# TAŞINMAZ DEĞERLEMEDE MAKİNE ÖĞRENME ALGORİTMALARININ KULLANIMI: PENDİK ÖRNEĞİ

E.C. Seyrek*, İ. Çölkesen, R. Bovkır, A.Ç. Aydınoğlu

Gebze Teknik Üniversitesi, Mühendislik Fakültesi,
Harita Mühendisliği Bölümü, 41400 Gebze-Kocaeli
(ecseyrek2018; icolkesen; rbovkir; aydinoglu @gtu.edu.tr)

**ANAHTAR SÖZCÜKLER:** Taşınmaz Değerleme, Makine Öğrenme, CBS, Genelleştirilmiş Lineer Model, Rastgele Orman

**ÖZET:**

Taşınmaz değerleme, taşınmaza ilişkin çevresel ve sosyal faktörlerin nitelik ve fayda bilgileri ile birlikte objektif olarak incelenerek söz konusu taşınmazın değerinin tespit edilmesi işlemidir. Taşınmaz değerleme, vergilendirme, kamulaştırma, özelleştirme, devletleştirme, sigortacılık, bankacılık vb. uygulamalarda aktif olarak kullanılmaktadır. Türkiye'de taşınmaz değerleme için nesnel olarak belirlenmiş değer kriterleri ve özelleşmiş bir yasal mevzuat mevcut değildir. Bunun sonucu olarak taşınmaz malların değerleri objektif olarak belirlenememektedir. Makine öğrenme algoritmaları yardımı ile çok büyük miktarlarda ve çok farklı türlerdeki verilerin analizi başarılı bir şekilde gerçekleştirilmektedir. Tıp, bilişim, optik, robotik, görüntü analizi gibi birçok alanda kullanılan makine öğrenmesi, son yıllarda veri madenciliği, istatistiki tahmin gibi alanlarda da aktif olarak kullanılmaktadır. Bu çalışmanın amacı taşınmaz malların değerlerinin objektif olarak belirlenebilmesi için taşınmaza ilişkin konumsal ve konumsal olmayan özelliklerin bir arada ele alınarak makine öğrenme algoritmalarıyla taşınmaz değer haritaları üretilmesidir. Bu amaçla, İstanbul ili Pendik ilçesi için Coğrafi Bilgi Sistemleri (CBS) yazılımları kullanılarak oluşturulan konumsal ve konumsal olmayan özniteliklere ait katmanlar, genelleştirilmiş lineer model (GLM) ve rastgele orman (RO) olmak üzere iki yaygın makine öğrenme algoritması kullanılarak taşınmaz değerlemeye ilişkin regresyon analizi yapılmıştır. Algoritmaların tahmin doğruluklarının karşılaştırılması için ortalama mutlak hata (MAE), kök ortalama karesel hata (RMSE), ortanca mutlak hata (MdAE), $R^2$ ve korelasyon katsayısı ölçütleri hesaplanmıştır. Algoritmaların performansı, tahmin doğruluğu, kullanıcı tabanlı parametrelerin tespiti ve işlem süresi yönünden karşılaştırılmıştır. Sonuçlar, bu çalışmada ele alınan taşınmaz değerleme problem için en düşük hata değerlerinin (yani en yüksek tahmin doğruluğunun) RO ile elde edildiğini göstermektedir. Bununla birlikte RO, GLM'ye göre daha fazla işlem üresi gerektirmesine rağmen, hassas bir parametre seçimi gerektirmemektedir ve performansı GLM'ye göre daha yüksektir. Çalışma sonuçları RO toplu öğrenme algoritmasının taşınmaz değerlemede kullanılabilir bir yöntem olduğunu göstermektedir.

**KEY WORDS:** Real Estate Valuation, Machine Learning, GIS, Generalised Linear Model, Random Forest

**ABSTRACT:**

Real estate valuation is the process of determining the value of the real property by examining environmental and social factors related to the real property with the quality and benefit information. Real estate valuation is actively used in taxation, expropriation, privatization, insurance and finances applications. In Turkey, legislation about real estate valuation is limited and there is no standard tool or approach for the real estate valuation. As a result, the value of the real property cannot be determined objectively. With the assistance of machine learning algorithms, the analysis of large amounts of data is performed successfully. Besides disciplines such as medicine, informatics, optical, robotics and image analysis, machine learning algorithms has been used actively in the fields of data mining and statistical estimation in recent years. The main goal of this study is to estimate the real estate values using spatial and non-spatial features of the real estates by applying machine learning algorithms. For this purpose, spatial and non-spatial features of sample real estates located in Pendik district of Istanbul were created with Geographic Information Systems (GIS) software, and regression analysis was performed for real estate valuation using two popular machine learning algorithms namely, generalized linear model (GLM) and random forest (RF). In order to analyze and compare the performance of algorithms, standard accuracy measures including mean absolute error (MAE), root mean squared error (RMSE), median absolute error (MdAE), $R^2$ and correlation coefficient were utilized. The performance of algorithms was compared in terms of prediction accuracy, determination of user-defined parameters and processing time. Results showed that the lowest error rates (i.e. highest prediction accuracy) was estimated by RF algorithm for the real estate valuation problem considered in this study. Furthermore, although required processing time for RF slightly higher than GLM algorithm, it does not require precise parameter setting and its performance was superior compared to GLM. All in all, the results of the study highlighted the potential usefulness of the RF ensemble algorithm in real estate valuation.

* E-Posta: ecseyrek2018@gtu.edu.tr

## 1. GİRİŞ

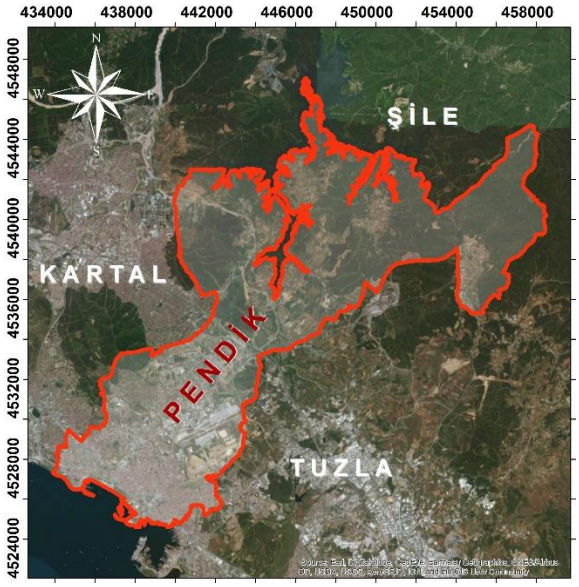
Taşınmaz değerine etki eden faktörlerin objektif olarak belirlenmesi ve bu faktörlere göre ilgili taşınmaz değerlerinin belirlenmesi ülke ekonomisi açısından oldukça önemlidir. Planlı şehirleşmenin yürütülmesi, yerleşme bölgelerinin seçimi, seçilen bu bölgeler arasında iç ve dış bağlantı giderlerinin karşılaştırılması, imar planlarının ekonomik olarak değerlendirilebilmesi, yapı iyileştirilmesi açısından büyük önem taşımaktadır (Karapınar vd., 2008). Ekonomik gelişimin yanı sıra, kentsel dönüşüm, vergilendirme, kamulaştırma, özelleştirme, kırsal ve kentsel arazi yönetimi gibi kamusal uygulamalarda ve bankacılık, kredilendirme, sigortacılık gibi özel sektör uygulamalarında objektif olarak belirlenmiş taşınmaz değerlerine yoğun bir şekilde ihtiyaç duyulmaktadır. 35 nolu Sermaye Piyasası Mevzuatına göre taşınmaz değerleme; bir taşınmazın, taşınmaz projesinin veya bir taşınmaza bağlı hak ve faydaların, belli bir tarihteki muhtemel değerinin bağımsız ve tarafsız olarak takdirini ifade eder (Resmi Gazete, 2001). Değerlemeye konu olan taşınmaza ilişkin nitelik, fayda, çevre, kullanım koşulları gibi faktörler değerlendirilerek söz konusu taşınmazın değeri tespit edilir (Güngör, 1999).

Taşınmaz değerinin saptanması için standart bir yaklaşım olmadığından değerlendirme yapılırken belli bir sisteme uyulmamaktadır. Gerçek anlamda, herhangi bir taşınmaza ait kesin değerin tespit edilmesi mümkün değildir. Bu nedenle değer; kullanma amacı, piyasa koşulları, faiz, beklenen yarar gibi faktörlerle değerlendirilerek oluşturulmaktadır. Söz konusu nedenlerden dolayı taşınmaz değer göreceli bir kavramdır. Konu toprak ise üstündeki taşınmaz (bina) ve toprağın konumu ile değişen imar koşulları; konu bina ise kurulu olduğu toprağın durumu ile binanın özellikleri bu değerde rol oynar (Arslan, 1997). Taşınmazların toplu olarak değerlemesinde klasik değerleme yöntemleri olan emsal, gelir, maliyet ve regresyon yöntemleri (Pagourtzi et al., 2003) yetersiz kalmaktadır. Taşınmaz değerini etkileyen birçok faktör mevcuttur ve bunlar oldukça karmaşıktır. Ayrıca, satış değerleri istatistiksel verilerdir. Bu nedenle taşınmaz değerlemenin ileri tahmin algoritmaları yardımıyla çok yönlü analiz edilmesi ve sonuçların objektif bir şekilde değerlendirilmesi gerekmektedir. Makine öğrenme algoritmaları olarak bilinen rasgele orman, yapay sinir ağları, bulanık mantık, lojistik regresyon gibi yöntemler son zamanlarda özellikle regresyon analizlerinin çözümünde sıklıkla kullanılan gelişmiş tahmin algoritmaları olarak kullanılmaktadır (Kontrimas and Verikas, 2011; Bogataj et al., 2011; Derinpınar, 2014; Güneş ve Yıldız, 2015; Demetriou, 2016; Bovkır vd., 2018).

Bu çalışmada öncelikle taşınmaz değerine etki eden konumsal ve konumsal olmayan faktörler Coğrafi Bilgi Sistemleri (CBS) yazılımlarının analiz yeteneği ile birlikte çalışma alanı olan Pendik ilçesi için oluşturulmuştur. Oluşturulan konumsal ve konumsal olmayan faktörlere ilişkin katmanlar, çalışma alanında yer alan taşınmazlara ait satış verileri temel veri seti olarak kullanılarak, genelleştirilmiş lineer model (GLM) ve rastgele orman (RO) algoritması yardımıyla taşınmaz değerlerinin tespiti ve değer haritasının oluşturulmasına yönelik uygulama yapılmıştır. Algoritmaların performansı, tahmin doğruluğu, kullanıcı tabanlı parametrelerin tespiti ve işlem süresi yönünden karşılaştırılmıştır.

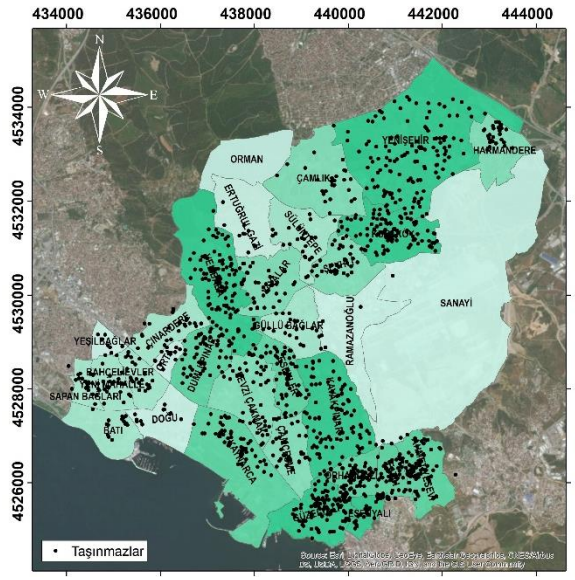
## 2. ÇALIŞMA ALANI VE VERİ SETİ

Çalışma alanı olarak İstanbul'un Pendik ilçesi seçilmiştir. Pendik ilçesi Marmara bölgesinde, İstanbul'un Anadolu Yakasında yer almaktadır (Şekil 1). İlçe doğuda Tuzla, kuzeyde Şile, batıda Kartal ve Sultanbeyli ilçelerine ve güneyde Marmara Denizi'ne komşudur. TÜİK'in 2018 yılı Adrese Dayalı Nüfus Kayıt Sistemine göre 693.599 nüfusa sahip Pendik, İstanbul'un dördüncü en büyük ilçesidir (URL 1). İlçede kara, hava ve deniz yolu ile ulaşım mümkündür.




Şekil 1. Pendik İlçe Sınırları

Çalışmada kullanılan veriler 116Y204 numaralı TÜBİTAK projesi kapsamında Pendik ilçesi için temin edilmiştir. Veri seti Pendik'in ilçe merkezindeki 31 mahallede satışta bulunan ve güncel satış değerleri mevcut olan 1.458 tane taşınmazı içermektedir (Şekil 2). İlçenin Kartal ilçesine sınır olan mahallelerinden Orman mahallesinde satışta olan bir konut bulunmadığından bu mahalle için örnek tespit edilememiştir. Taşınmaz satış değerleri hesap kolaylığı amacıyla normalize edilmiştir.




Şekil 2. Değerlemede kullanılan örnek taşınmazların dağılımı.

Taşınmazların 13 yapısal özniteliği içeren veri seti konumsal olmayan (KO) olarak adlandırılmıştır KO veri setinde yer alan öznitelikler Tablo 1'de verilmiştir. Tabloda görüleceği üzere, dairenin fiziksel özelliklerini içeren alan, banyo sayısı, cephe yönü, ısıtma türü, manzara, oda sayısı, bulunduğu kat öznitelikleri, dairenin bulunduğu binaya dair asansör, bina yaşı, binanın toplam kat sayısı, otopark alanı, site içinde bulunma gibi öznitelikler KO veri setinde yer almaktadır.

Tablo 1. Taşınmazların konumsal olmayan öznitelikleri.

| KO veri seti: Konumsal olmayan öznitelikler | |
|---|---|
| Alan | Manzara |
| Asansör | Oda Sayısı |
| Banyo Sayısı | Otopark Alanı |
| Bina Yaşı | Site İçinde Bulunma |
| Binanın Toplam Kat Sayısı | Taşınmazın Bulunduğu Kat |
| Cephe Yönü | Taşınmaz Durumu |
| Isıtma Türü | |

Değer tespiti yapılmak istenen taşınmazın şehirdeki konumu ve bu konumun topoğrafik özellikleri de büyük öneme sahiptir. Çalışmada ilçedeki ulaşım, kamu tesisleri gibi bazı önemli konumlara olan uzaklıkları ifade eden yüzeyler CBS ortamında Öklid mesafesi esas alınarak hesaplanmıştır. Tablo 2'de verilen özellikler içerisinde dini merkezlere uzaklık, alışveriş merkezlerine uzaklık gibi öznitelikler buna örnektir. Ayrıca taşınmazın bulunduğu muhitin sosyokültürel yapısı da taşınmaz değerleri üzerinde etkili olduğundan nüfus yoğunluğu, okuma yazma bilen kişilerin yoğunluğu ve üniversite mezunu kişi yoğunluğu gibi faktörleri belirlemek için CBS ortamında yoğunluk analizleri yapılmış ve ilgili faktörlerin yoğunluk dağılımlarını ifade eden yüzeyler oluşturulmuştur. Ayrıca bölgenin sayısal arazi modeli (SAM) kullanılarak eğim ve bakı karakteristiklerini ifade eden yüzeyler de oluşturulmuştur. CBS yazılımında yapılan analizlerle üretilen faktörlerin yer aldığı veri seti konumsal (K) olarak adlandırılmıştır ve oluşturulan toplam 21 öznitelik Tablo 2'de gösterilmiştir.

Tablo 2. Taşınmazların konumsal olmayan öznitelikleri.

| K veri seti: Konumsal analizlerle oluşturulan öznitelikler |
|---|
| Alışveriş Merkezlerine Uzaklık |
| Altyapı Tesislerine Uzaklık |
| Bakı |
| Caddeye Uzaklık |
| Dini Merkezlere Uzaklık |
| Eğim |
| Eğitim Kurumlarına Uzaklık |
| Havaalanına Uzaklık |
| İdari Tesislere Uzaklık |
| Kültürel Tesislere Uzaklık |
| Metroya Uzaklık |
| Mezarlığa Uzaklık |
| Nüfus Yoğunluğu |
| Okuma Yazma Bilenlerin Yoğunluğu |
| Otobüse Uzaklık |
| Otoparka Uzaklık |
| Otoyola Uzaklık |
| Sağlık Tesislerine Uzaklık |
| Sanayiye Uzaklık |
| Üniversite Mezunu Yoğunluğu |
| Yeşil Alana Uzaklık |

Tablodan görüleceği üzere, alışveriş merkezine uzaklık, altyapı tesislerine uzaklık, caddeye uzaklık gibi taşınmazın önemli konumlara olan uzaklıkları, bakı ve eğim gibi taşınmazın bulunduğu topoğrafyanın özellikleri, nüfus yoğunluğu, okuma yazma bilenlerin yoğunluğu gibi sosyokültürel öznitelikler K veri setinde yer almaktadır.

Taşınmaz değerlemede, değere etki eden konumsal ve konumsal olmayan kriterlerin bir arada kullanıldığı veri seti ise toplamda 34 özniteliğin bulunduğu konumsal ve konumsal olmayan (KKO) veri seti olarak adlandırılmıştır. Uygulama kapsamında, taşınmaz değer tespiti amacıyla üç ayrı veri seti ayrı ayrı değerlendirmeye alınmış ve değere etki eden özniteliklerin tahmin modelinin performansına olan etkileri analiz edilmiştir.

## 3. MAKİNE ÖĞRENME ALGORİTMALARI

Makine öğrenme, matematiksel ve istatistiksel yöntemlerin kullanılarak eldeki verilerden çıkarım yapıp, bu çıkarımlara dayanarak bilinmeyen verilere dair çıkarımlar yapılmasını sağlayan yöntemlerdir. Bu çalışmada temel regresyon yöntemlerinden genelleştirilmiş lineer model ve hem sınıflandırma hem de regresyon amacıyla yaygın olarak kullanılan rastgele orman algoritması kullanılmıştır.

### 3.1. Genelleştirilmiş Lineer Model

Regresyon analizi, iki veya ikiden fazla değişken arasındaki ilişkinin hangi matematiksel modelle ifade edileceğini araştırır. İkiden fazla sayıda bağımsız değişken bulunması durumunda modele çoklu lineer regresyon modeli ismi verilir (Özmen vd., 2013). Çoklu lineer regresyon modelinin genel formülü Eşitlik (1)'de verilmiştir.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \quad (1)$$

Lineer regresyon modeli, bağımlı ve bağımsız değişkenlerin doğrusal bir fonksiyonu olduğunu ve bağımlı değişkenin sürekli ve sabit varyansa sahip normal dağılımda olduğu kabulünü yapmaktadır (Dobson and Barnett, 2008). Kullanılan veri setinde bağımlı değişkenler normal dağılımda olmayabilir. Bu durumun üstesinden gelmek için Nelder ve Wedderburn (1972) bağımlı değişkenlerin normal dağılımda olmadığı regresyon modelleri için genelleştirilmiş lineer modelleri (GLM) geliştirmişlerdir. GLM, lojistik ve poisson regresyon modelleri durumunda, modelin formülasyonunda olasılık dağılımı ile ortalama ve varyans arasındaki ilişkiyi ve regresyon parametrelerinin tahminini içerir (Dobson and Barnett, 2008). GLM, bağımlı değişkenin koşullu ortalamasının bir fonksiyonunu, bir bağımsız değişkenler dizisinin doğrusal bir fonksiyonu olarak tahmin etmeyi içerir. Bu nedenle genelleştirilmiş modeller bağımlı değişkenin ortalamasının regresyon parametrelerinin doğrusal olmayan bir fonksiyonu olduğu ve normal dağılım göstermediği veri setleri ile çalışılmasına imkan sağlar. İlişki (link) fonksiyonu ve hata dağılımı, GLM ile regresyon modeli oluşumunda iki temel bileşendir. İlişki fonksiyonu bağımlı değişkenin ortalamasını regresyon parametrelerinin doğrusal bir fonksiyonu olacak şekilde dönüşümüne imkan sağlamaktadır. Diğer taraftan, hata dağılımı bağımlı değişkenin varyansının kendi ortalamasının bir fonksiyonu olmasını sağlar.

Taşınmaz değerleme probleminde bağımlı değişken olarak seçilen ve 0 ile 1 arasında normalize edilmiş satış değeri, normal dağılıma sahip olmayan sürekli bir dağılım

göstermektedir. Bu sebeple, çalışmada kullanılan veri setine en iyi uyumu sağlayacak "quasibinomial" hata dağılımı ve "logit" link fonksiyonu kullanılması uygun görülmüştür.

### 3.2. Rastgele Orman

RO algoritması, temel sınıflandırıcı olarak karar ağaçlarını kullanan popüler bir toplu öğrenme algoritmasıdır. Eğitim aşamasında birden çok karar ağacı kullanılması sebebiyle karar ağacı ormanı olarak tanımlanabilir (Breiman, 2001). RO, torbalama (bagging) yöntemi ile rastgele özellik seçiminin birleştirilmesi sonucu oluşturulmuş bir yöntemdir (Breiman, 2001). Her karar ağacı önyükleme (bootstrap) yöntemini kullanarak veri setinden örnekler seçilerek ve her düğümde tüm değişkenler arasında rastgele değişkenlerin saptanması ile oluşturulur (Özdemir, 2018).

RO algoritmasının oluşturulmasında kullanıcının ağaç sayısı (N) ve karar ağacı kurulurken her düğümde oluşturulacak değişken sayısı (m) olmak üzere iki parametre seçmesi gereklidir (Breiman, 2001). Seçilen parametreler doğrultusunda RO algoritması her karar ağacının eğitimi için veri setinden rastgele alt kümeler oluşturur. Oluşturulan veri alt kümelerinin 2/3'ü (in-bag) karar ağaçlarının eğitimi için ayrılırken geriye kalan 1/3'ü ise test verisi (out-of-bag) olarak ayrılır.

RO algoritmasında karar ağaçlarının oluşumunda budama olmaksızın maksimum boyutta ağaç geliştirmek için CART (Classification and Regression Trees) algoritması kullanılmaktadır (Breiman, 2001). CART algoritması GINI indeksini kullanarak karar ağacındaki en iyi dallanmayı belirler (Gislason, 2006). Bir $T$ eğitim setinden seçilen $C_i$ sınıfına ait rastgele bir örnek için GINI indeksinin hesaplanması Eşitlik (2)'deki formülde ifade edilmiştir. Bu formülde $(f(C_i,T)\backslash|T|)$ ifadesi, örneğin $C_i$ sınıfına ait olma olasılığını ifade etmektedir (Pal, 2005).

$$\sum\sum_{j\neq i}(f(c_i,T\backslash|T|)(f(c_i,T\backslash|T|) \qquad (2)$$

En küçük GINI değerine sahip sınıf, ağacın dallanmasını belirler. GINI değerinin artması sınıfın heterojenliğinin arttığını, azalması ise sınıfın homojenliğinin arttığını göstermektedir. GINI değeri sıfıra ulaşıncaya kadar dallanma devam eder (Watts, 2011).

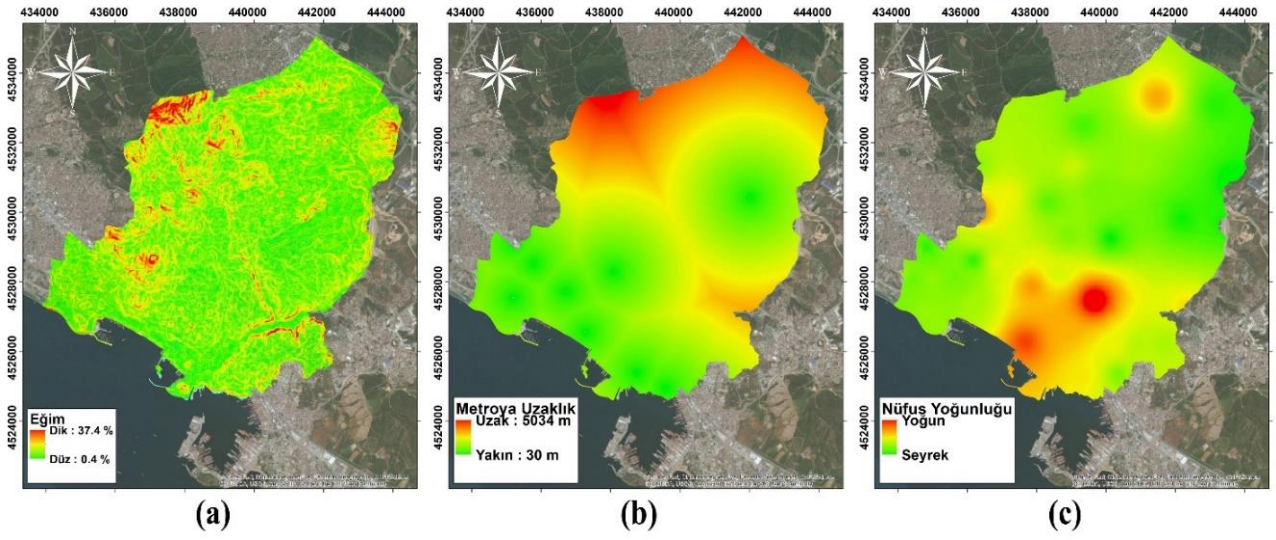
## 4. UYGULAMA

Değerlemede taşınmazın konumsal özniteliklerinin yer aldığı K veri seti CBS yazılımında yer alan analiz araçları ile üretilmiştir. Bölgenin Sayısal Arazi Modeli (SAM) kullanılarak eğim ve bakı katmanları oluşturulmuştur. Bölgenin nüfus ve eğitim düzeyini yansıtan katmanlar nokta yoğunluğu analizi ile; diğer katmanlar ise Öklid mesafe analizi ile üretilmiştir. Oluşturulan katmanlara örnek olarak eğim, metroya uzaklık ve nüfus yoğunluğuna ilişkin haritalar Şekil 3'te gösterilmiştir. Katmanlar ENVI 5.3 yazılımı kullanılarak katman istifleme işlemine tabi tutulup tek bir görüntü dosyasına dönüştürülmüştür. Değerlemede kullanılan KO veri seti doğrudan noktaların öznitelik tablosundan elde edilmiştir. Bütün özniteliklerin kullanıldığı KKO veri seti ise konuma dayalı olarak üretilen katmanların taşınmazlara ait noktalarla çakışan piksellerdeki değerlerinin öznitelik tablosuna eklenmesi ile oluşturulmuştur.

Taşınmaz değer regresyon analizi için R istatistik aracı kullanılmıştır. R yazılımı özgür ve açık kaynak kodludur. Gönüllülük esaslı olarak geliştirilen yazılım güçlü bir istatistiksel araçtır (de Micheaux, 2013). Ayrıca QGIS ve ArcGIS gibi popüler CBS yazılımlarına entegre olarak çalışabilmektedir.

Veri setlerinin %70'ini oluşturan 1.021 taşınmaza ait örnekler eğitim veri seti olarak seçilirken geriye kalan 437 örnek test verisi olarak ayrılmıştır. Bu çalışma kapsamında taşınmaz değer haritası oluşturulmasında GLM ve RO algoritmalarından yararlanılmıştır.

RO algoritmasıyla tahmin modeli oluşturulmasında kullanıcı tarafından belirlenmesi gereken 2 parametre (ağaç sayısı ($N$) ve her düğümde seçilecek özellik sayısı ($m$)) mevcuttur. Bu parametreler K veri seti için m=5 ve $N$=300; KO veri seti için $m$=4 ve $N$=300 ve KKO veri seti için $m$=6 ve $N$=300 olarak seçilmiştir.




Şekil 3. CBS ortamında oluşturulan (a) eğim, (b) metroya uzaklık ve (c) nüfus yoğunluğu katmanları

Oluşturulan GLM ve RO modellerinin geçerliliği test veri seti kullanılarak analiz edilmiştir. Tahmin edilen değerler ile gerçek değerler baz alınarak her veri seti için Eşitlik (3)'deki formüle göre hesaplanan mutlak hata (MAE) değerleri Tablo 3'te yer almaktadır. MAE değerinin sıfıra yakın çıkması, tahmin performansının yüksek olduğunu göstermektedir. Tablodaki hata değerleri incelendiğinde RO algoritmasının KKO veri seti ile oluşturduğu regresyon modelinin en düşük hata değerine sahip olduğu görülmektedir.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|A_i - P_i\right| \tag{3}$$

Tablo 3. Test veri seti için hesaplanan MAE hata ölçütü.

| Algoritmalar | Veri Setleri | | |
|---|---|---|---|
| | K | KO | KKO |
| LR | 0,095 | 0,111 | 0,100 |
| RO | 0,095 | 0,107 | **0,093** |

Uygulamada değerlemeye alınan algoritmalar ve veri setleri için Eşitlik (4)'teki formüle göre hesaplanan kök ortalama karesel hata (RMSE) değerleri Tablo 4'te yer almaktadır. RMSE hata ölçütünün sıfıra yakın değer alması tahmin performansının yüksek olduğunu gösterir. Tablodaki hata değerleri incelendiğinde, KKO veri seti kullanılarak oluşturulan RO modelinin hata değerinin en küçük olduğu tespit edilmiştir. RO algoritmasında K ve KKO veri setleri için hata değerleri birbirine yakın değerler aldığı, KO veri seti için hata miktarının daha yüksek olduğu görülmektedir.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(A_i - P_i\right)^2} \tag{4}$$

Tablo 4. Test veri seti için hesaplanan RMSE hata ölçütü.

| Algoritmalar | Veri Setleri | | |
|---|---|---|---|
| | K | KO | KKO |
| LR | 0,121 | 0,141 | 0,126 |
| RO | 0,120 | 0,138 | **0,118** |

Taşınmaz değerleme haritası için oluşturulan tahmin modelinin doğruluğunun belirlenmesinde Eşitlik (5)'e göre hesaplanan ortanca mutlak hata (MdAE) değerleri Tablo 5'te yer almaktadır. Aykırı değerlere karşı dayanıklı olan MdAE ölçütünün sıfıra yakın değer alması tahmin performansının yüksek olduğunu göstermektedir. Tabloda yer alan hata değerleri incelendiğinde KKO veri seti ile oluşturulan RO regresyon modelinin hata değerinin düşük olduğu ve modelin tahmin performansının daha yüksek olduğu anlaşılmaktadır. KO veri seti ile oluşturulan GLM regresyon modeli de en yüksek hata değerine sahiptir.

$$MdAE = \underset{i=1\ldots n}{Md}\left(\left|A_i - P_i\right|\right) \tag{5}$$

Tablo 5. Test veri seti için hesaplanan MdAE hata ölçütü.

| Algoritmalar | Veri Setleri | | |
|---|---|---|---|
| | K | KO | KKO |
| LR | 0,082 | 0,094 | 0,085 |
| RO | 0,078 | 0,092 | **0,077** |

$R^2$ değerleri algoritmalar için hesaplanarak Tablo 6'da gösterilmiştir. Eşitlik (6)'da yer alan formüle göre hesaplanan $R^2$ ölçütü 0 ile 1 arasında değer almaktadır. Ölçütün 1'e yakın değer alması modelin tahmin performansının yüksek olduğunu gösterir. Tablodaki değerler incelendiğinde K veri seti ile oluşturulan RO regresyon modelinin $R^2$ değerinin en yüksek değeri aldığı görülmektedir. KO veri seti ile kurulan her iki modelin de $R^2$ değeri düşüktür.

$$R^2 = \frac{\sum_{i=1}^{n}\left(P_i - A_i\right)^2}{\sum_{i=1}^{n}\left(A_i - \overline{A}\right)^2} \tag{6}$$

Tablo 6. Test veri seti için hesaplanan $R^2$ ölçütü.

| Algoritmalar | Veri Setleri | | |
|---|---|---|---|
| | K | KO | KKO |
| LR | 0,257 | 0,110 | 0,324 |
| RO | **0,373** | 0,084 | 0,219 |

Korelasyon katsayıları tahmin edilen değerler ile gerçek değerler arasındaki ilişkinin göstergesidir. Çalışmada korelasyon katsayıları Eşitlik (7)'deki formüle göre hesaplanmıştır. Korelasyon katsayısı -1 ile 1 arasında değer almaktadır. Değerin 1'e yakın olması tahmin değerleri ile gerçek değerler arasındaki ilişkinin pozitif yönde yüksek olduğunu; -1'e yakın olması ilişkinin negatif yönde yüksek olduğunu gösterir. Değerin 0 olması ise gerçek değerler ile tahmin edilen değerler arasında herhangi bir ilişki bulunmadığını gösterir. Tablo 7'de yer alan korelasyon katsayıları incelendiğinde KO veri seti ve RO algoritması kullanılarak oluşturulan modelin tahmin değerleriyle gerçek değerler arasında pozitif yönde en yüksek korelasyonun sağlandığı görülmektedir. Yalnızca konumsal olmayan özniteliklerin bulunduğu KO veri seti kullanılarak oluşturulan modellerde gerçek değerler ile tahmin değerleri arasındaki ilişkinin en düşük olduğu görülmektedir.

$$CorC = \frac{\dfrac{\sum_{i=1}^{n}\left(P_i - \overline{P}\right)\left(A_i - \overline{A}\right)}{n-1}}{\sqrt{\left(\dfrac{\sum_{i=1}^{n}\left(P_i - \overline{P}\right)^2}{n-1}\right)\left(\dfrac{\sum_{i=1}^{n}\left(A_i - \overline{A}\right)^2}{n-1}\right)}} \tag{7}$$

Tablo 7. Test veri seti için hesaplanan korelasyon katsayıları.

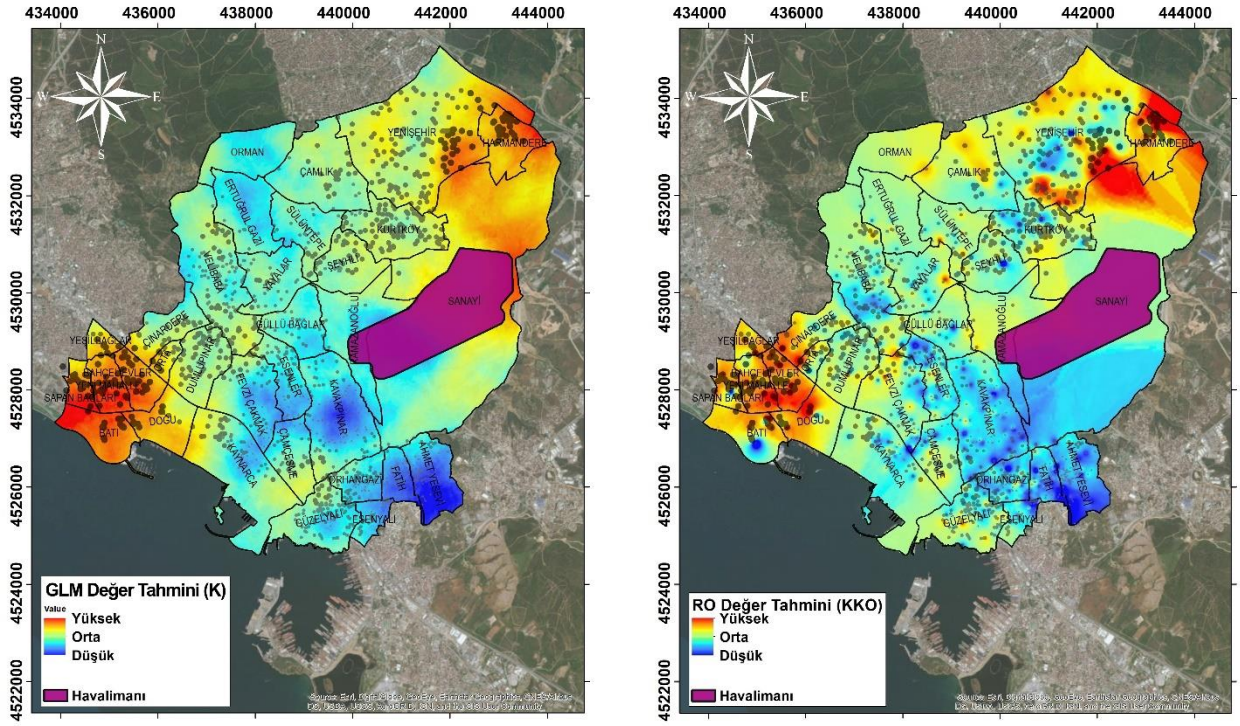
| Algoritmalar | Veri Setleri | | |
|---|---|---|---|
| | K | KO | KKO |
| LR | 0,453 | 0,028 | 0,399 |
| RO | 0,482 | 0,074 | **0,492** |

Algoritmaların performansı işlem süresi açısından da ele alınmış, model oluşumu (eğitim) ve test veri setinin tahmini için ihtiyaç duyulan süre her iki algoritma için de hesaplanmış ve karşılaştırılmıştır. GLM algoritmasının işlem süresi K veri seti için 0,10 saniye, KO veri seti için 0,22 saniye ve KKO veri seti için 0,28 saniye olarak tespit edilmiştir. RO algoritmasının işlem süresi ise K veri seti için 2,25 saniye, KO veri seti için 2,54 saniye ve KKO veri seti için 3,44 saniye olarak tespit

edilmiştir. Sonuçlardan da görüleceği üzere GLM algoritması model oluşumu ve tahmini için daha az süre gerektirmektedir.

GLM algoritması için hesaplanan hata ölçütleri incelendiğinde, algoritmanın en yüksek doğruluğu K veri seti ile sağladığı görülmüştür. R yazılımında raster verinin tamamına regresyon modeli uygulandığında Şekil 4a'da verilen değer haritası elde edilmiştir. Haritadaki siyah noktalar, değişen büyüklüklerine göre o konumdaki taşınmazın değerini, mavi alanlar bölgenin değerinin düşük olduğunu, kırmızı alanlar ise değerin yüksek olduğunu göstermektedir. Harita yorumlandığında, ilçenin kuzeybatı bölgesindeki Harmandere ve Yenişehir mahallelerinin, güneybatı bölgesinde ise Sapan Bağları, Yeni Mahalle ve Doğu mahallelerinin yüksek değerde; ilçenin güneydoğusundaki Kavakpınar, Ahmet Yesevi ve Fatih mahallelerinin ise düşük değerde olduğu gözlenmiştir.

RO algoritmasının üç veri seti ile oluşturduğu regresyon modellerinin performansları değerlendirildiğinde, KKO veri seti ile kurulan modelin performansının daha yüksek olduğu görülmüştür. RO tahmin modeline dayanarak bulunan taşınmaz değerleri, enterpolasyon analizi ile raster formata dönüştürülmüştür. Elde edilen taşınmaz değer haritası Şekil 4b'de yer almaktadır. Şekil incelendiğinde, ilçenin kuzeydoğu bölgesindeki Yenişehir mahallesinin bir kısmının, Sanayi mahallesinin ve ilçenin güneybatı bölgesindeki Yeni Mahalle ve Bahçelievler mahallelerinin yüksek değerli olduğu görülürken ilçenin güneydoğusundaki Ahmet Yesevi ve Fatih mahallelerinin değerinin düşük çıktığı görülmüştür.




(a) (b)

Şekil 4.Çalışma alanı için üretilen taşınmaz değer haritaları a) GLM algoritması ve K veri seti kullanılarak üretilmiş taşınmaz değer haritası ve b) RO algoritması ve KKO veri seti kullanılarak üretilmiş taşınmaz değer haritası

## 5. SONUÇLAR

Taşınmaz değerleme günümüzde çeşitli alanlarda kullanılan ve giderek önem kazanan bir uygulamadır. Günümüzün gelişen bilişim teknolojileri sayesinde büyük verilerin üretimi ve yönetimi yaygınlaşmıştır. Makine öğrenme algoritmaları matematiksel ve istatistiksel yöntemleri kullanarak büyük verilerin değerlendirilmesini sağlamaktadır. Bu çalışmada taşınmaz değerlemede makine öğrenme algoritmalarının uygulanabilirliği araştırılmıştır. Çalışma alanı olarak seçilen İstanbul-Pendik'te 1.458 taşınmaza ait konumsal ve konumsal olmayan özniteliklere dair katmanlar veri tabanı ve CBS analizleri kullanılarak üretilmiş, GLM ve RO algoritmaları ile taşınmaz değerleri tahmin edilmiştir.

GLM algoritması ile oluşturulan ve en iyi performansı gösteren K veri seti için MAE değeri 0,095; RMSE değeri 0,121; MdAE değeri 0,082 ve $R^2$ değeri 0,257 olarak hesaplanmıştır. Algoritmanın tahmin ettiği değerler ile gerçek satış değerleri arasında %45,2 korelasyon olduğu görülmektedir. RO algoritmasının KKO veri setini kullanarak oluşturduğu tahmin modeli incelendiğinde, tahmin edilen değerler ile gerçek satış değerleri baz alınarak, MAE değeri 0,093; RMSE değeri 0,118; MdAE değeri 0,078 ve $R^2$ değeri 0,219 olarak hesaplanmıştır. RO algoritmasının tahmin değerleri ile gerçek değerler arasındaki korelasyon ise %49,2 olarak hesaplanmıştır. Bu değerlere dayanarak, RO algoritmasının tahmin modelinin daha iyi performans verdiği sonucuna ulaşılabilir.

Tahmin modellerinin kurulmasında veri setlerinde bulunan özniteliklerin genel doğruluğa etkisi incelendiğinde, yalnızca konumsal olmayan verilerin kullanımının taşınmaz değerleri belirlenmesinde kurulan model için yeterli olmadığı görülmüştür. Konumsal özniteliklerin kullanımı, yalnızca konumsal olmayan özniteliklerin kullanımına oranla daha yüksek doğruluk göstermiştir. Dolayısıyla taşınmaz değerlemede konum kriterlerinin, tahmin modeli oluşturulmasında önemli derecede etkili olduğu söylenebilir. RO algoritması ile tahmin modeli kurulurken kullanıcının iki

tane parametre seçmesi gerekir. Parametrelerin seçimi modelin kurulmasında model doğruluğunu önemli ölçüde değiştirmemektedir. Bu durum RO algoritmasının önemli avantajlarından birisidir. Algoritmalar işlem süresi açısından karşılaştırıldığında GLM algoritmasının daha hızlı olduğu görülmektedir. RO algoritmasının performansı göz önünde bulundurulduğunda, işlem süresindeki fark büyük önem arz etmemektedir. RO algoritması ile üretilen taşınmaz değer haritası incelendiğinde Pendik'te yoğun yerleşim ve gelişimin düşük olduğu güneybatı ve kuzey kısımlarının yüksek değerli, güneybatı kısımlarının ise düşük değerli olduğu görülmektedir. Çalışma sonucunda RO algoritmasının taşınmaz değerleme probleminde uygulanabilir olduğu tespit edilmiştir. RO algoritması ile üretilen taşınmaz değer haritaları değer belirlemede referans olarak kullanılabilir.

## TEŞEKKÜR

Bu çalışma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 116Y204 nolu proje kapsamında desteklenmektedir.

## KAYNAKLAR

Arslan, R., 1997. Arazi Kullanış Ekonomisi, 1. baskı, *Yıldız Teknik Üniversitesi Basım-Yayın Merkezi Matbaası*, İstanbul.

Bogataj M., Tuljak Suban D., Drobne S., 2011. *Regression-Fuzzy Approach to Land Valuation. Central European Journal of Operations Research*, 19, 253-265.

Bovkır, R., Çölkesen, İ., Aydınoğlu, A.C., 2018. Calculating Land Values by Using Advanced Statistical Approaches in Pendik. *FIG Congress 2018*, 6–11 Mayıs, İstanbul, Türkiye.

Breiman, L., 2001. Random forests. Machine Learning, 45, 5-32, *Springer*.

de Micheaux, P. L., Drouilhet, R., & Liquet, B., 2013. The R software, *Springer*.

Demetriou D., 2016. The assessment of land valuation in land consolidation schemes: The need for a new land valuation framework. *Land Use Policy*, 54, 487-498.

Derinpınar, M.A., 2014. Bulanık Mantık ile Coğrafi Bilgi Teknolojilerini Kullanarak Taşınmaz Değerlemesi: Sarıyer-İstanbul Örneği, *Yüksek Lisans Tezi, İTÜ Bilişim Enstitüsü*

Dobson, Annette J., Barnett, Adrian G., 2008 An Introduction to Generalized Linear Models, Third Edition. *Texts in Statistical Science, 77. Chapman & Hall/CRC Press, Boca Raton, FL.*

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R., 2006. Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294-300.

Güneş T., Yıldız U., 2015. Mass Valuation Techniques Used in Land Registry and Cadastre Modernization Project of Republic of Turkey. *FIG Working Week 2015*.

Güngör, E., 1999. Gayrimenkul değerlemesi ve Türkiye'de sermaye piyasalarında gayrimenkul ekspertiz şirketlerine yönelik düzenlemeler yapılmasına ilişkin öneriler. *Yayımlanmamış SPK Yeterlik Etüdü. Sermaye Piyasası Kurulu,* Yatırımcılar Dairesi, Ankara.

Karapınar, A., Bayırlı, R., Bal, H., Altay, A., Bal, E. Ç., Torun, S., 2008. *SPK Lisanslama Sınavlarına Hazırlık – 7.Baskı,* Gazi Kitapevi, Ankara, Türkiye.

Kontrimas, V., Verikas, A. 2011. The mass appraisal of the real estate by computational intelligence, *Applied Soft Computing*, 11, 443-448.

Özdemir, S., 2018. Random Forest Yöntemi kullanılarak potansiyel dağılım modellemesi ve haritalaması: Yukarıgökdere Yöresi örneği, *Turkish Journal of Forestry* 19.1: 51-56.

Özmen A., Şıklar E., Durucasu H., Atlas M. ve Er F., 2013. İstatistik-II, *Anadolu Üniversitesi Web-Ofset Tesisleri*, Eskişehir.

Pagourtzi E. ve Assimakopoulos V., 2003. Real Estate Appraisal: A Review of Valuation Methods, *Journal of Property Investment & Finance,* 21, (4), 383-401.

Pal, M., 2005, Random Forest Classifier For Remote Sensing Classification, *International Journal Of Remote Sensing*, 26(1), 217-222.

Resmî Gazete, 2001. 35 nolu, Sermaye Piyasası Mevzuatı Çerçevesinde Gayrimenkul Değerleme Hizmeti Verecek Şirketler ile Bu Şirketlerin Kurulca Listeye Alınmalarına İlişkin Esaslar Hakkında Tebliğ, Sayı: 24491.

Watts, J. D., Powell, S.L., Lawrence, R. L., Hilker, T., 2011, Improved Classification of Conservation Tillage Adoption Using High Temporal And Synthetic Satellite Imagery, *Remote Sensing of Environment* 115 (2011) 66–75

URL 1. TÜİK Adrese Dayalı Nüfus Sistemi, www.tuik.gov.tr (Erişim tarihi: 02/04/2019)