

Toplu Taşınmaz Değerlemesi için Coğrafi Verinin İşlenmesi ve Rastgele Orman Makine Öğrenmesi Tekniğiyle Tahmin Modelinin Geliştirilmesi (Processing Geographical Data and Developing Prediction Model with Random Forest Machine Learning Technique for Mass Real Estate Valuation)

Süleyman ŞİŞMAN^{ID}, Arif Çağdaş AYDINOĞLU^{ID}

Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, 41400, Gebze, Kocaeli
ssisman@gtu.edu.tr, *aydinoglu@gtu.edu.tr

Geliş Tarihi (Received): 21.03.2022

Kabul Tarihi (Accepted): 13.01.2023

ÖZ

Taşınmazların değerinin yatırım planlamasından kredi düzenlenmelerine, vergilendirmeden çeşitli sektörel uygulamalara kadar kullanılmasından dolayı taşınmazlar ülke ekonomilerinin önemli bir bileşenidir. Taşınmazların değerinin güncel yaklaşımlar ile objektif olarak tespiti, planlı ve sürdürülebilir taşınmaz yönetim stratejilerinin geliştirilmesi için önemli bir gereksinim haline gelmiştir. Dolayısıyla tekil değerlendirme yaklaşımlarının yanı sıra toplu taşınmaz değerlendirme için gelişmiş makine öğrenme teknikleri kullanılmaya başlanmıştır. Çalışma kapsamında, konut tipindeki taşınmazların toplu değerlendirme için mahalli, konumsal ve yapısal gruplarda tanımlanan 120'den fazla kriter belirlenmiştir. Belirlenen çalışma alanında kriterleri temsil eden coğrafi veriler, çeşitli coğrafi analiz teknikleriyle uygulama için kullanılır hale getirilmiştir. Ardından Pearson korelasyon analizi ile 54 kriter model geliştirmede kullanılmak üzere seçilmiştir. Taşınmaz değer veri setinde, Boxplot grafik yaklaşımı ve Kümeleme & Aykırılık Analizi teknikleriyle verilerin genel dağılımından sapan 444 aykırı veri elimine edilmiştir. Makine öğrenmesine dayalı tahmin modelleri için çalışma alanındaki piyasa örneklemeleri ile %70 eğitim ve %30 test oranlarında veri setleri oluşturulmuştur. Rastgele Orman ve performans karşılaştırma için Çoklu Doğrusal Regresyon Analizi (ÇDR) makine öğrenme teknikleriyle tahmin modelleri geliştirilmiştir. Model performansları test veri seti üzerinden değerlendirilmiştir. Daha başarılı sonuçlar gösteren Rastgele Orman modeli ile kriter önem düzeyleri hesaplanmıştır. Konutun bulunduğu kat, konut alanı, bina yaşı, cephe yönü gibi yapısal ve otobüs durakları, eczane, marketlere yakınlık gibi konumsal kriterlerin ön plana çıktığı tespit edilmiştir. Ayrıca Rastgele Orman modeli ile CBS tabanlı üretilen taşınmaz değer haritasında mahallelerin taşınmaz değer dağılımları irdelenmiştir. Makine öğrenmesi teknikleri ve CBS'nin bütünsel kullanımıyla üretilen toplu taşınmaz değer haritaları taşınmaz yönetimi ihtiyaçları için etkin biçimde kullanılabilir.

Anahtar Kelimeler: CBS, ÇDR, Toplu Taşınmaz Değerleme, Makine Öğrenmesi, Rastgele Orman

ABSTRACT

Real estate is an essential component of the country's economy because real estate value is used from investment planning to loan arrangements, from

taxation to various sectoral practices. The objective determination of the real estate value with popular approaches has become an important requirement for the development of planned and sustainable real estate management strategies. Therefore, in addition to individual valuation approaches, advanced machine learning techniques have been used for mass real estate valuation. Within the scope of the study, more than 120 criteria defined in local, geographical and structural groups were determined for the mass valuation of residential real estate. Geographical data representing the criteria in the determined study area were made available for case study with various geographic analysis techniques. Then, with Pearson correlation analysis, 54 criteria were selected to be used in model development. In the real estate value dataset, 444 outlier data that deviated from the general distribution of the dataset were eliminated with the Boxplot graphical approach and Cluster & Outlier Analysis techniques. For the prediction models based on machine learning, datasets were prepared with 70% training and 30% testing rates of the market samples in the study area. Prediction models were developed with Random Forest and Multiple Linear Regression Analysis (MRA) machine learning techniques for performance comparison. Model performances were evaluated on the test dataset. Criteria importance levels were also calculated with the Random Forest model, which providing more successful results. It was determined that structural criteria such as the housing floor level, housing floor area, building age, direction of housing facade, and geographical criteria such as proximity to bus stops, pharmacies, and markets come to the fore. Besides, the real estate value distributions of the neighbourhoods were examined in the GIS-based real estate value map produced with the Random Forest model. Mass real estate value maps produced with the integrated use of machine learning techniques and GIS can be effectively used for real estate management needs.

Keywords: GIS, MRA, Mass Real Estate Valuation, Machine Learning, Random Forest

1. GİRİŞ

Taşınmazlar özel mülkiyetin ve kamu alanları yönetiminin temel alanlarından biri olarak kabul edilmektedir. Taşınmaz varlıkların uzun vadeli bir yatırım aracı olması, gelir getirme özelliği, bir güvence veya karşılık olarak kullanılabilmesi

taşınmazları oldukça cazip hale getirmektedir (Yalpir ve diğerleri, 2021). Taşınmazların yatırım amaçlı önemi tüm dünyada olduğu gibi ülkemizde de gittikçe artmaktadır. T.C. Cumhurbaşkanlığı Finans Ofisi tarafından 2019 yılında gerçekleştirilen Türkiye Hane Halkı Finansal Algı ve Tutum Araştırması sonuçlarına göre, taşınmazların diğer altın, döviz ve hisse senedi gibi finansal yatırım araçlarına kıyasla getiri miktarı ve güvenilirliğinin daha yüksek olması, uzun vadede değer kaybetmemesi gibi faktörlerden dolayı ortalama %81 ile en çok memnuniyet gören yatırım aracı olarak ifade edilmiştir (T.C-CBFO, 2020). Bu durum taşınmazların değer bakımından önemini ortaya koymakta olup, taşınmazların değerinin doğru tespit ve yönetiminin ihtiyacına işaret etmektedir. Taşınmazların güncel koşullarda alım satımına esas piyasa değerinin belirlenmesi ise taşınmaz değerlendirme disiplininin konusunu oluşturmaktadır. Bu kapsamda taşınmaz değerlendirme, değerlemenin yapıldığı günün koşullarına göre bir taşınmazın, taşınmaz üzerindeki hak ve faydalarının ya da taşınmaz üzerindeki projenin taşınmaza ilişkin fayda, nitelik, çevresel özellikler, konumsal özellikler, sosyo-demografik özellikler ve kullanım koşulları gibi kriterlerin birlikte değerlendirilmesi suretiyle olası piyasa değerinin objektif olarak belirlenmesi şeklinde ifade edilmektedir (Appraisal Institute, 2001, Açlar ve Çağdaş, 2008).

Günümüzde farklı uygulama alanlarında oldukça sık ihtiyaç duyulan taşınmaz değerinin belirlenmesi söz konusu olduğunda tekil ve toplu taşınmaz değerlendirme kapsamında iki kategori altında çeşitli yöntemler kullanılmaktadır. Tekil değerlendirme yöntemleri değer tespitinde yeterli objektifliği sağlayamamakta olup, değere etki eden kriterler ile değer arasındaki ilişkiler yeterli düzeyde değerlendirilememektedir (Pagourtzi ve diğerleri, 2003; Sisman ve diğerleri, 2021). Dolayısıyla çok sayıda taşınmazın eş zamanlı ve objektif olarak değerinin belirlenebildiği, bilimsel temellere dayalı kantitatif sonuçlar üretebilen, çok sayıda kriterin birlikte değerlendirilebildiği modern değerlendirme yaklaşımlarına gereksinim bulunmaktadır (Renigier-Bilozor ve diğerleri, 2019) Bu kapsamda, teknolojik gelişmeler, kamu ve özel sektörün hızlı ve daha düşük maliyetli değerlendirme isteği, taşınmaz değerinin farklı ihtiyaçlara altlık teşkil etmesi, farklı kaynaklardan gelen büyük verinin işlenerek değerlendirme faaliyetlerinde kullanılması gibi hususlar göz önüne alındığında; yapay zekânın bir alt dalı olan makine öğrenmesi teknikleri ve CBS'ye dayalı taşınmaz değer modellerinin üretilmesi ihtiyacı ortaya çıkmıştır.

Ülkemizde değerlendirme konusunda kurumsal anlamda yapılanma 2019 yılında Tapu Kadastro Genel Müdürlüğü (TKGM) bünyesinde Taşınmaz Değerleme Daire Başkanlığı'nın kurulması ile olmuştur (Resmi Gazete, 2018; Resmi Gazete, 2019). Kurulan Taşınmaz Değerleme Daire Başkanlığı, en başta ülke çapında taşınmazların değerinin tespit edilmesi ve bu sayede taşınmazlardan alınan vergilerin ve harç kayıplarının minimuma indirgenmesini hedeflemektedir. Ayrıca başkanlık tarafından taşınmazlara ilişkin fiyat endeksleri ve değer haritalarının üretilmesi ile kamulaştırma ve arazi toplulaştırma gibi uygulamalara altlık sağlamak gibi çeşitli amaçlarla bir değer bilgi bankasının oluşturularak esasta Toplu Değerleme Sistemi'nin kurulması amaçlanmaktadır. Kurulacak olan sistem ile uluslararası standartlarda bilimsel temellere uygun olarak güncel değer tahminleri gerçekleştirileği ön görülmektedir (TKGM, 2022).

Literatürde taşınmazların toplu değerlendirme için Hedonik Değerleme (Luo ve diğerleri, 2021), Çoklu Doğrusal Regresyon (ÇDR) (Sisman ve Aydınoglu, 2022), Destek Vektör Makineleri (DVM) (Wang ve diğerleri, 2014), Karar Ağaçları (Antipov ve Pokryshevskaya, 2012), Rastgele Orman (Çeh ve diğerleri, 2018), Yapay Sinir Ağları (YSA) (Yalpir, 2018; Yilmazer ve Kocaman, 2021), Nominal Değerleme (Mete ve Yomralioglu, 2021), Bulanık Mantık (Bovkir ve Aydınoglu, 2018) ve Çok Ölçütlü Karar Verme (ÇÖKV) (Del Giudice ve diğerleri, 2017; Unel ve Yalpir, 2019) gibi birçok teknik mevcuttur. Bu teknikler arasında makine öğrenmesi teknikleriyle üretilen modeller göz önüne alındığında karar ağaçları esaslı bir yöntem olan Rastgele Orman tekniğinin, farklı uygulamalarda optimum parametre seçimi kolaylığı, regresyon problemlerinin çözümünde kolay uygulanabilir olması, modele giren değişkenlerin önem düzeyi hakkında bilgi vermesi ve yüksek model performansı sağlamasından dolayı tercih sebebi olduğu ifade edilmiştir (Antipov ve Pokryshevskaya, 2012; Aydınoglu, 2019; Hong ve diğerleri, 2020; Aydınoglu ve diğerleri, 2021). Aydınoglu ve diğerleri, Rastgele Orman tekniği ile İstanbul ili Pendik ilçesinde bulunan 1459 konut tipindeki taşınmaz için 34 kriter ile değer tahminleri gerçekleştirerek, taşınmaz değer haritası ve kriter önem düzeyi bilgisi üretmişlerdir (Aydınoglu ve diğerleri, 2021). Yilmazer ve Kocaman ise Ankara ili Mamak ilçesinde Rastgele Orman ve ÇDR (aşamalı olarak) tekniklerini kullanarak toplu değerlendirme modelleri üretmişlerdir. Modellerde 1200 piyasa örnekleme kullanılmış olup, aykırı veri olarak nitelendirilen veriler manuel olarak veri setinden çıkarılmıştır. Çalışmada değeri etkileyen 96 kriter

tanımlanmakla birlikte, değeri etkileyen ve modellere dahil edilen 37 kriter uzman görüşlerine dayalı olarak belirlenmiştir. Kriter arası çoklu doğrusallık problemi ise Pearson korelasyon analizi ile değerlendirilmiştir. Fakat bu çalışmalarda doğrudan en yüksek model performansının elde edilmesi amaçlanmıştır. Modelleme öncesinde kriter seçimi, veri setleri içerisindeki konumsal ve konumsal olmayan aykırı verilerin istatistiksel veri analizi ile tespiti ve ayıklanması gibi hususlar yeterince ele alınmamıştır.

Bu çalışmada farklı sosyo-gelişmişlik özelliklere sahip yerleşim alanlarında konut tipindeki taşınmazların makine öğrenmesi teknikleriyle daha büyük piyasa örnekleme ve kriter grubu ile toplu taşınmaz değerlemesi için uygulama metodolojisi belirlenmesi amaçlanmaktadır. Bu kapsamda toplu değerlendirme sürecinde konumsal ve konumsal olmayan aykırı verilerin tespiti ve ayıklanması, başarılı sonuç veren Rastgele Orman tekniği ile çok sayıda coğrafi ve coğrafi olmayan kriterin analiziyle tahmin modellerin oluşturulması, model performans değerlendirmelerin gerçekleştirilmesi ve değeri etkileyen kriterlerin önem düzeylerinin belirlenmesi amaçlanmaktadır. Çalışma alanı veri setlerinde aykırı verilerin ayıklanmasının ardından, Pearson korelasyon analizi kullanılarak toplu değerlemede kullanılacak taşınmaz değeri ile ilişkili kriterler belirlenmektedir. Rastgele orman makine öğrenmesi tekniği ile çoklu doğrusal regresyon yöntemi performansları karşılaştırılmaktadır. Böylelikle sürdürülebilir toplu taşınmaz değerlemesi için coğrafi analiz ve makine öğrenme tekniklerine dayalı örnek yaklaşım geliştirilmektedir.

2. MATERYAL ve METOT

a. Taşınmaz Değerini Etkileyen Kriterler

Taşınmaza ilişkin piyasa değerinin pratikte tespit edilmesi; taşınmazların sahip olduğu hukuki anlamdaki yasal özellikler, konumu ve kullanımı ile kişiden kişiye nitelik ve nicelik olarak değişkenlik gösteren birçok farklı unsurdan dolayı pek mümkün değildir. Bununla birlikte taşınmaz değerinin ihtiyaç duyulduğu farklı uygulamalar için taşınmazların çeşitli özellikleri ile tahmini değerlerini belirlemek mümkündür. Taşınmazların değeri, çok boyutlu yapısı nedeniyle yasal, fiziksel, konumsal, mahalli, ekonomik ve sosyal faktörlerden etkilenebilmektedir (Bovkir ve Aydinoglu, 2018, Sisman ve diğerleri, 2021).

Çalışma kapsamında taşınmaz değerini

etkileyen kriterler, Avrupa Değerleme Standartları (TEGoVA, 2016), Uluslararası Değerleme Standartları (IVSC, 2016), Değerleme-Küresel Standartları (RICS, 2017), Taşınmazların Toplu Değerleme Standartları (IAOO, 2013), Ulusal Standartlar (TDUB, 2011), yönetmelikler (TKGM, 2011) ve akademik çalışmalar dikkate alınarak irdelenmiştir. Böylelikle taşınmaz değerini etkileyen kriterler konumsal, mahalli ve yapısal olmak üzere 3 tematik kategori içerisinde farklı alt kriter gruplarında tanımlanmıştır (Tablo 1).

Tablo 1. Taşınmaz değerini etkileyen kriterlerin belirlenmesi).

Mahalli Kriterler	Nüfus (İlçe Nüfusu, İlçe Nüfus Yoğunluğu (Kişi/km ²), Mahalle Nüfusu, Mahalle Nüfus Yoğunluğu (Kişi/km ²), Mahallede Yaşayan Çocuk (0-14), Genç (15-24), Yetişkin (25-65), Yaşlı (65+), Kadın, Erkek Yüzdesi) Gelir Düzeyi Dağılımı (A+, A, B, C, D Grubu Gelir Düzeyi Dağılımı, Mahallede Yaşayanların Kişi Başı Geliri, Mahallede Yaşayanların Kişi Başı Tasarrufu) Eğitim Durumu (Mahallede Yaşayan Okuma Yazma Bilmeyen, Okuması Olan Eğitimsiz, İlkokul, İlköğretim, Ortaokul Mezunu, Lise Mezunu, Üniversite Mezunu Kişi Yüzdesi) Medeni Durum (Mahallede Yaşayan Evli Çift, Bekar, Boşanmış Çift, Dul Yüzdesi) Yapı Yoğunluğu (Yapılaşma Yoğunluğu) İl & İlçe Sosyo-Ekonomik Gelişmişlik (İl Sosyo-Ekonomik Gelişmişlik Düzeyi, İlçe Sosyo-Ekonomik Gelişmişlik Düzeyi, İlçenin İl İçerisindeki Sosyo-Ekonomik Gelişmişlik Sıralaması), Mahalli Envanter (Hane Sayısı, Toplam Konut Sayısı, Özel iş yeri, Banka Sayısı, ATM Sayısı, Eczane Sayısı, Otomobil Sayısı, Araç Sayısı), Arsa Rayiç Değeri
Konumsal Kriterler	Eğitim Tesislerine Mesafe (Anaokullarına, İlk-Orta Okullara, Liselere, Üniversitelere Mesafe) Sağlık Tesislerine Mesafe (Aile Sağlık Merkezlerine, Hastanelere, Eczanelere Mesafe) Alışveriş&Ticaret Tesislerine Mesafe (Çarşı Merkezlerine, Alışveriş Merkezlerine, Marketlere, Pazar Alanlarına, Postane/Kargolara Şirketlerine Mesafe) Kültürel Tesislere Mesafe (Fuar Merkezler, Sinema ve Tiyatrolara, Müzelere, Kongre ve Kültür Merkezlerine, Kütüphanelere Mesafe) Kamu Hizmet Tesislerine Mesafe (İdari Tesislere, Adliye, Bankalara, ATM'lere, İtfaiye İstasyonlarına, Güvenlik Birimlerine Mesafe) Yeşil Alanlara Mesafe (Küçük Parklara, Büyük Parklara, Ormanlara Mesafe) Eğlence&Spor Tesislerine Mesafe

	(Sahile, Plajlara, Lunaparklara, Spor Tesislerine Mesafe) Konaklama Tesislerine Mesafe (Pansiyonlara, Otellere Mesafe) Sanayi Tesislerine Mesafe (Akaryakıt, Endüstri Tesislerine Mesafe) Dini Tesislere Mesafe (İbadethanelere, Mezarlıklara Mesafe) Çevresel Özellikler (Yıllık Ortalama Hava Kalitesi, Yıllık Ortalama Sıcaklık, Yıllık Ortalama Yağış), Topografik Özellikler (Eğim, Bakı), Raylı Sistem Fonksiyonlarına Mesafe (Metro, Marmaray İstasyonlarına Mesafe) Karayolu Fonksiyonlarına Mesafe (Otobüs Duraklarına, Taksi Duraklarına, Otoyollara, Caddeye, Sokaklara Mesafe, Otoparklara, Otogara Mesafe) Havayolu Fonksiyonlarına Mesafe (Havaalanlarına mesafe) Denizyolu Fonksiyonlarına Mesafe (Deniz Ulaşım İstasyonlarına Mesafe)
Yapısal Kriterler	Bağımsız Bölüm Özellikleri Konut Alanı, Oda Sayısı, Konut Oda Tipi, Bina Yaşı, Banyo Sayısı, Isıtma Sistemi, Konutun Bulunduğu Kat, Bina Toplam Kat Adedi, Manzara, Cephe Yönü, Balkon Mevcudiyeti Eklentiler Yüzme Havuzu, Asansör Mevcudiyeti, Otopark Mevcudiyeti, Site İçinde Olma, Çocuk Oyun Parkı Mevcudiyeti

b. Çalışma Alanı ve Veri Setleri

(1) Çalışma Alanının Belirlenmesi

İstanbul ilinin Anadolu yakasındaki Tuzla ilçesi ile Kocaeli ilinin Gebze, Çayirova ve Darıca ilçelerinde bulunan toplam 23 mahalle çalışma alanı olarak belirlenmiştir (Şekil 1). Bu çalışma alanı, Şişman (2021) tarafından yapılan araştırmaya göre mahalli sosyo-ekonomik değişkenlerin coğrafi kümeleme teknikleriyle analiz edilmesiyle tanımlanmıştır. Diğer taraftan yüzölçümü bakımından çalışma alanındaki en büyük ilçe Gebze, en küçük ilçe ise Darıca'dır. 2020 yılı itibarıyla, Tuzla ilçesi 273.608, Kocaeli'nin en kalabalık ilçesi Gebze 392.945, Çayirova ilçesi 140.274 ve Darıca ilçesi ise 214.796 nüfusa sahiptir (TÜİK, 2020). Ayrıca Tuzla ve Gebze ilçelerinde kırsal ve kentsel yerleşim alanları bulunmakla birlikte, farklı arazi kullanımlarına sahiptirler. Diğer taraftan başta Gebze olmak üzere Çayirova ve Darıca ilçeleri, kozmopolit yapısına rağmen önemli sanayi tesislerine sahiptir. İstanbul'un Tuzla ilçesi ise sınır ilçe özelliğinin yanı sıra, Tuzla-Gebze sınırı bölgesel sosyo-ekonomik gelişmişlik farklılıkları

barındırması ile (SEGE, 2022) toplu taşınmaz değerlemede bölgesel değer dağılımlarının değerlendirilmesi için doğru bir pilot uygulama alanı olarak kabul edilebilir.

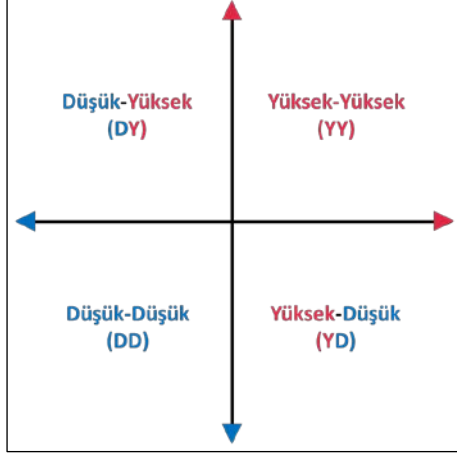
(2) Veri Setlerinin Temin Edilmesi

Çalışma alanı için Tablo 1'de tanımlanan mahalli, konumsal ve yapısal kriterleri temsil eden 3 farklı veri seti oluşturulmuştur. Mahalli kriterler veri seti, taşınmaz değerini etkileyen mahalli kriterlere ilişkin mahalle-ilçe-il idari birim düzeylerinde veriler içermektedir. Bu istatistik veri setleri, Pot-a Geohub veri portalı (GeoHub, 2020), TÜİK istatistik veri portalı (TÜİK, 2020) ve T.C. Sanayi ve Teknoloji Bakanlığı Kalkınma Ajansları Genel Müdürlüğü il ve ilçelerin Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırmaları raporları (SEGE, 2022) üzerinden temin edilmiştir. Konumsal kriterler veri seti doğrudan Tablo 1'de tanımlanan konumsal kriterlere ilişkin verileri içermektedir. Bu veri setleri çalışma kapsamında gerçekleştirilen coğrafi analizlerin girdisi niteliğinde olup, İstanbul Büyükşehir Belediyesi'nden ve Kocaeli Büyükşehir Belediyesi'nden temin edilmiş araştırma amaçlı kullanılan coğrafi veri altlıklarıdır. Yapısal kriterler veri seti, taşınmazların piyasa değerlerini ve Tablo 1'de tanımlanan yapısal kriterlere ilişkin verileri içermektedir. Ayrıca taşınmazların konum bilgileri de taşınmaz değer modelleri ve coğrafi kümeleme analizleri için girdi niteliğinde olup, REIDIN gayrimenkul bilgi şirketi tarafından tutulan doğrulanmış veri tabanlarından temin edilmiştir (REIDIN, 2022). Veri seti 01.01.2019 ile 09.03.2020 tarihleri arasında ait 32.315 piyasa örnekleme içermektedir. Veri seti içerisindeki aykırılıklar giderildikten sonra modelleme çalışmasında kullanılan 31.871 piyasa örnekleme ilişkin özet istatistikler Tablo 2'de verilmiştir.

Tablo 2. Veri seti tanımlayıcı istatistikleri.

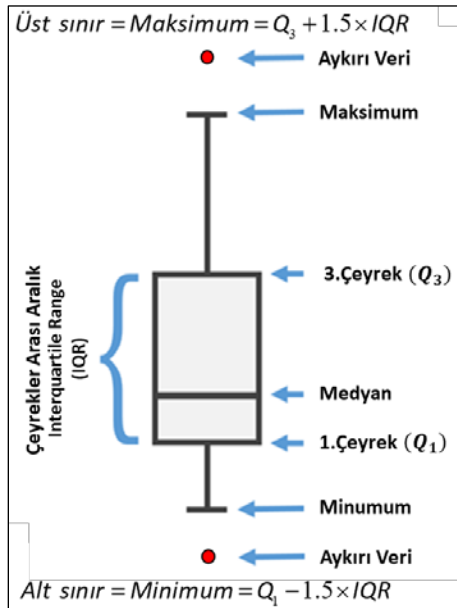
Veri seti Özet İstatistikleri	
Örneklem Sayısı	31.871
Maksimum Değer (₺)	640,000
Minimum Değer (₺)	85,000
Ortalama Değer (₺)	243524,90
Standart Sapma (₺)	71324,39
Ortalama Konut m ² Fiyatı (₺)	2232,06

düşük değerli verilerle çevrili, yüksek değerlere sahip verileri göstermektedir. Dolayısıyla Düşük-Yüksek ve Yüksek-Düşük içerisine düşen veriler coğrafi aykırı veriler olarak değerlendirilmektedir (Anselin ve diğerleri, 2007).



Şekil 2. Kümelenme ve Aykırılık analizi ile aykırı veri analizi.

Boxplot, grafik tabanlı aykırı veri tespit etme yaklaşımıdır (Dovoedo ve Chakraborti, 2015). Grafikte, veri seti içerisindeki değişkenler minimum, maksimum, medyan (ortanca), 1. çeyrek ve 3. çeyrek olmak üzere beş farklı istatistiki bilgi açısından değerlendirilmektedir (Şekil 3). Çeyrekler arası aralık dışındaki veriler aykırı kabul edilir.



Şekil 3. Boxplot yaklaşımı ile aykırı veri analizi.

Ç. Pearson Korelasyon Analizi

Pearson korelasyon analizi, veri kümesindeki iki sürekli değişken arasındaki ilişkileri değerlendiren kovaryans tabanlı bir istatistiksel analiz tekniğidir. Pearson korelasyon katsayısı Denklem-1 ile (Kalantar ve diğerleri, 2020) hesaplanır:

$$r_{xy} = \frac{\sum_{i=1}^n \frac{X_i - \bar{X}}{\sum_{k=1}^n (X_k - \bar{X})^2} \times \frac{Y_i - \bar{Y}}{\sum_{k=1}^n (Y_k - \bar{Y})^2}}{\sum_{k=1}^n (Y_k - \bar{Y})^2} \quad (1)$$

Eşitlikte r_{xy} Pearson korelasyon katsayısını temsil eder ve iki değişken arasındaki doğrusal bağımlılığın bir ölçüsüdür. -1 ile +1 arasında bir değer alan bu katsayının -1 değerine yakın olması, güçlü negatif ilişkiler anlamına gelirken, +1'e yakın değer ise, değişkenler arasında güçlü pozitif ilişkiler anlamına gelir. Ayrıca 0 değeri, değişkenler arasında korelasyon olmadığı anlamına gelir. X_i ve Y_i veri setindeki i'nci değişken için X ve Y'nin ilgili değerini temsil eden $\bar{X}$ ve $\bar{Y}$ ise sırasıyla X ve Y açıklayıcı değişkenlerinin ortalamasını ifade eder (Moon ve diğerleri, 2008, Egghe ve Leydesdorff, 2009, Bujang ve diğerleri, 2010). Pearson korelasyon analizi, makine öğrenmesine dayalı modelleme çalışmalarında çoklu doğrusallık probleminden kaçınmak için birbiri ile yüksek korelasyonlu kriterlerin tespitinde etkin olarak kullanılmaktadır. Korelasyon katsayısının $r_{xy} \geq 0,5$ 'ten büyük olması ile kriterler yüksek korelasyonlu olarak değerlendirilmektedir (Yilmazer ve Kocaman, 2020).

d. Çoklu Doğrusal Regresyon

Çoklu Doğrusal Regresyon (ÇDR), veri seti içerisinde bağımlı ve birden fazla bağımsız değişken arasındaki ilişkinin irdelenmesi için doğrusal formda matematiksel denklemler türeterek, istatistiksel temellere dayalı tahminler yapan denetimli bir makine öğrenme tekniğidir (Ünel, 2017, Akar ve Yalpir, 2021). ÇDR tekniğinde bağımlı değişkendeki değişim, kurulan regresyon modelindeki birden fazla bağımsız değişken ile eş zamanlı olarak açıklanmaya çalışılmaktadır. ÇDR tekniğinin yapısı Denklem-2 ile verilmektedir:

$$y_i = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \dots + \beta_n \times X_n + \varepsilon_i \quad (2)$$

Eşitlikte, y_i bağımlı değişkeni (örneğin taşınmaz değeri), $X_1, X_2, \dots, X_n$ modelde kullanılan bağımsız değişkenleri (alan, oda sayısı ve banyo sayısı gibi değeri etkileyen kriterler), β_0 model

sabit katsayısını, $\beta_1, \beta_2, \dots, \beta_n$ kriterlere ilişkin katsayıları, $\mathcal{E}_i$ hata terimini ifade etmektedir. Hata terimi, tahmin edilen değer ile piyasa değeri arasındaki, farkı belirlemektedir. Ayrıca ÇDR tekniğinde, karesel tahmin hatalarının toplamını en aza indirgenerek regresyon modeline ilişkin model sabiti β_0 ve bağımsız değişkenlere ilişkin β katsayıları optimize edilmektedir.

ÇDR tekniği taşınmaz değerlendirme uygulamaları kapsamında; taşınmaz değerini etkileyen kriterlerin belirlenmesi, değeri etkileyen kriterlerin değer üzerindeki etkisinin belirlenmesi, taşınmaz değer tahmin modeli geliştirme, taşınmaz değeri üzerindeki zamansal değişimlerin incelenmesi ve farklı algoritmalarla oluşturulan taşınmaz değer modellerinin karşılaştırılmasında kullanılmıştır (Yalpir, 2018; Yilmazer ve Kocaman, 2020).

e. Rastgele Orman Tekniği

Rastgele Orman (Random Forest) tekniği, Breiman (2001) tarafından sınıflandırma ve regresyon problemlerinin çözümü için geliştirilen bir topluluk öğrenme algoritmasıdır (Antipov ve Pokryshevskaya, 2012). Karar ağacı algoritması türü olan rastgele orman tekniği ile modellerin eğitimi gerçekleştirilirken birbirinden bağımsız olarak oluşturulan birçok karar ağacının bir araya getirilerek bir karar ormanı oluşturması amaçlanmaktadır (Breiman, 2001). Başka bir ifade ile rastgele orman algoritmasında korelasyonsuz birden çok karar ağacı yapısının bir araya getirilmesi sonucunda bir karar ormanı oluşturulması ve karar ağaçlarının her biri tarafından yapılmış olan tahminlerin birleşerek tek bir tahmin yapılması temel esastır. Bu yönüyle teknik, tahmin modellerinin performansını artıran bir makine öğrenmesi tekniğidir (Yilmazer ve Kocaman, 2020).

Rastgele orman tekniği, eğitim veri seti kullanılarak birbirinden bağımsız bir şekilde karar ağaçlarının oluşturulduğu torbalama tekniği (Breiman, 1996) ve veri setinde bulunan tüm öznitelikler içerisinde rastgele seçilen az sayıdaki öznitelik ile karar ağacı yapısındaki her düğüm noktasında en iyi dallanmayı gerçekleştiren özniteliğin seçildiği rastgele alt uzay (Ho, 1998) tekniğinin birleşimidir. Rastgele orman tekniği, dallanma için özniteliklerin belirlenmesinde Gini indeksi yaklaşımını kullanmaktadır (Pal, 2005). Veri seti içerisindeki özniteliklerin önem dereceleri ve etkileşimlerinin ortaya çıkması gibi avantajları, bu tekniği diğer tekniklere kıyasla ön plana çıkarmaktadır.

Rastgele orman tekniği, karar ormandaki her

bir karar ağacının eğitilebilmesi amacıyla orijinal eğitim veri seti içerisinde rastgele örnekleme ile alt kümeler oluşturur. Bu alt kümelerin 2/3'ü karar ağacı yapısı oluşturmada kullanılırken, torba dışı olarak da bilinen geri kalan 1/3'lük kısım ise ağaç yapısını test etmek için kullanılır (Çölkesen, 2015). Bu oransal dağılımın belirlenmesi ise modellerin eğitimi esnasında kullanılan ilgili yazılım kütüphanelerinde rastgele seçilen verilerle otomatize biçimde gerçekleştirilmektedir. Rastgele orman algoritmasında kullanıcı tarafından belirlenen iki önemli parametre model başarısı açısından oldukça önemlidir. İlk parametre ağaç yapısının oluşturulmasında her bir düğümde kullanılacak özniteliklerin sayısıdır. İkinci parametre ise karar ormanındaki toplam ağaç sayısıdır (Aydinoglu ve diğerleri, 2021).

f. Model Performans Değerlendirme Ölçütleri

Rastgele Makine öğrenmesi tekniklerine dayalı olarak gerçekleştirilen regresyon problemlerinde, farklı algoritmalarla ilişkin tahmin doğruluğunun değerlendirilmesi için çeşitli model performans analizi yaklaşımları bulunmaktadır. Bunlardan en çok kullanılanları; Ortalama Mutlak Hata (OMH), Ortalama Karesel Hata (OKH), Karesel Ortalama Hata (KOH), R^2 ölçütleridir. Başarılı toplu değerlendirme modellerinin geliştirilmesinde OMH, OKH ve KOH ölçütlerin mümkün olduğunca 0 değerine yakın olması beklenmektedir. R^2 ölçütü ise birimsiz olup, 1 değerine yakın olması yüksek model açıklama performansını temsil etmektedir. Bu ölçütlere ilişkin genel matematiksel denklemler sırasıyla aşağıda verilmiştir.

$$OMH = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

$$OKH = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$KOH = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

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

Diğer taraftan taşınmaz değerlendirme uygulamaları kapsamında IAAO tarafından önerilen model performans değerlendirme ölçütleri bulunmaktadır. Bunlar arasında en sık kullanılanları Dağılım Katsayısı (DKA) ve Fiyat İlişkili Farklılıklar (FİF) ölçütleri şeklindedir. DKA

değişkenlik veya tekdüzeliğin ölçüsü olarak tanımlanmakta olup değerlendirme oranlarının medyan orandan ortalama yüzde sapmasını değerlendirmektedir. IAAO oran standartlarına göre dört farklı grup içerisinde değerlendirilen DKA değeri Denklem-7 yardımıyla hesaplanmaktadır. Daha eski veya daha fazla heterojen bölgelerdeki müstakil evler ve kat mülkiyetleri için hesaplanan DKA değeri 5 ile 15 arasında bir değer olmalıdır. FİF ölçütü ise, Denklem-8 ile hesaplanmakta olup ortalama değerlendirme oranının ağırlıklı ortalama oranına bölünmesiyle hesaplanan 0,98 ile 1,03 arasında bir değer almalıdır. FİF tahmin modellerinde düşük ve yüksek değerli mülkler arasındaki değerlendirme-satış fiyatı oranlarındaki tekdüzelik düzeyini değerlendirmektedir.

$$DKA = \%100 \times \frac{\left(\sum_{i=1}^n |R_i - \bar{R}| \right) / n}{\bar{R}} \quad (7)$$

$$PRD = \frac{\bar{R}}{\sum_{i=1}^n \hat{y}_i / \sum_{i=1}^n y_i} \quad (8)$$

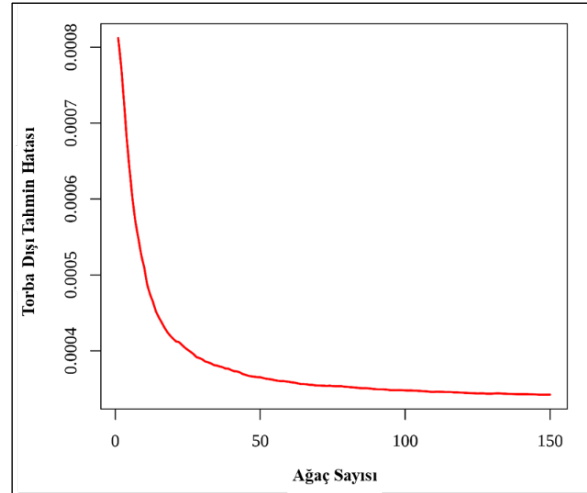
y_i taşınmazın piyasa değeri, $\hat{y}_i$ taşınmazın tahmin edilen değerini, $\bar{y}_i$ piyasa değerleri ortalamasını, $i = 1,2,3, \dots, n$ olmak üzere n toplam taşınmaz sayısını, R tahmin edilen ve piyasa değeri arasındaki değerlendirme oranını, $\bar{R}$ değerlendirme oranlarının ortalama değerini ve $\bar{R}$ değerlendirme oranlarının medyan değerini göstermektedir (IAAO, 2013, Levantesi ve Piscopo, 2020, Yilmazer ve Kocaman, 2020; Sun ve diğerleri, 2021, Yalpir ve diğerleri, 2021).

3. BULGULAR

a. Toplu Taşınmaz Değerleme Modeli Geliştirme

Tahmin modellerinin geliştirilmesi öncesinde taşınmaz değer veri seti Boxplot ve Kümelenme & Aykırılık Analizi teknikleriyle değerlendirilerek toplamda 444 adet aykırı veri tespit edilerek veri seti içerisinde çıkarılmıştır. Veri seti içerisindeki birim farklılıklarını gidermek için birçok makine öğrenimi modellerinin geliştirilmesinde olduğu gibi veri seti normalize edilmiştir. Ardından modelleme aşamasında çoklu doğrusallık probleminden kaçınmak için mahalli ve konumsal kriterler arasındaki ilişki Pearson korelasyon analizi ile değerlendirilmiştir. Kriterler ikili olarak değerlendirilerek, yüksek korelasyona ($r_{xy} \geq 0,5$) sahip kriterlerden birisi tercih edilmiştir. Bu

kapsamda 54 kriter modelleme için belirlenmiştir. Çalışma alanı için toplamda 31.871 piyasa örnekleme ve 54 kriter ile Rastgele Orman ve ÇDR teknikleri kullanılarak toplu taşınmaz değer tahmin modelleri üretilmiştir. Modellerin geliştirilmesinde açık kaynak kodu R yazılımı kullanılmıştır. Rastgele Orman modeli için “randomForest” kütüphanesi (randomForest, 2022) kullanılırken ÇDR modeli için R yazılımı içinde doğrudan tanımlı “lm” fonksiyonu (lm, 2022) kullanılmıştır. Örneklemelerin 22.310 tanesi (veri setinin yaklaşık %70’i) modellerin eğitimi, 9.561 tanesi (veri setinin yaklaşık %30’u) ise modellerin test edilmesi için kullanılmıştır. Modeldeki ağaç yapısının oluşmasında her düğüm noktasında kullanılacak öznelite/kriter sayısı 18 (54/3), kullanılacak ağaç sayısı ise 150 olarak belirlenmiştir. Eğitim esnasında farklı ağaç sayılarına karşılık gelen torba dışı tahmin hatası grafiği ise Şekil 4’te verilmiştir.

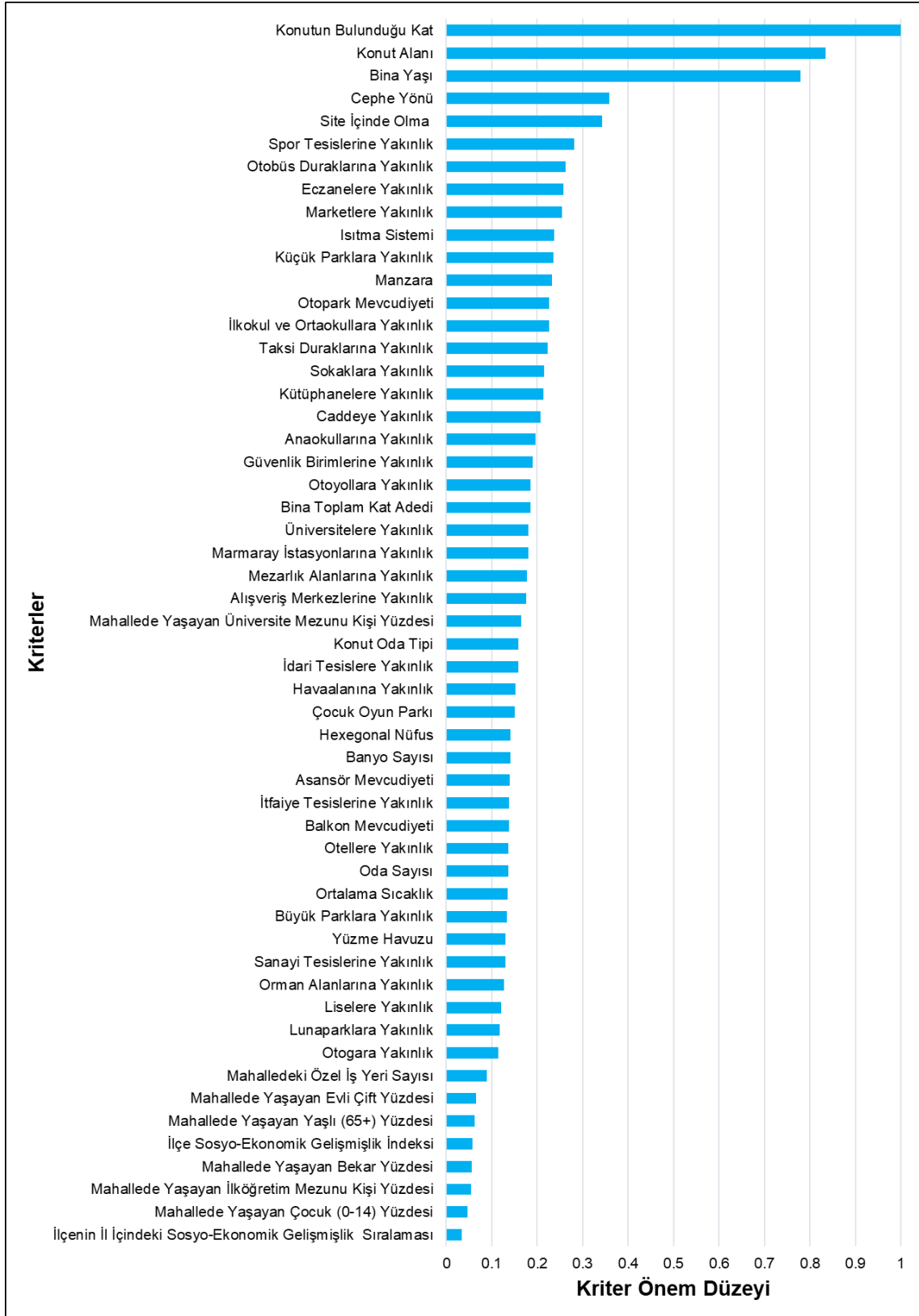


Şekil 4. Rastgele Orman modeli torba dışı tahmin hatası grafiği.

Not: Çalışma kapsamında normalize edilmiş veri seti kullanıldığından dolayı Rastgele Orman modeli torba dışı tahmin hatası grafiğindeki hata miktarlarının (düşey eksen) herhangi bir birimi yoktur.

b. Kriter Önem Düzeyleri ve Model Performanslarının İrdelenmesi

Çalışma alanı için üretilen Rastgele Orman modeli üzerinden kriter önem düzeyleri belirlenmiştir. Kriter önem düzeyleri irdelendiğinde yapısal ve konumsal kriterlerin ön plana çıktığı tespit edilmiştir. Bina yaşı, konut alanı ve konutun bulunduğu kat, cephe yönü, site içinde olma gibi kriterleri diğer kriterlere kıyasla ön plana çıkmıştır. Diğer taraftan mahalli kriter grubundaki ilçenin il içindeki sosyo-ekonomik gelişmişlik sıralaması ve mahallede yaşayan çocuk (0-14) yüzdesi kriterleri modelde en düşük öneme sahip kriterler olmuştur.



Şekil 5. Rastgele Orman modeli kriter önem düzeyleri

Modeli oluşturan kriterlerin tümüne ilişkin kriter önem düzeyleri Şekil 5'te verilmiştir. Kriter önem düzey sonuçlarına göre bina yaşı yüksek önem düzeyine sahip kriterler arasındadır. Bu durum farklı perspektiflerden değerlendirilebilir. Bina yaşının artması sonucu konut fiyatlarının düşmesi bir piyasa koşulu olarak değerlendirilebilir. İnşaat yapım maliyetlerinin artması da yeni yapılan binalar için maliyet unsuru olarak doğrudan değer üzerinde yansımaktadır. Yapım maliyetinden ve yapılaşma yoğunluğunun neredeyse doymuş seviyeye ulaşmasından kaynaklı yeni konut sayısının azlığı problemi bina yaşının önemli bir kriter olarak ortaya çıkmasına neden olmuştur. Diğer taraftan çalışma alanı AFAD tarafından 2018 yılında yayınlanan Türkiye Deprem Tehlike Haritası'na göre riskli alanlar içerisinde değerlendirilebilir (AFAD, 2018). 1999 yılından sonraki binalar deprem yönetmeliğine göre yapıldığı için, yeni yapılan binaların fiyatı maliyet kaynaklı yüksek olabilmektedir. Başka bir ifade ile daha kaliteli bina stoğunun oluşması gibi unsurlar, bina yaşı kriterinin önem düzeyinin yüksek olmasının nedenleri arasında düşünülebilir. Ayrıca yeni bina alım satımına banka kredilerinin uygun olması da bina yaşı kriterinin yüksek öneme sahip olmasının nedenleri arasında değerlendirilebilir.

Normalize edilmiş verilerle oluşturulan Rastgele Orman ve ÇDR modellerine ilişkin model performansları 9.561 test örnekleme üzerinden hesaplanarak kıyaslanmıştır. Sonuçlar makine öğrenmesi performans ölçütlerine göre karşılaştırıldığında Rastgele Orman modeli daha başarılı sonuçlar verirken, IAAO ölçütlerine göre kıyaslandığında ise her iki modelde başarı sağlanmaktadır (Tablo 3).

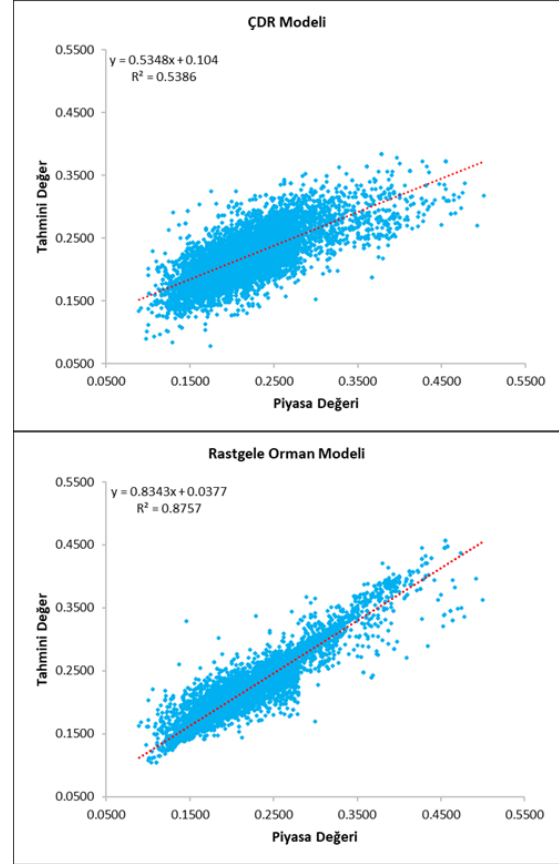
Tablo 3. Rastgele Orman ve ÇDR modeli performansları.

Performans Ölçütleri	Rastgele Orman	ÇDR
OMH	0,0112	0,0273
OKH	0,0004	0,0013
KOH	0,0190	0,0363
R ²	0,8757	0,5386
DKA	5,3798	12,4437
FİF	1,0102	1,0249

Not: Modelleme aşamasında normalize edilmiş veri seti kullanılmıştır. Dolayısıyla Rastgele Orman ve ÇDR modellerinin performanslarına ilişkin OMH, OKH ve KOH ölçütlerinin herhangi bir birimi yoktur. Başarılı toplu değerlendirme modelleri için bu ölçütlerin mümkün olduğunca 0 değerine yakın olması beklenmektedir. Benzer şekilde R² ölçütü de birimsiz olup, bu değer 1 değerine yakın olması yüksek model açıklama performansını temsil etmektedir. IAAO tarafından toplu değerlendirme modellerinin performanslarının değerlendirilmesi için önerilen DKA ve FİF ölçütleri ise birimsiz olup, doğruluk için belirli bir aralığı referans almaktadır.

Ayrıca piyasa değerleri ile Rastgele Orman ve ÇDR modellerinden tahmin edilen değerlerinin

veri saçılımları incelenmiştir. Rastgele orman modelinde veriler eğilim çizgisi etrafında kümelenmiş bir saçılım yapısı sergilerken, ÇDR modelinde verilerin daha dağınık bir saçılım yapısı sergilediği tespit edilmiştir (Şekil 6).



Şekil 6. Tahmin modelleri saçılım grafikleri.

Not: Çalışma kapsamında normalize edilmiş veri seti kullanıldığından dolayı saçılım grafiklerindeki yatay ve düşey eksenlerin herhangi bir birimi yoktur.

c. Taşınmaz Değer Tahmin Haritasının Üretilmesi

Değer tahmin modellerinin üretilmesi ve performanslarının karşılaştırılmasından sonra, daha başarılı tahmin sonuçları veren Rastgele Orman modeli üzerinden 30 m çözünürlüğe sahip CBS tabanlı taşınmaz birim değer haritası üretilmiştir. Haritaların üretiminde Ters Mesafe Ağırlıklandırma (TMA) enterpolasyon analiz tekniği kullanılmıştır (Şekil 7). Üretilen değer haritası incelendiğinde, bölgenin güneyinde Darıca ilçesinde yer alan Bayramoğlu mahallesinin geneli ile Emek mahallesinin belirli bir bölgesi, Çayırova ilçesinde bölgenin kuzeydoğusunda yer alan Yenikent ve Akse mahallelerinin belirli bölgeleri ve Gebze ilçesinde yer alan Köklü Çeşme mahallesinin belli bölgelerinin yüksek değer dağılımına sahip yerler

olmuştur. Diğer taraftan Çayırova ilçesinde yer alan Cumhuriyet ve Şekerpınar mahallerinin kuzeyinde yer alan organize sanayi bölgelerine yakın bölgeler düşük değer dağılımına yerler olmuştur. Ayrıca Tuzla ilçesinde bulunan Şifa ve

Mimar Sinan mahalleleri ile Çayırova ilçesinde bulunan Cumhuriyet Mahallesi orta-yüksek değer dağılımına sahip bölgeler olarak nitelendirilmektedir (Şekil 7).



Şekil 7. Rastgele Orman modeli taşınmaz değer haritası.

4. SONUÇ ve ÖNERİLER

Yatırım planlamaları için büyük öneme sahip olan taşınmazlar, proje geliştirme, kredi düzenlemeleri ve vergi uygulamaları ile ülke ekonomilerinin büyüme-küçülme hızını etkileyen önemli bir araçtır. Taşınmaz değerlerinin güncel yaklaşımlar ile objektif olarak toplu tespiti, günümüz planlı ve sürdürülebilir taşınmaz yönetim stratejilerinin geliştirilmesi bakımından oldukça önemli bir duruma gelmiştir.

Günümüzün değişen ihtiyaçları ve artan veri yoğunluğu göz önüne alındığında, taşınmaz değerlerinin tespiti için taşınmazların tekil olarak değerlendirilmesinin yerini çok sayıda taşınmazın eş zamanlı olarak değerlendirildiği toplu taşınmaz değerlendirme konusu almıştır. Dolayısıyla tekil değerlemede kullanılan yaklaşımların yanı sıra, toplu taşınmaz değerlendirme için gelişmiş tahmin yaklaşımları olarak bilinen makine öğrenme teknikleri kullanılmaya başlanmıştır.

Bu çalışma kapsamında ise konut tipindeki taşınmazların toplu değerlendirme için makine

öğrenme tekniklerine dayalı tahmin modelleri geliştirilmiştir. Uluslararası standartlar, akademik çalışmalar ve projeler incelenerek, değeri etkileyen kriterler konumsal, mahalli ve yapısal olmak üzere üç ana kategori altında tanımlanmıştır.

Belirlenen kriterlere ilişkin temin edilen veri setleri CBS ortamında düzenlenmiştir. Taşınmaz değer veri seti çeşitli veri madenciliği yöntemleri ile aykırılıklar açısından analiz edilmiştir. Toplamda 444 adet aykırı veri tespit edilerek veri seti içerisinde çıkarılmış olup, veri setine zamansal anlamda herhangi bir düzenleme yapılmamıştır. Aykırı veri analiz yaklaşımları veri seti içerisindeki genel dağılımdan yüksek ve düşük değere sahip örneklemelerin tespitine olanak sağlamıştır. Bu durum yüksek değerlere sahip kıymetli bilgi üretebilecek örneklemelerin veri seti içerisinde doğrudan çıkarılması olarak düşünülmemelidir. Kararlı tahmin modellerinin oluşturulmasında temel bir işlem adımı olarak değerlendirilmelidir. Veri seti içerisindeki aykırılıkların giderilmesinin ardından çalışma alanı içerisindeki piyasa örneklemelerinin %70'i eğitim, %30'u test olacak

şekilde oluşturulmuştur. Ayrıca modelleme aşamasında çoklu doğrusallık problemini engellemek için değer ve değerle ilişkili konumsal ve mahalli kriterler Pearson korelasyon analizi ile değerlendirilmiş olup, birbiri ile yüksek korelasyonlu kriterlerinden birisi tercih edilerek modellere dahil edilmiştir. Bu şekilde 54 kriter modelleme için belirlenmiştir. Rastgele Orman ve ÇDR teknikleri kullanılarak tahmin modelleri üretilmiştir. Model performansları test veri seti üzerinden değerlendirilmiştir. Model başarılarının değerlendirilmesinde ise hem makine öğrenmesi kapsamında en yaygın kullanılan performans değerlendirme ölçütleri hem de IAAO tarafından önerilen toplu taşınmaz değerlendirme model performans değerlendirme metrikleri kullanılmıştır. Üretilen modeller arasında Rastgele Orman modeli daha başarılı sonuçlar vermiştir. Rastgele Orman modeli üzerinden çalışma alanı için kriter önem düzeyleri hesaplanmıştır. Yapısal ve konumsal kriterlerin diğer kriterlere kıyasla ön plana çıktığı tespit edilmiştir. Daha başarılı model performansına sahip Rastgele Orman modeli üzerinden CBS tabanlı taşınmaz değer haritası üretilmiştir. Üretilen harita ile mahallelerin taşınmaz değer dağılımları irdelenmiştir.

Çalışma kapsamında toplu taşınmaz değerlendirme uygulamalarında CBS ve makine öğrenmesi tekniklerinin etkin biçimde kullanılabileceği sonuçlarla desteklenmiştir. Ayrıca modelleme çalışmalarında etkin kriterlerin ve model başarılarının bölgesel olarak tespit edilebileceği açıkça ortaya koyulmuştur. Diğer taraftan toplu değerlendirme tahmin modellerinin geliştirilmesi öncesinde, veri setleri içerisindeki konumsal ve konumsal olmayan aykırı verilerin ayıklanmasında literatüre bütünlük bir metodoloji sunulmaktadır.

Toplu taşınmaz değerlendirme için kullanılacak kriterleri temsil eden veri setlerinin Türkiye Ulusal CBS (TUCBS) veri temalarıyla birlikte çalışabilir yönetimi ve yeni nesil makine öğrenme tekniklerinin bütünlük kullanımıyla etkin değerlendirme faaliyetleri gerçekleştirilebilir. Ayrıca güncel makine öğrenmesi teknikleri kullanılarak farklı sektörel uygulama ihtiyaçları için taşınmaz değer tahmin modelleri ile dinamik CBS tabanlı taşınmaz değer haritaları üretilebilir.

TEŞEKKÜR

Araştırma kapsamında kullanılan coğrafi verilerin temininde Kocaeli Büyükşehir Belediyesi ve piyasaya ilişkin taşınmaz değer verilerinin sağlanmasında REIDIN firmasına, Hollanda Twente Üniversitesi, Büyük Veri Merkezi'nden

Doç. Dr. Serkan GİRĞİN'e katkılarından dolayı teşekkürlerimizi sunarız.

ORCID

Süleyman ŞİŞMAN 
<https://orcid.org/0000-0002-0924-1092>

Arif Çağdaş AYDINOĞLU 
<https://orcid.org/0000-0003-4912-9027>

KAYNAKLAR

- Açlar, A. ve Çağdaş V. (2008). Taşınmaz (Gayrimenkul) Değerlemesi (2. baskı). TMMOB Harita ve Kadastro Mühendisleri Odası, Ankara.
- AFAD. (2018). Afet ve Acil Durum Yönetimi Başkanlığı, Türkiye Deprem Tehlike Haritası. Erişim adresi: <https://deprem.afad.gov.tr/deprem-tehlike-haritasi> (24 Mayıs 2022)
- Aggarwal, C. C. (2017). Outlier Analysis (2. baskı), Springer.
- Akar, A. U. ve Yalpir, S. (2021). Using SVR and MRA Methods for Real Estate Valuation in the Smart Cities. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 46, 21-26. doi:10.5194/isprs-Archives-XLVI-4-W52021-21-2021
- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93–115.
- Anselin, L., Sridharan, S. ve Gholston, S. (2007). Using exploratory spatial data analysis to leverage social indicator databases: The discovery of interesting patterns. *Social Indicators Research*, 82(2), 287–309. doi:10.1007/s11205-006-9034-x
- Antipov, E.A. ve 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. doi:10.1016/J.ESWA.2011.08.077
- Appraisal of Real Estate. (2001). (12. baskı), Appraisal Institute.

- Aydinoğlu, A.Ç. (2019). *Akıllı Kentler için Büyük Veri Yönetimi ve Analizinde Birlikte Çalışabilirliğin Sağlanması* (Rapor No: 116Y204). Ankara: TÜBİTAK.
- Aydinoglu, A.C., Bovkir, R. ve Colkesen, I. (2021). Implementing a mass valuation application on interoperable land valuation data model designed as an extension of the national GDI. *Survey Review*, 53, 349–365. doi:10.1080/00396265.2020.1771967
- Bovkir, R. ve Aydinoglu, A. C. (2018). Providing land value information from geographic data infrastructure by using fuzzy logic analysis approach. *Land use policy*, 78, 46-60. doi:10.1016/j.landusepol.2018.06.040
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. doi:10.1007/BF00058655
- Breiman, L. (2001). Random forests, *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324
- Bujang, A. A., Zarin, H. A. ve Jumadi, N. (2010). The relationship between demographic factors and housing affordability. *Malaysian Journal of Real Estate*, 5(1), 49-58.
- Çölkesen İ. (2015). *Yüksek Çözünürlüklü Uydu Görüntüleri Kullanarak Benzer Spektral Özelliklere Sahip Doğal Nesnelerin Ayırt Edilmesine Yönelik Bir Metodoloji Geliştirme* (Doktora Tezi). İTÜ, Fen Bilimleri Enstitüsü, İstanbul.
- Del Giudice, V., De Paola, P. ve Cantisani, G. B. (2017). Valuation of real estate investments through Fuzzy Logic. *Buildings*, 7(1), 26. doi:10.3390/buildings7010026
- Dovoedo, Y. H. ve Chakraborti S. (2015). Boxplot-Based Outlier Detection for the Location-Scale Family, *Communications in Statistics. Simulation and Computation*, 44(6), 1492–1513. doi:10.1080/03610918.2015.1013037
- Egghe, L. ve Leydesdorff, L. (2009). The relation between Pearson's correlation coefficient r and Salton's cosine measure. *Journal of the American Society for information Science and Technology*, 60(5), 1027-1036. doi:10.1002/ASI.21009
- Fischer, M.M. ve Wang J. (2011). *Spatial Data Analysis* (1.Baskı). Springer.
- GeoHub. (2020). Erişim adresi: <https://geohubpota.yildiz.edu.tr/> (15 Ocak 2020)
- Ge, Z., Song Z., Ding S. X. ve Huang B. (2017). Data Mining and Analytics in the Process Industry: The Role of Machine Learning. *IEEE Access*, 5, 20590–20616. doi:10.1109/ACCESS.2017.2756872
- Ho, T.K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844. doi:10.1109/34.709601
- IAAO. (2013). *Standard on Ratio studies*. International Association of Assessing Officers (IAAO), Kansas City, Missouri, USA.
- IVSC (International Valuation Standards Council). (2016). *International Valuation Standards*, Londra, Birleşik Krallık.
- Kalantar, B., Ueda, N., Saeidi, V., Ahmadi, K., Halin, A.A. ve Shabani, F. (2020). Landslide Susceptibility Mapping: Machine and Ensemble Learning Based on Remote Sensing Big Data. *Remote Sensing*, 12, 1737. doi:10.3390/rs12111737
- Levantesi, S. ve Piscopo, G. (2020). The importance of economic variables on London real estate market: A random forest approach. *Risks*, 8, 1–17. doi:10.3390/risks8040112
- Liu, H., Shah, S. ve Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers and Chemical Engineering*, 28(9), 1635–1647. doi:10.1016/j.compchemeng.2004.01.009
- Lm. (2022). Erişim adresi: <https://www.rdocumentation.org/packages/stat/s/versions/3.6.2/topics/lm> (26 Mayıs 2022)
- Luo, H., Zhao, S. ve Yao, R. (2021). Determinants of Housing Prices in Dalian City, China: Empirical Study Based on Hedonic Price Model. *Journal of Urban Planning and Development*, 147(2), 05021017.

- Mete, M.O. ve Yomralioglu, T. (2021). Implementation of serverless cloud GIS platform for land valuation. *International Journal of Digital Earth*, 14(7), 836-850. doi:10.1080/17538947.2021.1889056
- Moon, B.Y., Kim, S.Y. ve Kang, G.J. (2008). Initial ship design using a Pearson correlation coefficient and artificial intelligence techniques. *International journal of modern physics B*, 22, 1801–1806. doi:10.1142/S0217979208047444
- Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T. ve French, N. (2003). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance*, 21(4), 383-401. doi:10.1108/14635780310483656
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222. doi: 10.1080/01431160412331269698
- randomForest. (2022). Erişim adresi: <https://www.rdocumentation.org/packages/randomForest/versions/4.7-1.1/topics/randomForest> (26 Mayıs 2022)
- Renigier-Bilozor, M., Janowski, A. ve d'Amato, M. (2019). Automated valuation model based on fuzzy and rough set theory for real estate market with insufficient source data. *Land use policy*, 87, 104021.
- REIDIN. (2022). Erişim adresi: <https://rebis.reidin.com/tr-TR> (30 Temmuz 2022)
- 4 Sayılı Cumhurbaşkanlığı Kararnamesi. (2018, 15 Temmuz). Resmi Gazete (Sayı: 30479)
- 30 Sayılı Cumhurbaşkanlığı Kararnamesi. (2019, 5 Şubat). Resmi Gazete (Sayı: 330677).
- RICS (Royal Institution of Chartered Surveyors). (2017). Professional Valuation Standards, Londra, Birleşik Krallık.
- SEGE. (2022). T.C. Sanayi ve Teknoloji Bakanlığı Kalkınma Ajansları Genel Müdürlüğü Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırmaları Raporları. Erişim adresi: <https://www.sanayi.gov.tr/merkez-birimi/b94224510b7b/sege> (20 Mayıs 2022)
- Sisman, S., Akar A.U. ve Yalpir S. (2021). The novelty hybrid model development proposal for mass appraisal of real estates in sustainable land management. *Survey Review*, 1-20. doi:10.1080/00396265.2021.1996797
- Sisman, S. ve Aydinoglu, A.C. (2022). Improving performance of mass real estate valuation through application of the dataset optimization and Spatially Constrained Multivariate Clustering Analysis. *Land Use Policy*, 119, 106167. doi:10.1016/j.landusepol.2022.106167
- Sun, H., Burton, H. V. ve Huang, H. (2021). Machine learning applications for building structural design and performance assessment: state-of-the-art review. *Journal of Building Engineering*, 33, 101816. doi:10.1016/j.jobbe.2020.101816
- Şişman, S. (2021). *CBS Tabanlı Makine Öğrenme Teknikleri ile Toplu Taşınmaz Değerlemesi* (Yüksek Lisans Tezi). GTÜ, Fen Bilimleri Enstitüsü, Kocaeli.
- TDUB (Türkiye Değerleme Uzmanları Birliği). (2011). TUGDES (Türkiye Değerleme Standartları) Çalışması.
- TEGoVA (The European Group of Valuers' Associations). (2016). European Valuation Standards (8. Basım). Belçika: Gillis Yayınevi.
- TKGM. (2011). TKMP 4. Bileşeni Gayrimenkul Değerinin Belirlenmesi ve Kayıt Altına Alınması, Teknik Rapor, Ankara, Türkiye.
- TKGM. (2022). Taşınmaz Değerleme Dairesi Başkanlığı Misyonu-Vizyonu, Tapu Kadastro Genel Müdürlüğü. Erişim adresi: <https://www.tkgm.gov.tr/tasimazd-db/misyon-ve-vizyon> (25 Mayıs 2022)
- Tobler, W.R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234.
- Unel, F. B. ve Yalpir, S. (2019). Valuations of building plots using the AHP method. *International journal of strategic property management*, 23(3), 197-212.
- TÜİK. (2020). Türkiye İstatistik Kurumu, Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları, 2020 Yılı Nüfus ve Demografi İstatistikleri.

- Ünel, F.B. (2017). *Taşınmaz Değerleme Kriterlerine Yönelik Coğrafi Veri Modelinin Geliştirilmesi* (Doktora Tezi). SÜ, Fen Bilimleri Enstitüsü, Konya.
- Wang, X., Wen, J., Zhang, Y. ve Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. *Optik*, 125(3), 1439-1443. doi:10.1016/j.ijleo.2013.09.017
- Wu L. (2009). *Mixed Effects Models for Complex Data* (1. Baskı). Chapman and Hall/CRC.
- Yalpir, Ş. (2018). Enhancement of parcel valuation with adaptive artificial neural network modeling. *Artificial intelligence review*, 49(3), 393-405. doi:10.1007/s10462-016-9531-5
- Yalpir, S., Sisman, S., Akar, A. U. ve Unel, F. B. (2021). Feature selection applications and model validation for mass real estate valuation systems. *Land Use Policy*, 108, 105539. doi:10.1016/j.landusepol.2021.105539
- Yilmazer, S. ve Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land use policy*, 99, 104889. doi:10.1016/j.landusepol.2020.104889
- Yilmazer, S. ve Kocaman, S. (2021). On the Prediction Performances of SVM and ANN Methods for Mass Appraisal Assessment: A Case Study from Ankara (Turkey). *FIG e-Working Week 2021, Smart Surveyors for Land and Water Management - Challenges in a New RealityVirtual*, 21–25 June 2021.
- T.C-CBFO.2020. T.C. Cumhurbaşkanlığı Finans Ofisi, Türkiye Hanehalkı Finansal Algı ve Tutum Araştırması. Erişim adresi: <https://www.cbfo.gov.tr/turkiye-hanehalki-finansal-algi-ve-tutum-arastirmasi> (9 Mart 2021)