

Toplu taşınmaz değerlemesine yönelik mekânsal yapay zekâ (GeoAI) analiz araç paketinin geliştirilmesi

Süleyman Şişman¹ , Büşra Demirci¹ , Arif Çağdaş Aydınoğlu^{1*} 

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

Öz: Günümüzde taşınmaz sektörünün hızlı değişimi, nüfus artışı, kentleşme ve ekonomik dinamiklerin çeşitlenmesi, kentsel alanlarda taşınmaz değerlerinin doğru, güncel ve hızlı biçimde belirlenmesini zorunlu kılmıştır. Bu gereklilik, geleneksel değerlendirme yaklaşımlarının ötesine geçilerek, istatistiksel, mekânsal ve yapay zekâ temelli yöntemlerin bütünsel biçimde kullanıldığı veri temelli toplu taşınmaz değerlendirme sistemlerinin önemini artırmıştır. Bu bağlamda, coğrafi bilgi teknolojileri ile makine öğrenmesi tekniklerinin bir arada kullanılması taşınmaz değerlendirme süreçlerinin otomasyonu açısından yenilikçi bir yaklaşım sunmaktadır. Bu çalışma kapsamında otomatize toplu taşınmaz değerlemesine yönelik analiz araçlarını içeren Mekânsal Yapay Zekâ (GeoAI) analiz araç paketi geliştirilmiştir. ArcGIS platformunda, Python tabanlı ArcPy modülü ve açık kaynak kütüphaneler (NumPy, Pandas, Scikit-learn, LightGBM vb.) kullanılarak geliştirilen bu araç paketi; veri ön işleme (aykırı değer belirleme, normalizasyon, değişken seçimi, eğitim-test ayrımı), regresyon tabanlı modelleme (Random Forest ve LightGBM) hiperparametre optimizasyonu ve model performans değerlendirme süreçlerini bütünsel biçimde gerçekleştirecek şekilde tasarlanmıştır. Geliştirilen analiz araçlarını test etmek amacıyla İstanbul ilinin Tuzla ve Pendik ilçeleri çalışma alanı olarak belirlenmiştir. Çalışma alanında 2019 Ocak-2023 Aralık dönemi için 37 268 adet piyasa örnekleme temin edilmiştir. Ardından bu veri seti coğrafi analizlerle zenginleştirilerek, toplamda 68 değişken içeren toplu taşınmaz değerlendirme veri seti oluşturulmuştur. Veri ön işleme araçları ile veri setinin aykırı veriler açısından optimizasyonu, normalizasyonu, eğitim/test veri setlerinin oluşturulması ve model değişkenlerinin belirlenmesi işlemleri gerçekleştirilmiştir. Modelleme araçları ile eğitim veri seti üzerinden toplu taşınmaz değerlendirme modellerinin eğitimleri gerçekleştirilmiştir. Performans değerlendirme analiz araçları ile test veri seti üzerinden modellerin performans analizleri MAE, MAPE, RMSE ve R² metriklerine göre gerçekleştirilmiştir. Sonuçlar coğrafi bilgi teknolojileri ve makine öğrenmesi tekniklerinin entegrasyonu ile geliştirilen analiz araç paketinin, toplu taşınmaz değerlendirme süreçlerinde hem doğruluk hem de otomasyon açısından yüksek bir potansiyel sunduğunu ortaya koymaktadır.

Anahtar Sözcükler: Yapay zekâ, Makine öğrenmesi, Mekânsal yapay zekâ (GeoAI), Toplu taşınmaz değerlendirme

Development of geospatial artificial intelligence (GeoAI) analysis tool package for mass real estate valuation

Abstract: Today, the rapid transformation of the real estate sector, population growth, urbanization, and the diversification of economic dynamics have made it essential to determine real estate values in urban areas accurately, up-to-date, and quickly. This necessity has increased the importance of data-driven mass real estate valuation systems, in which statistical, spatial, and artificial intelligence-based methods are integrated, going beyond traditional valuation approaches. In this context, the combined use of geographic information technologies and machine learning techniques offers an innovative approach to the automation of real estate valuation processes. Within the scope of this study, a Geospatial Artificial Intelligence (GeoAI) analysis tool package was developed, which includes analysis tools for automated mass real estate valuation. Developed on the ArcGIS platform using the Python-based ArcPy module and open-source libraries (such as NumPy, Pandas, Scikit-learn, LightGBM, etc.), this tool package is designed to perform data preprocessing (outlier detection, normalization, feature selection, train-test split), regression-based modeling (Random Forest and LightGBM), hyperparameter optimization and model performance evaluation steps in an integrated manner. To test the developed analysis tools, the districts of Tuzla and Pendik in Istanbul were determined as the study area. In the study area, 37 268 market samples were obtained for the period of January 2019 to December 2023. Then, this dataset was then enriched with spatial analyses, creating a mass real estate valuation dataset containing a total of 68 variables. Optimization of the dataset in terms of outliers, normalization, creation of training/test datasets and determination of model variables were performed using data pre-processing tools. Training of mass real estate valuation models was carried out on the training dataset using modelling tools. Model performance analyses were conducted on the test dataset using performance evaluation analysis tool, based on the MAE, MAPE, RMSE, and R² metrics. The results show that the analysis package developed through the integration of geographic information technologies and machine learning techniques offers high potential in mass real estate valuation processes in terms of both accuracy and automation.

Keywords: Artificial intelligence, Machine learning, Geospatial artificial intelligence (GeoAI), Mass real estate valuation

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

Geliş Tarihi/Received: 28.10.2025
Kabul Tarihi/Accepted: 08.01.2026



0000-0002-0924-1092, ssisman@gtu.edu.tr (Şişman S.)
0009-0005-3585-8992, b.demirci2021@gtu.edu.tr (Demirci B.)
0000-0003-4912-9027, aydinoglu@gtu.edu.tr (Aydınoğlu A.Ç.)*

1. Giriş

Günümüz kentleşme dinamikleri, değişen ekonomik koşullar ve artan nüfus yoğunluğu, kentlerdeki taşınmaz piyasalarının karmaşık, çok boyutlu ve dinamik bir işlev kazanmasına neden olmuştur. Bu dönüşüm, taşınmaz değerlemesinin yalnızca tekil olarak değil, aynı zamanda veri temelli olarak toplu ve sistematik biçimde gerçekleştirilmesini gerekli kılmaktadır. Özellikle kamu yönetimi, kentsel planlama, vergi politikaları geliştirme, yatırım planlama ve sürdürülebilir arazi yönetimi gibi taşınmaz değer bilgisine gereksinim duyulan alanlarda toplu taşınmaz değerlendirme süreçlerinin hızlı, şeffaf ve tutarlı biçimde gerçekleştirilmesi büyük öneme sahiptir (Ünel & Yalpir, 2023; Yalpir vd., 2021). International Association of Assessing Officers (IAAO) tarafından yapılan tanıma göre toplu taşınmaz değerlendirme; belirli bir tarih itibarıyla bir grup taşınmazın ortak veriler, standartlaştırılmış yöntemler ve istatistiksel testler kullanılarak değerlemesi işlemi olarak tanımlanmaktadır (IAAO, 2013). Bu tanım, toplu taşınmaz değerlemesinin yalnızca teknik bir faaliyet değil, aynı zamanda güçlü bir veri yönetimi ve analitik karar destek süreci olduğuna açıkça işaret etmektedir.

Toplu taşınmaz değerlendirme faaliyetlerinin etkin biçimde gerçekleştirilebilmesi, farklı kaynaklardan elde edilen büyük hacimli ve çok boyutlu veri setlerinin analiz edilebilmesi ve veri setlerinin coğrafi veriler kullanılarak zenginleştirilmesine bağlıdır. Bu gereksinimlere paralel olarak, son yıllarda yapay zekâ kapsamında makine öğrenmesi ve coğrafi bilgi teknolojileri gibi veri temelli yaklaşımlar, taşınmaz değerlendirme süreçlerine dahil edilerek etkin çözümler geliştirilmeye başlanılmıştır (Şişman & Aydınoglu, 2023). Örneğin, Soltani vd. (2022), bina ve bağımsız bölüm özellikleri, çeşitli kentsel fonksiyonlara erişebilirlik ve mahalli özellikleri temsil eden 38 değişken içeren bir veri seti ile toplu taşınmaz değerlendirme modelleri geliştirmişlerdir. Modellerin geliştirilmesinde Linear Regression (LR), Decision Tree (DT), Random Forest (RF) ve Gradient Boosting Machine (GBM) gibi makine öğrenmesi algoritmalarını kullanmışlardır. Genç vd. (2025), yapısal, konumsal ve demografik faktörlere ilişkin 22 değişken içeren bir veri seti ile LR, Gaussian Process Regression (GPR), Support Vector Machine (SVM), Artificial Neural Networks (ANN), RF gibi algoritmaları kullanarak taşınmaz değerlendirme modelleri geliştirmişlerdir. Modellerin eğitimi için kullanılan veri setindeki konumsal faktörler, Coğrafi Bilgi Sistemleri (CBS) tabanlı yakınlık analizleri yapılarak piyasa örneklemeleri ile ilişkilendirilmiştir. Ayrıca piyasa değerleri ile değerlendirme modellerinden elde edilen tahminler kullanılarak enterpolasyon analizleri gerçekleştirilmiş olup, taşınmaz değerlerinin mekânsal dağılımları irdelenmiştir. Truong vd. (2020), konut tipindeki taşınmazlar için yapısal ve konumsal 26 değişken içeren bir veri setini veri ön işleme ve keşifsel veri analiz yöntemleri ile değerlendirerek modelleme sürecine hazır duruma getirmişlerdir. Ardından RF, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Hybrid Regression Model (HRM) gibi algoritmalarla kullanarak değerlendirme modelleri geliştirerek, makine öğrenme algoritmalarının değerlendirme süreçlerinde etkin biçimde kullanılabileceğine dikkat çekmişlerdir. Jafary vd. (2024), arazi tipindeki taşınmazlar için arazi özellikleri, topoğrafik faktörler, konumsal faktörler, çevresel faktörler, sosyo-demografik faktörler, planlama faktörleri gibi 37 değişken içeren bir veri seti ile Support Vector Regression (SVR), RF, XGBoost ve Deep Neural Network (DNN) algoritmalarını kullanarak taşınmaz değerlendirme modelleri geliştirmişlerdir. Değerleme faktörlerinin geniş bir yelpazede değerlendirildiği çalışmada geliştirilen değerlendirme modelleri ile elde edilen tahminler kullanılarak CBS tabanlı arazi değer haritaları da üretilmiştir.

Makine öğrenmesi algoritmalarının değerlendirme veri setleri içindeki çok sayıda değişken için karmaşık örüntüleri ve ilişkileri öğrenebilme yeteneği sayesinde, taşınmaz değer tahmininde yüksek doğruluk sağlayan değerlendirme modelleri geliştirilerek karar destek süreçleri hızlandırılabilmektedir (Baldominos vd., 2018; Baur vd., 2023). Makine öğrenmesi algoritmaları ve CBS bir araya getirildiğinde, geleneksel tekil değerlendirme yaklaşımlarına (karşılaştırma, maliyet ve gelir yöntemleri) kıyasla daha nesnel, tekrarlanabilir ve veriye odaklı sonuçlar üretilebilmektedir (Mete & Yomrahoğlu, 2022). Bu bağlamda yapay zekâ teknikleri ile CBS'nin entegrasyonunu ifade eden Mekânsal Yapay Zekâ (Geospatial Artificial Intelligence, GeoAI)

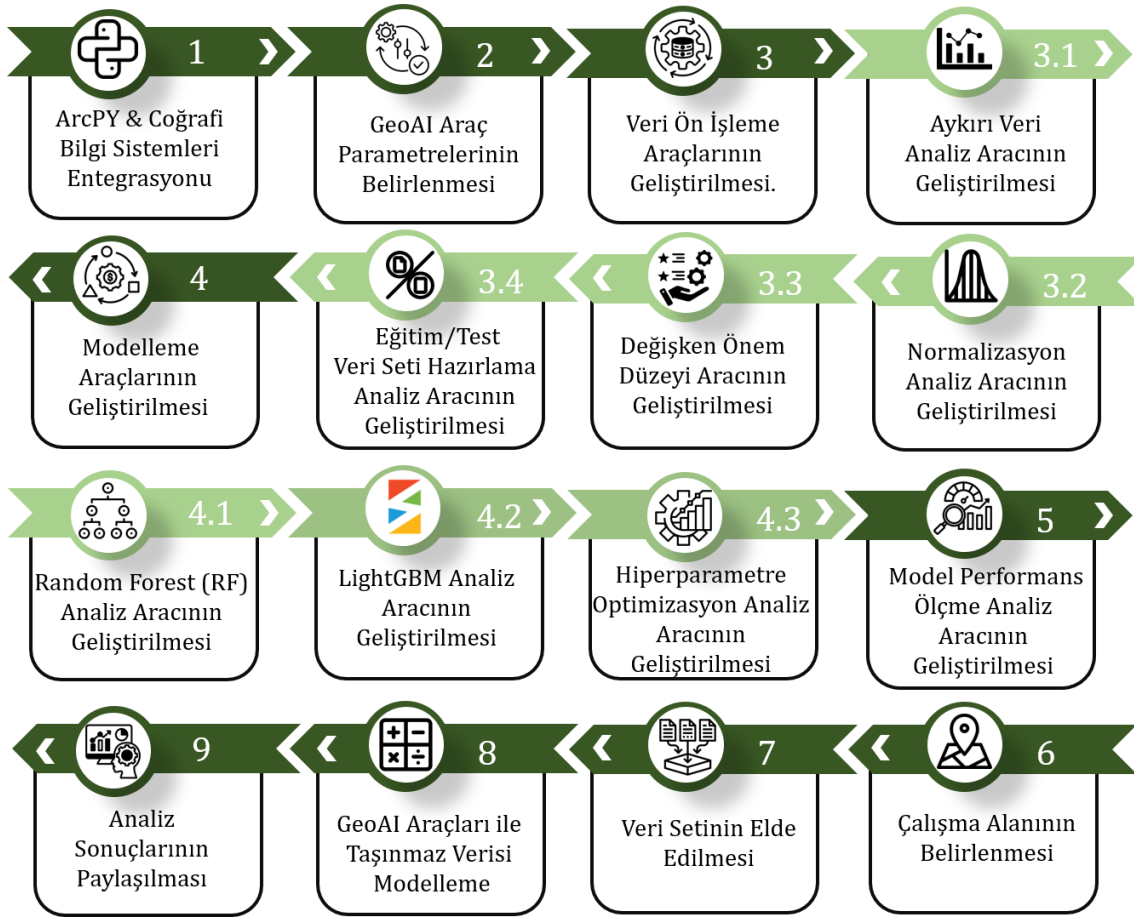
kavramı ön plana çıkmaktadır. GeoAI, mekânsal verilerin analizinde makine öğrenmesi, derin öğrenme ve istatistiksel modelleme gibi farklı yapay zekâ yöntemlerinin kullanılmasıyla ortaya çıkan disiplinler arası yeni bir yaklaşım olarak değerlendirilmektedir (Choi, 2023). GeoAI, büyük hacimli veri setlerinin işlenmesi, karmaşık mekânsal örüntülerin tanımlanması ve mekâna bağlı olarak değişen ilişkilerin modellenmesi gibi farklı alanlarda güçlü ve ölçeklenebilir çözümler sunmaktadır. Bu kapsamda GeoAI, toplu taşınmaz değerlendirme süreçlerinin otomasyonu ve değerlendirme modellerinin performanslarının iyileştirilmesi açısından önemli bir potansiyele sahiptir (Mete, 2025).

Makine öğrenmesi algoritmaları ile toplu taşınmaz değerlendirme süreci; aykırı veri analizi, normalizasyon, değişken seçimi gibi veri ön işleme, çeşitli algoritmalar kullanılarak tahmin modellerinin eğitildiği modelleme, MAE, MAPE, RMSE, R² gibi metriklerle performansın değerlendirildiği performans analizini içeren birbirine bağlı aşamalardan oluşan bir yapıya sahiptir (Şişman & Aydınoglu, 2022). Bu süreçte yer alan işlemlerin birbirinden bağımsız olarak gerçekleştirilmesi, özellikle büyük değerlendirme veri setleri söz konusu olduğunda hem zaman alıcı hem de modelleme sürecinde hata yapma riskini artıran bir durum olarak değerlendirilebilir. Bu nedenle söz konusu işlemlerin sistematik, bütünlük, tekrar üretilebilir ve hızlı biçimde gerçekleştirilebilmesi için CBS ortamında toplu taşınmaz değerlemesine yönelik süreçler için mekânsal yapay zeka analiz araçlarına gereksinim bulunmaktadır.

Bu çalışmada toplu taşınmaz değerlemesine yönelik makine öğrenmesi temelli GeoAI uygulama analiz araç paketinin geliştirilmesi hedeflenmektedir. Bu kapsamda toplu taşınmaz değerlendirme süreçlerinde veri ön işleme, modelleme ve model performans analizlerine ilişkin yöntemler detaylı biçimde irdelenerek geliştirilmesine gereksinim duyulan analiz araçları belirlenmiştir. Araç paketi içerisinde veri ön işleme, regresyon tabanlı modelleme (RF ve LightGBM), hiperparametre optimizasyonu ve model performans değerlendirme adımlarını bütünlük biçimde gerçekleştirecek analiz araçları geliştirilmiştir. Geliştirilen analiz araçlarını test etmek amacıyla İstanbul ilinin Tuzla ve Pendik ilçelerinden mahalleler içeren çalışma alanında pilot toplu taşınmaz değerlendirme uygulaması gerçekleştirilmiştir. Bu bağlamda analiz araçlarının geliştirilmesinden pilot uygulama alanında analiz araçlarının test edilmesi aşamasına kadar olan işlemler için izlenen metodolojik adımlar Şekil 1’de verilmiştir.

2. Analiz Araçlarına Genel Bakış

Çalışma kapsamında, makine öğrenmesi temelli regresyon analizlerinin özel olarak toplu taşınmaz değerlendirme süreçlerinde etkin şekilde kullanılmasını sağlayan GeoAI isimli bir CBS analiz araç paketi geliştirilmiştir (Şekil 2). Araç paketi, açık kaynaklı programlama dili olan Python ile ArcGIS Pro platformunda geliştirilmiştir. Paketin geliştirilmesi sırasında platformun mekânsal analiz yeteneklerini programlama dili ile birleştirmeyi sağlayan Arcpy modülü öncelikli kullanılmıştır. Kullanılan diğer Python modülleri ve kütüphaneleri ile kullanım amaçları ise sırasıyla şu şekildedir: NumPy kütüphanesi, matris ve dizi gibi sayısal hesaplamaların yapılmasında; Pandas kütüphanesi, verilerin yapılandırılması ve dönüştürülmesinde; Matplotlib ve Seaborn kütüphaneleri, verilerin görselleştirilmesi aynı zamanda makine öğrenmesi modellerinin çıktısı alınmasında kullanılmıştır. Makine öğrenmesi algoritmaları için, Scikit-learn kütüphanesinden RandomForestRegressor algoritması, lightGBM kütüphanesi üzerinden LGBMRegressor algoritmasına yer verilmiştir. Son olarak modellerin kaydedilip yeniden yüklenmesinde Joblib modülünden, sistem ve dizin işlemlerinde os modülünden, rastgele veri üretimi işlemlerinde ise random modülünden yararlanılmıştır (Pedregosa vd., 2011). Çalışmada paket geliştirme ortamı olan ArcGIS Pro platformunun 3.1, Python programlama dilinin 3.9.16 versiyonları kullanılmış diğer kütüphanelerin versiyon uyumluluğuna dikkat edilmiştir.



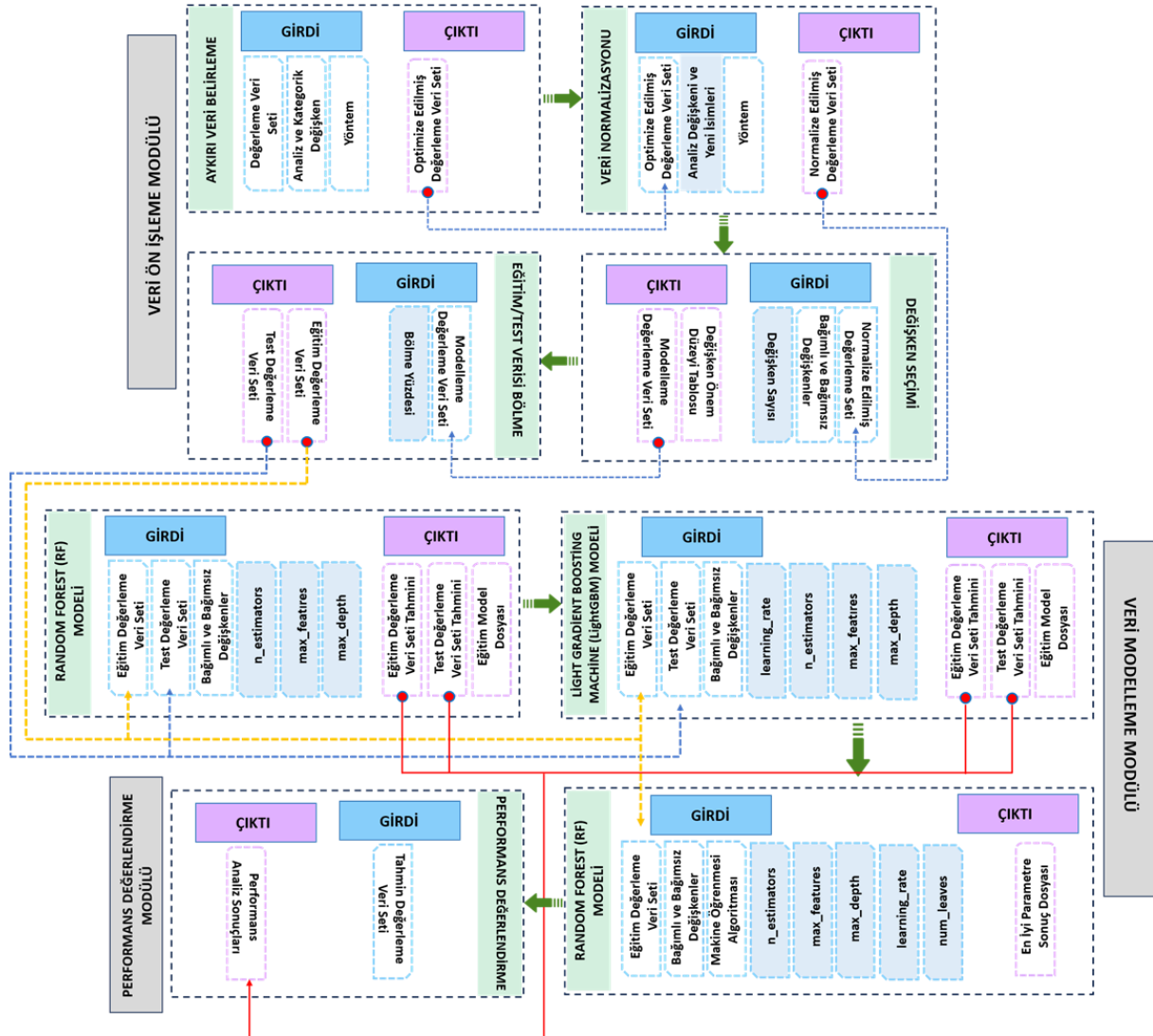
Şekil 1: Çalışma kapsamında izlenen metodoloji



Şekil 2: Mekânsal yapay zekâ (GeoAI) analiz araç paketine genel bakış

Geliştirilen araç paketi, veri bilimi sürecinin temel aşamalarını kapsayan veri ön işleme, veri modelleme ve performans değerlendirme olmak üzere üç temel modül olarak değerlendirilebilmektedir. Veri ön işleme modülü, modelleme öncesinde veri setinin analiz ve hazırlık süreçlerini gerçekleştiren dört araç kutusunu içermektedir. Aykırı Veri Belirleme Analiz Aracı, veri setindeki aykırı değerlerin tespit edilmesine olanak tanırken; Normalizasyon Analiz Aracı, veri seti içerisindeki değişkenlerin standartlaşmasını sağlamaktadır. Değişken Seçim Analiz Aracı modelde kullanılacak anlamlı bağımsız değişkenlerin belirlenmesini, Eğitim-Test-Validasyon Bölme Analiz Aracı ise veri setinin model eğitimi ve performans analizleri için bölünmesini sağlamaktadır.

Modelleme modülü, tahmin modellerinin geliştirilmesine yönelik RF ve LightGBM olmak üzere iki ayrı makine öğrenme algoritmasının entegre edildiği analiz araçları ile bu algoritmaların parametrelerinin iyileştirilmesine yönelik hiperparametre optimizasyon analiz aracını kapsamaktadır. Performans değerlendirme modülü ise, modelleme sürecinde elde edilen tahminlerin doğruluğunu ve güvenilirliğini çeşitli makine öğrenmesi performans metriklerine göre hesaplayan Performans Değerlendirme Analiz Aracı'ndan oluşmaktadır. Tüm analiz araçlarının etkileşimli olarak genel işleyişi ve girdi/çıkı veri akışları Şekil 3'te verilmiştir.



Şekil 3: Araç paketi geliştirme sürecinde kullanılan yöntemler ve izlenen iş akış şeması

2.1 Veri Ön İşleme Modülünün Geliştirilmesi

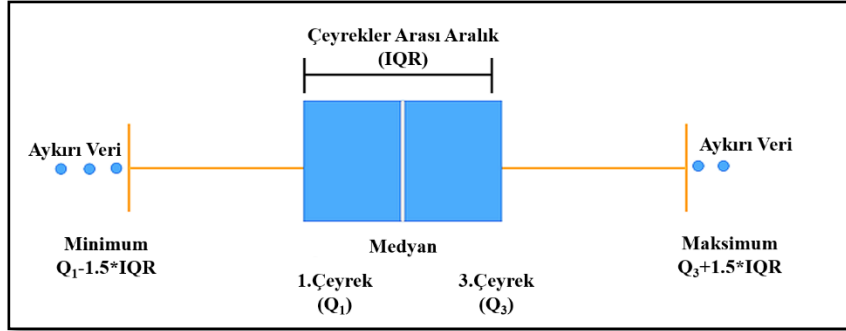
Bir veri kümesini, makine öğrenmesi tekniği kullanılarak analize daha uygun hale getirmek amacıyla gerçekleştirilen işlemler bütününe veri ön işleme denilmektedir. Bu işlem adımı genelde; aykırı veri belirleme, veri normalizasyonu, değişken düzeylerinin belirlenmesi ve son olarak veri setinin uygun yüzdelik oranlarda bölünerek eğitim ve test verisi elde edilmesi şeklinde gerçekleştirilmektedir.

2.1.1 Aykırı Veri Belirleme Analiz Aracı

Aykırı veri genel istatistiksel dağılımdan sapan, bu dağılımı uyumlu çoğu veriden bariz şekilde uzakta kalan, güvenilir sonuçları elde etmeyi zorlaştıran veri olarak tanımlanmaktadır. Oluşturulan bir tahmin modelinin daha doğru sonuç vermesi için bu değerlerin tespit edilmesi, veri setinden çıkartılarak verinin optimize edilmesi gerekir (Liu vd., 2004; Şişman &

(Aydınoglu, 2022). Literatür incelendiğinde Boxplot ve Median Absolute Deviation (MAD) yöntemlerinin yaygın biçimde kullanıldığı görülmüştür (Aydınoglu & Şişman, 2024; Çılgin vd., 2023). Bu kapsamda geliştirilen Aykırı Veri Belirleme Analiz Aracı'na kullanıcı seçimli olacak biçimde Boxplot ve MAD yöntemleri entegre edilmiştir.

Boxplot yöntemi, veri setindeki dağılımı, verilerin merkezi eğilimini ve yayılma özelliklerini inceleyen ve bu kavramları kutu grafikleriyle ifade eden yöntemdir. Bu yöntemin grafiğinde, veri seti içerisindeki değişkenler minimum, maksimum, medyan, 1. çeyrek ve 3. çeyrek olmak üzere beş farklı istatistik bilgi açısından değerlendirilmektedir (Şekil 4).



Şekil 4: Boxplot tekniği çalışma prensibi

MAD yöntemi ise aykırı veri belirlemede kullanılan istatistiksel tabanlı bir diğer yöntemdir. Bu yöntemde öncelikle veri setinin medyanı bulunur. Veri setini oluşturan her veriden (X), bulunan medyan değeri çıkartılarak mutlak değerler hesaplanır. Daha sonrasında tüm mutlak değerlerin medyanı, normal dağılımlı veri seti için belirlenen b çarpanı ile hesaplanmaktadır (Fitriyah & Budi, 2019). Son olarak aykırı verilerin tespitinde kullanılan minimum ve maksimum sınırların (kritik aralık) hesaplanması için k isimli eşik değeri belirlenir. Literatür incelendiğinde b çarpanının sayısal değeri genellikle 1.4826, k eşik değerinin ise 3 alındığı görülmektedir (Leys vd., 2013; von Davier & Bezirhan, 2023). Yöntem için kullanılan matematiksel formüller Eşitlik 1 ve Eşitlik 2'deki gibidir:

$$MAD = b * (medyan(|X - medyan(X)|)) \quad (1)$$

$$medyan(X) - (k \times MAD) < X < medyan(X) + (k \times MAD) \quad (2)$$

Şekil 5a'da kullanıcı ara yüzü verilen Aykırı Veri Belirleme Analiz Aracı için parametre kısmında ayrıca bir kategorik değişken seçimi eklenmiştir. Böylelikle analiz edilmek istenen sürekli değişkeni, belirli bir kategorik değişkene göre alt gruplara ayırarak değerlendirilebilmektedir. Bu değerlendirme her kategorik grubun kendi taşınmaz aykırı değerlerinin tespit edilmesi ve veri setinden ayrıştırılması şeklindedir. Örneğin toplu taşınmaz değerlendirme veri seti içerisinde fiyat ve oda tipi (1+1, 2+1 vb.) değişkenlerin olduğu varsayıldığında, tanımlanan bu parametre ile her bir oda tipi için aykırı veriler tespit edilebilmektedir. Aykırı Veri Belirleme Analiz Aracı'nda girdi parametre olarak kullanıcıdan başlangıç veri setini almak için Değerleme Veri Seti, veri seti içerisinde aykırı değerler açısından analiz edilecek Analiz Değişkeni, analiz edilmek istenen sürekli değişkeni, belirli bir kategorik değişkene göre alt gruplara ayırarak değerlendirilebilmek için Kategorik Değişken ve aykırı veri analizinde kullanılacak yaklaşım için Yöntem parametreleri eklenmiştir. Araç çıktısı olarak ise seçilen aykırı veri yöntemine göre aykırı verilerden temizlenmiş ".shp" formatında Optimize Edilmiş Veri Seti, tespit edilen aykırı verilerin bir araya getirildiği Aykırı Veri Seti ve kullanıcının Boxplot yöntemi seçmesi durumunda, belirlenen aykırı değerlerin grafik olarak görselleştirilmesi ve kaydedilmesini sağlayan Boxplot Grafiği çıktıları tanımlanmıştır.

2.1.2 Normalizasyon Analiz Aracı

Veri setindeki aykırı değerlerin tespit edilerek çıkartılması, veri ön işlemenin bir sonraki adımı olan normalizasyon işlemi için veri setinin hazır hale gelmesi anlamını taşımaktadır. Makine öğrenmesi süreçlerinde farklı birimlerde olan verilerin aynı ölçek aralığına getirilerek daha doğru işlenmesini amaçlayan yöntemlere veri normalizasyonu denmektedir. Modelleme algoritmasının tahmin doğruluğunu arttırmak amacıyla yaygın olarak kullanılan Z-Skor, Mutlak Maksimum, Minimum-Maksimum ve Robust gibi çeşitli normalizasyon yöntemleri geliştirilen Normalizasyon Analiz Aracı'na entegre edilmiştir (Şekil 5b).

Z-skor: Veri setini oluşturan her bir verinin, aritmetik ortalamadan ne kadar uzak olduğunu standart sapmalarına bölerek açıklayan metottur. Formülde x' standartlaştırılmış değer, x orijinal değer, $\bar{x}$ aritmetik ortalama ve σ_x standart sapmadır (Kappal, 2019).

$$x' = \frac{x - \bar{x}}{\sigma_x} \quad (3)$$

Minimum-Maksimum: Veri setindeki her bir veriyi kullanıcının belirlediği yeni minimum ve maksimum değere göre yeniden ölçeklendiren bu yöntem orijinal veri değerleri arasındaki ilişkiyi korur. Formüldeki $\min(x)$ verinin minimumunu, $\max(x)$ verinin maksimumunu, a kullanıcı tarafından belirlenen minimum değeri b ise kullanıcı tarafından belirlenen maksimum değeri ifade etmektedir.

$$x' = a + \frac{(x - \min(x))(b - a)}{\max(x) - \min(x)} \quad (4)$$

Mutlak Maksimum: Normalleştirme yapılacak sütundaki her bir değeri o sütunun en büyük değerine (mutlak maksimum) bölerek veriyi standardize eden yöntemdir. Normalleştirme aralığı -1 ile 1 arasında olan bu metodun formülü aşağıdaki gibidir. Formüldeki $\max(|x|)$ verilerin mutlak değerlerin maksimumudur.

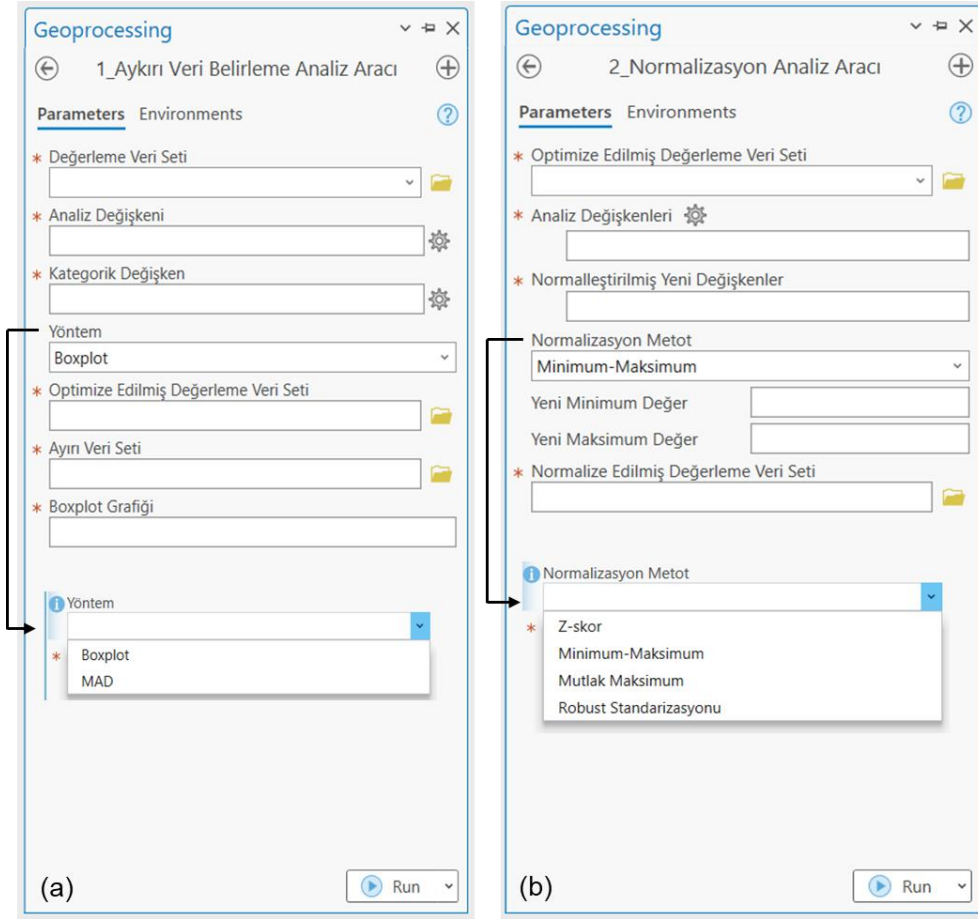
$$x' = \frac{x}{\max(|x|)} \quad (5)$$

Robust Normalizasyonu: Veri setini çeyrekler arası açıklık (IQR) ve medyana göre normalleştiren bu metot veri setindeki aykırı değerlerden çok fazla etkilenmez. Formüldeki $medyan(x)$ verilerin medyanı ve $IQR(x)$ verilerin çeyrekler arası aralığıdır.

$$x' = \frac{x - medyan(x)}{IQR(x)} \quad (6)$$

Şekil 5b'de kullanıcı ara yüzü verilen Normalizasyon Analiz Aracı, bir önceki aşamada aykırı değerlerden arındırılarak optimize edilen veri setini, Optimize Edilmiş Değerleme Veri Seti adlı parametre aracılığıyla girdi olarak kabul etmektedir. Analiz Değişkenleri isimli parametre ise normalize edilecek birden fazla sütunu girdi verisi içerisinde kullanıcının seçmesine olanak sağlamaktadır. Bu sütunların seçilmesinin ardından araç kutusuna, her bir sütun isminin değiştirilmesini ve normalleştirilmiş değerlerin yer almasını sağlayan Normalleştirilmiş Yeni Değişkenler parametresi eklenmiştir. Bu sayede, her bir değişken normalize edildikten sonra veri setinde orijinal sütunun yerine kullanıcı tarafından belirlenen isimle yeni bir sütun oluşturulur. Aracın desteklediği yöntemler bir önceki analiz aracında olduğu gibi Yöntem isimli parametre ile kullanıcıyla etkileşimli hale getirilmiş, kullanıcının analiz sürecinde tercih edeceği yöntemi seçmesi sağlanmıştır. Bu yöntemlerin gereksinimleriyle paralel olarak araca gizli parametreler eklenmiştir. Kullanıcı normalizasyon metotlarından Minimum-Maksimum'u seçtiği zaman, bu yöntemdeki yeni değer aralıklarını gizli olan Yeni Maksimum ve Yeni Minimum

Değer parametreleri ile belirleyebilmektedir. Aksi halde bu parametreler diğer yöntemlerin seçilmesi durumunda ara yüzde görülmemektedir. Son olarak araca yeni değişkenlerden oluşan ve “.shp” veri formatında olan Normalleştirilmiş Veri Seti parametresi eklenmiştir.



Şekil 5: Aykırı veri belirleme (a) ve normalizasyon analiz aracı (b) ara yüzleri

2.1.3 Değişken Seçim Analiz Aracı

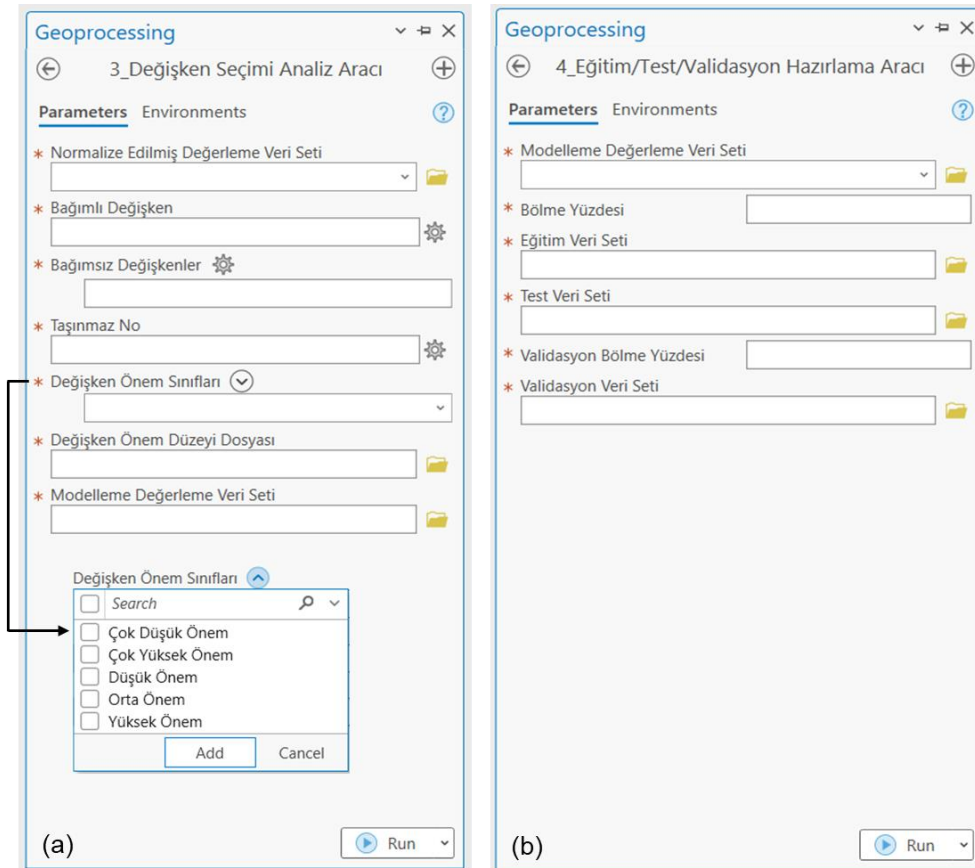
Veri ön işleme modülünün önemli bir aracı olarak Değişken Seçim Analiz Aracı geliştirilmiştir. Bu araç, çok sayıda değişken içeren büyük boyutlu veri setlerinin boyutunu indirmek, model için anlamlı değişkenleri belirleyerek model performansını arttırmak amacıyla geliştirilmiştir (Venkatesh & Anuradha, 2019). Verilerin önem düzeyini belirlemek amacıyla analiz aracı arka planda literatürde de yaygın biçimde kullanılan RF algoritmasını kullanmaktadır (Wang, 2023). Bu algoritmanın kullanılmasının temel sebebi, değişkenlerin önem düzeyinin belirlenmesinde sunduğu yerleşik mekanizmadır (Dewi & Chen, 2019). Algoritmadaki feature_importances_ özelliği, her bir özelliğin modeldeki katkısını değerlendirir ve kullanıcının hangi özelliklerin daha önemli olduğunu anlamasını sağlamaktadır.

Şekil 6a’da kullanıcı ara yüzü verilen Değişken Seçim Analiz Aracı, Kullanıcının Normalizasyon Analiz Aracı sayesinde standardize duruma dönüştürdüğü veri setini girdi olarak kabul etmekte olup, Bağımlı ve Bağımsız Değişkenler parametrelerine ihtiyaç duymaktadır. Bağımlı değişken algoritmanın tahmin etmeye veya açıklamaya çalıştığı hedef olarak açıklanıp tek bir değişkenden oluşmaktadır. Bağımsız değişkenler ise veri seti içerisinde bağımlı değişken dışında kalan ve genellikle bağımsız değişkenlerin tahmini için kullanılan bir ya da daha fazla değişkenden oluşan girdiler olarak tanımlanmaktadır. Bu analiz aracında RF algoritmasıyla bağımlı değişken parametresinin tahminine bulundukları katkılara göre bağımsız değişkenlerin önem düzeyleri hesaplanmaktadır. Araçta yer alan Taşınmaz No isimli parametre, veri setinin benzersiz sütununu girmek için oluşturulmuştur. Ayrıca, araçta değişkenlerin hesaplanan önem düzeylerini kategorize etmek

amacıyla kullanıcı tarafından seçilebilen Değişken Önem Sınıfları parametresi tanımlanmıştır. Bu kapsamda bu parametre için Çok Düşük Önem, Düşük Önem, Orta Önem, Yüksek Önem ve Çok Yüksek Önem kategorileri seçilebilmekte olup, değişkenlerin bu kategorilere göre ataması K-Means kümeleme tekniği kullanılarak gerçekleştirilmektedir. Bu parametre sayesinde kullanıcılar, istedikleri önem düzeyine göre modellere dahil edecekleri değişkenleri seçebilmektedir. Geliştirilen araçta Değişken Önem Düzeyi ve Modelleme Veri Seti isimli iki çıktı tanımlanmıştır. Değişken Önem Düzeyi çıktısı, kullanıcının seçtiği tüm bağımsız değişkenlerin önem düzeylerini hesaplayıp sonuçların “.xlsx” formatında kaydedilmesini sağlamaktadır. Bu dosya içerisinde değişkenlerin önem skorlarının yanı sıra her bir değişkenin önem derecesi (Çok Yüksek, Yüksek vb.) de yer almaktadır. Modelleme Veri Seti çıktısı ise, “.shp” veri formatında üretilen ve çalışma ortamına katman olarak eklenen bir coğrafi veri setini ifade etmektedir.

2.1.4 Eğitim-Test-Validasyon Veri Seti Bölme Analiz Aracı

Makine öğrenmesi algoritmaları ile tahmin modelleri geliştirilebilmesi ve geliştirilen modellerin performans analizleri için eğitim, test ve validasyon veri setlerine gereksinim duyulmaktadır. Bu işlem ise genellikle başlangıçta mevcut olan veri setinin belirli oranlarda bölünmesi ile gerçekleştirilmektedir. Bu kapsamda veri ön işlem modülü için son analiz aracı olarak Eğitim-Test-Validasyon Veri Seti Bölme Analiz Aracı geliştirilmiştir. Geliştirilen analiz aracı, veri setini kullanıcı tarafından belirlenen yüzdelik oranlara göre rastgele biçimde eğitim, test ve validasyon veri seti olarak üçe ayırmaktadır. Söz konusu işlem, modelin yalnızca eğitim verisine aşırı uyum sağlamasını engelleyerek, genelleme yeteneğini artırmayı hedeflemektedir. Şekil 6b’de kullanıcı ara yüzü verilen Eğitim-Test-Validasyon Veri Seti Bölme Analiz Aracı, Değişken Seçim Analiz Aracı ile üretilen Modelleme Değerleme Seti’ni girdi olarak almakta olup veri setini kullanıcının Bölme Yüzdesi/Validasyon Bölme Yüzdesi parametrelerinde belirttiği oranlara göre eğitim, test ve validasyon olmak üzere üçe ayırmaktadır. Analiz araç çıktılarında ise “.shp” veri formatında Eğitim Veri Seti, Test Veri Seti ve Validasyon Veri Seti tanımlanmıştır.



Şekil 6: Değişken seçimi (a) ve Eğitim-Test-Validasyon veri seti bölme analiz aracı (b) ara yüzleri

2.2 Modelleme Modülünün Geliştirilmesi

Makine öğrenmesi modelleri genel anlamda regresyon, sınıflandırma ve kümeleme modelleri olmak üzere üç kısımda incelenmektedir (Aydın, 2018; Basheer vd., 2022; Peker vd., 2017). Bu çalışmada makine öğrenmesinin regresyon kısmıyla ilgili modelleme araçları oluşturulmuş literatürde yaygın biçimde kullanılan ve yüksek modelleme performansı sağlayan RF ve LightGBM algoritmaları analiz araçlarına entegre edilmiştir.

2.2.1 Random Forest ve LightGBM Analiz Araçları

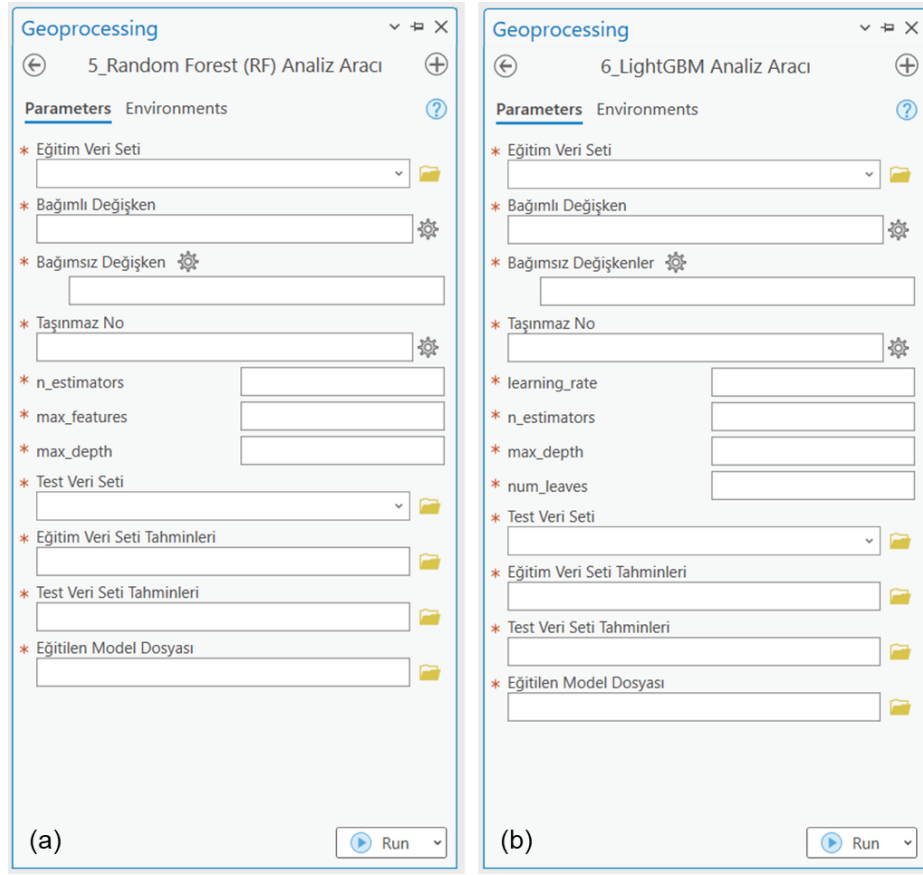
Algoritmaların çalışma prensiplerine bakıldığında; temelinde karar ağaçları bulunan ve bir topluluk öğrenmesi yöntemi olan RF, Breiman (2001) tarafından ortaya atılmıştır. RF algoritmasında, her bir karar ağacı veri setinin rastgele seçilen bir alt kümesi kullanılarak eğitilir ve ağaçlar arasında mümkün olduğunca düşük korelasyona sahip, bağımsız yapılar oluşturulmaya çalışılır (Aydınoglu & Şişman, 2024; Ho, 1995). Karar ağaçları, veri setindeki değişkenler aracılığıyla dallanır ve bu dallanma, değişkenlerin belirli eşik değerlerine göre gerçekleşir. Ancak, tek bir karar ağacı ile çalışıldığında, özellikle veri seti küçükse modelin aşırı öğrenme riski artar (Liao vd., 2016; Veranyurt vd., 2020). RF algoritması karmaşık veriler için birden fazla karar ağacının ortalamasını alarak aşırı öğrenme durumunu önler. Son zamanlarda toplu taşınmaz değerlendirme alanında veri seti içerisindeki özniteliklerin önem düzeylerini ve ilişkilerini detaylı incelemesinden dolayı bu eğitim modeli ön plana çıkmaktadır (Aydınoglu vd., 2023).

LightGBM algoritması ise Microsoft tarafından 2017 yılında geliştirilmiştir (Gan vd., 2021; Ke vd., 2017). GBM tabanlı olup karar ağaçlarını eğitirken yaprak bazlı büyüme stratejisini kullandığından ötürü klasik algoritmadan ayrılan LightGBM, büyük ve karmaşık veri setlerinde çalışırken daha az bellek kullanılmasını sağlar (Fan vd., 2019). Seviye bazlı büyüme, karar ağaçlarını her seviyede tüm yaprakları eş zamanlı olacak şekilde dallandırarak dengeli ve simetrik bir yapı oluşturur, bu durum aşırı öğrenme riskini azaltır. Ancak bu yöntem, yapının genişliğini yavaşça arttırmasından, ağaçları oluştururken daha fazla bellek kullanmasından ve esnekliği sınırlı olmasından ötürü bazı durumlarda doğruluk açısından yetersiz kalmaktadır (Sun vd., 2020; Wang vd., 2024). Yaprak bazlı büyümede ise dallanma, kaybı en çok azaltan yapraktan devam edecek şekilde ilerler dolayısıyla bazı dallar daha çok derinleşirken bazıları daha kısa kalır. Bu durum ağaçların dengesiz bir şekilde genişlemesine sebep olsa da ağaçların daha hızlı ve daha düşük hata oranında oluşmasını sağlamaktadır (Sun vd., 2023). Bu amaçla çalışmada Şekil 7’de kullanıcı ara yüzleri verilen Random Forest ve LightGBM Analiz araçları geliştirilmiştir.

Geliştirilen Random Forest Analiz Aracı’nın parametreleri incelendiğinde; Eğitim-Test Veri Seti Bölme Analiz Aracı ile elde edilen Eğitim ve Test Veri Setlerini girdi olarak kabul etmektedir. Aynı zamanda algoritmanın ihtiyaç duyduğu bağımlı ve bağımsız değişkenler, veri setinin benzersiz değişkeni (Taşınmaz No) kullanıcı tarafından girilmektedir. RF algoritmasının daha iyi bir eğitilmiş model elde etmesi ve model performansının artırılması amacıyla araca modelde kullanılacak ağaç sayısı (n_estimators), maksimum özellik (max_features), oluşturulacak ağacın derinliği (max_depth) gibi model parametreleri kullanıcılar tarafından girilecek şekilde yapılandırılmıştır. Geliştirilen analiz aracı için Eğitim Veri Seti Tahminleri, Test Veri Seti Tahminleri ve Eğitilen Model Dosyası olmak üzere üç çıktı tanımlanmıştır. Eğitim Veri Seti Tahminleri, eğitilmiş model tarafından eğitim veri setinin tekrar tahmin edilmesi sonucunda üretilen tahminleri içeren veri setini, Test Veri Seti Tahminleri eğitilmiş model tarafından test veri setinin tahmin edilmesi sonucunda üretilen tahminleri içeren veri setini ve Eğitilen Model Dosyası ise analiz aracının eğitim veri seti ile ürettiği modeli ifade etmektedir (Şekil 7a).

Random Forest Analiz Aracı’na benzer biçimde geliştirilen LightGBM Analiz Aracı’nda girdi parametreler olarak Eğitim ve Test Veri Setleri, veri setinin benzersiz değişkeni (Taşınmaz No) ve LightGBM algoritmasının kendine özgü parametreleri tanımlanmıştır. Bu kapsamda Random Forest Analiz Aracı’ndan farklı olarak, iteratif öğrenme süreçlerinde önemli bir rol oynayan learning_rate parametresi eklenmiştir. Bu parametre, modelin öğrenme sürecinde her adımda ne kadar ilerleyeceğini

kontrol eder ve modeldeki ağırlık güncellemelerinin hızını veya büyüklüğünü belirlemek için kullanılmaktadır. Aynı zamanda num_leaves parametresi eklenerek karar ağaçlarında oluşacak yaprak sayısının kontrol edilmesi amaçlanmıştır. Geliştirilen analiz aracı için çıktı olarak Random Forest Analiz Aracı ile aynı şekilde Eğitim Veri Seti Tahminleri, Test Veri Seti Tahminleri ve Eğitilen Model Dosyası tanımlanmıştır (Şekil 7b).

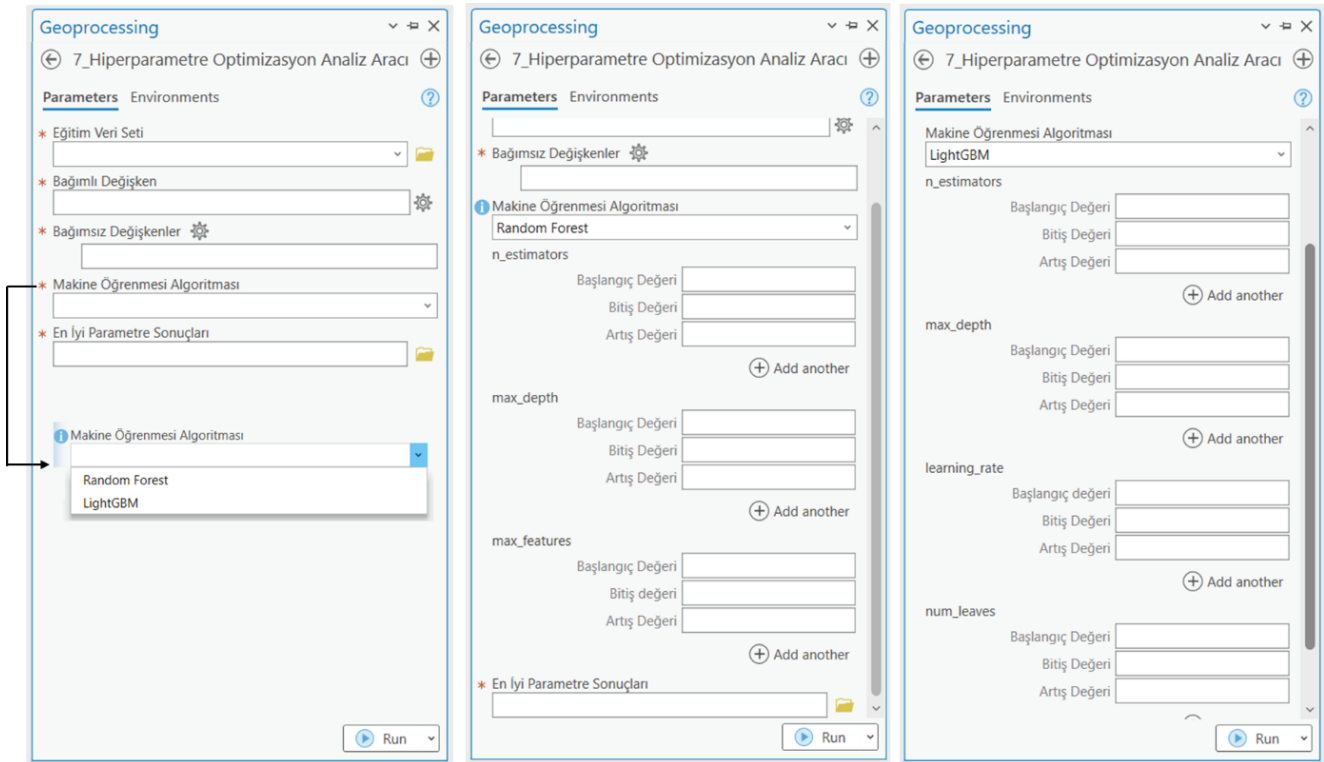


Şekil 7: Random Forest (a) ve LightGBM analiz aracı (b) ara yüzleri

2.2.2 Hiperparametre Optimizasyon Analiz Aracı

Makine öğrenmesinde hiperparametre optimizasyonu, bir modelin performansını doğrudan etkileyen algoritmaya özgü hiperparametrelerin en uygun değerlerinin sistematik olarak belirlenmesi süreci olarak ifade edilmektedir. Hiperparametreler aracılığıyla modelin karmaşıklığı, öğrenme hızı, genelleme kapasitesi ve hesaplama maliyeti gibi modelleme sürecine ilişkin temel unsurlar kontrol edilmektedir. Optimizasyon sürecinde, tanımlanan bir parametre değer aralığı içerisinde farklı hiperparametre kombinasyonları denenmekte ve her bir kombinasyon için geliştirilen modelin performansı belirli performans ölçütleri kullanılarak değerlendirilmektedir (Nguyen vd., 2025; Quan, 2024). Elde edilen sonuçlar doğrultusunda, en yüksek performans sağlayan hiperparametre kombinasyonu belirlenerek modelin genelleme yeteneği artırılmakta; aşırı ve yetersiz öğrenme gibi modelleme problemlerinin giderilmesi amaçlanmaktadır. Bu bağlamda hiperparametre optimizasyonu, güvenilir ve yüksek performans sağlayan makine öğrenmesi modellerinin geliştirilmesinde kritik bir aşama olarak değerlendirilmektedir (Şişman vd., 2025). Hiperparametre optimizasyonu için çeşitli yöntemler bulunmakla birlikte, önceden tanımlanan parametre aralıkları içerisindeki tüm olası kombinasyonların sistematik olarak değerlendirilmesine olanak tanıyan Grid Search yöntemi yaygın biçimde kullanılmaktadır (Ogunsanya vd., 2023). Bu doğrultuda, çalışmada Grid Search yöntemine dayalı olarak en uygun hiperparametre değerlerini belirleyen ve kullanıcı arayüzü Şekil 8'de verilen Hiperparametre Optimizasyon Analiz Aracı geliştirilmiştir.

Geliştirilen analiz aracı Eğitim Veri Setini girdi olarak kabul etmekte olup, makine öğrenmesi algoritmalarının ihtiyaç duyduğu bağımlı ve bağımsız değişkenler kullanıcı tarafından girilmektedir. Makine Öğrenmesi Algoritması parametresi kapsamında kullanıcı seçimli olacak biçimde RF ve LightGBM algoritmaları tanımlanmıştır. RF algoritması için `n_estimators`, `max_depth` ve `max_features` hiperparametreleri; LightGBM algoritması için ise `n_estimators`, `max_depth`, `learning_rate` ve `num_leaves` hiperparametreleri dikkate alınmıştır. Analiz aracında her bir hiperparametre için arama uzayını belirlemek amacıyla Başlangıç Değeri, Bitiş Değeri ve Artış Değeri tanımlanmış olup, bu sayede en iyi model performansının elde edilmesine yönelik sistematik bir hiperparametre aralığı belirlenebilmektedir. Geliştirilen aracın çıktısı olarak En İyi Parametre Sonuçları tanımlanmıştır. Bu çıktı, kullanıcının seçtiği makine öğrenmesi algoritmasına göre belirlenen aralıklarda parametre kombinasyonlarının tümünü değerlendirerek en yüksek performans sağlayan model parametrelerinin “.txt.” formatında kaydedilmesini sağlamaktadır. Bu analiz aracı sayesinde modellerin performansı hiperparametreler için en uygun değerlerin belirlenmesi ile iyileştirilmekte ve yüksek performans sağlayan final model elde edilmektedir.



Şekil 8: Hiperparametre optimizasyon analiz aracı ara yüzü

2.3 Performans Değerlendirme Modülünün Geliştirilmesi

Makine öğrenmesi algoritmalarına dayalı olarak gerçekleştirilen regresyon problemlerinde, farklı algoritmalara ilişkin tahmin doğruluğunun değerlendirilmesi için çeşitli hata ölçütleri ve model performans analizi yaklaşımları bulunmaktadır. Bunlardan en çok kullanılanları; Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) ve R^2 (R-squared) ölçütleridir (Alburshaid & Al-Alawi, 2024; Alizamir vd., 2020; Aydınoğlu vd., 2025; Dimopoulos & Bakas, 2019; Hong vd., 2020; Mayer vd., 2019; Yılmaz & Kocaman, 2020). Bu model performans metriklerine ilişkin genel eşitlikler aşağıda verilmiştir:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

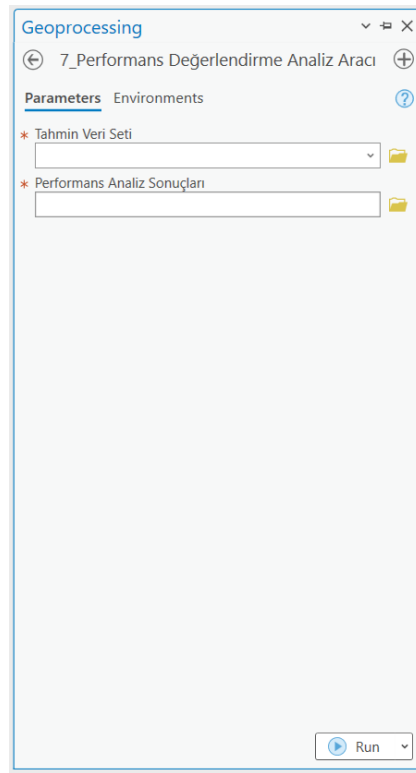
$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

Eşikliklerde y_i : bağımlı değişkenin gerçek değerini, $\hat{y}_i$: bağımlı değişkenin tahmin edilen değerini, $\bar{y}$: gerçek değerlerin ortalamasını ve n ise model performans analizinde kullanılacak veri sayısını ifade etmektedir.

Çalışma kapsamında son araç olarak tahmin modellerinin performansını Eşitlik 7-10 arasında verilen metriklerle değerlendirmek amacıyla Performans Değerlendirme Analiz Aracı geliştirilmiştir. Şekil 9'da kullanıcı ara yüzü verilen Performans Değerlendirme Analiz Aracı, Random Forest ve LightGBM Analiz araçları ile üretilen eğitim veya test verisi için oluşturulan Tahmin Veri Seti'ni girdi olarak kabul etmektedir. Çıktı olarak ise kullanıcılar tarafından girdi olarak verilen tahminler için Modellerin Performans Sonuçları tanımlanmıştır.



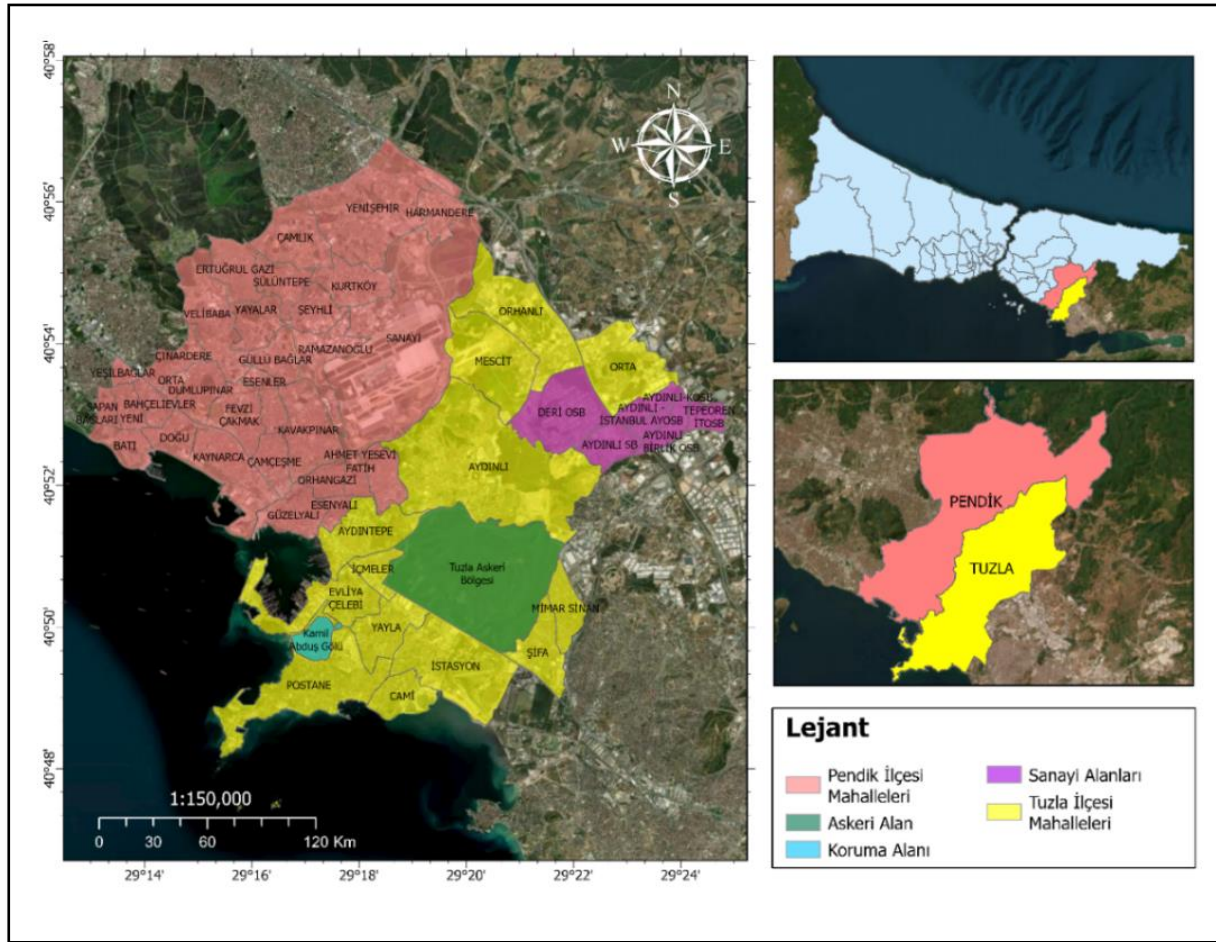
Şekil 9: Performans değerlendirme analiz aracı ara yüzü

3. Mekânsal Yapay Zekâ (GeoAI) Analiz Araçları ile Toplu Taşınmaz Değerleme

3.1 Çalışma Alanının Belirlenmesi ve Veri Setlerinin Elde Edilmesi

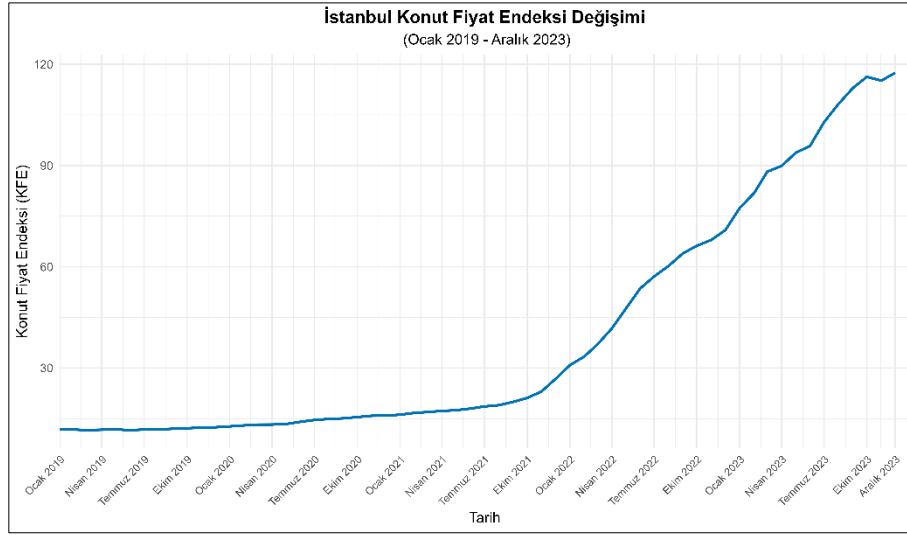
Geliştirilen analiz araçları ile otomatize biçimde toplu taşınmaz değerlendirme modelleri geliştirmek ve model performansını test etmek amacıyla pilot uygulama gerçekleştirilmiştir. Çalışma alanı olarak İstanbul ilinin Pendik ve Tuzla ilçelerinden mahalleler belirlenmiştir. İstanbul'un doğusunda yer alan bu iki ilçe hem kentsel dönüşüm dinamikleri hem de aktif konut piyasaları ile dikkat çekmektedir. 2024 yılı nüfus verilerine göre, Pendik ilçesinin nüfusu 749 356, Tuzla ilçesinin ise 301 400'dür (TÜİK, 2025a). Aynı yıl içinde Pendik'te 10 654, Tuzla'da ise 7460 konut satışı gerçekleşmiş olup bu ilçeler, nüfus artışıyla paralel olarak konut sektöründe hareketlilik göstermektedir (TÜİK, 2025b). Her iki ilçede de kentsel ve kırsal yerleşim alanları bir arada bulunmakta, arazi kullanımı açısından çeşitlilik gözlemlenmektedir. Ancak bu çalışma doğrudan

konut tipindeki taşınmazlara odaklandığı için kentsel yerleşim alanlarına odaklanılmıştır. Askeri alanlar, kırsal bölgeler ve sanayi bölgeleri çalışma alanından hariç tutulmuş olup, Pendik ilçesinden 31, Tuzla ilçesinden 13 olmak üzere toplam 44 mahalle çalışma alanını oluşturmuştur (Şekil 10).



Şekil 10: Çalışma alanının belirlenmesi

Çalışma alanında 2019 Ocak – 2023 Aralık dönemine ait 37 268 konut tipindeki piyasa örneklemini içeren veri seti, Endeksa firmasından temin edilmiştir (Endeksa, 2025). Temin edilen veri seti taşınmaz fiyat değişkeninin yanı sıra konut tipindeki taşınmazların değerini etkileyen birçok bina ve bağımsız bölüm özelliklerini temsil eden değişken içermektedir. Fiyat değişkeninin zamansal olarak güncellenmesi amacıyla, T.C. Merkez Bankası tarafından konut piyasasındaki fiyat değişimlerini izlemek üzere oluşturulan ve aylık olarak yayımlanan Konut Fiyat Endeksi (KFE) değerleri, İstanbul ili için 2019 Ocak – 2023 Aralık dönemini kapsayacak biçimde Şekil 11’de sunulduğu üzere temin edilmiştir. Ardından KFE değerleri ile $KFE \text{ Güncelleme Oranı} = KFE_{Aralık2023} / KFE_{Güncelleme \text{ Tarihi}}$ şeklinde zamansal olarak getirilecek düzeltme oranı hesaplanmıştır. Hesaplanan güncelleme oranları kullanılarak veri setindeki taşınmaz fiyat değişkeni güncellenmiştir. Ayrıca temin edilen veri seti konumsal analizlerle zenginleştirilmiş olup, Tablo 1’de ifade edilen mahalli, konumsal, bina ve bağımsız bölüm özellikleri kategorilerinde toplam 68 adet değişkeni içermektedir.



Şekil 11: 2019 Ocak-2023 Aralık döneminde İstanbul konut fiyat endeksi değişimi (TCMB, 2025)

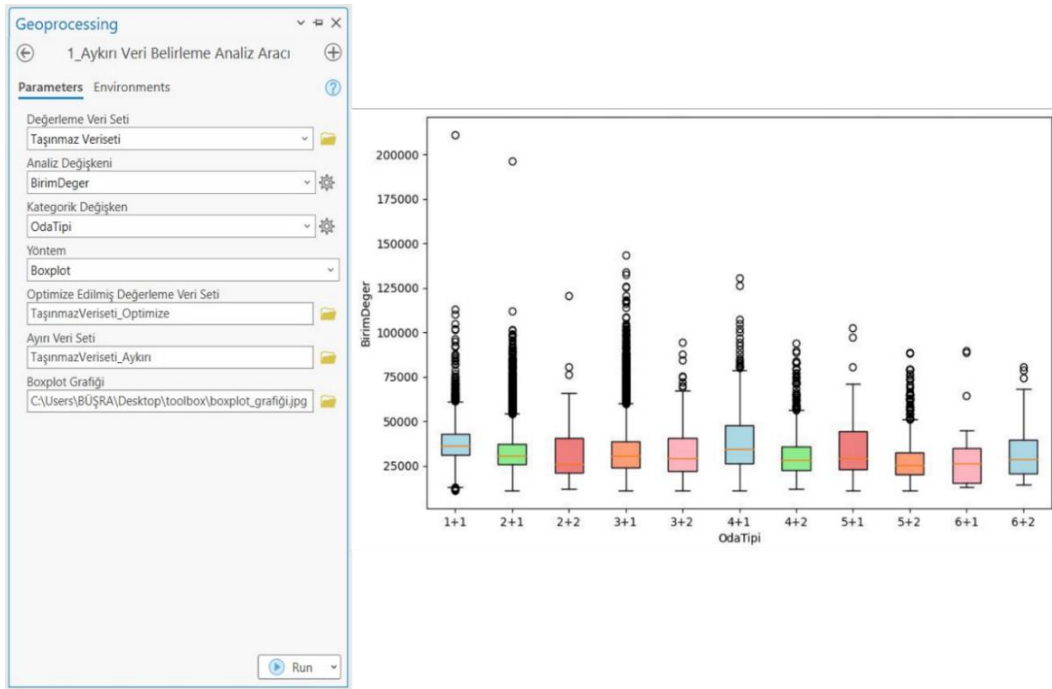
Tablo 1: Taşınmaz değerini etkileyen değişkenlere ilişkin veri seti içeriği

Mahalli Özellikler	Nüfus (Mahalle Nüfus Yoğunluğu (kişi/km ²), Mahallede Yaşayan Çocuk (0-14), Genç (15-29), Yetişkin (30-44), Yaşlı (60+), Kadın, Erkek Yüzdesi), Gelir Düzeyi (A, A+), Eğitim Durumu (Mahallede Yaşayan Lisans Mezunu ve Lisansüstü Mezunu Kişi Yüzdesi), Yapı Yoğunluğu, Mahalli Sosyo-Ekonomik Gelişmişlik Düzeyi, Arsa Rayiç Değeri
Konumsal Özellikler	Eğitim Tesislerine Mesafe, Sağlık Tesislerine Mesafe (Devlet Hastaneleri, Üniversite Hastaneleri, Özel Hastanelere Mesafe), Alışveriş & Ticaret Tesislerine Mesafe (Alışveriş Merkezlerine, Çarşı Merkezlerine Mesafe), Kültürel Tesislere Mesafe, Kamu Hizmet Tesislerine Mesafe, Yeşil Alanlara Mesafe, Konaklama Tesislerine Mesafe, İbadethanelere Mesafe, Raylı Sistem Fonksiyonlarına Mesafe (Metro, Marmaray İstasyonlarına Mesafe), Karayolu Fonksiyonlarına Mesafe (Otoyollara, Caddeye Mesafe), Havaalanlarına Mesafe, Denizyolu Ulaşım İstasyonlarına Mesafe, Otobüs Duraklarına Mesafe
Bina Özellikleri	Bina Yaşı, Bina Toplam Kat Adedi, İmkanlar (Yüzme Havuzu, Asansör Mevcudiyeti, Otopark Mevcudiyeti, Site İçinde Olma, Çocuk Oyun Parkı Mevcudiyeti)
Bağımsız Bölüm Özellikleri	Konut Alanı, Oda Sayısı, Oda Tipi, Konutun Bulunduğu Kat, Manzara, Cephe Yönü, Balkon Mevcudiyeti, Banyo Sayısı, Isıtma Sistemi

3.2 Alan Odaklı Aktarımlı Öğrenme ile Tek Aşamalı Ön Eğitim Yaklaşımı

3.2.1 Aykırı Verilerin Belirlenmesi

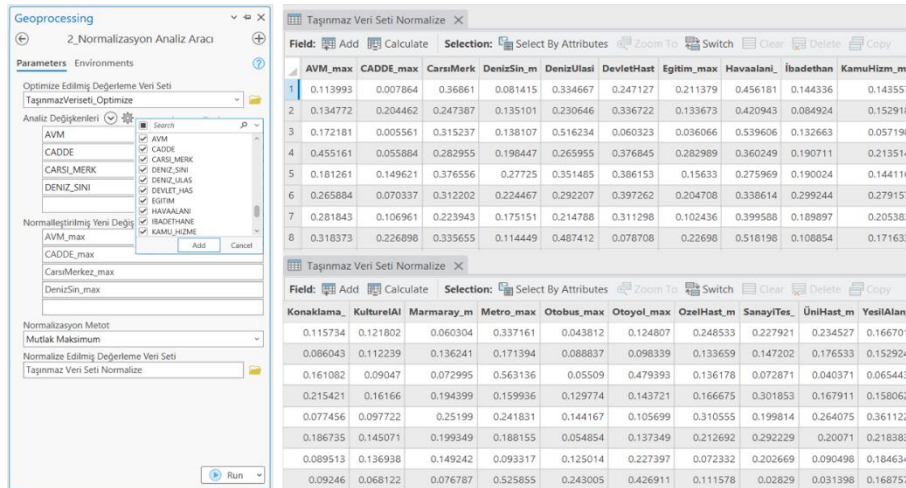
Oluşturulan 37 268 adet piyasa örneklemini içeren değerlendirme veri seti geliştirilen analiz araçları ile öncelikle veri ön işleme sürecine tabi tutulmuştur. Bu kapsamda ilk olarak veri seri içerisindeki Birim Değer (₺/m²) değişkeni için aykırı veriler Aykırı Veri Belirleme Analiz Aracı ile Şekil 12'deki gibi değerlendirilmiştir. Boxplot yöntemi ile oda tipi kategorik değişkenine göre toplam da 1413 adet aykırı veri tespit edilerek veri seti içerisinde çıkarılarak optimize edilmiş değerlendirme veri seti elde edilmiştir.



Şekil 12: Aykırı veri belirleme analiz aracı-Boxplot yöntemi ile aykırı verilerin belirlenmesi

3.2.2 Veri Seti Normalizasyonu

Aykırı verilerin tespiti ve veri seti içerisinde çıkarılmasıyla optimize edilmiş bir veri seti elde edilmiştir. Bu veri seti farklı birim ve büyüklüklerde değişkenler içerdiğinden Normalizasyon Analiz Aracı kullanılarak standardize bir duruma dönüştürülmüştür. Bu süreçte çalışmada kullanılan veri setinin dağılım özellikleri dikkate alınarak en uygun ölçekleme yöntemi olarak Mutlak Maksimum Normalizasyon yöntemi belirlenmiş tüm değişkenler bu yöntem aracılığıyla 0 ile 1 değerleri arasına getirilerek standart bir forma dönüştürülmüştür (Şekil 13).

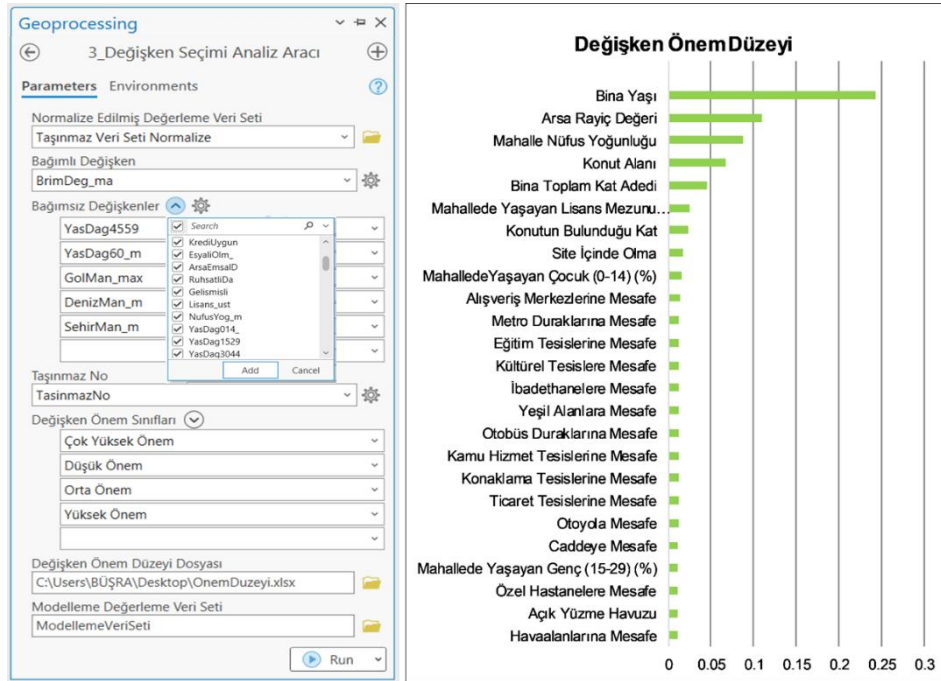


Şekil 13: Normalizasyon analiz aracı- Mutlak Maksimum normalizasyon yöntemi ile değişkenlerin standardize edilmesi

3.2.3 Toplu Taşınmaz Değerleme Modelleri Değişkenlerinin Belirlenmesi

Yüksek performans sağlayan toplu taşınmaz değerlendirme modellerinin geliştirilebilmesi amacıyla Değişken Önem Düzeyi Analiz Aracı ile veri seti içerisindeki 68 değişkenin tamamı için değişken önem düzeyi değerlendirilmiştir. Analiz aracının Değişken Önem Sınıfları parametresinde Çok Düşük Önem kategorisi dışındaki tüm sınıflar dikkate alınmış olup, analiz aracı ile model geliştirme süreçlerinde kullanılmak üzere toplam 25 değişken belirlenmiştir (Şekil 14). Analiz aracında bu değişkenler ile yeni bir modelleme veri seti oluşturulmuştur. Yeni oluşturulan veri seti, modelleme sürecinde kullanılacak

değişken sayısının azaltılması yoluyla hem işlem hızının artırılması hem de modelin genel doğruluk performansının iyileştirilmesi için kritik bir adım olarak değerlendirilmektedir. Bina Yaşı, Arsa Rayiç Değeri, Mahalle Nüfus Yoğunluğu, Konut Alanı ve Bina Toplam Kat Adedi gibi faktörlerin önem düzeylerinin yüksek olduğu gözlemlenmiştir.



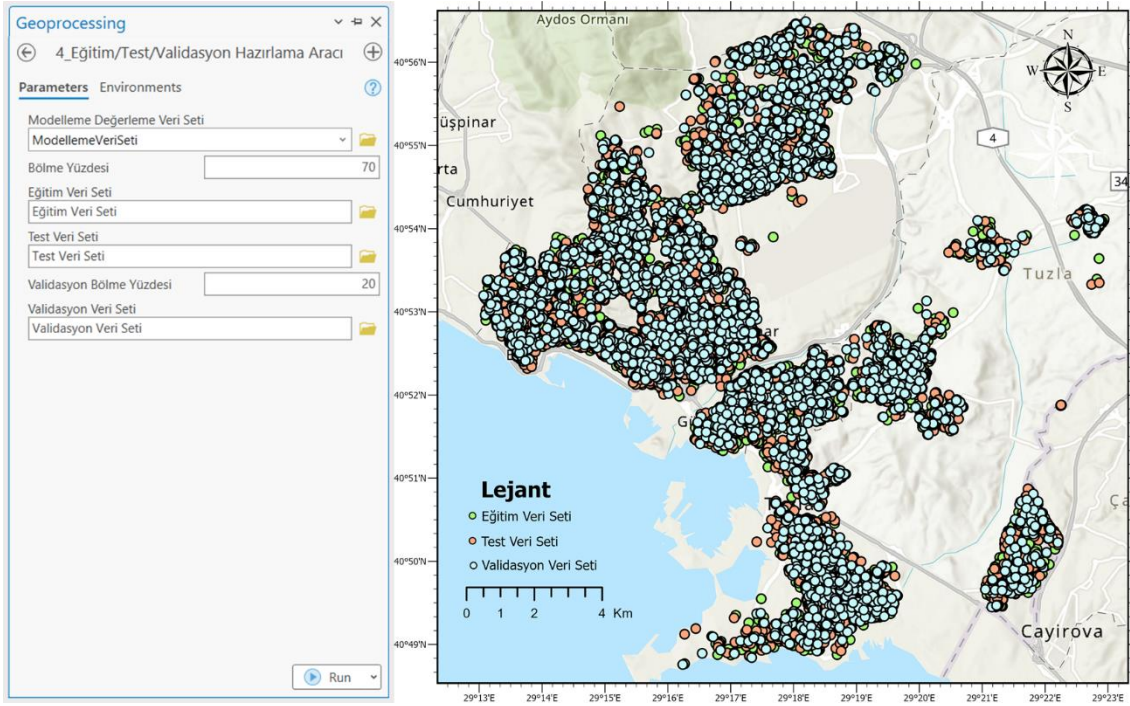
Şekil 14: Değişken seçim analiz aracı ile modelleme değişkenlerinin belirlenmesi

3.2.4 Eğitim, Test ve Validasyon Veri Setlerinin Oluşturulması

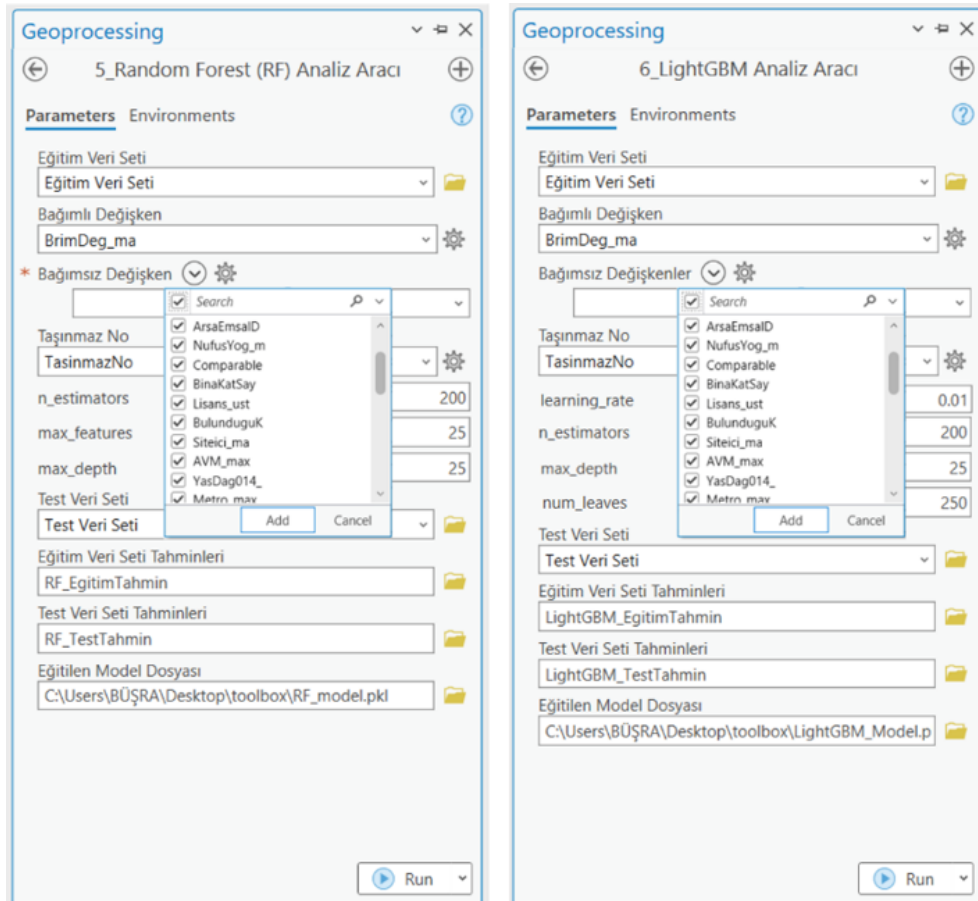
Önem düzeyi yüksek değişkenlerle modelleme veri setinin oluşturulmasının ardından, Eğitim-Test Veri Seti Bölme Analiz Aracı ile toplu taşınmaz değerlendirme modellerinin eğitim ve test süreçlerinde kullanılmak üzere veri setleri oluşturulmuştur. Bu analiz aracı ile 35 855 adet verinin %70'i eğitim verisi (25 098 adet), %30'u ise test verisi (10 757 adet) ve eğitim verisinin %20 si validasyon (5020 adet) olacak şekilde rastgele şekilde bölünmüştür (Şekil 15). Eğitim, test ve validasyon veri setleri ayrı katmanlar olarak kaydedilmiş olup, modelleme ve performans analiz aşamalarında kullanılmak üzere hazır duruma getirilmiştir.

3.3 Toplu Taşınmaz Değerleme Modellerinin Geliştirilmesi

Geliştirilen Random Forest ve LightGBM Analiz araçlarında eğitim veri seti girdi olarak kullanılarak iki farklı toplu taşınmaz değerlendirme modelinin eğitimi gerçekleştirilmiştir. Modellerde Birim Değer (₺/m²) bağımlı değişken, Değişken Seçim Analiz Aracı ile belirlenen Şekil 14'teki 25 değişken ise bağımsız değişkenler olarak tanımlanmıştır (Şekil 16). Modellerin geliştirilmesinde Random Forest Analiz Aracı'nda oluşturulacak karar ağacı sayısını ifade eden n_estimator: 200, oluşturulan karar ağaçlarının derinliğini sınırlandırmak amacıyla kullanılan max_depth:25 ve her ağaçta göz önünde bulundurulacak maksimum öznelik sayısını belirleyen max_features: 25 olarak ayarlanmıştır. Benzer şekilde LightGBM Analiz Aracı'nda da learning_rate: 0.01, n_estimator: 200, max_depth:25 ve num_leaves: 250 olarak tanımlanmıştır.

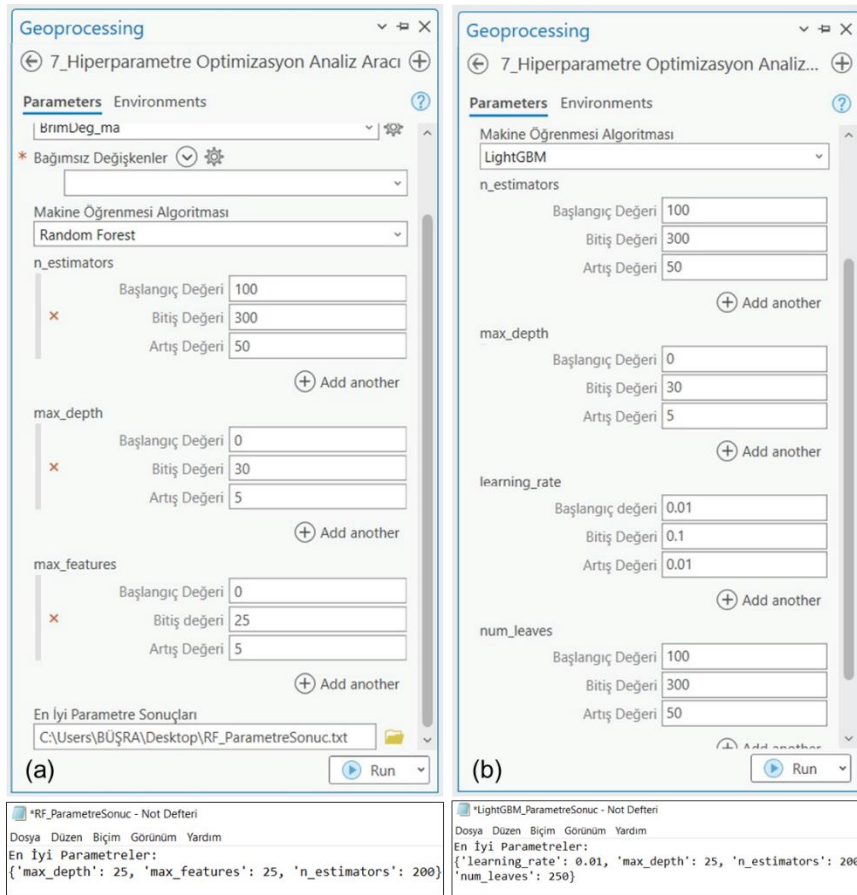


Şekil 15: Eğitim-Test-Validasyon veri seti bölme analiz aracı ile eğitim, test ve validasyon veri setlerinin oluşturulması



Şekil 16: Random Forest ve LightGBM analiz araçları ile toplu taşınmaz değerleme modellerinin geliştirilmesi

Geliştirilen Hiperparametre Optimizasyon Analiz Aracı kullanılarak en uygun parametre değerleri belirlenmiştir. Buna göre, RF algoritması için max_depth: 25, max_features:25 ve n_estimators:300 olarak tespit edilmiştir. LightGBM algoritması için ise learning_rate: 0.01, n_estimators:200, max_depth:25 ve num_leaves: 250 değerleri en uygun parametreler olarak belirlenmiş olup, final tahmin modelleri bu parametreler kullanılarak geliştirilmiştir (Şekil 17).



Şekil 17: Hiperparametre optimizasyon analiz aracı ile Random Forest (a) ve LighGBM (b) modellerinin optimizasyonu

3.4 Model Performans Ölçümü ve Değişken Önem Düzeylerinin Değerlendirilmesi

Modelleme sürecinde geliştirilen final RF ve LightGBM modellerinin performansları, test veri seti üzerinden Performans Değerlendirme Analiz Aracı ile Şekil 18'deki gibi değerlendirilmiştir. Bu kapsamda, her iki modelin tahmin başarısı MAE, MAPE, RMSE ve R^2 ölçütleri ile değerlendirilmiş olup, model performans sonuçları Tablo 2'de özetlenmiştir.

Figure 18 shows the Performance evaluation analysis tool interface. The tool displays a table of results for the Random Forest (a) and LightGBM (b) models. The table has columns for OBJECTID, Shape, TasinmazNo, Satis_Degeri, and Tahmini_Deger. The results are shown for 11 objects, with the first 10 objects having a Shape of 'Point' and the 11th object having a Shape of 'Polygon'.

OBJECTID	Shape	TasinmazNo	Satis_Degeri	Tahmini_Deger
1	Point	2	0.57862	0.575167
2	Point	6	0.279231	0.37866
3	Point	8	0.345231	0.350705
4	Point	9	0.196154	0.189084
5	Point	10	0.561391	0.458344
6	Point	11	0.341716	0.301766
7	Point	13	0.552115	0.487269
8	Point	17	0.568966	0.423768
9	Point	26	0.661058	0.594072
10	Point	33	0.185503	0.17803
11	Polygon	34	0.550995	0.445258

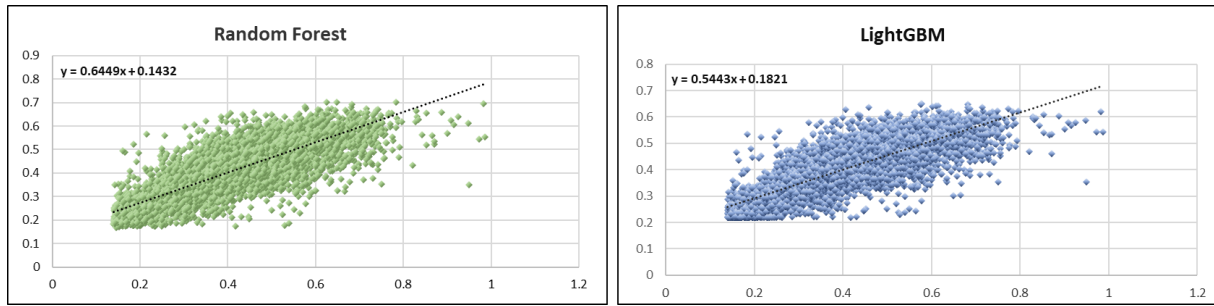
Şekil 18: Performans değerlendirme analiz aracı ile model performanslarının değerlendirilmesi

Model performans analiz sonuçları değerlendirildiğinde, her iki model de genel anlamda makul tahmin sonuçları üretmiş olsa da, RF algoritmasının hata metrikleri açısından nispeten daha düşük değerler sunması ve R^2 değerinin daha yüksek olması,

bu algoritmanın çalışmada kullanılan veri seti üzerinde daha iyi bir öğrenme başarısı gösterdiğini ortaya koymaktadır. Özellikle R^2 skorunun 0.664 olması, bağımlı değişkenin varyansının %66.4'ünün model tarafından açıklandığını ifade etmektedir. Bu oran, modelin veri içindeki ilişkileri etkili bir biçimde öğrendiğini göstermektedir. Geliştirilen modellerin tahmin sonuçlarına ait veri saçılımları Şekil 19'da sunulmuş olup, her iki modelde de saçılımın eğilim çizgisi etrafında toplandığı gözlemlenmiştir. Öte yandan, LightGBM algoritması literatürde özellikle büyük ve yüksek boyutlu verilerde avantajlı bir yöntem olarak bilinmekle birlikte, bu çalışmada kullanılan veri setinin boyutu ve değişken dağılımı dikkate alındığında RF algoritması daha uygun ve istikrarlı sonuçlar vermiştir. Elde edilen sonuçlar, model seçiminde veri setinin yapısal özelliklerinin dikkate alınmasının önemini ortaya koymakta ve benzer çalışmalarda kullanılacak algoritmaların seçiminde yol gösterici nitelik taşımaktadır.

Tablo 2: Taşınmazlara ait zemin, imar durum ve değer bilgileri

Model	MAE	MAPE	RMSE	R^2
RF	0.051	13.598	0.072	0.664
LightGBM	0.055	14.905	0.074	0.639

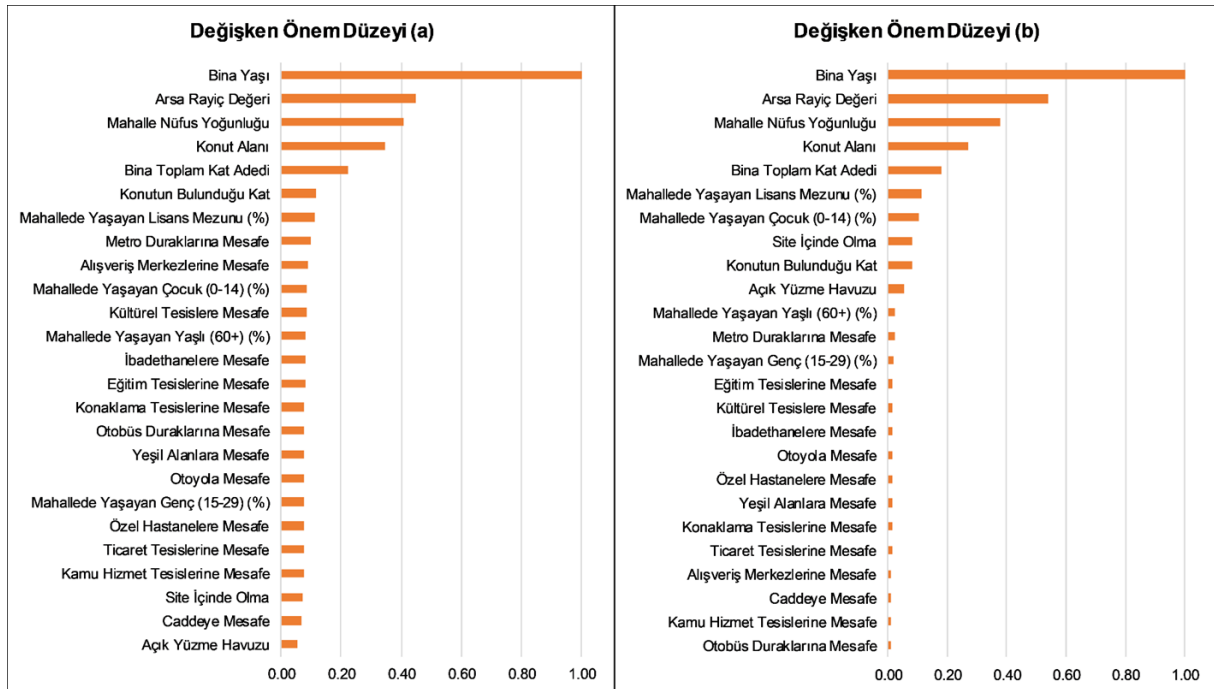


Şekil 19: Tahmin modellerinin saçılım grafikleri

Modellerin performans analizlerinin ardından, modellere dâhil edilen değişkenlerin göreceli önem düzeyleri değerlendirilmiş ve sonuçlar Şekil 20'de verilmiştir. Her iki modelde de bina yaşı, arsa rayiç değeri, mahalle nüfus yoğunluğu, konut alanı, bina toplam kat adedi, konutun bulunduğu kat ve mahallede yaşayan lisans mezunu (%) gibi yapısal ve mahalli değişkenlerin taşınmaz değer tahmininde ön plana çıktığı gözlemlenmiştir. Bu durum, taşınmaz değerinin yalnızca bina ve bağımsız bölümün fiziksel özellikleriyle değil, aynı zamanda bulunduğu çevrenin sosyo-ekonomik ve mekânsal karakteristikleriyle birlikte şekillendiğini; kullanılan veri setinin yapısı da dikkate alındığında, çalışma alanında değer büyük ölçüde bu bileşenlerin etkileşimi sonucunda oluştuğunu ortaya koymaktadır. Diğer taraftan, ulaşım tesisleri, alışveriş tesisleri, eğitim tesisleri, kültürel tesisler, kamusal hizmet tesisleri ve ibadethanelere olan mesafeler gibi konumsal özellikleri temsil eden değişkenlerin her iki modelde de orta düzeyde önem düzeyine sahip olduğu görülmektedir. Bu durum söz konusu değişkenlerin taşınmaz değerini etkilediğini, ancak bina ve bağımsız bölüm özelliklerini temsil eden değişkenlere kıyasla daha ikincil bir rol üstlendiğini göstermektedir. Mahalle düzeyindeki yaş grupları ve eğitim düzeyine ilişkin demografik özelliklerin modellere katkısı ise, özellikle taşınmazların bulunduğu bölgedeki sosyo-ekonomik yapının taşınmaz değeri üzerindeki açıklayıcı gücünü ortaya koymaktadır. Elde edilen bu sonuçlar, taşınmaz değerinin yapısal, fiziksel, konumsal ve sosyo-ekonomik bileşenlerin birlikte değerlendirilmesi gerektiğini ortaya koyan ve makine öğrenmesi temelli çalışmaların bulgularıyla paralellik göstermektedir (Aydınoğlu vd., 2023; Chen vd., 2020; Genç vd., 2025; Iban, 2022; Jafary vd., 2024; Tchuente, 2024; Yilmazer & Kocaman, 2020).

Değişken önem düzeyleri modeller arasında karşılaştırıldığında, RF algoritmasının değişken önemlerini daha dengeli bir biçimde değerlendirdiği, LightGBM modelinin ise bazı değişkenleri daha baskın şekilde öne çıkardığı görülmektedir. Bu

farklılık, algoritmaların karar ağaçlarını oluşturma ve değişkenler arasındaki karmaşık ilişkileri öğrenme yaklaşımları ile ilişkilendirilebilir. RF modelinde karar ağaçlarının birbirinden bağımsız ve rastgele alt örneklemeler üzerinden oluşturulması, değişken önemlerinin daha homojen bir dağılım göstermesine olanak tanımaktadır. Bu yaklaşım, orta ve nispeten düşük önem düzeyine sahip değişkenlerin de modele katkısının korunmasını sağlamakta ve genel eğilimlerin daha dengeli biçimde yansıtılmasına imkân vermektedir. LightGBM modeli ise gradyan tabanlı artırma (boosting) yaklaşımı ile tahmin hatalarını minimize etmeye odaklanmakta; bu yaklaşım özellikle bina yaşı, arsa rayiç değeri, mahalle nüfus yoğunluğu ve konut alanı gibi yüksek açıklayıcılığa sahip değişkenlere daha yüksek ağırlıklar atayarak, diğer değişkenlerin marjinal katkılarını görece olarak azaltma eğilimi göstermektedir. LightGBM modeli ise gradyan tabanlı artırma (boosting) yaklaşımıyla tahmin hatalarını minimize etmeye odaklanmakta; bu doğrultuda bina yaşı, arsa rayiç değeri, mahalle nüfus yoğunluğu ve konut alanı gibi yüksek açıklayıcılığa sahip değişkenlere daha yüksek ağırlıklar atayarak, diğer değişkenlerin marjinal katkılarını görece olarak azaltma eğilimi göstermektedir. Algoritmalar arasındaki bu farklılıklar, özellikle daha düşük önem düzeyine sahip değişkenlerde belirginleşmektedir. LightGBM modelinin yaprak bazlı büyüme stratejisi, verideki karmaşık ve doğrusal olmayan ilişkileri daha agresif biçimde yakalayabilmesini sağlamaktadır; bu nedenle “Site İçinde Olma” ve “Açık Yüzme Havuzu” gibi kukla/kategorik değişkenler RF modeline kıyasla daha yüksek göreceli önem düzeyleri ile değerlendirilmiştir. Buna karşılık RF modeli, seviye bazlı büyüme yapısı sayesinde değişkenler arasında daha kademeli ve dengeli bir önem düzeyi ayrımı sunmaktadır. Bu bağlamda, RF modeli değişkenlerin önem düzeylerini dengeli bir çerçevede sunarken, LightGBM modeli taşınmaz değerini doğrudan etkileyen belirli niteliklerin ayırt edici gücünü daha keskin bir hiyerarşi ile ortaya koymaktadır. Sonuçlar, toplu taşınmaz değerlendirme uygulamalarında yalnızca model performansının değil aynı zamanda değişken önem düzeyleri ve model yorumlanabilirliğinin de önemli bir değerlendirme ölçütü olduğunu göstermektedir.



Şekil 20: RF (a) ve LightGBM (b) modelleri ile elde edilen değişken önem düzeyleri

4. Sonuçlar ve Öneriler

Günümüzde kentleşme dinamiklerinin hızla değişmesi, taşınmaz piyasalarının daha kompleks bir yapıya dönüşmesi ve taşınmaz değerine gereksinim duyan karar vericilerin/sektörel paydaşların doğru değer bilgisine hızlı biçimde erişim ihtiyacının artması, tekil taşınmaz değerlendirme yöntemlerinin yetersiz kaldığı noktaları görünür kılmıştır. Özellikle büyük taşınmaz değerlendirme veri setlerinin analizi ile bu verilerden anlamlı sonuçlar çıkarılması, klasik değerlendirme yöntemlerinin

ötesinde çözümler gerektirmektedir. Bu gereksinim, yapay zekâ ve CBS gibi teknolojilerin entegrasyonunu içeren GeoAI gibi yenilikçi yaklaşımlarla otomatize biçimde toplu taşınmaz değerlendirme süreçlerinin gerçekleştirilmesinin önemini artırmaktadır.

Bu çalışma kapsamında otomatize toplu taşınmaz değerlemesine yönelik analiz araçlarını içeren GeoAI analiz araç paketi geliştirilmiştir. ArcGIS platformuna entegre biçimde geliştirilen bu araç paketi; veri ön işleme (aykırı değer belirleme, normalizasyon, değişken seçimi, eğitim-test ayrımı), regresyon tabanlı modelleme, hiperparametre optimizasyonu ve model performans değerlendirme süreçlerini bütünsel biçimde gerçekleştirilmesini mümkün kılmaktadır. Geliştirilen analiz araç paketini test etmek amacıyla İstanbul ilinin Tuzla ve Pendik ilçeleri çalışma alanı olarak belirlenmiştir. Çalışma alanında 2019 Ocak-2023 Aralık dönemi için 37 268 adet piyasa örnekleme temin edilmiş olup, piyasa örneklemelerinin coğrafi analizlerle zenginleştirilmesiyle toplamda 68 değişken içeren toplu taşınmaz değerlendirme veri seti oluşturulmuştur. Araç paketi içerisindeki veri ön işleme araçları ile veri setinin aykırı veriler açısından optimizasyonu, normalizasyonu, eğitim/test veri setlerinin oluşturulması ve model değişkenlerinin belirlenmesi işlemleri gerçekleştirilmiştir. Modelleme araçları ile eğitim veri seti üzerinden toplu taşınmaz değerlendirme modelleri geliştirilmiştir. Son aşamada ise performans değerlendirme analiz aracı ile test veri seti üzerinden modellerin performans analizleri gerçekleştirilmiştir. Sonuçlar çalışma alanı için değerlendirildiğinde RF ve LightGBM algoritması için benzer ve tutarlı sonuçlar elde edilmekle birlikte; MAE: 0.051, MAPE: 13.598, RMSE: 0.072 ve R^2 : 0.664 değerlerine göre RF algoritmasının nispeten daha iyi performans sergilediği gözlemlenmiştir. Bu durum, model performansının yalnızca algoritma seçimine değil, aynı zamanda veri setinin büyüklüğüne ve değişkenler arasındaki ilişkilere de büyük ölçüde bağlı olduğunu göstermektedir. Performans sonuçları, literatürde benzer makine öğrenmesi algoritmaları kullanılarak gerçekleştirilen toplu taşınmaz değerlendirme çalışmalarında elde edilen sonuçlarla karşılaştırılabilir niteliktedir (Adetunji vd., 2022; Aydınoglu vd., 2025; Ho vd., 2021; Kang vd., 2021; Mora-Garcia vd., 2022). Ayrıca modellerin R^2 değerlerinin orta-yüksek bir seviyede olması, taşınmaz değerini etkileyen çok boyutlu karmaşık ilişkilerin modeller tarafından etkin biçimde öğrenilebildiğini göstermektedir. Bu doğrultuda elde edilen sonuçlar, geliştirilen GeoAI analiz araç paketinin toplu taşınmaz değerlendirme süreçlerinde uygulanabilir ve güvenilir bir çözüm sunduğunu açıkça ortaya koymaktadır.

Modelleme sonuçları değişken önem düzeyleri açısından değerlendirildiğinde, RF ve LightGBM modellerinin her ikisinde de bina yaşı, arsa rayiç değeri, mahalle nüfus yoğunluğu, konut alanı, bina toplam kat adedi, konutun bulunduğu kat ve mahallede yaşayan lisans mezunu (%) gibi yapısal ve mahalli değişkenlerin ön plana çıktığı görülmektedir. Bu durum, taşınmaz değerinin fiziksel bina ve bağımsız bölüm özelliklerinin yanı sıra, taşınmazın bulunduğu çevrenin sosyo-ekonomik ve mekânsal karakteristikleri tarafından da etkilendiğini göstermektedir. Ayrıca ulaşım, alışveriş, eğitim ve kültürel tesislere mesafe gibi konumsal erişilebilirliği temsil eden değişkenlerin her iki modelde de orta düzeyde önem değerlerine sahip olması, bu faktörlerin taşınmaz değerini etkilediğini ancak doğrudan yapısal özelliklere kıyasla daha ikincil bir rol üstlendiğini göstermektedir. Bu bulgu, söz konusu unsurların taşınmaz değerini birlikte ve tamamlayıcı biçimde şekillendirdiğini ortaya koyan önceki çalışmalarla genel olarak uyumludur (Aydınoglu vd., 2023; Aydınoglu & Şişman, 2024; Chen vd., 2023; Mete, 2025; Teoh vd., 2023; Trindade Neves vd., 2024).

Geliştirilen GeoAI analiz araç paketi, toplu taşınmaz değerlendirme süreçlerinde büyük veri setleri ile hızlı ve tekrarlanabilir analizlerin yapılabilmesine imkan sağlayarak, mekânsal karar destek sistemleri açısından önemli bir potansiyele sahiptir. Ayrıca geliştirilen araç paketinin sağladığı bütünsel yapı, toplu taşınmaz değerlendirme süreçleri için sadece modelleme aşamasında değil; veri ön işleme ve performans analizi süreçlerinde de standartlaşmış, tekrarlanabilir ve otomatize bir yaklaşım sunmaktadır. Özellikle yoğun kentleşme olan ve taşınmaz piyasasının hareketli olduğu alanlarda, piyasa verilerinin sistematik biçimde işlenmesi ve yapay zekâ yöntemleriyle analiz edilmesi, hem kamu kurumlarının hem de özel sektörün

daha bütüncül ve veri temelli kararlar almasına olanak sağlayabilir. Sonraki aşamalarda ise geliştirilen GeoAI analiz araç paketinin farklı mekânsal ölçeklerde çeşitli türdeki taşınmazların toplu değerlemesinde kullanımı ve modelleme aşamasında kullanılan algoritmaların çeşitlendirilmesi planlanmaktadır.

Teşekkür

Bu çalışma TÜBİTAK tarafından desteklenen 122R021 no'lu 1001 projesi kapsamında gerçekleştirilmiştir. Bu çalışma kapsamında kullanılan gayrimenkul değer verilerinin sağlanmasında Endeksa firmasına katkılarından dolayı teşekkürlerimizi sunarız.

Yazar Katkısı

Süleyman Şişman: Düzenleme, Kavramsallaştırma, Veri analizi, Modelleme, Yazım. **Büşra Demirci:** Yazılım, Görselleştirme, Literatür taraması, Yazım. **Arif Çağdaş Aydınoglu:** Fikir, Kavramsallaştırma, Metodoloji.

Çıkar Çatışması Beyanı

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

Kaynaklar

- Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2022). House price prediction using random forest machine learning technique. *Procedia Computer Science*, 199, 806-813.
- Alburshaid, E. A., & Al-Alawi, A. I. (2024). Forecasting Methods for the Real Estate Market: A Review. *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)*, 472-477.
- Alizamir, M., Kim, S., Kisi, O., & Zounemat-Kermani, M. (2020). A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the USA and Turkey regions. *Energy*, 197, 117239.
- Aydın, C. (2018). Makine öğrenmesi algoritmaları kullanılarak itfaiye istasyonu ihtiyacının sınıflandırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (14), 169-175.
- Aydınoglu, A. Ç., Bovkır, R., & Çölkesen, İ. (2023). Toplu taşınmaz değerlemede makine öğrenme algoritmalarının kullanımı ve konumsal/konumsal olmayan özniteliklerin tahmin doğruluğuna etkilerinin karşılaştırılması. *Jeodezi ve Jeoinformasyon Dergisi*, 10(1), 63-83.
- Aydınoglu, A. Ç., & Şişman, S. (2024). Comparing modelling performance and evaluating differences of feature importance on defined geographical appraisal zones for mass real estate appraisal. *Spatial Economic Analysis*, 19(2), 225-249.
- Aydınoglu, A. Ç., Şişman, S., & Girgin, S. (2025). Examining the market sample size for machine learning-based mass appraisal: a case study in Pendik district of Istanbul. *Journal of Property Research*, 1-23.
- Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. *Applied sciences*, 8(11), 2321.
- Basheer, S., Wang, X., Farooque, A. A., Nawaz, R. A., Liu, K., Adekanmbi, T., & Liu, S. (2022). Comparison of land use land cover classifiers using different satellite imagery and machine learning techniques. *Remote Sensing*, 14(19), 4978.
- Baur, K., Rosenfelder, M., & Lutz, B. (2023). Automated real estate valuation with machine learning models using property descriptions. *Expert Systems with Applications*, 213, 119147.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Chen, L., Yao, X., Liu, Y., Zhu, Y., Chen, W., Zhao, X., & Chi, T. (2020). Measuring impacts of urban environmental elements on housing prices based on multisource data—a case study of Shanghai, China. *ISPRS International Journal of Geo-Information*, 9(2), 106.

- Chen, Y., Jiao, J., & Farahi, A. (2023). Disparities in affecting factors of housing price: A machine learning approach to the effects of housing status, public transit, and density factors on single-family housing price. *Cities*, 140, 104432.
- Choi, Y. (2023). GeoAI: Integration of artificial intelligence, machine learning, and deep learning with GIS. *Applied Sciences*, 13(6), 3895.
- Çılgın, C., Gökşen, Y., & Gökçen, H. (2023). The effect of outlier detection methods in real estate valuation with machine learning. *İzmir Sosyal Bilimler Dergisi*, 5(1), 9-20.
- Dewi, C., & Chen, R. C. (2019). Random forest and support vector machine on features selection for regression analysis. *Int. J. Innov. Comput. Inf. Control*, 15(6), 2027-2037.
- 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.
- Endeksa. (2025). <https://www.endeksa.com/tr/analiz/turkiye/endeks/satilik/konut>, (Erişim Tarihi: 10 Mart 2025).
- Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., & Zeng, W. (2019). Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agricultural water management*, 225, 105758.
- Fitriyah, H., & Budi, A. S. (2019). Outlier detection in object counting based on hue and distance transform using median absolute deviation (MAD). *2019 international conference on sustainable information engineering and technology (SIET)*, 217-222.
- Gan, M., Pan, S., Chen, Y., Cheng, C., Pan, H., & Zhu, X. (2021). Application of the machine learning lightgbm model to the prediction of the water levels of the lower columbia river. *Journal of Marine Science and Engineering*, 9(5), 496.
- Genç, N., Çolak, H. E., & Özbilgin, F. (2025). Spatial Performance Approach to Machine Learning Algorithms: A GIS-Based Comparison Analysis for Real Estate Valuation. *Transactions in GIS*, 29(1), e13303.
- Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd international conference on document analysis and recognition*, 1, 278-282.
- Ho, W. K., Tang, B. S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of property research*, 38(1), 48-70.
- 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 Ratio studies*. International Association of Assessing Officers (IAAO), Kansas, Missouri, ABD.
- Iban, M. C. (2022). An explainable model for the mass appraisal of residences: The application of tree-based Machine Learning algorithms and interpretation of value determinants. *Habitat International*, 128, 102660.
- Jafary, P., Shojaei, D., Rajabifard, A., & Ngo, T. (2024). Automated land valuation models: A comparative study of four machine learning and deep learning methods based on a comprehensive range of influential factors. *Cities*, 151, 105115.
- Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., & Ratti, C. (2021). Understanding house price appreciation using multi-source big geo-data and machine learning. *Land use policy*, 111, 104919.
- Kappal, S. (2019). Data normalization using median median absolute deviation MMAD based Z-score for robust predictions vs. min-max normalization. *London Journal of Research in Science: Natural and Formal*, 19(4), 39-44.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146-3154.
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*, 49(4), 764-766.
- Liao, Z., Ju, Y., & Zou, Q. (2016). Prediction of G protein-coupled receptors with SVM-prot features and random forest. *Scientifica*, 2016(1), 8309253.
- Liu, H., Shah, S., & Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers & chemical engineering*, 28(9), 1635-1647.
- Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European real estate research*, 12(1), 134-150.
- Mete, M. O., & Yomralıoğlu, T. (2022). Mass valuation of Real Estate Using GIS-based nominal valuation and machine learning methods. *European Real Estate Society, ERES*, 1-7.
- Mete, M. O. (2025). Developing GeoAI Integrated Mass Valuation Model Based on LADM Valuation Information Great Britain Country Profile. *Transactions in GIS*, 29(1), e13273.

- Mora-Garcia, R. T., Cespedes-Lopez, M. F., & Perez-Sanchez, V. R. (2022). Housing price prediction using machine learning algorithms in COVID-19 times. *Land*, 11(11), 2100.
- Nguyen, H. N., Tran, Q. T., Ngo, C. T., Nguyen, D. D., & Tran, V. Q. (2025). Solar energy prediction through machine learning models: A comparative analysis of regressor algorithms. *PloS one*, 20(1), e0315955.
- Ogunsanya, M., Isichei, J., & Desai, S. (2023). Grid search hyperparameter tuning in additive manufacturing processes. *Manufacturing Letters*, 35, 1031-1042.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Duborg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Peker, M., Özkaraca, O., & Kesimal, B. (2017). Enerji tasarruflu bina tasarımı için ısıtma ve soğutma yüklerini regresyon tabanlı makine öğrenmesi algoritmaları ile modelleme. *Bilişim Teknolojileri Dergisi*, 10(4), 443-449.
- Quan, S. J. (2024). Comparing hyperparameter tuning methods in machine learning based urban building energy modeling: A study in Chicago. *Energy and Buildings*, 317, 114353.
- Soltani, A., Heydari, M., Aghaei, F., & Pettit, C. J. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. *Cities*, 131, 103941.
- Sun, X., Liu, M., & Sima, Z. (2020). A novel cryptocurrency price trend forecasting model based on LightGBM. *Finance Research Letters*, 32, 101084.
- Sun, D., Wu, X., Wen, H., & Gu, Q. (2023). A LightGBM-based landslide susceptibility model considering the uncertainty of non-landslide samples. *Geomatics, Natural Hazards and Risk*, 14(1), 2213807.
- Şişman, S., & Aydınoglu, A. Ç. (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.
- Şişman, S., & Aydınoglu, A. Ç. (2023). Toplu Taşınmaz Değerlemesi için Coğrafi Verinin İşlenmesi ve Rastgele Orman Makine Öğrenmesi Tekniğiyle Tahmin Modelinin Geliştirilmesi. *Harita Dergisi*, 169, 13-27.
- Şişman, S., Kara, A., & Aydınoglu, A. Ç. (2025). Leveraging spatial data infrastructure for machine learning based building energy performance prediction. *Plos one*, 20(10), e0335531.
- Tchuente, D. (2024). Real estate automated valuation model with explainable artificial intelligence based on Shapley values. *The Journal of Real Estate Finance and Economics*, 1-39.
- TCMB (2025). Türkiye Cumhuriyet Merkez Bankası (TCMB)-Konut Fiyat Endeksi İstatistikleri, <https://www.tcmb.gov.tr/wps/wcm/connect/tr/tcmb+tr/main+menu/istatistikler/reel+sektor+istatistikleri/konut+fiyat+endeksi/>, (Erişim Tarihi: 15 Nisan 2025).
- Teoh, E. Z., Yau, W. C., Ong, T. S., & Connie, T. (2023). Explainable housing price prediction with determinant analysis. *International Journal of Housing Markets and Analysis*, 16(5), 1021-1045.
- Trindade Neves, F., Aparicio, M., & de Castro Neto, M. (2024). The impacts of open data and eXplainable AI on real estate price predictions in smart cities. *Applied Sciences*, 14(5), 2209.
- Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing price prediction via improved machine learning techniques. *Procedia Computer Science*, 174, 433-442.
- TÜİK (2025a). Nüfus ve Demografi, <https://data.tuik.gov.tr/Kategori/GetKategori?p=Nufus-ve-Demografi-109>, (Erişim Tarihi: 18 Nisan 2025).
- TÜİK (2025b). Konut Satış İstatistikleri, <https://data.tuik.gov.tr/Bulten/Index?p=Konut-Satis-Istatistikleri-Haziran-2024-53760>, (Erişim Tarihi: 18 Nisan 2025).
- Ünel, F. B., & Yalpir, Ş. (2023). Sustainable tax system design for use of mass real estate appraisal in land management. *Land Use Policy*, 131, 106734.
- Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. *Cybern. Inf. Technol*, 19(1), 3-26.
- Veranyurt, Ü., Deveci, A., Esen, M. F., & Veranyurt, O. (2020). Makine öğrenmesi teknikleriyle hastalık sınıflandırması: random forest, k-nearest neighbour ve adaboost algoritmaları uygulaması. *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi*, 6(2), 275-286.
- von Davier, M., & Bezirhan, U. (2023). A robust method for detecting item misfit in large-scale assessments. *Educational and*

Psychological Measurement, 83(4), 740-765.

Wang, Z. (2023). Research on feature selection methods based on random forest. *Tehnički vjesnik*, 30(2), 623-633.

Wang, W., Xue, C., Zhao, J., Yuan, C., & Tang, J. (2024). Machine learning-based field geological mapping: A new exploration of geological survey data acquisition strategy. *Ore Geology Reviews*, 166, 105959.

Yalpır, Ş., Şişman, 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.

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.