

Toplu taşınmaz değerlendirilmede makine öğrenme algoritmalarının kullanımı ve konumsal/konumsal olmayan özneliklerin tahmin doğruluğuna etkilerinin karşılaştırılması

Arif Çağdaş Aydınoglu^{1*}, Rabia Bovkır¹, İsmail Çölkesen¹

¹Gebze Teknik Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, Gebze, Kocaeli, Türkiye.

Öz: Taşınmaz değerinin güncel teknolojik yaklaşımlarla objektif olarak belirlenmesi, etkin ve sürdürülebilir taşınmaz yönetim planlaması kapsamında önemli bir role sahiptir. Toplu taşınmaz değerlendirilme, taşınmazların tekel olarak değerlendirilmesi yerine zaman ve maliyet anlamında kayıplarını azaltarak çok sayıda taşınmazı eş zamanlı olarak değerlendirilerek değer tespitlerinin yapılması işlemidir. Toplu taşınmaz değerlendirilmede objektif, doğru ve hızlı sonuçlar elde için gelişmiş tahmin yaklaşımları olarak bilinen makine öğrenme tekniklerinin kullanımına başvurulmaktadır. Bu yöntemlerle geliştirilen modeller hem yüksek performans hem de değerlendirilme sürecinde objektiflik sağlamaktadır. Ayrıca bu yöntemler ile değeri etkileyen kriterlerin ve kriterlerin önem düzeylerinin bütüncül biçimde değerlendirilmesi mümkün olmaktadır. Bu çalışmada Çoklu Doğrusal Regresyon (ÇDR), Genelleştirilmiş Doğrusal Model (GDM), Destek Vektör Makineleri (DVM), Karar Ağaçları (KA) ve Rastgele Orman (RO) algoritmaları kullanılarak toplu taşınmaz değerlendirilmedeki model başarıları irdelenmiştir. Veri setleri Konumsal (K), Konumsal Olmayan (KO) ve Konumsal ve Konumsal Olmayan (KKO) olarak 3 gruba ayrılarak modelleme için ayrı ayrı uygulanmıştır. Model performans ölçütleri olarak Ortalama Mutlak Hata (Mean Absolute Error, MAE), Karesel Ortalama Hata (Root Mean Squared Error, RMSE), Ortalama Karesel Hata (Mean Squared Error, MSE) ve R^2 hesaplanmıştır. Uygulama alanı olarak seçilen İstanbul ili Pendik ilçesi için örnek taşınmaz satış değerlerini temsil eden 1475 örnekleme noktası kullanılarak, 3 farklı veri seti (K, KO, KKO) ve farklı metotlar ile kurulan modellerin performansları irdelenmiştir. Buna göre en yüksek doğrulukta yöntemin RO, en düşük doğrulukta yöntemlerin ise KA ve GDM olduğu görülmüştür. Farklı veri setlerinin model doğruluğuna etkileri irdelendiğinde tek başına K veri seti ile KKO veri setinin kullanımı arasında büyük bir fark görülmemiştir.

Anahtar Sözcükler: Toplu taşınmaz değerlendirilme, Makine öğrenme algoritmaları, Kriter analizi, Model tahmin doğruluğu

Using machine learning algorithms in mass valuation and comparing the effects of geographical/non-geographical features on prediction accuracy

Abstract: The objective determination of real estate values with current technological approaches has an important role in effective and sustainable real estate management plans. Mass appraisal is the process of valuing a large number of real estate simultaneously instead of evaluating the real estate individually for reducing the loss in terms of time and cost. Machine learning methods, known as advanced estimation approaches, are used to obtain more objective, accurate and fast results in mass valuation processes. In addition, besides sufficient objectivity and accuracy in value determination, these methods can evaluate the relations between the value and the criteria affecting the value holistically. In this context, model successes in mass valuation were examined using Multiple Linear Regression (MLR), Generalized Linear Model (GLM), Support Vector Machines (SVM), Decision Trees (DT) and Random Forest (RF) algorithms. The datasets were divided into 3 groups as Geographic (G), Non-Geographic (NG) and Geographic + Non-Geographic (GNG) and applied separately for modeling with different methods. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE) and R^2 were calculated for determining the model measures. Pendik district of Istanbul province was chosen as the application area. By applying different methods with 1475 sampling points representing the real estate sales values for the application area, the performances of the models established were examined with 3 different data sets (G, NG, GNG). Accordingly, RF is the method that gives the highest accuracy while DT and GLM were found as the methods with the lowest accuracy. When the effects of different datasets on the model accuracy were examined, big difference was not observed between the use of the G and the GNG datasets.

Keywords: Mass valuation, Machine learning algorithms, Criteria analysis, Model prediction accuracy

* Sorumlu Yazar/Corresponding Author: Tel: +90 262 605 2061

Geliş Tarihi/Received: 08.04.2022

Kabul Tarihi/Accepted: 24.06.2022



0000-0003-4912-9027, aydinoglu@gtu.edu.tr (Aydınoglu A.Ç.)*
0000-0002-9527-1350, rbovkir@gtu.edu.tr (Bovkır R.)
0000-0001-9670-3023, icolkesen@gtu.edu.tr (Çölkesen İ.)

1. Giriş

Sürdürülebilir kalkınma ve gelişimin sağlanmasında, akıllı şehirleşme faaliyetleri ile taşınmaz ve ilişkili tüm bilginin etkin şekilde yönetimi oldukça önemlidir. Taşınmaz mal değerlemesi, akıllı şehirler kapsamında planlı şehirleşmenin yürütülmesi, yerleşme bölgelerinin seçimi, bu bölgeler arasında bağlantı giderlerinin karşılaştırılması, imar planlarının ekonomik olarak değerlendirilebilmesi, yapı iyileştirilmesi, arsa ve arazi düzenlemesi açısından büyük önem taşımaktadır (Karapınar vd., 2008). Taşınmaz değeri için standartlaşmış bir yaklaşım bulunmama ile birlikte, kullanma amacı, piyasa koşulları, faiz ve beklenen yarar gibi faktörlerle değerlendirilerek göreceli bir yaklaşım uygulanabilir. 35 nolu Sermaye Piyasası Mevzuatı'na göre taşınmaz değerlendirme 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. Vergilendirme, kamulaştırma, özelleştirme, devletleştirme, sigortacılık ve bankacılık gibi uygulamalarda aktif olarak kullanılmakta olan taşınmaz değerlendirme, ülkemizde ve dünyada son zamanlarda ciddi bir önem arz etmeye başlamıştır (Bovkır & Aydınoglu, 2018; Demetriou, 2016; Güneş & Yıldız, 2015, 2016; Yalpir, 2007).

Taşınmaz değerinin saptanması için matematiksel olarak objektif ve standartlaşmış bir teknik olmadığından değer; kullanma amacı, piyasa koşulları, faiz ve beklenen yarar gibi faktörlerle değerlendirilerek oluşturulur (Aydınoglu, Çölkesen, Şenbil, Bovkır, & Yomrahoğlu, 2020; Güngör, 1999; Karapınar vd., 2008; Şişman, 2021; Yalpir, 2007). Söz konusu nedenlerden ötürü değer göreceli bir kavramdır. Değerlemeye konu toprak ise üstündeki taşınmaz (bina) ve toprağın konumu ile değişen imar koşulları; eğer değerlendirme konusu bina ise kurulu olduğu toprağın durumu ile binanın özellikleri bu değerde rol oynamaktadır (Arslan, 1997). Taşınmazların değerlendirilmesi için birçok yöntem kullanılsa da karşılaştırma (emsal), gelir ve maliyet yöntemleri en çok bilinen ve kullanılan yöntemlerdir. Bu yöntemlerin dışında stokastik yöntemler olarak adlandırılan ve genel anlamda olasılık ve istatistiksel yöntemlere dayanan nominal değerlendirme yöntemi, yapay zekâ ve makine öğrenme tekniklerine dayanan bulanık mantık, yapay sinir ağları, karar ağaçları (KA) ve rastgele orman (RO) gibi gelişmiş tahmin yöntemleri de kullanılabilmektedir (Aydınoglu vd., 2020; Candaş, 2012; Şişman, 2021; Ünel & Yalpir, 2019; Yalpir, 2007).

Toplu değerlendirme, birden fazla taşınmazın grup halinde eş zamanlı olarak istatistiksel yöntemler ve testler kullanılarak değerlendirme süreci olarak tanımlanabilir (IAAO, 2013). Taşınmazların toplu olarak hızlı, objektif ve doğru olarak değerlendirilmesi, kentsel dönüşüm, vergilendirme, kamulaştırma, imar uygulamaları, kredilendirme, sigortacılık ve arazi kullanım planlama gibi birçok uygulama için oldukça önem arz etmektedir (Aydınoglu vd., 2020; Dawidowicz & Żróbek, 2017). Bu nedenle toplu taşınmaz değerlemede ileri istatistiksel yöntemler ve makine öğrenmesi yöntemleri kullanılarak taşınmazlara ait tüm bilgiler, daha doğru ve şeffaf sonuçlar elde etmek için bütünlük olarak hızlı bir şekilde analiz edilip yorumlanabilir (Güneş & Yıldız, 2015; Hong, Choi, & Kim, 2020). Toplu taşınmaz değerlendirme konusunun öneminin artması ile Tapu ve Kadastro Genel Müdürlüğü (TKGM), Tapu ve Kadastro Modernizasyonu Projesi kapsamında Dünya Bankası'nın (World Bank) desteği ile pilot projeler başlatmıştır (Güneş & Yıldız, 2015, 2016). 2018 yılında TKGM bünyesinde Taşınmaz Değerleme Dairesi Başkanlığı, taşınmaz değerlendirme işinin ülke genelinde standartlaştırılması ve ulusal değer veri tabanı oluşturulması kapsamında "Toplu Değerleme Sistemi" kurularak değerlemeye ait bilgi ve işlemleri tek bir merkezden yönetmeyi amaçlamıştır (Resmi Gazete, 2018).

Değerleme sürecinde taşınmazın fiziksel ve yasal özelliklerinin yanı sıra, bulunduğu bölgenin eğimi ve bakışı gibi topoğrafik özellikleri, cazibe merkezlerine ve kamu hizmeti alınabilecek merkezlere olan mesafeleri, taşınmazın bulunduğu çevredeki nüfus yoğunluğu ve mevcut sosyo-kültürel yapılar bir arada analiz edilip değerlendirilerek değer tespitinin yapılması büyük

önem arz etmektedir (Kara, van Oosterom, Çağdaş, Işıkdag, & Lemmen, 2020; Şişman, 2021; The Appraisal Institute, 2001; Yazıcı, 1997). Bu kapsamda ele alındığında taşınmaz değerlendirme işleminin başarısı, söz konusu faktörlerin objektif bir şekilde değerlendirilmesi ile doğrudan ilişkilidir. Bu amaca yönelik olarak geleneksel değerlendirme yaklaşımları (örneğin, emsal yöntemi, gelir yöntemi ve maliyet yöntemi) yerine son yıllarda gelişmiş tahmin yaklaşımları olarak bilinen ve makine öğrenmesi esasına dayanan destek vektör makineleri (DVM), KA, hibrit sınıflandırma modelleri (örneğin, fonksiyonel model ağaçlar ve lojistik model ağaçlar) ve toplu öğrenme algoritmaları (örneğin, RO ve rotasyon orman) kullanılmaktadır (Aydinođlu vd., 2020; Ćeh, Kilibarda, Lisec, & Bajat, 2018; Dimopoulos & Moulas, 2016; Manasa, Gupta, & Narahari, 2020; Mayer, Bourassa, Hoesli, & Scognamiglio, 2019; Wang X., Wen, Zhang, & Wang Y., 2014; Yalprı, 2018). Bu algoritmaların genel çalışma prensibi, bilinen örnek veriler ile karmaşık veri setlerinin modellenerek tahmin modeli oluşturularak tahminlerin yapılmasıdır. Özellikle toplu taşınmaz değerlemesinde birden fazla taşınmazın grup halinde eş zamanlı olarak irdelenmesi için ileri istatistiksel ve makine öğrenmesi yöntemleri kullanarak taşınmazlara ait tüm bilgiler daha doğru ve şeffaf sonuçlarla elde edilebilir.

Bu çalışmada, toplu taşınmaz değerlemede farklı makine öğrenme algoritmalarının kullanımı ve tahmin doğruluklarının incelenerek uygulamalara entegre edilebilir en başarılı yöntemin belirlenmesi amaçlanmaktadır. Ayrıca taşınmaz değer tespitinde konumsal ve konumsal olmayan özniteliklerin tahmin doğruluğuna olan etkileri de analiz edilecektir. Bu kapsamda ikinci bölümde çalışmada kullanılacak taşınmaz değer kriterleri ve makine öğrenme algoritmaları açıklanmıştır. Üçüncü bölümde çalışma alanı ve veri setleri, performans ve doğruluk ölçütleri açıklanarak 5 farklı makine öğrenmesi algoritma modeline ait hata ölçütleri listelenmiştir. Son bölümde ise bulgular ve değerlendirmelere yer verilmiştir.

2. Yöntem

2.1 Taşınmaz Kriterlerinin Belirlenmesi

Başarılı bir toplu taşınmaz değerlendirme işlemi için değere etki eden tüm yasal ve tematik faktörlerin belirlenmesi ve değerlendirme sürecine dâhil edilmesi gerekmektedir. Bir taşınmazın değeri yerel veya ülke düzeyinde yasal faktörlerden, fiziksel faktörlerden, ekonomik ve sosyal faktörlerden etkilenebilmektedir (Candaş, 2012). Bu kapsamda taşınmaz değerine etki eden en önemli ve uygulamalarda en sık kullanılan faktörleri belirlemek amacıyla konu ile ilgili yayınlanmış ulusal ve uluslararası standart dokümanları olan Avrupa Değerleme Standartları (TEGoVA, 2016), Uluslararası Değerleme Standartları (IVSC, 2020), Taşınmaz Değerleme – Evrensel Standartları (RICS, 2017), Toplu Değerleme Standardı (IAAO, 2013), Türkiye Değerleme Standardı (TDUB, 2017) ve konu ile ilgili mevzuat (SPK, 2006; TKGM, 2011) ile birlikte yayınlanmış akademik çalışmalar ve projeler incelenmiştir (Antipov & Pokryshevskaya, 2012; Candaş, 2012; Chen, Ong, Zheng, & Hsu, 2017; Dimopoulos & Bakas, 2019; Ünel & Yalprı, 2019; Yalprı & Bünyan Ünel, 2016; Yılmaz & Kocaman, 2020; Yomralođlu, 1993). Literatür araştırmasına göre erişilebilir veriler dikkate alınarak belirlenen kriterler Tablo 1’de gruplanarak listelenmiştir.

2.2 İleri İstatistik ve Makine Öğrenme Algoritmaları ile Taşınmaz Değerlemesi

Toplu taşınmaz değerlendirme kapsamında literatür incelendiğinde değerlendirme birçok istatistik tabanlı yöntemler ve makine öğrenmesi tekniğinin kullanıldığı görülmektedir. Bunlar arasında, Çoklu Doğrusal Regresyon (ÇDR) analizleri (Božić, Milićević, Pejić, & Marošan, 2013; Güneş & Yıldız, 2015; Şişman, Akar, & Yalprı, 2021; Yalprı & Özkan, 2008; Yalprı, Durduran, Bünyan Ünel, & Yolcu, 2014; Yalprı, Şişman, Akar, & Bünyan Ünel, 2021), DVM (Bulut Nas, 2011; Kontrimas & Verikas, 2011), KA (Reyes-Bueno, García-Samaniego, & Sánchez-Rodríguez, 2018) ve RO (Aydinođlu, Bovkır, & Çölkesen, 2021; Lasota, Łuczak, Niemczyk, Olszewski, & Trawiński, 2013, Şişman, 2021) algoritmalarının regresyon analizlerinde sıklıkla kullanıldığı görülmektedir. Bu kapsamda çalışmada ÇDR, Genelleştirilmiş Doğrusal Model (GDM),

DVM, KA ve RO algoritmaları kullanılarak toplu taşınmaz değerlemedeki model başarıları irdelenmiştir.

Tablo 1: Taşınmaz değer kriterleri

Parsel Özellikleri	Çevresel Özellikler	Bağımsız Bölüm Özellikleri
• Alan	• Kamu Hizmetlerine (eğitim, sağlık, resmi ve idari tesisler vb.) Mesafeler	• Yaşam Alanı
• Pozisyon (Köşe, Orta, Ön)	• İlgi Noktalarına (AVM, çarşı, kültür ve spor tesisleri vb.) Mesafeler	• Yapım Yılı
• İmar Durumu	• Zararlı Alanlara (sanayi, arıtma, atık, akaryakıt tesisleri vb.) Mesafeler	• Kat Sayısı
• Planlama	• Yeşil Alanlara Mesafeler	• Oda Sayısı
• Yasal Hak/ Kısıtlama/ Sorumluluklar	• Ulaşım/Toplu Taşıma İmkânları	• Isıtma Sistemi
• Mülkiyet Durumu	• Sosyo-Ekonomik (nüfus yoğunluğu, eğitim durumu, gelir vb.) Özellikler	• Kullanım Türü
• Topografik Özellikler	• Gürültü Durumu	• Yapı Türü/Malzemesi
• Vergi		• Mevcut Araçlar (balkon, asansör, garaj, havuz vb.)
		• Manzara

2.2.1 Çoklu Doğrusal Regresyon

Bir hedef değişkenle o değişkeni tanımlayan diğer değişkenler arasındaki ilişkinin modellenmesi, tüm istatistiksel teknikler içerisinde kullanılan en yaygın tekniktir. Regresyon analizi genel olarak, bağımsız değişken veya değişkenler (x) ile bağımlı değişken (y) arasındaki ilişkinin matematiksel bir fonksiyonu şeklinde tanımlanmaktadır. Regresyon modeli, kullanıcıya bağımlı değişken ve bu değişkeni açıklayan bağımsız değişkenlerden hangisinin hedef değişken üzerinde bir etkiye sahip olduğunu belirlemesine izin veren fonksiyonel bir ilişki kurmaktadır (Ott & Longnecker, 2010). Diğer bir ifadeyle, regresyon analizine konu olan bağımsız değişkenlerin bağımlı değişken ile birlikte ele alındığı model olarak ifade edilebilir. ÇDR analizi iki ya da daha fazla sayıdaki bağımsız değişken ile bağımlı değişkenin varyansının açıklanmasını esas alan istatistiksel bir yöntemdir. Matematiksel olarak ÇDR denklemi aşağıda verilen Eşitlik 1 ile ifade edilebilir.

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

Bu eşitlikte, n bağımsız değişken sayısını, y bağımlı değişkeni, β_0 sabit değeri, x bağımsız değişkeni, β_i bağımsız değişkenlerin regresyon katsayılarını ve ε regresyon modelindeki hata varyansını ifade etmektedir.

2.2.2 Genelleştirilmiş Doğrusal Modeller

GDM klasik doğrusal regresyon modelinin gelişmiş bir hali olarak ifade edilebilir. Klasik doğrusal regresyon modeli, bağımlı değişkenin bağımsız değişkenlerin doğrusal bir fonksiyonu olduğunu ve bağımlı değişkenin sürekli ve sabit varyansa sahip normal dağılımda olduğu kabulüne dayanmaktadır (Dobson & Barnett, 2008). Ancak uygulamaların birçoğunda bağımlı değişkenler normal dağılımda olmayıp, varyansları da değişkenlik gösterebilmektedir. Bu problemin üstesinden gelebilmek için Nelder ve Wedderburn (1972), bağımlı değişkenlerin normal dağılımda olmadığı regresyon modelleri için GDM'yi ortaya atmışlardır. GDM, lojistik ve poisson regresyon modelleri kullanıldığında, olasılık dağılımı ile ortalama ve varyans arasındaki ilişkiyi ve regresyon parametrelerinin tahmin edilmesini esas almaktadır (Dobson & Barnett, 2008). GDM, bir 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 imkân sağlamaktadır. İlişki fonksiyonu ve hata dağılımı, GDM ile regresyon modeli oluşumunda iki temel bileşendir. İlişki fonksiyonu, bağımlı

değişkenin ortalamasını regresyon parametrelerinin doğrusal bir fonksiyonu olacak şekilde dönüşümüne imkân sağlamaktadır. Diğer taraftan, hata dağılımı bağımlı değişkenin varyansının kendi ortalamasının bir fonksiyonu olmasını sağlar.

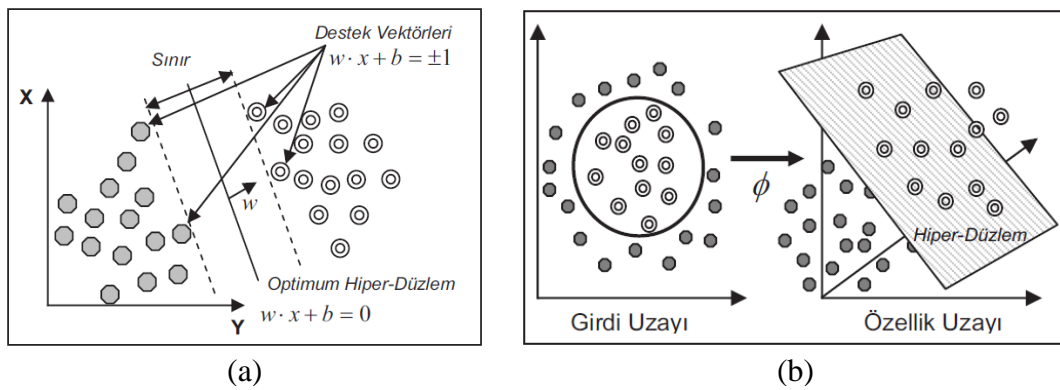
2.2.3 Destek Vektör Makineleri

DVM algoritması, istatistiksel öğrenme teorisine dayalı bir kontrollü sınıflandırma ve regresyon algoritmasıdır (Boser, Guyon, & Vapnik, 1992; Cortes & Vapnik, 1995; Vapnik, 1982). DVM'nin sahip olduğu matematiksel algoritmalar başlangıçta iki sınıflı doğrusal verilerin sınıflandırılması problemi için tasarlanmış daha sonra kullanılan algoritmalar doğrusal olmayan verilerin ve çok sınıflı verilerin sınıflandırılması problemlerinin çözümü için genelleştirilmiştir. Veri setindeki benzer özellikteki örnekler sonsuz sayıda düzlem ile birbirinden ayrılabilir. DVM'deki amaç ise Şekil 1a'da gösterildiği gibi veri kümelerini ayıran düzlemlerden Eşitlik 2'deki formülü kullanarak maksimum sınıra sahip olan hiper düzlemi bulmaktır.

$$f(x) = \langle w, x \rangle + b \quad (2)$$

Eşitlik 2'de $x \in R^N$ olup N -boyutlu bir uzayı, w ağırlık vektörünü (hiper düzlemin normali) ve b eğilim değerini göstermektedir (Kavzoğlu & Çölkesen, 2010; Osuna, Freund, & Girosi, 1997). Veri setinin en uç kısmında yer alan ve hiper düzlemi belirleyen örnekler destek vektörleri olarak adlandırılmaktadır. Destek vektörlerinin belirlenmesi için karar sınırlarının oluşturulması yeterli olduğundan, veri setindeki örnek sayısının az olması durumunda da DVM algoritması başarılı sonuçlar verebilmektedir.

Günlük hayatta kullanılan veriler genellikle doğrusal olmadığı için bu tip verilerde DVM sınıflandırıcısının kullanılabilmesi için kernel (çekirdek) fonksiyonlarına ihtiyaç duyulmaktadır. Doğrusal özellik göstermeyen veriler kernel fonksiyonları kullanılarak yüksek boyutlu özellik uzayına dönüştürülür (Şekil 1b). Literatürde kernel fonksiyonu olarak doğrusal (linear), polinom (polynomial), radyal tabanlı (RBF) ve sigmoid kernelleri yaygın şekilde uygulanmaktadır (Kavzoğlu & Çölkesen, 2010).

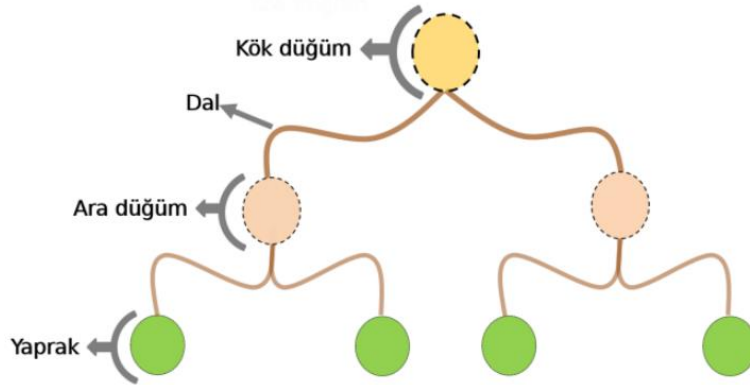


Şekil 1: Destek vektörleri algoritmasının temel çalışma prensibi (a) doğrusal olarak ayrılabilen veri setleri için hiper-düzlemin belirlenmesi ve (b) doğrusal olmayan veri setleri için kernel fonksiyonu ile hiper-düzlemin belirlenmesi (Kavzoğlu & Çölkesen, 2010)

2.2.4 Karar Ağaçları

KA, yaygın kullanımı olan sınıflandırma ve regresyon algoritmalarından birisidir. Bu yöntemin yaygın olarak kullanımının nedeni, ağaç yapılarının oluşturulmasında kullanılan kuralların anlaşılabilir ve sade olmasıdır. Bir ağaç yapısı düğüm, dal ve yaprak olarak adlandırılan üç temel kısımdan oluşur (Şekil 2). Bu ağaç yapısında her bir öznelik bir düğüm tarafından temsil edilir. Dallar ve yapraklar ağaç yapısının diğer elemanlarıdır. Ağaçta en son kısım yaprak, en üst kısım kök ve bu kısımlar arasında kalan kısımlar ise dal olarak ifade edilir (Quinlan & Cameron-Jones, 1995). Karar ağacındaki her bir düğüm sadece

bir ana düğüme ve iki veya daha fazla alt düğüme sahiptir. Karar ağaçları ile sınıflandırma problemlerinde veri setine ait sınıf sayısı iki olduğunda ikili sınıflandırıcılar, sınıf sayısı ikiden fazla olması durumunda ise çoklu sınıflandırıcılar söz konusudur. Çoklu sınıflandırıcılar arasında CART (Classification and Regression Tree) (Breiman, Friedman, Olshen, & Stone, 1984; Breiman, 1996, Çölkesen & Kavzoğlu, 2017) algoritması örnek olarak verilebilir ve bu algoritma hem sınıflandırma hem de regresyon amaçlı kullanılabilir.



Şekil 2: Karar ağacı algoritmasının temel yapısı (Çölkesen & Kavzoğlu, 2017)

2.2.5 Rastgele Orman

RO algoritması, temel sınıflandırıcı olarak eğitim aşamasında birden çok karar ağacını kullanmakta ve bu yapıyla karar ormanı olarak tanımlanabilmektedir (Breiman, 2001). Temel çalışma prensibi Şekil 3'te ifade edilen RO algoritması, ormandaki her bir karar ağacının eğitimi için orijinal eğitim veri seti içerisinde rastgele alt kümeler oluşturur. Söz konusu alt kümelerin 2/3'ü karar ağacı yapısını oluşturmak için kullanılırken, diğer kısmı *out-of-bag* (OOB) olarak bilinen veri setidir ve ağaç yapısının tahmin geçerliliğini test etmek için kullanılır. Ormandaki her bir karar ağacı sınıflandırma sonucu bir oy alır ve orman içerisindeki tüm ağaçlardan en çok oy alan (en düşük hata oranına sahip) belirlenerek sınıflandırmaya esas ağaç yapısı tespit edilir. Sınıf etiketi bilinmeyen herhangi bir örnek, tüm ağaç tahminlerinde en fazla oy alan sınıfa atanması suretiyle sınıflandırılır.

RO algoritması, karar ağacı oluşumunda dallanma için öznelitliklerin seçiminde Gini indeksi yöntemini kullanır. RO algoritması ile toplu öğrenme modeli oluşumunda kullanıcı tarafından belirlenen iki temel parametre mevcuttur. Bunlar her bir düğüme kullanılacak örneklerin sayısı (t) ve oluşturulacak ağaç sayısıdır (n). RO algoritması, her bir düğümden en iyi dallanma için sadece seçilen özellikleri dikkate alır. RO sınıflandırıcısı bu n sayıda ağaçtan elde edilen n sayıda oy arasından en fazla oya sahip olan sınıfı seçer ve bu şekilde yeni veri setine ait sınıf etiketini belirler (Kavzoğlu & Çölkesen, 2013; Pal, 2005).

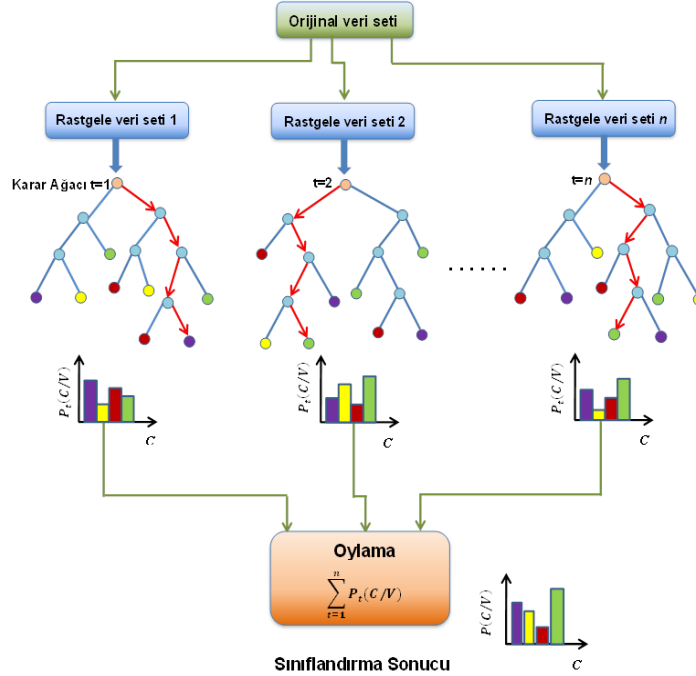
2.3 Korelasyon Katsayısı

Korelasyon katsayısı ($CorC$), örnekler için bilinen taşınmaz değerleri ile tahmin edilen taşınmaz değerleri arasındaki istatistiksel korelasyonu ölçer, diğer bir ifadeyle değerler arasında doğrusal bir ilişkinin olup olmadığının bir göstergesidir. $CorC$, Eşitlik 3 yardımı ile hesaplanmaktadır.

$$CorC = \frac{\frac{\sum_{i=1}^n (P_i - \bar{P})(A_i - \bar{A})}{n-1}}{\sqrt{\left(\frac{\sum_{i=1}^n (P_i - \bar{P})^2}{n-1}\right) \left(\frac{\sum_{i=1}^n (A_i - \bar{A})^2}{n-1}\right)}} \quad (3)$$

Eşitlik 3'te, A_i örnek taşınmaza ait bilinen değeri, \bar{A} bilinen değerlerin ortalamasını, P_i örnek taşınmaza ilişkin tahmin edilen

deđeri, \bar{P} tahmin edilen deđerlerin ortalamasını ve n örnek sayısını ifade etmektedir. Hesaplanan deđer tam bir korelasyon olduđunda yani dođrusal bir iliřki olduđunda 1 deđerini alırken, korelasyon olmadıđında ise 0 deđerini almaktadır. Katsayı deđerinin -1 ile 0 arasında deđer alması, model tahmin deđerleri ile gerçek satış deđerleri arasında negatif yönde iliřki olduđunu göstermektedir. Bu ölçüt taşınmazların deđer tahmininin mahalle bazında deđerlendirilmesi aşamasında kullanılmıştır.



Şekil 3: RO algoritmasının temel çalışma prensibi (Çölkese 2015)

2.4 Performans Analizi ve Doğruluk Ölçütleri

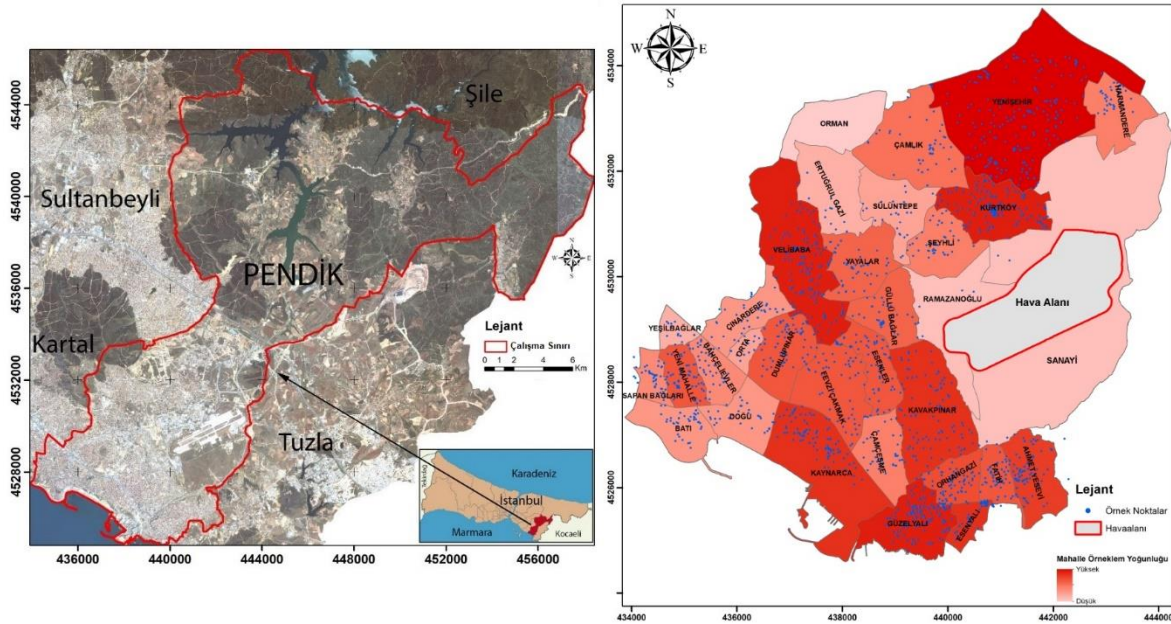
Çalışma kapsamında toplu taşınmaz deđerlerinin tespitinde kullanılan tahmin algoritmalarının performans analizi ve tahmin dođruluđunun deđerlendirilmesinde Tablo 2’de temel eşitlikleri verilen ortalama mutlak hata (Mean Absolute Error, MAE), karesel ortalama hata (Root Mean Squared Error, RMSE), ortalama karesel hata (Mean Squared Error, MSE) ve R^2 ölçütleri deđerlendirmeye alınmıştır. Tablodaki eşitliklerde n örnek sayısını, y tahmin edilen deđer ve \hat{y} bilinen deđer ifade etmektedir. MAE, bilinen taşınmaz deđerleri ile tahmin edilen deđerlerin farklarının dikkate alındıđı bir hata metriđidir. MAE’nin sifıra yakın deđer alması tahmin edilen ve gerçek deđerlerin birbirine çok yakın olduđunu, yani regresyon modelinin başarılı olduđunun bir göstergesidir. RMSE ölçütü ise tahmin ve bilinen deđerler arasındaki farkların karelerinin ortalamasının karekökü alınarak hesaplanmaktadır. Diđer bir ifadeyle, RMSE tahmin hatalarının standart sapmasıdır. RMSE hatasının sifıra yakın çıkması modelin tahmin gücünün yüksek olduđunu göstermektedir. MSE, regresyon eđrisinin bir dizi noktaya ne kadar yakın olduđunu belirten bu dođruluk ölçütü olup, algoritmanın performansını ölçer ve her zaman pozitif deđer alır. Aldıđı deđerin sifıra yakın olması algoritmanın iyi performans sergilediđinin bir göstergesidir. R^2 , açıklanabilen varyasyonun toplam varyasyona oranı řeklinde ifade edilen bir ölçüt olup, 0 ile 1 arasında deđer almaktadır. Hesaplanan deđerin yüksek olması regresyon modelinin performansının iyi olduđu anlamına gelmektedir. Veri sayısı arttıkça R^2 ’nin güvenilirliđi artmaktadır.

Tablo 2: Performans analizinde kullanılan ölçütler ve temel eşitlikleri

Doğruluk Ölçütü	Temel Eşitlik
Ortalama Mutlak Hata (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
Karesel Ortalama Hata (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
Ortalama Karesel Hata (MSE)	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
R ² Ölçütü	$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

3. Uygulama: Çalışma Alanı ve Veri Setleri

Çalışma alanı olarak İstanbul ilinin Pendik ilçesi seçilmiştir. Pendik ilçesi doğuda Tuzla, batıda Kartal ve Sultanbeyli, kuzeyde Şile, güneyde Marmara Denizi'ne komşudur (Şekil 4). 7.5 km'lik kıyı şeridine ve 190 km² yüzölçümüne sahiptir. Bölgenin geniş yüzölçümü ve kentsel-kırsal farklılıklara sahip olması uygulama alanı olarak seçilmesinin en büyük nedenidir. Çalışma kapsamında ilçe sınırları içerisindeki 36 mahalle kontrol noktası olarak kullanılabilir taşınmazlar açısından analiz edilmiştir. Çalışma alanında E-80 otoyolunun kuzeyinde bazı mahallelerde kullanılabilir uygun taşınmaz bulunmadığından çalışma alanı kapsamına alınmamıştır (Şekil 4).



Şekil 4: Çalışma alanı (sol) ve örneklem dağılımı (sağ)

Çalışmada toplu taşınmaz değerlemede farklı algoritma doğruluklarının incelenmesinin yanında, değer tespitinde konumsal ve konumsal olmayan özniteliklerin tahmin doğruluğuna etkilerinin incelenmesi amacıyla 3 farklı veri seti oluşturulmuştur. 116Y204 nolu TÜBİTAK araştırma projesi kapsamında Tablo 3'teki CBS analizleriyle oluşturulan 21 öznitelikten oluşan konumsal (K), 13 öznitelikten oluşan konumsal olmayan (KO) ve bütün özniteliklerin kullanıldığı ve toplamda 34

öznitelikten oluşan konumsal ve konumsal olmayan (KKO) öznitelik veri setleri modelleme için kullanılmıştır. Konumsal öznitelikler, kriterleri temsil eden ve ulaşım, sosyo-kültürel, kamu hizmetleri ve kullanım koşullarını temsil eden coğrafi veri altlıklarının coğrafi analizi sonucu elde edilmiştir:

- **İlgi noktalarının (POI) oluşturulması:** Çokgen (polygon) geometride tanımlanan eğitim, sağlık, kültürel, dini, idari, alışveriş, altyapı tesisleri; otopark, sanayi, mezarlık, yeşil alanlar ve nüfus yoğunluğu ve eğitim düzeyi verilerinin nokta (point) geometrisindeki ilgi noktalarına (POI) dönüştürülmesi ile oluşturulur.
- **Tematik yüzeylerin elde edilmesi:** Tematik yüzeylerin oluşturulması için Sayısal Arazi Modeli (SAM) kullanılarak eğim ve bakı analizleri; nüfus ve eğitim düzeyini temsil eden yüzeylerin oluşturulabilmesi için yoğunluk analizi ve diğer tüm tematik yüzeylerin oluşturulabilmesi için de Öklid Mesafe analizi (Euclidean Distance) kullanılmıştır.

Konumsal olmayan öznitelik bilgileri, Pendik ilçe genelinde satışta olan ve güncel satış değerleri mevcut konutlardan elimine edilerek oluşturulan 1475 örnekleme noktası için mevcuttur. Tespit edilen örnek noktalar, Şekil 4’te görüleceği üzere ilçe genelinde tahmin modelinin beklenen doğruluğunun artırılması amacıyla tüm çalışma alanına dağılacak şekilde tespit edilmiştir. 1475 adet örnekten 1032 tanesi tekrarlama olmaksızın rastgele seçilerek eğitim verisi olarak kullanılmış ve 3 ayrı tahmin modeli oluşturulmuştur. Geriye kalan 443 tane örnek ise oluşturulan tahmin modellerinin performanslarının değerlendirilmesi amacıyla doğrulama verisi olarak kullanılmıştır. Veri setlerinin içerdikleri öznitelikler Tablo 3’te yer almaktadır.

Tablo 3: Taşınmaz değerlendirme modellerinde kullanılan özellikler

K	KO	KKO	Özellik	K	KO	KKO	Özellik
○	●	●	Cephe Yönü	●	○	●	Dini Tesislere Uzaklık
○	●	●	Isıtma	●	○	●	Eğitim
○	●	●	Otopark Alanı	●	○	●	Eğitim Durumu
○	●	●	Manzara	●	○	●	Havaalanına Uzaklık
○	●	●	Site İçi	●	○	●	İdari Tesislere Uzaklık
○	●	●	Toplam Kat	●	○	●	Kültürel Tesislere Uzaklık
○	●	●	Bulunduğu Kat	●	○	●	Metroya Uzaklık
○	●	●	Banyo Sayısı	●	○	●	Mezarlığa Uzaklık
○	●	●	Yaş	●	○	●	Nüfus Yoğunluğu
○	●	●	Oda Sayısı	●	○	●	Okuma Yazma Oranı
○	●	●	Durum	●	○	●	Otobüs Durağına Uzaklık
○	●	●	Asansör	●	○	●	Otoparka Uzaklık
○	●	●	Taşınmaz Alanı	●	○	●	Otoyola Uzaklık
●	○	●	Alışveriş Merkezine Uzaklık	●	○	●	Sağlık Tesislerine Uzaklık
●	○	●	Altyapı Tesislerine Uzaklık	●	○	●	Sanayiye Uzaklık
●	○	●	Bakı	●	○	●	Üniversite Mezunu Yoğunluğu
●	○	●	Caddeye Uzaklık	●	○	●	Yeşil Alanlara Uzaklık

4. Bulgular ve İrdeleme

4.1 Toplu Taşınmaz Değerlerinin Tespiti

4.1.1 ÇDR Algoritması

Taşınmaz değerlemede ÇDR modeli R yazılımında “lm” fonksiyonu kullanılarak kurulmuştur. ÇDR modeline ilişkin sabit (intercept) ve öznelik katsayıları (coefficient), standart hatalar, student-t istatistik değerleri ve bu istatistik değerlere karşılık gelen p değerleri hesaplanmıştır. Ele alınan 3 eğitim veri seti ile oluşturulan ÇDR modelleri test veri setine uygulandığında, değerlendirmeye alınan farklı hata ölçütleri için hesaplanan istatistiki değerler Tablo 4’te verilmiştir. K veri setinin kullanılması durumunda, diğer veri setleri ile oluşturulan modellere göre daha düşük hata değerlerinin elde edildiği ve en yüksek doğruluğa ulaşıldığı görülmüştür. KKO veri seti ile R^2 ölçütü en yüksek değerler alırken, bu değerler K veri seti ile oluşturulan modele oldukça yakındır. Elde edilen sonuçlar bir arada ele alındığında, konumsal özelliklerin yanında konumsal olmayan özelliklerin kullanılması çoklu doğrusal regresyon algoritmasının performansında önemli derecede bir iyileşme ortaya çıkarmamıştır. ÇDR algoritmasında modelin kurulması ve test tahmin süresi karşılaştırıldığında K ve KO veri seti için 0.01 saniye, KKO veri seti için 0.03 saniye olduğu görülmüştür. Bu yaklaşım ÇDR algoritmasının işlem süresi açısından kullanılabilirliğini destekler niteliktedir.

Tablo 4: ÇDR algoritmasıyla elde edilen hata ölçütleri

Veri	MAE	RMSE	MSE	R^2
K	0.1012	0.1300	0.0169	0.2151
KO	0.1105	0.1422	0.0202	0.0116
KKO	0.1029	0.1317	0.0173	0.2344

4.1.2 GDM Algoritması

R programlama yazılımında GDM yardımıyla taşınmaz değer tespitinde geliştirilmiş lineer model “glm” fonksiyonu kullanılmıştır. Söz konusu fonksiyonun uygulanmasında ve model oluşumunda bağımsız değişkenin dağılımı “quasibinomial” ve ilişki fonksiyonu olarak “logit” fonksiyonu kullanılmıştır. Logit fonksiyonunu ilişki fonksiyonu olarak kullanan binomial dağılımda bağımsız değişken 0 ve 1 değerleri alırken, quasibinomial dağılımda 0 ile 1 aralığında değerler alır. Binom ve quasibinomial modeller arasındaki temel farklardan en önemlisi, quasibinomial dağılım bağımsız değişkende varyansın çok farklı olduğu durumlarda tercih edilmesidir (Plant, 2012). GDM ile model oluşumunda tüm değişkenler modele dahil edilmiştir. GDM modeline ilişkin sabit ve öznelik katsayıları, standart hatalar, student-t istatistik değerleri ve bu istatistik değerlere karşılık gelen p değerleri hesaplanmıştır.

Ele alınan 3 eğitim veri seti ile oluşturulan GDM regresyon modelleri test veri setine uygulandığında, değerlendirmeye alınan hata ölçütleri için hesaplanan sonuçlar Tablo 5’te verilmiştir. K veri setinin kullanılması durumunda, hesaplanan ölçütlerde oluşturulan modelle diğer veri setleri ile oluşturulan modellere göre daha düşük hata değerlerinin elde edildiği ve en yüksek doğruluğu verdiği görülmüştür. KKO veri seti ile R^2 ölçütü en yüksek değerler alırken, bu değerler K veri seti ile oluşturulan modele oldukça yakındır. Elde edilen sonuçlar bir arada ele alındığında, ÇDR ile elde edilen sonuçlara paralel olarak, konumsal özelliklerin yanında konumsal olmayan özelliklerin kullanılması GDM algoritmasının performansında önemli derecede bir iyileşme ortaya çıkarmamıştır. GDM algoritmasında modelin kurulması ve test tahmin süresi karşılaştırıldığında K ve KKO veri seti için 0.05 saniye; KO veri seti için 0.03 saniye olduğu görülmüştür. Bu yaklaşım GDM algoritmasının işlem süresi açısından kullanılabilirliğini destekler niteliktedir.

Tablo 5: GDM algoritmasıyla elde edilen hata ölçütleri

Veri	MAE	RMSE	MSE	R ²
K	0.1013	0.1300	0.0169	0.2152
KO	0.1104	0.1422	0.0202	0.0116
KKO	0.1028	0.1316	0.0173	0.2349

4.1.3 DVM Algoritması

Taşınmaz değerlendirilmede DVM modellerinin R yazılımında kurulmasında “LibSVM” tabanlı “e1071” kütüphanesi ve “epsilon regression” modeli kullanılmıştır. Parametrelerin optimizasyonu işlemi yine aynı kütüphane dahilinde bulunan optimizasyon (tune) fonksiyonu kullanılmıştır. Doğrusal, polinom ve radyal tabanlı kerneller kullanılarak DVM modelleri kurulmuş ve parametre optimizasyonu gerçekleştirilmiştir. Tablo 6’da farklı kernel fonksiyonları ile kurulmuş DVM modellerinin hata ölçütleri verilmiştir. Kernel fonksiyonu seçimine göre değerlendirmeye alınan üç veri seti içerisinde en düşük hata düzeyine ulaşılan sonuçlar ise Tablo 7’de listelenmiştir. Karşılaştırılan kernel fonksiyonlarında veri setlerinde yalnızca konumsal verilerin kullanıldığı durumda, veri setinde konumsal ve konumsal olmayan verilerin birlikte kullanıldığı durumdan daha düşük hata değerleri elde edilmiştir. K veri setini kullanan kernel fonksiyonlarından radyal tabanlı fonksiyonun en düşük hata değerlerine sahip olduğu gözlenmiştir. Polinom kernelinin parametre sayısının fazla olması, parametrelerin optimizasyonunun diğer kernel fonksiyonlarına göre çok uzun sürmesi, polinom derecesi parametresinin arttıkça model kurulumunu karmaşılaştırması ve taşınmaz değerlemede kullanılan veri setlerinde daha büyük hata değerlerine sahip olması sebebiyle taşınmaz değerlendirilmede kullanılmasının uygun olmayacağı tespit edilmiştir. Radyal tabanlı fonksiyon ele alındığında, diğer kernel fonksiyonlarına göre daha düşük hata değerlerine sahip olması, daha güçlü bir tahmin modeli oluşturması ve işlem süresinin diğer kernel fonksiyonlarına göre önemli ölçüde kısa sürmesi bu fonksiyonları diğer kernel fonksiyonlarından daha üstün kılmaktadır. Böylelikle taşınmaz değer haritası üretiminde DVM algoritmasının en uygun kernel fonksiyonunun radyal tabanlı kernel olduğuna karar verilmiştir.

Tablo 6: Farklı kernel fonksiyonları ile kurulmuş DVM algoritmaları ile elde edilen hata ölçütleri

DVM LİNEER KERNEL	MAE	RMSE	MSE	R ²
K	0.10135	0.13198	0.01742	0.13274
KO	0.10935	0.14227	0.02024	0.00195
KKO	0.10197	0.13264	0.01759	0.14119
DVM POLİNOM KERNEL				
K	0.10078	0.13209	0.01745	0.16531
KO	0.10835	0.14140	0.01999	0.01691
KKO	0.10177	0.13360	0.01785	0.16907
DVM RADYAL KERNEL				
K	0.09890	0.13036	0.01699	0.18779
KO	0.10914	0.14206	0.02018	0.00429
KKO	0.10949	0.14178	0.02010	0.00133

Tablo 7: DVM algoritmasında en düşük hata değerlerine sahip kernellerin karşılaştırılması

V+KERNEL	MAE	RMSE	MSE	R ²
K, LİNEER	0.10135	0.13198	0.01742	0.13274
K, POLİNOM	0.10078	0.13209	0.01745	0.16531
K, RADYAL	0.09890	0.13036	0.01699	0.18779

4.1.4 KA Algoritması

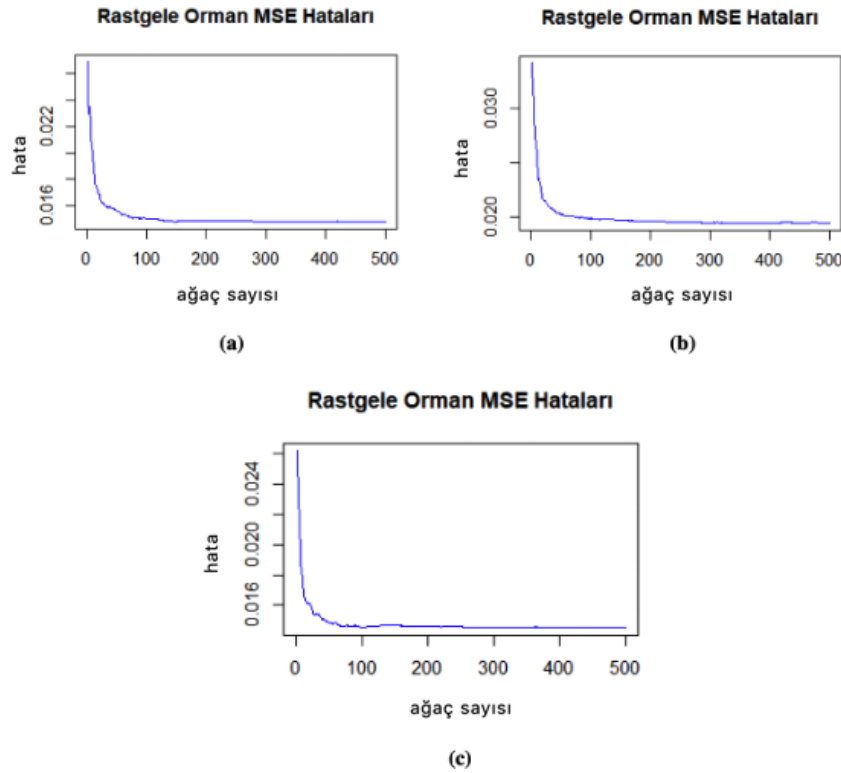
Taşınmaz değerlendirme haritasının üretimi için KA modelinin kurulmasında R yazılımındaki “rpart” kütüphanesi kullanılmıştır. 3 veri seti için de oluşturulan modellere budama işlemi uygulanmamıştır. KA değer tahminleri ile test verilerinin güncel piyasa değerleri kullanılarak hesaplanan hata ölçütleri Tablo 8’de yer almaktadır. Hata miktarları incelendiğinde KKO veri seti kullanılan KA modelinin bütün hata ölçütlerinde en düşük hata miktarlarına sahip olduğu tespit edilmiştir. KA modellerinin oluşturulması ve test veri setinin tahmin süreleri K veri için 0.31 saniye, KO veri seti için 0.03 saniye ve KKO veri için 0.07 saniye olarak tespit edilmiştir. Karar ağacı modellerinin oluşturulma sürelerinin çalışma kapsamında incelenen algoritmalar arasında en hızlısı olduğu tespit edilmiştir.

Tablo 8: KA algoritmasında en düşük hata değerlerine sahip kernellerin karşılaştırılması

Veri	MAE	RMSE	MSE	R ²
K	0.10130	0.13018	0.01695	0.18907
KO	0.11185	0.14525	0.02110	0.03692
KKO	0.09707	0.12499	0.01562	0.18907

4.1.5 RO Algoritması

RO algoritması ile tahmin modelinin R yazılımında oluşturulması için “randomForest” kütüphanesi kullanılmıştır. Modeller kurulurken başta 500 ağaç ile oluşturulan modellerin hata grafikleri (Şekil 5) incelenerek tüm modeller için 150 ağaç kullanılmasının uygun olduğu belirlenmiştir. Dğümlerde kullanılacak örnek sayısını belirleyen t parametresi ise veri setlerinde kullanılan öznitelik sayısının karekökünün tam sayıya yuvarlanması suretiyle K veri seti için 5, KO veri seti için 4 ve KKO veri seti için 6 olarak seçilmiştir.

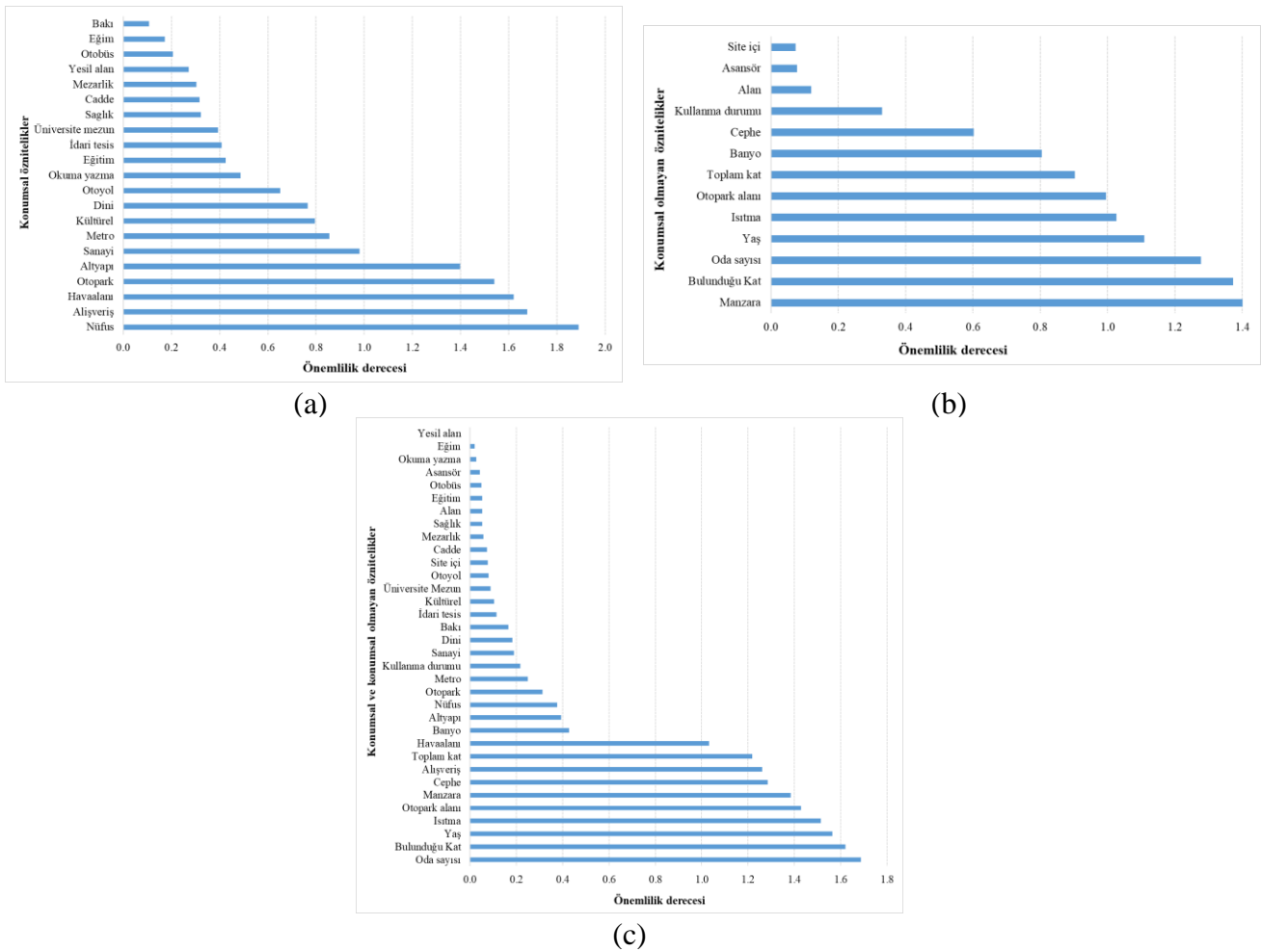


Şekil 5: RO algoritmasında K (a), KO (b) ve KKO (c) veri setlerine ait MSE hata grafikleri

K veri seti için önemlilik düzeyleri (Şekil 6a) incelendiğinde alınan konumsal öznitelikler içerisinde nüfus yoğunluğu, alışveriş merkezlerine olan uzaklık ve havaalanına olan uzaklık öznitelikleri RO algoritması tarafından önemlilik derecesi en

yüksek ilk üç konumsal öznelik olarak belirlenmiştir. Diğer taraftan örnek noktaların bulunduğu konumlara ilişkin topoğrafik özelliklerden bakı ve eğim algoritma tarafından önemlilik derecesi en düşük olan konumsal özneliklerdir. Böylelikle topoğrafik faktörlerden ziyade, taşınmazın bulunduğu bölgedeki nüfus yoğunluğu, taşınmazın havaalanı ve alışveriş merkezlerine olan uzaklığı başta olmak üzere, otopark alanlarına yakınlık, altyapı tesislerine yakınlık, dini tesislere yakınlık ve kültürel tesislere yakınlıklar gibi değerlendirmeye alınan diğer özneliklerin de taşınmaz değer tespitinde etkili olduğu ifade edilebilir.

KO veri seti için önemlilik düzeyleri (Şekil 6b) incelendiğinde; manzara, taşınmazın bulunduğu kat, ısıtma durumu, sahip olduğu oda sayısı ve taşınmazın yaşı RO algoritması tarafından önemlilik derecesi en yüksek öznelikler arasındadır. Diğer taraftan, taşınmazın site içinde oluşu ve asansör durumu algoritma tarafından en düşük önemlilik derecesine sahip olan öznelikler olarak tespit edilmiştir. KKO veri seti için önemlilik düzeyleri (Şekil 6c) incelendiğinde, önemlilik derecesi yüksek bulunan ilk 10 öznelik içerisinde taşınmazın konumsal özneliklerinden alışveriş merkezlerine olan uzaklık, nüfus yoğunluğu ve havaalanına olan uzaklıklar yer almaktadır. Konumsal olmayan özneliklerden ise taşınmazın bulunduğu kat, taşınmazın yaşı, bulunduğu cephe, ısıtma sistemi, toplam kat sayısı, manzarası ve otopark durumu ilk 10 öznelik içerisinde yer almaktadır. Şekil incelendiğinde, otobüs durağına olan uzaklık, taşınmazın site içinde oluşu, yeşil alana ve sağlık merkezlerine olan uzaklıklar önem derecesi en düşük öznelikler olarak tespit edilmiştir. Önemlilik derecesine ilişkin sonuçlar, konumsal ve konumsal olmayan özelliklerin ağaç oluşumunda etkili olduğu, dolayısıyla değer tahmininde taşınmaza ilişkin her iki öznelik grubunun da etkili olabileceğini gösterir niteliktedir.



Şekil 6: RO algoritmasında K (a), KO (b) ve KKO (c) veri setleri için özneliklerin katsayıları

RO algoritması için veri setlerine göre hesaplanan hata değerleri Tablo 9'da yer almaktadır. Tablo incelendiğinde bütün

özniteliklerin yer aldığı KKO veri setinin en düşük hata değerlerini aldığı gözlenmiştir. K veri setinin aldığı hata değerleri de KKO veri setinin hata değerlerine yakınlık göstermektedir. RO algoritmasında modelin kurulması ve test tahmin süresi karşılaştırıldığında, K veri seti için 4.61 saniye, KO veri seti için 2.56 saniye ve KKO veri seti için 6.15 saniye olduğu görülmüştür. Veri setlerinin boyutu arttıkça, işlem süresinin doğru orantılı olarak arttığı söylenebilir. İşlem süreleri diğer algoritmalar ile karşılaştırıldığında, algoritmanın işleyiş sürecindeki işlemlere göre (eğitim ve test verilerinin ayrılması, OOB hatalarının hesaplanması vb.) kısa bir sürede gerçekleştiği kabul edilebilir.

Tablo 9: RO algoritması ile elde edilen hata ölçütleri

Veri	MAE	RMSE	MSE	R ²
K	0.09842	0.12694	0.01611	0.34226
KO	0.11156	0.14391	0.02071	0.08789
KKO	0.09786	0.12633	0.01596	0.28448

4.2 Mahalle Bazında Taşınmaz Piyasa ve Hesap Değerleri Arasındaki Hatalar ve İlişkiler

Çalışma kapsamında kullanılan algoritmaların en yüksek doğruluğu veren veri setleriyle oluşturdukları modellerin mahalle bazında hata ve ilişkileri incelenmiştir. Pendik ilçesindeki her mahalle için taşınmazların piyasa değerleri ile tahmin modelleri arasındaki RMSE ve MAE hataları hesaplanmıştır. Ayrıca her mahalledeki taşınmazın piyasa değerleri ile model tahminlerinin arasında ilişki bulunup bulunmadığı da korelasyon katsayı hesaplanarak belirlenmiştir.

Taşınmaz değer haritalarının RMSE ve MAE hatalarının mahalle bazında gösterildiği grafikler Şekil 7’de verilmiştir. ÇDR algoritması ile kurulan modelin hata değerleri incelendiğinde, mahalle bazında hata değerlerinin genellikle birbirine yakın olduğu görülürken Doğu, Harmandere ve Orta Mahalleleri’nin hata değerlerinin yüksek olduğu görülmüştür. Ahmet Yesevi, Ertuğrul Gazi ve Sapan Bağları Mahalleleri’nde ise en düşük hata değerleri elde edilmiştir. GDM tahmin modelinde ÇDR algoritmasının hata değerlerine benzerlik göstermektedir. DVM algoritması ile kurulan model incelendiğinde özellikle Çınardere, Doğu ve Esenyalı Mahalleleri’nde hata değerlerinin diğer mahallelere göre daha fazla çıktığı görülmüştür.

Mahalleler arasındaki hata değerlerinin değişken olduğu gözlenen DVM modelinde Ahmet Yesevi, Çamlık ve Ertuğrul Gazi Mahalleleri’nin hata değerlerinin düşük; dolayısıyla tahmin performansının yüksek olduğu görülmüştür. KA algoritması ile kurulan model incelendiğinde hata miktarlarının genellikle birbirlerine yakın olarak seyrettikleri görülürken Doğu, Harmandere ve Orta Mahalleleri’nde hata miktarlarının diğer mahallelere göre daha büyük miktarda olduğu gözlenmiştir. RO algoritması ile kurulan modelin mahalle bazlı hata miktarlarının diğer modellerdeki hatalar kadar eşit dağılım göstermediği görülmektedir. Bahçelievler, Doğu ve Harmandere Mahalleleri’nin hata miktarlarının yüksek olduğu görülürken Ahmet Yesevi, Çamlık, Çınardere, Fatih ve Güllübağlar Mahalleleri’nde tahmin performansının daha yüksek olduğu gözlenmiştir.

Taşınmaz değerlemede kullanılan algoritmaların tahmin modellerinin güncel piyasa satış değerleri ile aralarındaki ilişkiyi incelemek amacıyla hesaplanan korelasyon katsayıları Şekil 8’deki grafiklerde gösterilmiştir. ÇDR algoritmasının K veri seti kullanarak oluşturduğu tahmin modelinde taşınmazlara ait piyasa değerleri ile tahmin edilen değerler arasındaki korelasyonun genellikle düşük olduğu gözlenmiştir. Orta, Ramazanoğlu ve Sanayi Mahalleleri’nde korelasyon katsayılarının %40’ın üzerinde çıktığı görülürken Çamlık, Esenler ve Batı Mahalleleri’nde korelasyonun negatif yönde yüksek çıktığı belirlenmiştir. Mahallelere ait korelasyon katsayıları genel olarak incelendiğinde ise korelasyon katsayılarının genellikle düşük olduğu gözlenmiştir. GDM algoritmasının K veri seti ile oluşturduğu tahmin modeli incelendiğinde Çamçeşme, Fevzi Çakmak ve Ramazanoğlu Mahalleleri’nde korelasyon değerinin %35’in üzerinde çıktığı görülmektedir. Çamlık, Esenler ve Kavakpınar Mahalleleri’nde negatif yönde korelasyon olduğu görülürken, diğer mahallelerin korelasyon değerlerinin %25’in

altında olduğu görülmüştür. Bu bilgilere dayanarak GDM algoritmasının tahmin modelinin taşınmaz değerlendirme probleminde başarılı olmadığı söylenebilir.



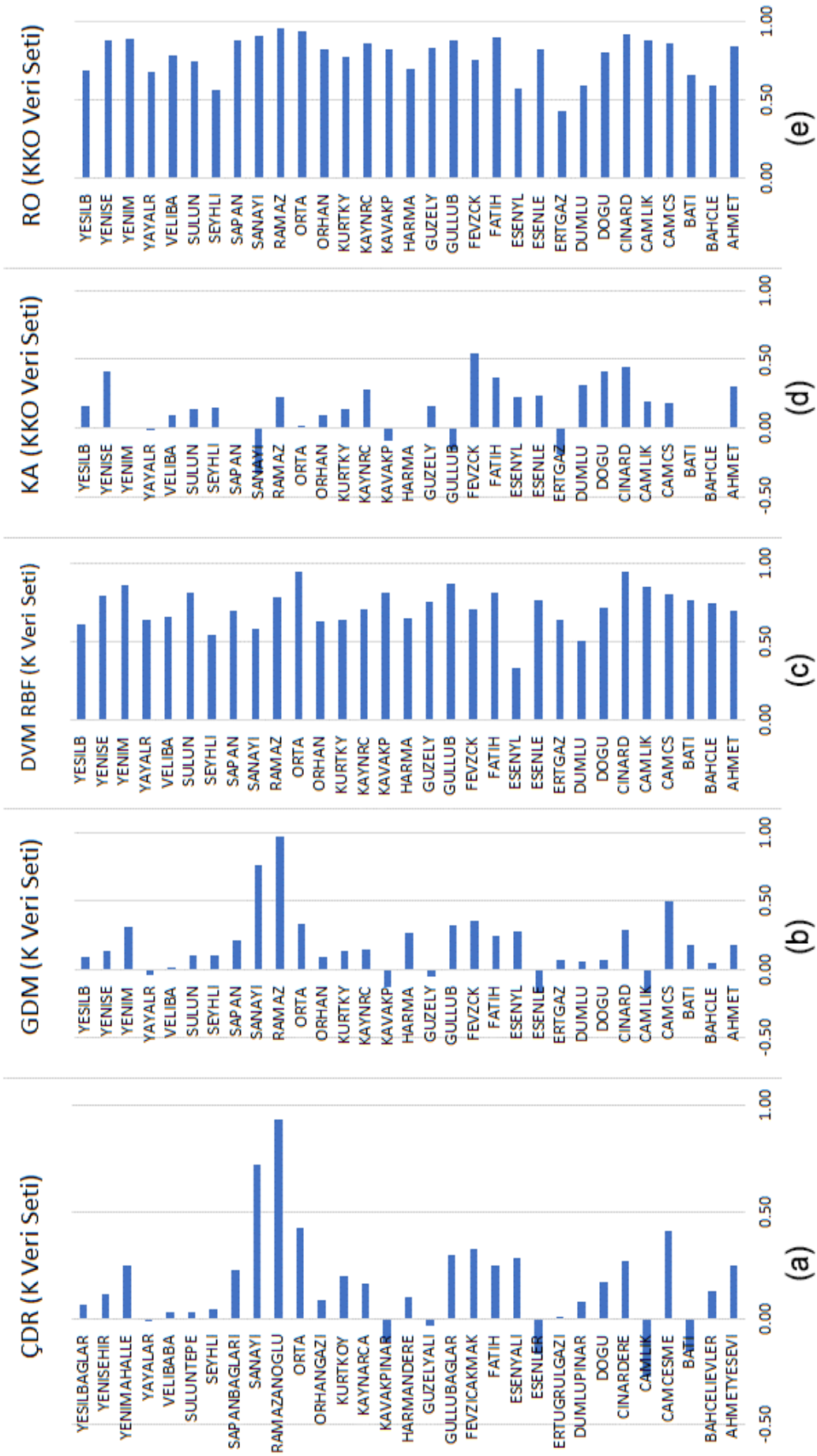
Şekil 7: ÇDR (a), GDM (b), DVM (c), KA(d) ve RO(e) algoritmaları ile oluşturulan değer haritalarının mahalle bazında RMSE ve MAE hata değerleri

DVM'nin radyal tabanlı kerneli ve K veri seti kullanılarak oluşturulan tahmin modelinde, mahalle bazında hesaplanan korelasyon katsayıları incelendiğinde, korelasyon katsayılarının tamamının pozitif değer aldığı, Çınardere, Güllübağlar ve Orta Mahalleleri'nin taşınmaz tahmin değerlerinin güncel piyasa satış değerleriyle aralarında pozitif yönde yüksek korelasyon olduğu gözlenmiştir. Dumlupınar ve Esenyalı Mahalleleri'nin korelasyon katsayıları diğer mahallelerin korelasyon katsayılarına oranla daha düşük çıkmıştır. Mahalleler için hesaplanan korelasyon katsayıları genel olarak değerlendirildiğinde, ortalama korelasyonun %71.5 olduğu görülmüştür ve DVM regresyon modelinin tahmin gücünün iyi olduğu söylenebilir. KA algoritması ve KKO veri seti kullanılarak oluşturulan tahmin modelinde mahalle bazında hesaplanan korelasyon katsayıları incelendiğinde, mahallelerin büyük çoğunluğu için tahmin modelinin bulduğu değerler ile piyasa değerleri arasındaki korelasyonun düşük olduğu görülmektedir. Pendik'in Çınardere, Fevzi Çakmak ve Yenışehir Mahalleleri'nde pozitif yönde yüksek korelasyon olduğu, Ertuğrul Gazi, Güllübağlar ve Sanayi Mahalleleri'nde korelasyonun negatif yönde yüksek olduğu gözlenmiştir. İlçenin Bahçelievler, Batı, Harmandere, Sapan Bağları ve Yeni Mahalle Mahalleleri'nde korelasyon katsayısının sıfır çıktığı tespit edilmiştir. Bu mahallelerin korelasyon katsayılarının sıfır veya sıfıra çok yakın çıkmasının sebebi, modelin mahalledeki bütün taşınmazların değerini birbirine çok yakın olarak tahmin etmesi sebebiyle varyansın düşük olmasıdır. Tahmin modelinde 31 mahallenin 23 tanesinin korelasyon katsayısının %25'in altında çıkması, KA regresyon modelinin taşınmaz değerlendirme probleminde yeterli performansı gösteremediğinin göstergesidir.

RO algoritması ve KKO veri seti kullanılarak oluşturulan tahmin modelinde mahalle bazında hesaplanan korelasyon katsayıları incelendiğinde Çınardere, Fatih ve Orta Mahalleleri'nin korelasyonunun %90'ın üzerinde olduğu görülürken yalnızca Ertuğrul Gazi Mahallesi'nin %50'nin altında korelasyona sahip olduğu görülmüştür. Mahallelerin korelasyon katsayılarının ortalaması %78 standart sapması ise %12 olarak hesaplanmıştır. Ayrıca grafik görsel olarak yorumlandığında korelasyon katsayılarının birbirlerine yakın değerlerde olduğu ve negatif yönde korelasyon bulunmadığı görülmüştür. Bu çıkarımlara dayanarak RO regresyon modelinin taşınmaz değerlendirme problemi için iyi bir performans sağladığı sonucuna varılabilir.

5. Sonuçlar

Son yıllarda gelişmiş tahmin yaklaşımları olarak bilinen ve makine öğrenme/veri madenciliğine dayanan DVM, KA ve toplu öğrenme algoritmaları toplu taşınmaz değerlendirme sıklıkla kullanılmaktadır. Ancak bu çok farklı modellerin toplu taşınmaz değerlendirme optimal kullanımları ve elde edilen doğruluklarının incelenmesi gerekmektedir. Ayrıca taşınmaz değer tespitinde konumsal ve konumsal olmayan kriterlerin tahmin doğruluğuna etkilerinin incelenmesi, optimal yaklaşımların belirlenmesinde önemlidir. Çalışmada ÇDR, GDM, DVM, KA ve RO algoritmaları kullanılarak toplu taşınmaz değerlendirme model başarıları irdelenmiştir. Uygulama alanı olarak seçilen İstanbul ili Pendik ilçesi için güncel taşınmaz satış değerlerini temsil eden 1475 örnekleme noktası ile 3 farklı veri seti (K, KO, KKO) ile farklı metotlar ile kurulan modellerin performansları irdelenmiştir. Buna göre en yüksek doğruluk veren yöntem RO; en düşük doğruluk veren yöntemlerin ise KA ve GDM olduğu görülmüştür.



Şekil 8: ÇDR (a), GDM (b), DVM (c), KA(d) ve RO(e) algoritmaları ile oluşturulan değer haritalarının mahalle bazında korelasyon katsayıları

RO algoritması için bütün özniteliklerin yer aldığı KKO veri setinin en düşük hata değerlerini aldığı ancak K veri setinin aldığı hata değerleri ile de KKO veri setinin hata değerlerinin oldukça yakın olduğu gözlenmiştir. RO algoritması ve KKO veri seti kullanılarak oluşturulan tahmin çoğu mahallenin korelasyonunun %90 civarında ve ortalamasının %78 olduğu; ortalama standart sapmanın ise %12 olduğu görülmüştür. Bu çıkarımlara dayanarak RO modelinin taşınmaz değerlendirme problemi için iyi bir performans sağladığı görülmektedir. KA ve GDM modelleri süre bakımından en hızlı modeller olsalar da taşınmaz değerlendirme probleminde diğer yöntemlere kıyasla başarılı olmadıkları söylenebilir. KA algoritması ve KKO veri seti kullanılarak oluşturulan tahmin modelinde mahalle bazında hesaplanan korelasyon katsayıları incelendiğinde mahallelerin büyük çoğunluğu için tahmin değerleri ile güncel piyasa değerleri arasındaki korelasyonun düşük olduğu görülmüştür. 31 mahallenin 23 tanesinin korelasyon katsayısının %25'in altında; bazı mahallelerin ise sıfır olarak hesaplandığı tespit edilmiştir. GDM algoritmasının K veri seti ile oluşturduğu tahmin modeli incelendiğinde birkaç mahallenin korelasyon değerinin %35'in üzerinde çıktığı diğer mahallelerin korelasyon değerlerinin ise %25'in altında olduğu görülmüştür. Korelasyon katsayılarının sıfır veya sıfıra çok yakın çıkmasının, modelin bütün taşınmazların değerini birbirine çok yakın olarak tahmin etmesi sebebiyle varyansların düşük olmasıdır.

Üretilen modellerin tümünde konutların toplu taşınmaz değerlendirme sürecinde RO modelleri diğer modellere kıyasla daha başarılı sonuçlar vermiştir. Bu anlamda RO algoritmasının geliştirilecek vergilendirme, bankacılık, sigortalama, kamulaştırma gibi uygulamalara entegre edilebilir en başarılı yöntem olduğu belirlenmiştir. Değeri etkileyen konumsal ve konumsal olmayan kriterlerin genel anlamda model performansında büyük fark yaratmadığı görüldüğünden toplu taşınmaz değerlendirme sürecinde öne çıkan yapısal ve konumsal kriterlerin tüm çalışma alanı için ayrı ayrı kriter önem düzeyleri hesaplanmıştır. İlgili değerlendirme ile ileride kurulacak uygulama modelleri için ayırt edici kriterlerin belirlenerek model performanslarının artırılması hedeflenmektedir. Kriter önem düzeyleri irdelendiğinde K veri seti için nüfus yoğunluğu, alışveriş merkezlerine olan uzaklık ve havaalanına olan uzaklık; KO veri seti için ise manzara, taşınmazın bulunduğu kat ve oda sayısı ile bina yaşı önemlilik derecesi en yüksek öznitelikler olarak belirlenmiştir. İlgili öznitelikler KKO veri seti içinde bütüncül olarak irdelendiğinde de yüksek önemlilik derecesine sahiptirler. Ancak konumsal ve konumsal olmayan öznitelikler bütüncül değerlendirildiğinde taşınmazın oda sayısı ve bulunduğu kat ile bina yaşı gibi taşınmaza ait yapısal kriterlerin önemlilik derecesi konumsal kriterlere oranla biraz daha yüksek çıktığı görülmüştür. Bütüncül olarak değerlendirildiğinde önemlilik derecesine ilişkin sonuçlar, konumsal ve konumsal olmayan özelliklerin ağaç oluşumunda etkili olduğu, dolayısıyla değer tahmininde taşınmaza ilişkin her iki öznitelik grubunun da etkili olabileceğini göstermektedir.

Teşekkür

Bu çalışma TÜBİTAK tarafından desteklenen 116Y204 nolu 1001 projesi kapsamında gerçekleştirilmiştir.

Yazar Katkısı

Arif Çağdaş Aydınoglu: Fikir, Tasarım, Denetleme, Yazım. **Rabia Bovkır:** Literatür taraması, Veri toplama ve işleme, Yazım. **İsmail Çölkesen:** Tasarım, Analiz ve yorumlama, Makale değerlendirme.

Çıkar Çatışması Beyanı

Yazarlar, bu çalışmada bilinen ilgili herhangi bir finansal veya finansal olmayan çıkar çatışması olmadığını beyan eder.

Kaynaklar

- Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications*, 39(2), 1772-1778.
- Arslan, R. (1997). *Arazi Kullanış Ekonomisi*. İstanbul, Türkiye: Yıldız Teknik Üniversitesi Basım Yayın Merkezi.
- Aydinođlu, A. Ç., Çölkesen, İ., Şenbil, M., Bovkır, R., & Yomralođlu, T. (2020). *116Y204 no'lu TÜBİTAK Projesi Sonuç Raporu*. TÜBİTAK ÇAYDAG, Ankara, 04.2020.
- Aydinođlu, A. Ç., Bovkır, R., & Çölkesen, İ. (2021). Implementing a mass valuation application on interoperable land valuation data model designed as an extension of the national GDI. *Survey review*, 53(379), 349-365.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*.
- Bovkır, R., & Aydinođlu, A. Ç. (2018). Providing land value information from geographic data infrastructure by using fuzzy logic analysis approach. *Land use policy*, 78, 46-60.
- Božić, B., Milićević, D., Pejić, M., & Marošān, S. (2013). The use of multiple linear regression in property valuation. *Geonauka*, 1(1), 41-45.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C.J. (1984). *Classification And Regression Trees*. Routledge.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Bulut Nas, B. (2011). *YSA ve DVM yöntemleri ile taşınmaz değerlemesi için bir yaklaşım geliştirme (Yüksek Lisans Tezi)*. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya, Türkiye.
- Candaş, E. (2012). *Taşınmaz Değerlemesi İçin Mevzuat Altyapısının Modellenmesi (Yüksek Lisans Tezi)*. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, Türkiye.
- Čeh, M., Kilibarda, M., Liseć, A., & Bajat, B. (2018). Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS international journal of geo-information*, 7(5), 168.
- Chen, J. H., Ong, C. F., Zheng, L., & Hsu, S. C. (2017). Forecasting spatial dynamics of the housing market using support vector machine. *International Journal of Strategic Property Management*, 21(3), 273-283.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- Çölkesen, İ. (2015). *Yüksek Çözünürlüklü Uydu Görüntüleri Kullanarak Benzer Spektral Özelliklere Sahip Doğal Nesnelerin Ayırt Edilmesine Yönelik Bir Metodoloji Geliştirme* (Doktora tezi). İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, Türkiye
- Çölkesen, İ. & Kavzođlu, T. (2017). Kanonik Korelasyon Orman Algoritması ile Uzaktan Algılanmış Görüntülerin Sınıflandırılması [Özel Sayı]. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 17(1), 102-111.
- Dawidowicz, A., & Źróbek, R. (2017). Land administration system for sustainable development—case study of Poland. *Real Estate Management and Valuation*, 25(1), 112-122.
- 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.
- Dimopoulos, T., & Moulas, A. (2016). A proposal of a mass appraisal system in Greece with CAMA system: Evaluating GWR and MRA techniques in Thessaloniki Municipality. *Open geosciences*, 8(1), 675-693.
- Dimopoulos, T., & Bakas, N. (2019). Sensitivity analysis of machine learning models for the mass appraisal of real estate. Case study of residential units in Nicosia, Cyprus. *Remote sensing*, 11(24), 3047.
- Dobson, A. J., & Barnett, A. G. (2008). *An Introduction to Generalized Linear Models (3rd Edition)*. Chapman & Hall/ CRC Press.
- Güneş, T., & Yıldız, U. (2015). Mass valuation techniques used in land registry and cadastre modernization project of Republic of Turkey. *FIG working week: From the Wisdom of the Ages to the Challenges of the Modern World*.
- Güneş, T., & Yıldız, Ü. (2016). Property valuation and taxation for improving local governance in Turkey. *Land tenure journal*, 15(2), 141-160.
- Güngör, E. (1999). *Gayrimenkul değerlemesi ve Türkiye'de sermaye piyasalarında gayrimenkul ekspertiz şirketlerine yönelik düzenlemeler yapılmasına ilişkin öneriler*. Yeterlik Etüdü. T.C. Başbakanlık Sermaye Piyasası Kurulu Kurumsal Yatırımcılar Dairesi, Ankara, Türkiye.

- Hong, J., Choi, H., & Kim, W. S. (2020). A house price valuation based on the random forest approach: the mass appraisal of residential property in South Korea. *International Journal of Strategic Property Management*, 24(3), 140-152.
- IAAO (2013). *Standard on Mass Appraisal of Real Property*. International Association of Assessing Officers, Kansas, ABD.
- IVSC (2020). *International Valuation Standards*. International Valuation Standards Council, Londra, Birleşik Krallık.
- Kara, A., van Oosterom, P., Çağdaş, V., Işıklıdağ, Ü., & Lemmen, C. (2020). 3 Dimensional data research for property valuation in the context of the LADM Valuation Information Model. *Land use policy*, 98, 104179.
- Karapınar, A., Bayırlı, R., Bal, H., Altay, A., Bal, E. Ç., & Torun, S. (2008). *SPK Lisanslama Sınavlarına Hazırlık*. Ankara: Gazi Kitabevi.
- Kavzoğlu, T., & Çölkesen, İ. (2010). Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi. *Harita Dergisi*, 144(7), 73-82.
- Kavzoğlu, T., & Çölkesen, İ. (2013). An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. *International journal of remote sensing*, 34(12), 4224-4241.
- Kontrimas, V., & Verikas, A. (2011). The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing*, 11(1), 443-448.
- Lasota, T., Łuczak, T., Niemczyk, M., Olszewski, M., & Trawiński, B. (2013). Investigation of property valuation models based on decision tree ensembles built over noised data. *International Conference on Computational Collective Intelligence*. Springer, Berlin, Heidelberg.
- Manasa, J., Gupta, R., & Narahari, N. S. (2020). Machine learning based predicting house prices using regression techniques. *2020 2nd International conference on innovative mechanisms for industry applications (ICIMIA)*.
- Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research*.
- Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370-384.
- Osuna, E. E., Freund, R., & Girosi, F. (1997). *Support Vector Machines: Training and Applications*. AIM-1602CBCL-144, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Massachusetts, ABD.
- Ott, R. L., & Longnecker M. (2010). *An introduction to statistical methods and data analysis*. Belmont (CA): Brooks/Cole.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International journal of remote sensing*, 26(1), 217-222.
- Plant, R. E. (2012). *Spatial Data Analysis in Ecology and Agriculture using R*. Boca Raton: CRC Press.
- Quinlan, J., & Cameron-Jones, R. (1995). Oversearching and layered search in empirical learning. *breast cancer*, 286, 2-7.
- Resmi Gazete. (2001). 35 nolu, Sermaye Piyasası Mevzuatı Çerçevesinde Gayrimenkul Değerleme Hizmeti Verecek Şirketler İle Bu Şirketlerin Kurulca Listeye Alınmalarına İlişkin Esaslar Hakkında Tebliğ Sayı: 24491.
- Resmi Gazete, (2018). Bakanlığa bağlı kamu kurumlarının teşkilatına ilişkin Cumhurbaşkanlığı Değişikliği Kararı. Resmi Türkiye Cumhuriyeti Gazetesi, No: 30479.
- Reyes-Bueno, F., García-Samaniego, J. M., & Sánchez-Rodríguez, A. (2018). Large-scale simultaneous market segment definition and mass appraisal using decision tree learning for fiscal purposes. *Land Use Policy*, 79, 116-122.
- RICS (2017). *Professional Valuation Standards*. Royal Institution of Chartered Surveyors, Londra, Birleşik Krallık.
- SPK (2006). *Sermaye Piyasasında Uluslararası Değerleme Standartları Hakkında Tebliğ*. Sermaye Piyasası Kurulu, Seri: VIII, No: 45, Ankara, Türkiye.
- Şişman, S. (2021). *CBS Tabanlı Makine Öğrenme Teknikleri ile Toplu Taşınmaz Değerlemesi* (Yüksek Lisans Tezi), Gebze Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Kocaeli, Türkiye.
- Şişman, S., Akar, A. U., & Yalprı, S. (2021). The novelty hybrid model development proposal for mass appraisal of real estates in sustainable land management. *Survey Review*, 1-20.
- TDUB (2017). *Türkiye Değerleme Standartları (TUGDES) Çalışması*, Türkiye Değerleme Uzmanları Birliği.
- TEGoVA (2016). *European Valuation Standards*. 8. Baskı. The European Group of Valuers' Associations, Belçika: Gillis Yayınevi.
- TKGM (2011). *TKMP 4. Bileşeni Gayrimenkul Değerinin Belirlenmesi ve Kayıt Altına Alınması*, Teknik Rapor, Ankara, Türkiye.
- The Appraisal Institute (2001). *The Appraisal of Real Estate*, 12. Ed. Appraisal Institute, Chicago, ABD.
- Ünel, F. B., & Yalprı, Ş. (2019). Valuations of building plots using the AHP method. *International Journal of Strategic Property Management* 23(3), 197-212.
- Vapnik, V. N. (1982). *Estimation of Dependences Based on Empirical Data*. New York: Springer-Verlag.

- Wang, X., Wen, J., Zhang, Y., & Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. *Optik*, 125(3), 1439-1443.
- Yalprı, Ő. (2007). *Bulanık mantık metodolojisi ile taşınmaz deęerleme modelinin geliřtirilmesi ve uygulaması: Konya örneęi* (Doktora Tezi). Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya, Türkiye
- Yalprı, Ő., & Özkan, G. (2008). The usage of artificial intelligence in determining the residential real-estate prices in urban areas and the comparison of valuation methods. *Integrating Generations FIG Working Week 2008*, 14-19.
- Yalprı, Ő., Durduran, S. S., Ünel, F. B., & Yolcu, M. (2014). Creating A Valuation Map in GIS Through Artificial Neural Network Methodology: A Case Study. *Acta Montanistica Slovaca*, 19(2), 79-89.
- Yalprı, Ő., & Bünyan Ünel, F. (2016). Türkiye'de ve Uluslararası çalışmalarda arsa deęerlemede kullanılan kriterlerin irdelenmesi ve Faktör Analizi ile azaltımı. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 16(2), 303-322.
- Yalprı, Ő. (2018). Enhancement of parcel valuation with adaptive artificial neural network modeling. *Artificial intelligence review*, 49(3), 393-405.
- Yalprı, Ő., Őiřman, S., Akar, A. U., & Ünel, F. B. (2021). Feature selection applications and model validation for mass real estate valuation systems. *Land use policy*, 108, 105539.
- Yazıcı, K. (1997). *Özelleřtirmede Deęerleme Yöntemleri ve Deęerleme Kriterleri* (Uzmanlık Tezi). T.C. Bařbakanlık Devlet Planlama Teřkilatı Müsteřarlıęı, Yıllık Programlar ve Konjonktür Deęerlendirme Genel Müdürlüęü.
- Yılmaz, S., & Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy*, 99, 104889.
- Yomraloęlu, T. (1993). *A nominal asset value-based approach for land readjustment and its implementation using geographical information systems* (Doktora Tezi). University of Newcastle upon Tyne, Department of Surveying, Newcastle, İngiltere.