de algoritma performansını arttırdığı söylenebilmektedir. Ayrıca elde edilen R^2 değerinin Davis (1971) ve Lewis (1982) önerisine göre

değerlendirilmesi durumunda elde edilen eşitliğin ($R^2= 0.70-1.0$ aralığında) çok güçlü bir korelasyon olduğu söylenebilmektedir. Diğer metrik değerlerin sıfıra yaklaşması ise korelasyonun R^2 değeri ile orantılı bir şekilde, iyi bir tahmin performansı sunduğunun göstergesidir. Şekil 8' de tanımlanan ve modele girdi olarak verilen parametrelerin genel dağılım özellikleri Çizelge 1'de sunulmuştur.

Çizelge 1. Girdi parametrelerinin genel dağılım özellikleri

Parametre	Maximum	Minimum	Ortalama	Medyan	Standart sapma
w_L (%)	49.2	29.5	40.35	40.35	6.46
w_n (%)	29.3	18.8	22.6	22.6	2.75
γ_n (t/m ³)	2.000	1.425	1.860	1.883	0.11
γ_k (t/m ³)	1.669	1.400	1.540	1.537	0.07
I_p -gerçek (%)	28.2	14.1	19.06	18.25	4.07
I_p -tahmin (%)	26.195	12.99	19.17	19.557	3.88

Çizelge 1'e göre girdi parametresi olarak tanımlanan tüm değişkenlerde, kendi içlerinde medyan ve ortalama değerlerinin çok yakın olduğu ve normal bir dağılımla ifade edilebileceği görülmektedir. Şekil 8'den de anlaşılacağı gibi tahmin edilen ve gerçek I_p - w_L dağılımları birbirine çokça benzerlik göstermektedir ve bu durum modelin performansı açısından olumlu bir göstergedir. Ayrıca, w_n ve γ_k gibi çok değişkenli regresyon analizi girdileri ile gerçek ve tahmin edilen I_p değerleri arasında standart fonksiyonlar ile (doğrusal, polinom, üssel, logaritmik vb.) açıklanabilecek bir ilişki görülmemektedir. Bu durum, dağınık veri setlerinin ilişkilendirilmesinde ve yüksek performanslı bir tahmin modeli elde edilmesinde RO regresyon metodunun avantajını bir kez daha vurgulamaktadır.

5. Tartışma ve Sonuç

Bu çalışma kapsamında RO regresyon yönteminin düşük plastisiteli killi zeminlerin (CL) plastisite indisinin tahminindeki etkinliği sunulan tek ve çok değişkenli analizler ile incelenmiştir. Tahminleme sürecinde kullanılmak üzere belirlenen zeminlere ait

veriler, İstanbul Valiliği Yatırım İzleme ve Koordinasyon Müdürlüğü'nden temin edilmiş ve ağırlıklı olarak İstanbul İli Avrupa yakasının güneybatısında uygulanan zemin etüt raporları değerlendirmeye alınmıştır. İlgili veri seti kullanılarak, tek değişkenli doğrusal regresyon analizi, tek değişkenli RO yöntemi analizi, 1. veri setinin kullanımı ile çok değişkenli RO yöntemi analizi, 2. veri setinin kullanımı ile çok değişkenli RO yöntemi analizi olmak üzere 4 temel analiz uygulanmıştır. Bu bağlamda, makale kapsamında uygulanan analizlerin sonucunda elde edilen metriklerin değerleri Çizelge 2'de sunulmuştur.

Çizelge 2. Analizler sonucunda elde edilen metrikler

Metrik	Tek değişkenli analiz		Çok değişkenli analiz	
	DR*	RO	RO (1.veri seti)	RO (2. veri seti)
R	0.804	0.865	0.964	0.969
R^2	0.647	0.749	0.930	0.940
SSE	5371	709	93.150	21.510
MSE	8.070	5.330	1.700	0.890
RMSE	2.840	2.300	1.300	0.940
MAPE	11.240	8.830	4.960	4.320
SSE	2.210	1.780	0.950	0.810

*Doğrusal regresyon analizi

Elde edilen metrik değerlerin karşılaştırılması ve geoteknik mühendisliği bakış açısı ile yorumlanması sonucunda elde edilen değerlendirmeler aşağıda maddeler halinde özetlenmiştir:

1. I_p değerlerinin yalnızca w_L değerlerinin kullanımı ile tahminini amaçlayan literatür çalışmalarından Seed vd. (1964), Laskar ve Pal (2012), Sen ve Pal (2014), Spagnoli vd. (2018), Sharma ve Sridharan (2018) kullanmış oldukları farklı yöntemler ile I_p - w_L ilişkisini doğrusal bir şekilde elde etmişlerdir. Ancak söz konusu ilişkilerin, göz önüne alınan tüm kil içerikli zeminler için tatminkâr sonuçlar ortaya koyamadığı Akbay Arama vd. (2021) tarafından yapılan karşılaştırmalar ile tartışılmıştır. Bu bakış açısından yola çıkılarak, bu çalışma kapsamında da, kullanılan özel bir veri setinden ve kil içerikli bir zemin grubundan faydalanılarak, doğrusal bir ilişki kullanımı

- ile I_p - w_L davranışını yorumlamanın yeterliliği sorgulanmıştır.
2. Metrik değerler bir bütün olarak göz önüne alındığında tek değişkenli RO yöntemi uygulamalarının doğrusal regresyon analizlerinden daha iyi performans sergilediği görülmektedir. Ancak söz konusu metriklerden özellikle SSE değerinin büyüklüğü, her iki yöntemde de, korelasyon tahmin gücünün yetersizliği sonucunu ortaya çıkarmaktadır.
 3. Çok değişkenli analizlerdeki bağımsız değişken parametrelerin artırılması durumunda RO regresyonunun vereceği tepkinin, analiz doğruluğunu hangi yönde etkilediği incelenmiştir. Çok değişkenli analiz uygulamasının RO yönteminin performansını daha da arttığı söylenebilmektedir. I_p değerinin tahmininde analizlere girdi olarak dâhil edilen w_n , γ_n , ve sonrasında γ_k parametrelerinin RO regresyonu performansını arttığı metrik değerlerde gözlenen iyileşme ile belirlenebilmektedir. Bu durumun en önemli nedeni olarak, RO regresyonu yönteminin kolektif bir makine öğrenmesi algoritması olması düşünülebilir. RO regresyonu, sayısının kullanıcı tarafından belirlenebileceği karar ağaçlarının tahmin ettiği değerleri toparlayıp ortalamasını almakta ve bu şekilde doğruluğu yüksek olan çıktılar sağlamaktadır. Ayrıca, RO regresyonu, sıklıkla kullanılan diğer analiz yöntemlerindeki gibi herhangi bir doğrusal, polinom, üssel, logaritmik vb. fonksiyonlara bağlı çalışmamaktadır. Bu tür çözümler, denklemin verdiği çıktılar ile sınırlı kalmaktayken, RO regresyonu yöntemi ise diğer yöntemlerle çözümlenemeyen aykırı ve/veya dağınık verilerin tahminine ve/veya çözümüne ulaşabilmektedir.
 4. Çalışma kapsamında kullanılan dağınık veri setlerinde (2. veri seti), standart kullanımdaki analiz yöntemlerine göre daha başarılı performans elde edilmiştir.

5. Literatürde parametre tahmin süreci ile ilişkili çalışmaların birçoğundan farklı olarak bu çalışmada, çok değişkenli analizlerin daha tatminkâr performanslar sunduğu görülmüştür. Çok değişkenli analizlerde kullanılan ilave parametreler likit limit deney sürecinde kolaylıkla belirlenebilen ve hesaplanabilen (w_n , γ_n , γ_k), ilave bir deney süreci veya süresi gerektirmeyen deneyler veya hesaplamalardır. Bu doğrultuda, daha fazla geoteknik parametre tanımı ile şekillendirilen RO regresyonu analizleri, uygulamalarda da kullanılabilir ve hızlı bir çözümleme süreci sunmaktadır.
6. Sonuçlarda sunulan bu değerlendirmelerin, çalışma kapsamında kullanılan veri seti için geçerli olduğu ve farklı veri setleri ile RO metodu uygulamalarının analiz sonucunda elde edilecek metrik değerler doğrultusunda yorumlanması gerektiği unutulmamalıdır.
7. Bu çalışma farklı araştırmaların yürütülmesinde, sağladığı bakış açısı ile faydalı olabilecek ve kullanılan veri açısından tatminkâr bir sayı sunarak yeni çalışmaların önünü açabilecek niteliktedir. Bu doğrultuda, yapay sinir ağları veya derin öğrenme teknikleri gibi daha ileri düzey bilgisayar teknolojileri kullanılarak, ilgili tekniklerin geoteknik tasarımda kullanılabilirlikleri sorgulanabilir.

Teşekkür

Bu çalışma kapsamında değerlendirilen verilerin erişimine sağlamış oldukları katkıdan dolayı İstanbul Valiliği Yatırım İzleme ve Koordinasyon Müdürlüğü'ne teşekkürlerimizi sunarız.

5. Kaynaklar

- Ahmed, Z., Mohamed, K., Zeeshan, S., Dong, X., 2020. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, **2020**, 1-35.
- Akar, Ö., Güngör, O., Akar, A., 2010. Rastgele Orman Sınıflandırıcısı ile arazi kullanım alanlarının belirlenmesi. *III. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu, Gebze, Kocaeli*, **1**, 142-152.

- Akar, Ö., Güngör, O., 2012. Rastgele Orman Algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. *Jeodezi ve Jeoinformasyon Dergisi*, **1**, 2, 139-146. DOI: 10.9733/jgg.241212.1t.
- Akbay Arama, Z., Yücel, M., Akın, M.S., Dalyan, İ., 2021. A comparative study on the application of artificial intelligence networks versus regression analysis for the prediction of clay plasticity. *Arabian Journal of Geosciences*, **14**, 534.
- Archer, K.J., 2008. Empirical characterization of Random Forest variable importance measure, computational statistical data analysis. *Computational Statistics & Data Analysis*, **52**, 4, 2249-2260.
- ASTM D-4318. Standard test methods for liquid limit, plastic limit, and plasticity index of soils.
- ASTM D-422. Standard test method for particle-size analysis of soils.
- Breiman, L., 2001. Random Forests, Machine Learning. *2001 Kluwer Academic Publishers*, **45**, 1, 5-32.
- Çinicioğlu, F., Öser, C., Uzman, E., Kutu, S., Güler, M., 2002. Killerde kıvam parametrelerinin birbirleriyle ilişkilendirilmesi. *Zemin Mekaniği ve Temel Mühendisliği Dokuzuncu Ulusal Kongresi*, Eskişehir, Türkiye.
- Das, M. D., 2010. Principals of Foundation Engineering. SI Edition, 15-17.
- Davis, J.A. 1971. Elementary Survey Analysis. Prentice Hall, New Jersey.
- Ermias, B., Vishal, V., 2020. Application of artificial intelligence for prediction of swelling potential of clay-rich soils. *Geotech Geol Eng*, **38**, 6189–6205.
- Gislason, P.O., Benediktsson, J. A, Sveinsson, J.R., 2004. Random Forest Classification of multi-source remote sensing and geographic data. *IEEE International Geoscience and Remote Sensing Symposium IGARSS '04 Proceedings*, **2**, 1049 – 1052.
- Grömping, U., 2009. Variable importance assessment in regression: linear regression versus Random Forest. *Am Stat*, **63**, 4, 308–319. DOI:10.1198/tast.2009.08199.
- Iyisan, R., 1993. Geoteknik özelliklerin belirlenmesinde sismik ve penetrasyon deney sonuçlarının karşılaştırılması (Doktora Tezi), İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Geoteknik Mühendisliği, 255.
- Kandpal, M., Kandpal, A., 2014. Establishing correlation between size estimation metrics and effort-A statistical approach. *International Journal of Computer Applications*, **95**, 21, 1-6.
- Kayabası, A., 2020. Geotechnical properties of fine-grained soils in Ankara/Turkey: an assessment of the existing empirical equations. *Environ Earth Sci*, **79**, 282.
- Laskar, A., Pal, S. K., 2012. Geotechnical characteristics of two different soils and their mixture and relationships between parameters. *EJGE*, **17**, 2821–2832.
- Lewis, C. D., 1982. Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting. Butterworths Scientific, London.
- Ly, H., Nguyen, T., Pham, B. T., 2021. Estimation of soil cohesion using machine learning method: A Random Forest approach. *Advances in Civil Engineering*, 8873993, 14. <https://doi.org/10.1155/2021/8873993>
- Liaw, A., Wiener, M., 2002. Classification and regression by Random Forest. *R News*, **2**, 3, 18–22.
- Naveena, N., Sanjay, S. J., 2018. Establishing relationship between plasticity index and liquid limit by simple linear regression analysis. *Int J Res Appl Sci Eng Technol*, **6**, 6, 1975–1978.
- Nuray, S. E., Gençdal, H. B., Akbay Arama, Z., 2021. Zeminlerin kıvam ve kompaksiyon özelliklerinin tahmininde Rastgele Orman Regresyonu yönteminin uygulanabilirliği. *Journal of Engineering Sciences and Design*, **9**, 1, 265-281. DOI: 10.21923/jesd.804446.
- Ouedraogo, I., Defourny, P., Vanclooster, M., 2019. Application of Random Forest Regression and comparison of its performance to multiple linear regression in modeling groundwater nitrate concentration at the African Continent scale. *Hydrogeology Journal*, **27**, 1081–1098. <https://doi.org/10.1007/s10040-018-1900-5>.

- Pal, M., 2003. Random Forest for land cover classification. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS '03Proceedings*, **6**, 3510-3512.
- Pham, T.B., Qi, C., Ho, L.S., Thoi, T.N., Ansari, N.A., Nguyen, M.D., Nguyen, H.D., Ly, H.B., Le, H.V., Prakash, I., 2020. A novel hybrid soft computing model using Random Forest and particle swarm optimization for estimation of undrained shear strength of soil. *Sustainability*, **12**, 2218; DOI:10.3390/su12062218.
- Prakash, S., Jain, P.K., 2002. Engineering Soil Testing, Nem Chand & Bros, Roorkee.
- Seed, H. B., Woodward, R.J., Lundgren, R., 1964. Fundamental aspects of the atterberg limits. *J Soil Mech Found Div*, **90**, 6, 75–106.
- Segal, M.R. 2003. Machine Learning Benchmarks and Random Forest Regression. <https://escholarship.org/uc/item/35x3v9t4>.
- Sen, B., Pal, S. K., 2014. Index properties of soils collected from different locations and correlations of parameters. *EJGE*, **19**, 3443–3452.
- Sharma, B., Sridharan, A., 2018. Liquid and plastic limits of clays by cone method. *Geo-Engineering*, **9**, 22–31.
- Shimobe, S., Spagnoli, G., 2020. Fall cone tests considering water content, cone penetration index, and plasticity angle of fine-grained soils. *J Rock Mech Geotech Eng*, **12**, 1347–1355. <https://doi.org/10.1016/j.jrmge.2020.02.005>.
- Spagnoli, G., Sridharan, A., Oreste, P., Bellato, D., Matteo, L. D., 2018. Statistical variability of the correlation plasticity index versus liquid limit for smectite and kaolinite. *Appl Clay Sci*, **156**, 152–159.
- Srinath, K.R. 2017. Python – The fastest growing programming language. *International Research Journal of Engineering and Technology*, Volume **4**, Issue 12, 354-357.
- Swamidass, P.M., 2000. Mean absolute percentage error (MAPE). *Encyclopedia of Production and Manufacturing Management*. Springer, Boston, MA. https://doi.org/10.1007/1-4020-0612-8_580.
- TS 1900-1. Türk Standardı. İnşaat mühendisliğinde zemin laboratuvar deneyleri-fiziksel özelliklerin tayini.
- Waske, B., Heinzl, V., Braun, M., Menz, G., 2007. Random Forests for classifying multi-temporal sar data, Proc. 'Envisat Symposium Montreux, Switzerland. <http://envisat.esa.int/envisatsymposium/proceedings/sessions/3D3/461589wa.pdf>.
- Watts J. D., Lawrence R. L., 2008. Merging Random Forest classification with an object-oriented approach for analysis of agricultural lands. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII(B7).
- Whyte, I. L., 1982. Soil plasticity and strength-a new approach using extrusion. *Ground. Eng.*, **15**, 1, 16-24.
- Yogatama, B. A., Tirta, B. A., 2021. Python application in geotechnical engineering practices. *Simposium Nasional Teknologi Infrastruktur Abad ke-21*.
- Zhang, W., Wu, C., Zhong, H. Li, Y., Wang, L., 2021. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geoscience Frontiers*, 469-477.

İnternet Kaynakları

<https://veribilimcisi.com/mse-rmse-mae-mape-metrikleri-nedir/> (14.07.2017).

<https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d> (14.07.2017).