

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ
KULLANILARAK GAYRİMENKUL DEĞER
TAHMİNİNDE COĞRAFİ REFERANSLI VERİNİN
KATKISI

Mustafa Deniz AYAR

YÜKSEK LİSANS TEZİ
Harita Mühendisliği Anabilim Dalı
Uzaktan Algılama ve CBS Programı

Danışman
Doç. Dr. Alper ŞEN

Aralık, 2024

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ KULLANILARAK
GAYRİMENKUL DEĞER TAHMİNİNDE COĞRAFİ
REFERANSLI VERİNİN KATKISI

Mustafa Deniz AYAR tarafından hazırlanan tez çalışması 24.12.2024 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Harita Mühendisliği Anabilim Dalı Uzaktan Algılama ve CBS Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Alper ŞEN
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. Alper ŞEN, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. Ali Melih BAŞARANER, Üye
Yıldız Teknik Üniversitesi

Prof. Dr. Nesibe Necla ULUĞTEKİN, Üye
İstanbul Teknik Üniversitesi

Danışmanım Doç. Dr. Alper ŞEN sorumluluğunda tarafımca hazırlanan Makine Öğrenmesi Yöntemleri Kullanılarak Gayrimenkul Değer Tahmininde Coğrafi Referanslı Verinin Katkısı başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Mustafa Deniz AYAR

İmza

Sevgili eşime...



TEŞEKKÜR

Yüksek lisans öğrenimim boyunca desteklerini benden esirgemeyen, mesleğime bakış açımı değiştirip ufkumu açan, başarıları ile bana ilham veren, yüksek lisans tezimin ortaya çıkmasında rehberliğini benden esirgemeyen danışman hocam Doç. Dr. Alper ŞEN'e;

Yüksek lisans öğrenimim ve tez yazım sürecinde sevgisini ve sabrını hiç yitirmeyen, süreç boyunca beni cesaretlendiren en büyük destekçim sevgili eşim Tuğba GÜNGÖR AYAR'a;

Son olarak çalışmamın gerçekleştirilmesinde katkıda bulunan personeli olduğum BAYRAMPAŞA BELEDİYE BAŞKANLIĞI'na;

Teşekkür ederim.

Mustafa Deniz AYAR

İÇİNDEKİLER

SİMGE LİSTESİ	vii
KISALTMA LİSTESİ	viii
ŞEKİL LİSTESİ	x
TABLO LİSTESİ	xi
HARİTA LİSTESİ	xii
ÖZET	xiii
ABSTRACT	xv
1 GİRİŞ	1
2 LİTERATÜR ARAŞTIRMASI	5
3 GAYRİMENKUL DEĞERLEME	9
3.1 Gayrimenkul Değerlemesi	9
3.2 Geleneksel Gayrimenkul Değerleme Yöntemleri	10
3.2.1 Emsal Karşılaştırma Yöntemi	11
3.2.2 Gelir Yöntemi	11
3.2.3 Maliyet Yöntemi	12
3.3 Gayrimenkul Değerlemesine Etki Eden Faktörler	12
3.4 Mevzuatta Gayrimenkul Değerleme	13
4 YÖNTEMLER	16
4.1 Makine Öğrenmesi Nedir?	16
4.2 Makine Öğrenmesi Türleri	19
4.2.1 Denetimli Öğrenme	19
4.2.2 Denetimsiz Öğrenme	20
4.2.3 Yarı Denetimli Öğrenme	20
4.2.4 Pekiştirmeli Öğrenme	20
4.3 Metodoloji	21

4.4	Veri Önişleme	22
4.5	Makine Öğrenmesi Algoritmaları	25
4.5.1	Çoklu Doğrusal Regresyon	26
4.5.2	Coğrafi Ağırlıklı Regresyon	28
4.5.3	Rastgele Orman	29
4.5.4	Adaptive Boosting (AdaBoost)	30
4.5.5	Extreme Gradient Boosting (XGBoost)	31
4.5.6	CatBoost	33
4.6	Model Değerleme Kriterleri	34
5	UYGULAMA	36
5.1	Çalışma Alanı	36
5.2	Veri Seti	37
5.2.1	Veri Seti Özellikleri	37
5.2.2	Veri Önişleme Uygulaması	38
5.3	Bulgular	40
6	SONUÇ	53
	KAYNAKÇA	55
	TEZDEN ÜRETİLMİŞ YAYINLAR	61

SİMGE LİSTESİ

y	Bağımlı Değişken
x	Bağımsız Değişken
R^2	Belirleme Katsayısı
m^2	Metrekare
σ	Standart Sapma
TL	Türk Lirası

KISALTMA LİSTESİ

ADNKS	Adrese Dayalı Nüfus Kayıt Sistemi
AdaBoost	Adaptive Boosting
CART	Classification and Regression Tree
CHAID	Chi-square Automatic Interaction Detection
CRISP-DM	Cross-Industry Standard Process for Data Mining
GBM	Gradient Boosting Machine
IBM	International Business Machine
IVSC	International Valuation Standards Committee
KAKS	Kat Alanı Kat Sayısı
KNN	K-Nearest Neighbors
LGBM	Light Gradient Boosting Machine
LOF	Local Outlier Factor
LideBİR	Lisanslı Değerleme Şirketleri Birliği
MAE	Ortalama Mutlak Hata
MAPE	Ortalama Mutlak Yüzde Hata
MLP	Multi Layer Perceptron
MSE	Ortalama Mutlak Karesel Hata
OBB	Out of-Bag
OLS	Ordinary Least Squares
RBF	Radial Basis Function
RICS	The Royal Institution of Chartered Surveyors
RMSE	Ortalama Kök Karesel Hata
SPK	Sermaye Piyasası Kurulu

SPL	Sermaye Piyasası Lisanslama Sicil ve Eğitim Kuruluşu
TAKS	Taban Alanı Kat Sayısı
TDUB	Türkiye Değerleme Uzmanları Birliği
TEGoVA	The European Group of Valuers' Association
TUGDES	Türkiye Gayrimenkul Değerleme Standartları
TÜİK	Türkiye İstatistik Kurumu
XGBoost	Extreme Gradient Boosting
YSA	Yapay Sinir Ağları



ŞEKİL LİSTESİ

Şekil 4.1	Yapay zeka hiyerarşisi	16
Şekil 4.2	Klasik programlama akış şeması	17
Şekil 4.3	Makine öğrenmesine dayalı programlama akış şeması	18
Şekil 4.4	Makine öğrenmesi türleri	19
Şekil 4.5	İş akış şeması	22
Şekil 4.6	Basit doğrusal regresyon grafiği	26
Şekil 4.7	Çoklu doğrusal regresyon grafiği	27
Şekil 4.8	Karar ağacı yapısı	29
Şekil 4.9	Rastgele orman algoritması	29
Şekil 5.1	Korelasyon Matrisi	41
Şekil 5.2	Coğrafi referanslı Rastgele Orman modelindeki özniteliklerin önem grafiği	43
Şekil 5.3	Coğrafi referanssız Rastgele Orman modelindeki özniteliklerin önem grafiği	43
Şekil 5.4	Coğrafi referanslı CatBoost modelindeki özniteliklerin önem grafiği	44
Şekil 5.5	Coğrafi referanssız CatBoost modelindeki özniteliklerin önem grafiği	44
Şekil 5.6	Coğrafi ağırlıklı regresyon modelinin tahminleri ile gerçek fiyatlar arasındaki farkların standart sapma dağılım grafiği	46

TABLO LİSTESİ

Tablo 1.1	Ülke nüfusunun yıllara göre dağılımı	2
Tablo 1.2	Yıllara göre konut satış sayısı	2
Tablo 3.1	Ülkemiz mevzuatında gayrimenkul değerlendirme	14
Tablo 4.1	One-hot encoding örneği	24
Tablo 5.1	Veri setinin ilk 5 satırı	37
Tablo 5.2	Değişken özellikleri	38
Tablo 5.3	Fiyat değişkeninin özet istatistiği	38
Tablo 5.4	Fiyat değişkeninin frekans tablosu	39
Tablo 5.5	Kat değişkeninin frekans tablosu	39
Tablo 5.6	Oda Sayısı değişkeninin güncel frekans tablosu	40
Tablo 5.7	Veri ön işleme sonrası güncel veri seti	42
Tablo 5.8	Modellerin tahmin hataları	46

HARİTA LİSTESİ

Harita 5.1	Çalışma alanı haritası	47
Harita 5.2	İlan dağılım haritası	48
Harita 5.3	XGBoost regresyon modelinin tahminleri ile gerçek fiyatlar arasındaki farkların standart sapma dağılım haritası	49
Harita 5.4	CatBoost regresyon modelinin tahminleri ile gerçek fiyatlar arasındaki farkların standart sapma dağılım haritası	50
Harita 5.5	Rastgele orman regresyon modelinin tahminleri ile gerçek fiyatlar arasındaki farkların standart sapma dağılım haritası . . .	51
Harita 5.6	Coğrafi ağırlıklı regresyon modelinin tahminleri ile gerçek fiyatlar arasındaki farkların standart sapma dağılım haritası . . .	52

Makine Öğrenmesi Yöntemleri Kullanılarak Gayrimenkul Değer Tahmininde Coğrafi Referanslı Verinin Katkısı

Mustafa Deniz AYAR

Harita Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Doç. Dr. Alper ŞEN

Gayrimenkul değerlemesi, mülkiyetin doğru bir şekilde değerlendirilmesi ve tahmini piyasa değerinin belirlenmesi sürecidir. Fakat doğru bir gayrimenkul değerlemesi, insan uzman bilgisi gerektirir ve geleneksel gayrimenkul değerlendirme süreçleri zaman alıcıdır. Geleneksel yöntemler genellikle manuel değerlendirme tekniklerini içerirken, makine öğrenmesi teknikleri gayrimenkul değerlendirme alanında zaman, maliyet, iş gücü ve objektiflik konularında büyük bir potansiyel sunmaktadır. Son yıllarda emlak piyasasındaki gayrimenkullerin fiyatlarını tahmin etmek için makine öğrenmesi teknikleri ve algoritmalarının kullanımı artmıştır. Gayrimenkul değerlendirme için makine öğrenmesi yöntemleri, mülkün değerini etkileyen faktörleri analiz ederek ve tahmin modelleri oluşturarak çalışır. Bu faktörler arasında konum, yapısal özellikler, emlak piyasası verileri, sosyal ve ekonomik faktörler yer alabilir. Çok sayıda faktör bulunduğundan geleneksel yöntemlerde, çok sayıda değişken ve karmaşık ilişkiler göz ardı edilebilir. Bu gibi sebeplerden ötürü son yıllarda makine öğrenmesi yöntemleri ile gayrimenkul fiyat tahmini konusu üzerine birçok çalışma yapılmıştır.

Bu çalışmada, gayrimenkul değerlemesinde makine öğrenmesi yöntemlerinin karşılaştırmalı kullanımı incelenmiş ve coğrafi referanslı verilerin tahmin doğruluğuna katkısı araştırılmıştır. Çalışmada 2023 yılı Mayıs ve Haziran aylarına ait İstanbul ili Bayrampaşa ilçesinde ki satılık 1267 adet konut ilanı veri seti için kullanılmıştır. Veri setinde konutlara ait yapısal ve mülki bilgilere (m^2 (Brüt), m^2 (Net), oda sayısı, bina yaşı, toplam kat sayısı, bulunduğu kat, ısıtma

türü, banyo sayısı, balkon bilgisi, eşya bilgisi, kullanım durumu bilgisi, site bilgisi, kredi durumu, aidat miktarı, tapu türü, fiyat) ek coğrafi referanslı (enlem, boylam) bilgileri de bulunmaktadır. Bu veri setine ağaç temelli makine öğrenmesi yöntemlerinden Rastgele Orman, Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), CatBoost ile Çoklu Doğrusal Regresyon ve Coğrafi Ağırlıklı Regresyon yöntemleri uygulanarak veri eğitilmiş, test verileri tahmin edilmiştir. Makine öğrenmesi modelleri Belirleme Katsayısı (R^2), Ortalama Mutlak Hata (MAE), Ortalama Mutlak Yüzde Hata (MAPE), Ortalama Kök Karesel Hata (RMSE) ölçütleri ile karşılaştırılmıştır. Karşılaştırma sonucunda Coğrafi Ağırlıklı Regresyon ile yapılan tahmin modeli 0,87 R^2 skoru ile en başarılı model olmuştur. Coğrafi Ağırlıklı Regresyonu, 0,86 R^2 skoru ile CatBoost ve Rastgele Orman, 0,85 R^2 skoru ile XGBoost, 0,76 ve 0,74 R^2 skoru ile Çoklu Doğrusal Regresyon ve AdaBoost yöntemleri takip etmiştir. Coğrafi Ağırlıklı Regresyon modelinin en başarılı sonucu verdiği görülmüştür. CatBoost, XGBoost ve Rastgele Orman algoritmalarının da başarılı tahmin sonuçları verdiği görüldükten sonra veri setinden coğrafi referanslı veriler (enlem, boylam) çıkarılarak mevcut modeller ile tahminler yapılmış, coğrafi referanslı özelliklerin model başarılarına etkisi gözlemlenmiştir. Coğrafi referanslı veriler olmadan hazırlanan bu yeni modeller ile sırasıyla CatBoost 0,82 R^2 , Rastgele Orman 0,81 R^2 , XGBoost 0,78 R^2 , Çoklu Doğrusal Regresyon 0,75 R^2 ve AdaBoost 0,73 R^2 skoru elde edilmiştir. Böylece çalışma sonunda Coğrafi Ağırlıklı Regresyon, CatBoost, XGBoost ve Rastgele Orman yöntemlerinin gayrimenkul değer tahminlerinde kullanılabileceği gözlemlenmiş, coğrafi referanslı verilerin değer tahminlerinin doğruluğunu arttırdığı anlaşılmıştır.

Anahtar Kelimeler: Gayrimenkul değerleme, Makine öğrenmesi, Rastgele orman, AdaBoost, XGBoost, CatBoost, Çoklu doğrusal regresyon, Coğrafi ağırlıklı regresyon.

ABSTRACT

Contribution Of Geo-Referenced Data In Real Estate Value Prediction Using Machine Learning Methods

Mustafa Deniz AYAR

Department of Geomatic Engineering
Master of Science Thesis

Supervisor: Assoc. Prof. Alper ŞEN

Real estate valuation is the process of accurately assessing property and determining its estimated market value. However, accurate real estate valuation requires human expertise, and traditional valuation processes are time-consuming. While traditional methods often involve manual evaluation techniques, machine learning techniques offer great potential in the field of real estate valuation in terms of time, cost, labor and objectivity. In recent years, the use of machine learning techniques and algorithms to predict the prices of properties in the real estate market has increased. Machine learning methods for real estate valuation analyze factors influencing property values and build predictive models. These factors include location, structural attributes, real estate market data, and social and economic conditions. Due to the large number of factors, traditional methods may ignore the large number of variables and complex relationships. For these reasons, there have been many studies on real estate price prediction with machine learning methods in recent years.

In this study, the comparative use of machine learning methods in real estate valuation is examined and the contribution of geo-referenced data to prediction accuracy is investigated. In the study, 1267 housing advertisements for sale in Bayrampaşa district of Istanbul province for the months of May and June 2023 are used for the data set. The dataset includes structural and property details (gross area, net area, number of rooms, building age, total number of floors, floor level, heating type, number of bathrooms, balcony presence, furniture status, usage status, site details, credit eligibility, dues amount, deed type, and price) as well

as geo-referenced details (latitude, longitude). Random Forest, Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), CatBoost, Multiple Linear Regression and Geographically Weighted Regression methods were applied to this dataset. Machine learning models were compared with Coefficient of Determination (R^2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE). As a result of the comparison, the prediction model with Geographically Weighted Regression was the most successful model with an R^2 Score of 0,87. Geographically Weighted Regression was followed by CatBoost and Random Forest with an R^2 Score of 0,86, XGBoost with an R^2 Score of 0,85, Multiple Linear Regression and AdaBoost with R^2 Scores of 0,76 and 0,74. The Geographically Weighted Regression model was found to be the most successful. After it was observed that CatBoost, XGBoost and Random Forest algorithms also gave successful prediction results, geo-referenced features (latitude, longitude) were removed from the dataset and predictions were made with the existing models and the effect of geo-referenced features on model success was observed. With these new models prepared without geo-referenced data, CatBoost 0,82 R^2 , Random Forest 0,81 R^2 , XGBoost 0,78 R^2 , Multiple Linear Regression 0,75 R^2 and AdaBoost 0,73 R^2 scores were obtained respectively. Thus, at the end of the study, it was observed that Geographically Weighted Regression, CatBoost, XGBoost and Random Forest methods can be used in real estate value estimations, and it was understood that geo-referenced data increases the accuracy of value estimations.

Keywords: Real estate valuation, Machine learning, Random forest, AdaBoost, XGBoost, CatBoost, Multiple linear regression, Geographically weighted regression.

1 GİRİŞ

İnsanlar yaşamlarını yaşadıkları çevrelere, o çevrelerin arasındaki ilişkiler ve ihtiyaçlara, dönemin teknolojik, sosyal ve ekonomik durumlarına göre şekillendirmek ve yönlendirmek isterler. Dünya nüfusunun artışı, tarım faaliyetlerinin azalmasıyla köyden kente göçün artması, endüstrinin gelişmesi ve bunlara bağlı birtakım gelişmeler hızlı kentleşmeye sebep olmuştur. Kentler, bu şekillendirme, yönlendirme ve gelişmeler neticesinde bugünkü halini almıştır.

Konutlar, insanların kent hayatları içindeki en önemli alanlarıdır. Bu alanlar sadece barınma ihtiyacını karşılamakla kalmaz, aynı zamanda güvenlik, konfor ve sosyalleşme gibi birtakım temel ihtiyaçlara da cevap verirler. İyi bir konut, insanların yaşam kalitesini artırır ve mutlu, sağlıklı bir hayat sürmelerine de katkıda bulunur.

Ülkemizdeki nüfus artışı, hızlı endüstrileşme ve buna bağlı teknolojik gelişmeler köyden kente göçü hızlandırmış, konut problemini de beraberinde getirmiştir. Türkiye İstatistik Kurumu (TÜİK)'na ait Adrese Dayalı Nüfus Kayıt Sistemi (ADNKS)'nin 2013-2023 tarihleri arasındaki Tablo 1.1'de gösterilen sonuçlarına göre toplam nüfusun %90'ından fazlasının kentsel alanlarda ikamet ettiği görülmektedir [1]. Bu yüzdeler sonu kentsel alanlardaki konut talebini ve dolayısıyla taşınmazların ve gayrimenkullerin değerlerini de arttırmıştır. Kentsel alanlardaki nüfus artışı yıllar içinde milyonlarca konut satışını da beraberinde getirmiştir. TÜİK verilerine göre 2013-2023 yılları arasında her yıl 1.000.000 ile 1.500.000 arası konut satışı yapılmıştır (Tablo 1.2) [2].

Kent nüfusunun artması konuta olan talebi arttırmış, arsa ve konut gayrimenkulleri önemli bir yatırım aracı haline gelmiş ve günümüzde yılda milyonlarca alım-satım yapılan bir konut piyasası oluşturmuştur. Konut piyasasının güvenli bir şekilde işleyebilmesi, arsa ve konut fiyatlarının doğru, objektif ve güvenli bir şekilde belirlenmesi ile mümkündür. Doğru, objektif ve güvenilir olmayan bir değerlendirme, piyasaya ve taşınmazlar üzerinden elde edilen emlak vergisi ve tapu harcı gibi

ülke ekonomisine katkı sağlayan gelirlere olumsuz etki eder. Bu bilgiler ışığında, konut ve konuta ilişkin birçok alanda taşınmaz değerlendirme konusu önemli bir unsur olmuştur.

Tablo 1.1 Ülke nüfusunun yıllara göre dağılımı [1]

Yıl	Kentsel Alan		Kırsal Alan		Toplam Nüfus
	Nüfus	Payı (%)	Nüfus	Payı (%)	
2013	70.034.413	91,3	6.633.451	8,7	76.667.864
2014	71.286.182	91,8	6.409.722	8,2	77.695.904
2015	72.523.135	92,1	6.217.919	7,9	78.741.053
2016	73.671.748	92,3	6.143.123	7,7	79.814.871
2017	74.161.132	91,8	6.049.393	8,2	80.810.525
2018	75.666.497	92,3	6.337.385	7,7	82.003.882
2019	77.151.280	92,8	6.003.717	7,2	83.154.997
2020	77.736.041	93,0	5.878.321	7,0	83.614.362
2021	78.908.631	93,2	5.771.642	6,8	84.680.273
2022	79.613.279	93,4	5.666.274	6,6	85.279.553
2023	79.399.292	93,0	5.973.085	7,0	85.372.377

Taşınmaz değerlemesine kamulaştırma, vergilendirme, özelleştirme, tescile esas işlemler, irtifak hakkı tesisi gibi kamusal uygulamalar ile bankacılık, sigortacılık, kredilendirme gibi özel sektör uygulamalarında ihtiyaç duyulur [3]. Gayrimenkul değerlendirme; sermaye piyasası, para piyasası ve gayrimenkul sektörü olmak üzere geniş bir kesimi ilgilendiren ve bununla birlikte ülke ekonomisine de önemli ölçüde etkileri olabilen bir faaliyettir [4]. Bu nedenle devletlerin, taşınmaz değerlerinin belirlendiği sistemlerini oluşturmaları ve böylece taşınmaz pazarlarının şeffaflığını sağlamaları gerekmektedir [5].

Tablo 1.2 Yıllara göre konut satış sayısı [2]

Yıl	Toplam Satış
2013	1.157.190
2014	1.165.381
2015	1.289.381
2016	1.341.453
2017	1.409.314
2018	1.375.398
2019	1.348.729
2020	1.499.316
2021	1.491.856
2022	1.485.622
2023	1.225.926

Ülkemizde gayrimenkul değerlendirme uzmanları tarafından yapılan değerlemeler genellikle geleneksel değerlendirme yöntemleri olan emsal karşılaştırma, gelir ve

maliyet yöntemleriyle yapılmaktadır [6]. Bu yöntemler gayrimenkullerin tek başlarına değerlendirilmesine neden olmaktadır ve zaman alıcı yöntemlerdir. Ancak günümüz teknolojisindeki gelişmeler ile bilgisayar işlemcileri hızlanmış, büyük veri ambarları oluşturulmuş, açık kaynak sistemlerle bilgiye ulaşım kolaylaşmış, yapay zekâ ve makine öğrenmesi alanlarında ilerlemeler kaydedilmiştir. İşte bu gelişmeler gayrimenkullerin toplu değerlendirilmesine olanak sağlamaktadır. Gayrimenkullerin toplu değerlendirilmesi, zaman, maliyet, iş gücü ve objektiflik konularındaki olumlu sonuçları sebebiyle geleneksel yöntemlerden ayrılmaktadır.

Çalışmamızın bu kısmında taşınmaz değerlemesinin istatistiksel veriler ışığında sosyal ve ekonomik yönüne vurgu yapılmış, toplu taşınmaz değerlemesinin, geleneksel gayrimenkul değerlendirme yöntemlerine karşı önemine değinilmiştir.

İkinci kısımda bu çalışmanın konusu olan toplu gayrimenkul değerlemesi konusu hakkında bilim insanları tarafından daha önce yapılan ve yayınlanan çalışmalar ele alınmış, literatür taraması sunulmuştur.

Üçüncü kısımda taşınmaz değerlemesinin tanımı ve türleri açıklanmış, değerlendirme süreci, değerlemeye etki eden faktörler ile taşınmaz değerlemesinin ülkemiz mevzuatındaki yeri hakkında bilgi verilmiştir.

Dördüncü kısımda makine öğrenmesinin temellerine, çalışmamızın metodolojisine ve çalışmada uygulanan Rastgele Orman, Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), CatBoost, Çoklu Doğrusal Regresyon ve Coğrafi Ağırlıklı Regresyon tekniklerinin açıklamalarına yer verilmiş, veri ön işleme konusu ile çalışmada kullanılan model değerlendirme kriterleri açıklanmıştır.

Beşinci bölümde gayrimenkul değerlemesinde makine öğrenmesi yöntemlerinin karşılaştırmalı kullanımı incelenmiş ve coğrafi referanslı verilerin tahmin doğruluğuna katkısı araştırılmıştır. Bu çalışmada, 2023 yılı Mayıs ve Haziran aylarına ait İstanbul ili Bayrampaşa ilçesindeki satılık 1267 adet konut ilanı veri setine veri ön işleme adımları uygulanmış ve veri seti makine öğrenmesi algoritmalarına hazır hale getirilmiştir. Bu veri setine ağaç temelli makine öğrenmesi yöntemlerinden Rastgele Orman, Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), CatBoost ile Çoklu Doğrusal Regresyon ve Coğrafi Ağırlıklı Regresyon yöntemleri uygulanarak veri eğitilmiş, test verileri tahmin edilmiştir. Makine öğrenmesi modelleri Belirleme Katsayısı (R^2), Ortalama Mutlak Hata (Mean Absolute Error, MAE), Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error, MAPE), Ortalama Kök Karesel Hata (Root Mean Square Error, RMSE) ölçütleri ile karşılaştırılmış, yüksek tahmin başarısına sahip modellerin tahmin fiyatları ile gerçek fiyatlar arasındaki farklar görselleştirilmiştir.

Son bölümde ise sonuçlar ve öneriler kısmı bulunmaktadır. Çalışmamızın sonucu, geliştirilmesi gereken kısımlar ile farklı metotlara ve önerilere bu bölümde yer verilmiştir.



2 LİTERATÜR ARAŞTIRMASI

Gayrimenkul değerinin bilinmesi, tahmini ve analizi bilim insanlarının son yıllarda üzerinde durduğu bir konu olmuştur. Literatür incelendiğinde makine öğrenmesi yöntemleriyle gayrimenkul değerlemesinin birçok çalışmaya konu olduğu görülmüştür. Ayrıca bu çalışmada kullanılan Rastgele Orman, Adaboost, XGBoost, CatBoost, Çoklu Doğrusal Regresyon ve Coğrafi Ağırlıklı Regresyon makine öğrenmesi yöntemleri yine bu çalışmalarda kullanılan modellerdir.

[7] çalışmalarında, Türkiye'nin Antalya ilinin Alanya ilçesinde turizm sektöründen etkilenen emlak piyasasını dengelemek ve değer değişimlerini objektif ve bilimsel olarak incelemek için makine öğrenmesi teknikleri kullanılmıştır. Matematiksel modelleme yerine makine öğrenmesi algoritmaları (Destek Vektör Makineleri, k-En Yakın Komşuluk, Rastgele Orman) kullanılarak algoritma tabanlı değerlendirme yapmış, $0,73 R^2$ skoru ile Destek Vektör Makineleri yönteminin en başarılı tahminleri yaptığını elde etmiştir.

[8], gayrimenkulün toplu değerlemesi için makine öğrenmesi yöntemlerinin karşılaştırılması amacıyla aralarında Yapay Sinir Ağları, Destek Vektör Makinesi ve Rastgele Orman yöntemlerinin bulunduğu beş makine öğrenmesi tekniğini karşılaştırmayı amaçlamış, en iyi sonucun Yapay Sinir Ağları tekniğinin verdiğini görmüştür.

[9], gayrimenkul fiyat tahmininde konum özelliklerinin uygunluğunu yüksek ve ince ayrıntı düzeyleriyle ölçen farklı bir yaklaşım önererek yedi popüler makine öğrenmesi yöntemini (Yapay Sinir Ağları, Rastgele Orman, Adaptive Boosting (AdaBoost), Gradyan Arttırma, Destek Vektör Makineleri, k – En Yakın Komşuluk, Doğrusal Regresyon) karşılaştırmıştır. Bu çalışmada, $0,79 R^2$ skoru ile Rastgele Orman Algoritması en başarılı tahminleri yapmıştır.

[10] çalışmalarında, emlak toplu değerlemesi için Rusya – Saint Petersburg'da bulunan 2848 adet konuttan oluşan ve 15 değişkene sahip veri setine Rastgele

Orman, Chi-square Automatic Interaction Detection (CHAID), Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Tree, CART), K-En Yakın Komşu Algoritması (K-Nearest Neighbors, KNN), Çok Katmanlı Algılayıcılar (Multi Layer Perceptron, MLP), Radyal Tabanlı Fonksiyon (Radial Basis Function, RBF), Çoklu Regresyon Analizi ve Güçlendirilmiş Ağaçlar (Boosted Trees) yöntemlerini kullanmış, çalışma sonunda Rastgele Orman algoritmasının m^2 başına fiyatın tahmin edilmesi ve toplam satış fiyatının hesaplanması durumunda tüm algoritmalarından daha iyi performans ortaya koyduğu görülmüştür.

[11], Kanada – Ottawa’da 4334 adet mülk içeren gerçek dünya veri kümesini kullanarak emlak fiyatlarını önceden belirlemek için Rastgele Orman ve Çoklu Doğrusal Regresyon modellerini kullanmış, bu veri setinin %80’ini eğitim, %20’sini test için ayırmıştır. Sonuçlar Ortalama Karesel Hata (MSE), RMSE ve R^2 kullanılarak değerlendirilmiş, 0,99 R^2 skoru ile Çoklu Doğrusal Regresyon yöntemi daha başarılı tahmin sonuçları vermiştir.

[12] yaptıkları çalışmada Hollanda’nın 5 büyük belediyesinin (Rotterdam, Amsterdam, Eindhoven, Amersfoort ve Groningen) gayrimenkul değerlendirme değerlerini modellemek amacıyla Doğrusal Regresyon, Coğrafi Ağırlıklı Regresyon ve XGBoost algoritmalarını kullanmışlardır. 0,84 R^2 skoru ile XGBoost algoritması kullanılan tahmin modeli diğer iki algoritmanın kullanıldığı tahmin modellerinden daha iyi performans göstermiştir.

[13] çalışmasında, Amerika Birleşik Devletleri’nin Virjinya eyaletinin Fairfax County bölgesindeki 5359 konut verisinin bulunduğu veri kümesi kullanarak Repeated Incremental Pruning to Produce Error Reduction (RIPPER), C4.5 Ağacı, Naive Bayes ve AdaBoost algoritmaları ile tahmin modelleri oluşturmuş, sırası ile 0,2488 ve 0,2570 ortalama hata ile RIPPER ve AdaBoost algoritmaları ile oluşturulan tahmin modelleri diğer yöntemler ile oluşturulan tahmin modellerinden daha başarılı olmuşlardır.

[14], Cross-Industry Standard Process for Data Mining (CRISP-DM) metodolojisi ile emlak piyasası için bir değerlendirme modeli oluşturmak amacıyla İstanbul’daki 11 adet değişkene sahip 36.759 adet konut verisine Lineer Regresyon, Polinomal Regresyon, Karar Ağaçları, Rastgele Orman ve XGBoost algoritmaları uygulamıştır. Yaklaşık 0,20 MAPE ile Rastgele Orman ve XGBoost algoritmaları en başarılı algoritmalar olmuştur.

[15], Güney Kore’de 2006–2017 tarihlerine ait konut işlem verilerini kullanarak emlak vergisinin hesaplanmasında kullanılan Sıradan En Küçük Kareler (Ordinary Least Squares, OLS) regresyon yöntemine karşı Rastgele Orman algoritması

yöntemine dayalı konut fiyat tahmin modeli oluşturmuşlardır. Ortalama yüzde sapma OLS için yaklaşık %20, Rastgele Orman algoritması için %5,5 civarında ölçülmüştür.

[16], Sakarya'nın Adapazarı, Arifiye, Karasu ve Serdivan ilçelerindeki toplam 213 satılık konut ilanından oluşan veri kümesine Derin Öğrenme, Polinomal Regresyon ve Rastgele Orman algoritmaları kullanarak fiyat tahmin modeli oluşturmuş ve 0,86 R^2 soru ile Derin Öğrenme metodu en başarılı model olmuştur.

[17], Gradyan arttırma algoritmaları arasında bir çalışma yaparak Light Gradient Boosting Machine (LGBM) ve XGBoost algoritmaları ile hibrit bir model oluşturmuş, bu hibrit modeli temel gradyan arttırma modelleri olan LGBM, XGBoost, AdaBoost ve Gradient Boosting Machine (GBM) ile karşılaştırmıştır. Amerika Birleşik Devletleri'nin Kaliforniya eyaletine ait 20.640 adet konutlara ait enlem ve boylam bilgilerini de içeren veri seti kullanılmıştır. Çalışma sonucunda hibrit modelin 0,193 MSE, 0,285 MAE ve 0,156 MAPE değerleri ile diğer temel gradyan arttırma modellerine göre daha iyi sonuçlar verdiği anlaşılmıştır. Hibrit model dışında temel modeller arasında 0,198 ve 0,201 MSE, 0,290 ve 0,295 MAE, 0,161 ve 0,163 MAPE değerleri ile sırasıyla LGBM ve XGBoost modelleri en başarılı sonuçları vermiştir.

[18], 34.857 adet emlak ilanının bulunduğu 21 değişkenli Melbourne emlak veri setine Lineer Regresyon, Rastgele Orman ve GBM algoritmalarını kullanarak tahmin modelleri oluşturmuş, 0,83 ve 0,82 R^2 skorları ile sırasıyla GBM ve Rastgele Orman algoritmaları modeller 0,64 R^2 skorlu Doğrusal Regresyon modeline karşı üstünlük kurmuşlardır.

[19], konut fiyat tahmini için XGBoost, Destek Vektör Makineleri, Rastgele Orman, Çok Katmanlı Algılayıcılar ve Çoklu Doğrusal Regresyon algoritmalarını karşılaştırmışlardır. Bu çalışma için Amerika Birleşik Devletleri'nin Iowa eyaletindeki Ames şehrinin veri seti kullanılmıştır. Bu veri seti 82 adet değişken içeren 2930 ev ilanından oluşmaktadır. Bu 5 algoritma arasından 0,92 R^2 skoru ve 88,94 çapraz doğrulama puanı ile en başarılı algoritma XGBoost algoritması olmuştur.

[20], King County – Washington veri setine aralarında CatBoost, Rastgele Orman ve K-En Yakın Komşuluk algoritmalarının da bulunduğu 12 algoritma ile ev fiyatlarını tahmin etmek için bir yöntem geliştirmişlerdir. Bu algoritmaları MSE, R^2 skoru, MAE, RMSE gibi çeşitli ölçütlerle karşılaştırmışlardır. Bu 12 model arasından 0,915 R^2 skoru ile CatBoost algoritması diğer modellere kıyasla iyi bir performans sergilemiştir.

[21], çalışmalarında bir önceki çalışmada olduğu gibi King County – Washington veri setini kullanmıştır. Bu veri seti ile Çoklu Doğrusal Regresyon, CatBoost, XGBoost, Rastgele Orman olmak üzere 8 model kullanarak konut fiyat tahmini için uygulanabilir bir yöntem formüle etmeye odaklanmışlardır. CatBoost'un tüm modeller arasında en iyi performansı gösterdiğini ortaya koymuşlardır.

[22], [23], [24], [25] toplu emlak fiyat değerlemesi ve analizi çalışmaları da Coğrafi ağırlıklı regresyon modelini kullanan çalışmalardır.

Bu çalışmada literatürdeki çalışmalardan farklı olarak Rastgele Orman, Adaboost, XGBoost, CatBoost, Çoklu Doğrusal Regresyon ve Coğrafi Ağırlıklı Regresyon yöntemleri karşılaştırılmış ve coğrafi referanslı değişkenler olan enlem ve boylamın tahmin doğruluğuna etkisi karşılaştırmalı olarak araştırılmıştır.



3

GAYRİMENKUL DEĞERLEME

3.1 Gayrimenkul Değerlemesi

Özel mülkiyet hakkının olduğu gelişmiş ülkelerde, gayrimenkul değerlerinin belirlenmesi, bilim insanlarının yaklaşık bir yüz yıldır yoğunlaştığı, ülke koşullarına göre çeşitli çözüm önerileri getirdikleri ve sürekli olarak yenilikler yaptıkları önemli bir çalışma alanıdır.

Taşınmazların doğru, güvenilir ve tarafsız bir şekilde değerinin belirlenmesi hem bireyler hem de toplum açısından büyük önem taşır. Bu değerlendirme, taşınmaz malikleri, alıcılar ve satıcılar için kritik bir bilgi sağlamakla kalmaz, aynı zamanda toplumun ekonomik kaynaklarının büyük bir kısmını oluşturan taşınmazların gerçek değerinin anlaşılması açısından da büyük bir anlam ifade eder.

Taşınmaz değerlendirme için literatürde birçok tanım mevcuttur. Genel anlamda, bir taşınmazın, taşınmaza bağlı hakların, sorumlulukların ve kısıtlılıkların ya da taşınmaz projesinin değerlendirme günündeki tanımlı değerinin, bağımsız, yansız ve nesnel ölçütlere dayanarak kestirimidir [26]. Taşınmaz değerlendirme objektif ve tarafsız bir şekilde, bir taşınmaza ilişkin nitelik, fayda, çevre, kullanım koşulları gibi faktörlerin değerlendirilmesi suretiyle söz konusu taşınmaz değerinin tespit edilmesi için gerekli işlemlerin bütünü olarak da tanımlanabilir [27].

Bir gayrimenkulün değerini etkileyen ve belirleyen birçok özellik, standart ve yöntem bulunmaktadır. Gayrimenkulün değerini etkileyen; sosyal, kültürel, ekonomik ve konumsal özellikler başlıkları altında birçok özellik bulunması değerlendirme standartlarının oluşturulması gerekliliğini de beraberinde getirmiştir. Gayrimenkul değerlendirme konusunun bu denli özel bir alan olması, teknik detaylarının oldukça fazla olması ve bu konuda mesleki etik kurallarının oluşturulması önem arz etmiştir. Bu konuda uluslararası değerlendirme standartlarının oluşturulması, geliştirilmesi, yaygınlaştırılması ve gayrimenkul değerlendirme mesleğinin güvenilirlik ve şeffaflık ilkelerinin sağlanması, mesleki örgütlerin tek

bir çatı altında toplanarak veri alışverişinin kolaylaştırılması adına ulusal ve uluslararası birçok gayrimenkul değerleme kuruluşu faaliyete geçirilmiştir [28]. Uluslararası faaliyete geçirilmiş kuruluşlara; Avrupa Değerlemeci Birliği Grubu (The European Group of Valuers' Association, TEGoVA), Uluslararası Değerleme Standartları Komitesi (International Valuation Standards Committee, IVSC) ve Lisanslı Ölçmeciler Kraliyet Enstitüsü (The Royal Institution of Chartered Surveyors, RICS) örnek verilebilir [29].

Ülkemizde ise 28.07.1981 tarihli 2499 sayılı Sermaye Piyasası Kanunu ile Sermaye Piyasası Kurulu (SPK) kurulmuş, taşınmaz değerleme konusu kurulun görevleri arasında kabul edilmiştir. Sermaye Piyasası Lisanslama Sicil ve Eğitim Kuruluşu (SPL), SPK tarafından sermaye piyasası kurumlarında ve halka açık ortaklıklarda çalışanlara veya çalışacaklara lisans verme, lisans sahibi kişilerin sicilini tutma ve lisanslarla ilgili eğitim programları düzenlemekle yetkilendirilmiş diğer bir kuruluştur. Sermaye Piyasası Kanunu'na dayandırılarak kurulmuş olan Türkiye Değerleme Uzmanları Birliği (TDUB), gayrimenkul piyasası ile değerleme faaliyetlerinin gelişmesini sağlamak ve değerleme standartlarını oluşturmak göreviyle Sermaye Piyasası Kanunu'na dayandırılarak kurulmuş bir meslek birliğidir. SPK, SPL ve TDUB dışında Lisanslı Değerleme Şirketleri Birliği Derneği (LideBİR), Tapu ve Kadastro Genel Müdürlüğü Taşınmaz Değerleme Dairesi Başkanlığı ülkemizdeki diğer kurum, kuruluş ve organizasyonlardır.

Ülkemizde çalışan değerleme uzmanlarının izlemesi gereken yöntemler, değerleme tanımları ve standartları konusunda kılavuz olması amacıyla TDUB tarafından Eylül 2011 tarihinde Türkiye Değerleme Standartları (TUGDES) çalışması taslak olarak yayınlanmıştır. Çalışma gayrimenkul değerinin doğru olarak belirlenmesi için kullanılması gereken parametreleri ve standartları detaylı olarak incelemiştir. Çalışma özünde değerleme raporlarının ortak bir içerikte hazırlanması benimsenmiştir [30].

3.2 Geleneksel Gayrimenkul Değerleme Yöntemleri

Ülkemizde gayrimenkul değerleme uzmanları tarafından yapılan değerlemeler genellikle geleneksel değerleme yöntemleri olan emsal karşılaştırma, gelir ve maliyet yöntemleriyle yapılmaktadır [6]. Ancak bu yöntemler dışında yapay sinir ağları (YSA), hedonik fiyatlandırma yöntemi, mekânsal analiz yöntemleri, bulanık mantık gibi ileri değerleme yöntemleri de bulunmaktadır [31].

Değerleme için seçilecek yöntem gayrimenkulün konumuna, kullanım amacına göre seçilir. Yapılı veya yapısız imar parselleri, konutlar için daha önce yapılmış

emsal satışlar varsa seçilebilecek en uygun yöntem emsal karşılaştırma yöntemidir. İşyeri, kiralık konutlar gibi taşınmazlarda gelecekte getireceği gelirler önemli olduğundan gelir yöntemi kullanılmalıdır. Sanayi sitesi, fabrika, otel, hastane gibi çok özel niteliklere sahip ve emsalinin bulunmasının zor olduğu taşınmazlarda ise maliyet yöntemi değerlendirme yöntemi olarak seçilmelidir.

3.2.1 Emsal Karşılaştırma Yöntemi

Bu yöntem, değeri tespit edilecek gayrimenkul özellikleri (konum, büyüklüğü, şekli, yüksekliği, taban alanı kat sayısı (TAKS), kat alanı kat sayısı (KAKS) gibi imar özellikleri, yapı hakkı, kentsel işlevi, zemin ve altyapı özellikleri vb.) ile benzerlik taşıyan ve yakın zamanda satışı gerçekleşen gayrimenkullerin değerinin karşılaştırılmasıyla rayiç değerlerin tespit edilmesi temeline dayanmaktadır [32].

Bu yöntemin uygulanabilmesinin en önemli koşulu değeri saptanacak taşınmazın emsallerinin olması gerektiğidir. Ayrıca emsal seçilirken emsal taşınmazların özellikleri ile değeri saptanacak taşınmazın özelliklerinin de aynı olması gerekmektedir. Değeri saptanacak taşınmaz konut ise konut emsalleri, arsa ise arsa emsalleri seçilmelidir.

Gayrimenkulün kullanım amacı değerlemeye etki eden en önemli faktör olduğundan emsal özellikleri de kullanım amacına göre değişiklik gösterir. Bir konut taşınmazın özellikleri ile sanayi alanındaki bir taşınmazın özellikleri aynı değildir. Sanayi bölgesindeki taşınmaz için birim hacim önemli olduğundan TL/m³ hesaplanırken, konutlar için TL/m² birim alan hesaplanır.

Bu yöntemde emsal nitelikteki k sayıda emsal taşınmazın bulunduğu bir kümede her emsal niteliklerine göre güncel birim değerleri ayrı ayrı hesaplanır. Eğer değerler güncel değilse TÜİK verileri ile güncellenir. Hesaplanan güncel birim değerlerin ortalaması alınarak emsallerle karşılaştırılır, emsallerin birim değerlerinin ortalama birim değerden sapma miktarı $\pm\%15$ standart sapma kabul edilerek alt ve üst sınırlar arasında bulunmayan emsaller kümeden çıkarılır. Standart sapma içinde kalan emsallerin birim değerlerinin ortalaması ile değeri belirlenecek taşınmazın değeri hesaplanır.

3.2.2 Gelir Yöntemi

Taşınmazların değeri, sağladıkları gelirler doğrultusunda belirlenebiliyorsa, piyasa değeri tespitinde gelir yönteminin kullanılması gerekmektedir. Gelir yönteminde, bir gayrimenkulün değerinin hesaplanmasında temel ölçüt, elde edilebilecek net

gelir olarak kabul edilir. Bu net gelir, gayrimenkulden elde edilen kira gelirleri, yapıya ilişkin diğer tesislerin getirileri ve arsa payını kapsar. Arsalar genellikle sürekli olarak kullanılabilirken, yapıların kullanım süreleri aşınma nedeniyle sınırlıdır. Bu nedenle, net gelir hesaplanırken arsa, yapı ve diğer yapısal unsurların değerleri ayrı ayrı değerlendirilir.

Bu yöntemde piyasa değerinin tespitinde temel ölçüt taşınmazların sağladıkları gelir olduğundan, bu yöntem genellikle ticari, sanayi ve endüstriyel kullanım amacıyla kullanılan iş merkezi, iş hanı, fabrika gibi kira getirisi olan taşınmazların değerlendirilmesinde kullanılır.

3.2.3 Maliyet Yöntemi

Üzerinde yapı bulunan, kira geliri bilinmeyen ve emsal satış bilgisi bulunmayan taşınmazların değerlendirilmesinde maliyet yöntemi kullanılmaktadır. Bu yöntem bir yapının değerlendirme gününde birebir aynısının yeniden inşa edilmesi durumundaki maliyetinin hesaplaması yöntemidir.

Maliyet yöntemi, diğer iki yöntemin dışında piyasada ticareti sık rastlanmayan taşınmazların değerlemesinde, emsal satış bilgisi bulunmayan durumlarda, fabrika, otel, hastane, sanayi sitesi gibi özel kullanıma sahip yapıların değerlemesinde, taşınmazın kira gerisinin bilinmemesi durumunda kullanılır.

Yapının yeniden inşa edilmesi durumunda maliyeti hesaplanırken, değerlemesi yapılacak yapının değerlendirme tarihindeki aynı malzeme, standart, tasarım, işçilik kullanılarak yeniden inşası durumunda oluşacak maliyete arsa değeri maliyeti eklenir. Değerlemesi yapılacak olan yapının amortisman ve yıpranma payı hesaplanarak bu maliyetten düşülür. Böylece maliyet yöntemi ile değerlendirme yapılacak yapının değeri hesaplanır.

3.3 Gayrimenkul Değerlemesine Etki Eden Faktörler

Bir gayrimenkule ait değer tespiti gerçek anlamda mümkün değildir. Çünkü her taşınmazın konumsal özelliği ve kullanım şekliyle kendine has birçok özelliği bulunmaktadır. Bu sebeple uygulamada gerçek değer tespiti imkansızdır, gayrimenkullerin gerçek değeri ancak tahmin edilebilir.

Herhangi bir gayrimenkulün değerlendirilmesinde dikkate alınması gereken hususlar şunlardır; gayrimenkulün kullanım amacı, gayrimenkul mahallinin özellikleri, gayrimenkulün konumu, mevzi özellikleridir [33].

Gayrimenkulün kullanım amacı değerlemeye etki eden faktörlerden biridir. Örnek vermek gerekirse konut olarak kullanılan bir gayrimenkulün değerlendirme kriterleri ile ticari amaçla kullanılan bir gayrimenkulün değerlendirme kriterleri farklı olacaktır. Konutlar genellikle piyasa talebi, çevresel faktörlere ve emsal satışlara göre değerlendirilirken, ticari alanlar genellikle potansiyel gelir miktarına ve kira getirilerine göre değerlendirilirler. Gayrimenkulün bulunduğu bölgede yapılan imar planları ve uygulamaları da gayrimenkulün kullanım amacını belirleyen ve dolayısıyla değerine etki eden bir diğer husustur.

Gayrimenkul değerlemesinde dikkate alınması gereken hususlardan biri de gayrimenkul mahallinin özellikleridir. Taşınmazın konumunun topoğrafyası, çevresinin kırsal veya kentsel bölge olma durumu, bitki örtüsü, yatırımlar ve faydalanılan hizmetler, çevredeki diğer gayrimenkul sahiplerinin sosyo-kültürel durumları gayrimenkul değerlemesine etki eden mahall özellikleridir.

Gayrimenkulün konumu değerlemeye etki eden bir diğer faktördür. Gayrimenkulün konumunun yeşil alanlara, sosyal donatı alanlarına, çekim merkezlerine yakınlığı-uzaklığı ve ulaşım hizmetleri açısından yeterli olup-olmaması değerlemeye doğrudan etki eder. Gayrimenkullerin kullanım amacına göre alışveriş merkezleri, pazar yerleri gibi ticari çekim merkezlerine, okul, otopark, çocuk bahçesi, hastane gibi kamusal ve sosyal donatı alanlarına, metro, metrobüs gibi ulaşım hizmetlerine yakın olması sosyo-ekonomik anlamda insanlara kazanım sağlar. Ayrıca havaalanı, otoyol, çöp toplama merkezi gibi insan sağlığına doğrudan etki edebilecek gürültülü ve sağlıksız alanlara uzaklıkta gayrimenkulün değerini artırır.

Gayrimenkulün değerini etkileyen bir diğer hususta gayrimenkulün mevzi özellikleridir. Burada kastedilen gayrimenkulün bulunduğu özel konumuyla ilgili detaylar ve özelliklerdir. Gayrimenkulün üzerinde bulunduğu arazinin düz, eğimli veya engebeli oluşu, zemininin kayalık veya bataklık olması, altyapı hizmetlerinin mevcudiyeti, cephe yönü, manzarası gibi özellikleri de mevzi özellikler olarak sınıflandırılabilir.

3.4 Mevzuatta Gayrimenkul Değerleme

Kamulaştırma, arazi-arsa düzenlemeleri, kentsel dönüşüm, vergilendirme, veraset, miras ve ipotek gibi konularda gayrimenkullerin değerlemesine ihtiyaç duyulmaktadır. Bu ihtiyaçlar çerçevesinde ülkemiz mevzuatında gayrimenkul değerlendirme konularına değinilen bazı kanun, tüzük ve yönetmelikler ilgili maddeleri ile Tablo 3.1’de belirtilmiştir [34].

Tablo 3.1 Ülkemiz mevzuatında gayrimenkul değerleme [34]

Kanun Adı	Tarih	İlgili Maddesi
Köy Kanunu	1924	Madde 44(2)
Vergi Usul Kanunu	1961	Madde 309
Emlak Vergisi Kanunu	1970	Madde 29
Anayasa	1982	Madde 46, Madde 47
Kamulaştırma Kanunu	1983	Madde 11
İmar Kanunu	1985	Madde 17
Kadastro Kanunu	1987	Madde 36
Türk Medeni Kanunu	2001	Madde 889, Madde 904
Afet Riski Altındaki Alanların Dönüştürülmesi Hakkında Kanun	2012	Madde 6
Tapu Sicil Tüzüğü	2013	Madde 38
Arazi ve Arsa Düzenlemeleri Hakkında Yönetmelik	2020	Madde 16, Madde 37

Ülkemizde gayrimenkul değerlemesiyle ilgili başta Anayasa olmak üzere birçok kanun, yönetmelik ve tebliğ olmasına rağmen hiçbirinde değerleme standartları ve kuralları bulunmamaktadır [35]. Bu sebeple ülkemiz için değerleme faaliyetlerinin belli standartlarda ve etkin bir biçimde gerçekleştirilmesi için bir yasal düzenlemeye ihtiyaç duyulmaktadır.

Köy Kanunu 44. Maddesinin 2. Bendinde köy sınırı içinde bulunan gayrimenkulleri İhtiyar Meclisi'nin değer pahasıyla satın almasına olanak tanımıştır. Vergi Usul Kanunu 309. Maddesinde ise arazinin kıymet takdirinde ele alınacak esaslardan söz edilmiştir. Bu esaslar; benzer satışlar, arazinin mevki, arazinin yüzölçümü, arazinin önemli yerlere uzaklığı ve tarım arazisinin verimi gibi esaslardır. Emlak Vergisi Kanunu'nun 29. Maddesinde ise vergi değerinin arsa ve araziler için Vergi Usul Kanunu'na göre takdir komisyonları tarafından, binalar içinse metrekare inşaat maliyetleri ve arsaların değerleri esas alınarak hesaplanacağı belirtilmiştir.

Anayasamızda taşınmaz değerlemesi; kamulaştırmaya ilişkin kanun maddelerine konu olmuştur. Anayasamızın 46. Maddesinde “Devlet ve kamu tüzel kişileri; kamu yararının gerektirdiği hallerde, gerçek karşılıklarını peşin ödemek şartıyla, özel mülkiyette bulunan taşınmaz malların tamamını veya bir kısmını, kanunla gösterilen esas ve usullere göre, kamulaştırmaya ve bunlar üzerinde idarî irtifaklar kurmaya yetkilidir.” ve 47. Maddesinde “Devletleştirme gerçek karşılığı üzerinden yapılır. Gerçek karşılığın hesaplanma tarzı ve usulleri kanunla düzenlenir.” denilmektedir. Bu iki madde de kamulaştırmanın taşınmazların gerçek değeriyle yapılması gerektiğini vurgulanmıştır.

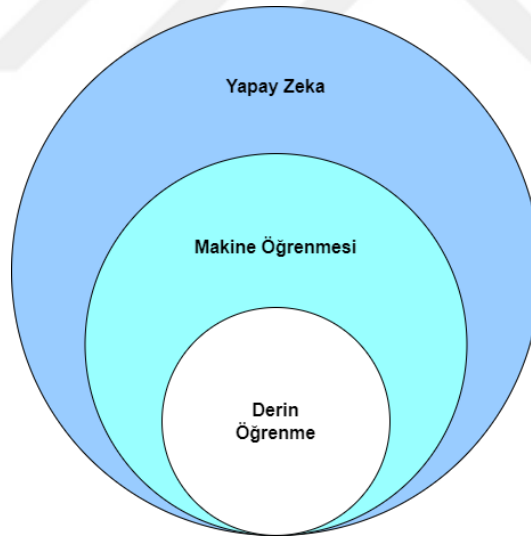
2942 sayılı Kamulaştırma Kanunu 11. Maddesi ülkemizdeki taşınmaz değerleme mevzuatları arasında taşınmaz değerinin belirlenmesi konusunda en fazla ayrıntıya sahip kanun maddesidir. Bu maddeye göre “15 inci madde uyarınca oluşturulacak bilirkişi kurulu, kamulaştırılacak taşınmaz mal veya kaynağın bulunduğu yere

mahkeme heyeti ile birlikte giderek, hazır bulunan ilgilileri de dinledikten sonra taşınmaz mal veya kaynağın; a) Cins ve nevini, b) Yüzölçümünü, c) Kıymetini etkileyebilecek bütün nitelik ve unsurlarını ve her unsurun ayrı ayrı değerini, d) Varsa vergi beyanını, e) Kamulaştırma tarihindeki resmi makamlarca yapılmış kıymet takdirlerini, f) Arazilerde, taşınmaz mal veya kaynağın mevkii ve şartlarına göre ve olduğu gibi kullanılması halinde getireceği net gelirini, g) Arsalarda, kamulaştırma gününden önceki özel amacı olmayan emsal satışlara göre satış değerini, h) Yapılarda resmi birim fiyatları ve yapı maliyet hesaplarını ve yıpranma payını, her bir ölçünün etkisi açıklanmak kaydıyla bedelin tespitinde etkili olacak diğer objektif ölçüleri, esas tutarak düzenleyecekleri raporda bütün bu unsurların cevaplarını ayrı ayrı belirtmek suretiyle ve ilgililerin beyanını da dikkate alarak Sermaye Piyasası Kurulu tarafından kabul edilen değerlendirme standartlarına uygun, gerekçeli bir değerlendirme raporuna dayalı olarak taşınmaz malın değerini tespit ederler. Taşınmaz malın değerinin tespitinde, kamulaştırmayı gerektiren imar ve hizmet teşebbüsünün sebep olacağı değer artışları ile ilerisi için düşünülen kullanma şekillerine göre getireceği kâr dikkate alınmaz.” denilmektedir.

Kadastro Kanunu 36. Maddesinde yargılama giderleri ve kadastro harcının hesaplanmasında ilgili taşınmaza ait emlak vergisi değerinin esas alınacağı belirtilmiştir. Türk Medeni Kanunu 899. ve 904. Maddelerinde ipotek borç senedi ve irat senedi gibi rehin ve alacak haklarının miktarının taşınmazın değerine göre konulabileceği hükmü yer alır. 6306 sayılı Afet Riski Altındaki Alanların Dönüştürülmesi Hakkındaki Kanunun 6. Maddesinde arsa paylarının Bakanlıkça rayiç değerinin tespit edilerek, bu rayiç değerden az olmaması koşuluyla satışından bahsedilmiştir. Tapu Sicili Tüzüğü’nde de Kamulaştırma Kanunu gibi taşınmazın değerini etkileyecek kriterler belirtilmiştir. Tapu Sicili Tüzüğü’nün 38. Maddesinde taşınmazın değerinin; cinsi, yüzölçümü, emlak vergisi beyan değeri, resmi makamlarca yapılmış değer takdiri, getireceği gelir, emsal satış değerleri, maliyet değeri ve yıpranma payı ve değeri etkileyecek diğer nitelikler ile belirleneceği hükmedilmiştir.

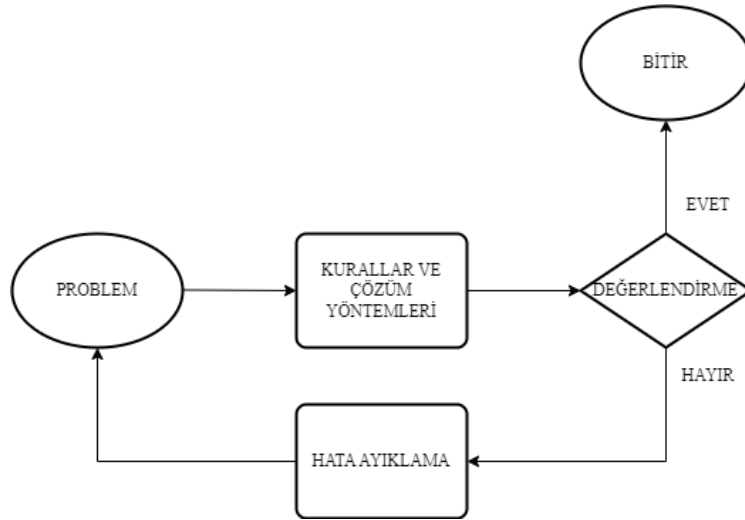
4.1 Makine Öğrenmesi Nedir?

Makine öğrenmesi, veriler kullanarak performansı arttırmak veya doğru tahminler/sınıflandırmalar yapmak için kullanılan hesaplama yöntemlerinin tümüdür. Makine öğrenmesi, genel olarak bir makinenin/bilgisayarın insan davranışını taklit etme yeteneği olarak tanımlanan yapay zekânın bir alt dalıdır (Şekil 4.1). Bilgisayar oyunları ve makine öğrenmesi alanının öncülerinden Arthur Lee Samuel 1959 yılında makine öğrenmesini “*bilgisayarları programlamadan onlara öğrenme yeteneği kazandıran çalışma alanı*” olarak tanımlamıştır [36].



Şekil 4.1 Yapay zekâ hiyerarşisi [37]

Klasik bir programlama algoritmasında problemin değerlendirilmesi sonrası problemin çözümü için kurallar ve çözüm yöntemleri belirlenir. Kurallar ve çözüm yöntemleri problemin çözümü için yeterli ise problem çözülmüş olur, ancak yeterli değilse hatalar analiz edilir ve kurallar ile çözüm yöntemleri analiz edilen hatalara göre tekrar düzenlenir. Problem çözülene kadar bu işlem tekrarlanır ve program kullanılmaya hazır hale getirilir (Şekil 4.2).

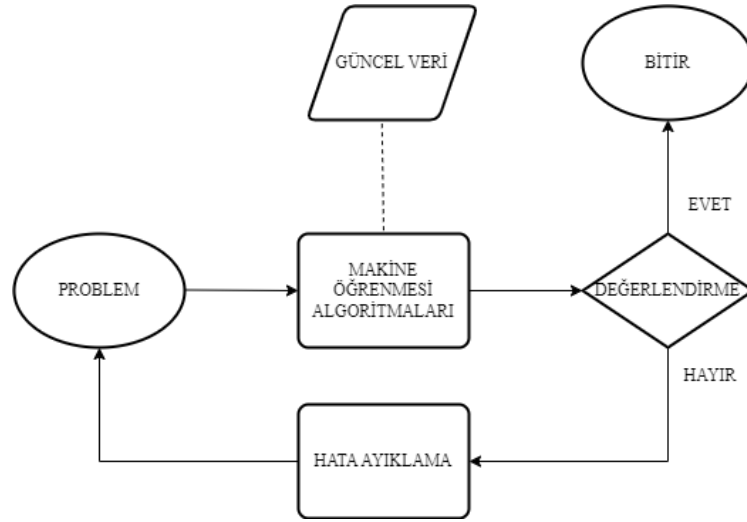


Şekil 4.2 Klasik programlama akış şeması [38]

Ancak problem zor olduğunda problemin çözümü için belirlenen kurallar ile çözüm yöntemleri karmaşık ve uzun olacaktır. Ayrıca karmaşık ve uzun bir algoritmanın zaman içinde güncellenmesi de zor olacaktır. Aksine makine öğrenmesi tekniklerine dayanan bir algoritmanın kurallar ve çözüm yöntemleri sadece eğitim verisi denilen veri setleri ile rahatça güncellenebilir, test veri seti ile algoritmanın sonucu değerlendirilir (Şekil 4.3). Ayrıca makine öğrenmesi algoritmaları klasik programlama algoritmaları ile çözülemeyecek ses tanıma, dil işleme, görüntü sınıflandırma gibi problemleri de çözebilir.

Makine öğrenmesinin tarihi [39] çalışmalarında sunduğu sinir ağlarının ilk matematiksel modeli ile başlar. Ardından [40]'de davranışın sinir ağları ve beyin aktivitesiyle nasıl ilişkili olduğuna dair teorileri ortaya atmıştır. [41]'de bir bilgisayarın gerçek bir zekâyâ sahip olup olamayacağını test etmek için Turing Testini ortaya çıkarmıştır. Testi geçebilmek için bir bilgisayarın bir insanı kendisinin de insan olduğuna inandırması gerekmektedir. [36], ilk makine öğrenmesi programını yazmıştır. Bu programda, makine dama oyununu oynamakta ve program dama oynadıkça kendini geliştirmektedir.

Makine öğrenmesinde bir sonraki en önemli adım, [43] çalışmalarında en yakın komşuluk algoritmasını yayınlamasıyla atılmıştır. Bu algoritma, bir örneğin hangi kümede olduğunun ve en yakın komşusunun belirlendiği bir sınıflandırma algoritmasıdır. 1979'da Stanford Üniversitesi öğrencileri sandalyelerle dolu odayı insan müdahalesi olmadan geçebilen bir araç ürettiler [44]. [45] Explanation-Based Learning üzerine yaptığı çalışmasında bir sistemin belli bir durumu analiz ederek genel kurallar çıkarmasını ve bu kuralları yeni durumlara uygulamasını sağlayan bir yaklaşım geliştirmiştir. 1992 yılında American Telephone and Telegraph



Şekil 4.3 Makine öğrenmesine dayalı programlama akış şeması [42]

Company ses tanıma çağrı işleme hizmeti ile çağrıları uygun şekilde yönlendirmek ve ele almak için otomatik konuşma tanıma teknolojisini kullanarak makinelerle ses işleme işlemini gerçekleştirmiştir [46]. [47], sınıflandırma için Destek Vektör Makineleri olarak bilinen ve sınıflandırma, regresyon ve diğer uygulamalar için bu alandaki son teknoloji temsil eden bir algoritma önermişlerdir.

1997 de Intenational Business Machines (IBM) şirketinin ürettiği bilgisayar Deep Blue satranç oyununda dünya şampiyonunu yenmiştir. “Derin Öğrenme” kavramı [48] tarafından bilgisayarların görüntü ve videolardaki nesneleri ve metinleri görmesini ve ayırt etmesini sağlayan yeni algoritmalar önermesiyle ortaya çıkmıştır. 2014 yılında Facebook fotoğraftaki kişileri insan gibi tanıyabilen veya doğrulayabilen bir yazılım algoritması olan DeepFace’i geliştirmiştir. 2016 yılında Google’ın yapay zekâ algoritması AlphaGo, satrançtan daha zor bir oyun olan Go oyununda profesyonel bir oyuncuyu mağlup etmiştir. 2017 yılında Waymo şirketi tarafından Amerika Birleşik Devletleri’nde otonom araçları test etmeye başladı ve aynı yıl otonom taksileri tanıttı. 2020 yılında ise OpenAI firması tarafından bir istem verildiğinde insan benzeri metin üretme yeteneğine sahip bir doğal dil işleme algoritması olan GPT-3 duyurulmuştur.

Makine öğrenmesinin tarihi gelişimine baktığımızda birçok farklı alanda kullanıldığı görülmektedir. Günümüzde üretim hattındaki ürünleri otomatik sınıflandırma, beyin tarama görüntülerinden tümör tespiti, uzun belgeleri otomatik özetleme, kişisel asistan oluşturma, performans ölçütüne dayalı gelir tahmini, sesli komutlara tepki verme, kredi kartı sahtekarlığı tespiti, müşterilere ürün önerme, oyun botu oluşturma gibi sanayi, ticaret, sağlık, güvenlik ve oyun sektörü başta olmak üzere birçok sektörde makine öğrenmesi kullanılmaktadır [42].

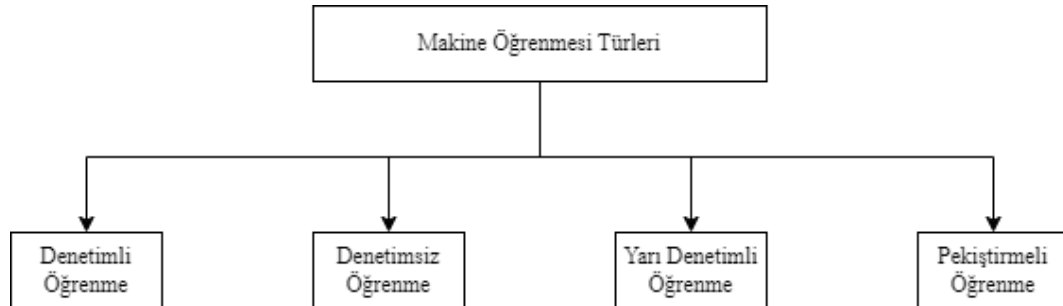
4.2 Makine Öğrenmesi Türleri

Makine öğrenmesi algoritmalarının sayısı ve türleri oldukça fazladır. Fakat en temelde makine öğrenmesi algoritmaları eğitim sırasında insan gözetiminde aldıkları denetime göre Şekil 4.4'te belirtildiği gibi şu şekilde sınıflandırılmaktadır [49]:

- Denetimli öğrenme
- Denetimsiz öğrenme
- Yarı denetimli öğrenme
- Pekiştirmeli öğrenme

Makine öğrenmesi algoritmaları bu sınıflandırmanın dışında şu kategorilerde de sınıflandırılmaktadır:

- Gelen veri akışıyla kademeli olarak öğrenip öğrenememesine göre; Yığın, çevrimiçi öğrenme,
- Veri örneklerine veya veri içindeki örüntülere dayalı öğrenmesine göre; Örnek tabanlı, model tabanlı öğrenme.



Şekil 4.4 Makine öğrenmesi türleri [49]

4.2.1 Denetimli Öğrenme

Denetimli öğrenmede, makine öğrenmesi modeli eğitim verisi olarak bir dizi etiketli veriyi işleyerek, etiketsiz verileri tahmin etmeye çalışır. Bu tür öğrenme modelleri sınıflandırma, regresyon veya sıralama problemlerinde kullanılır. Denetimli öğrenme algoritmalarına şu örnekler verilebilir:

- K-en Yakın Komşular
- Doğrusal Regresyon
- Lojistik Regresyon
- Destek Vektör Makineleri
- Karar Ağaçları ve Rastgele Orman
- Sinir Ağları

4.2.2 Denetimsiz Öğrenme

Makine öğrenmesi her zaman etiketli verilere ihtiyaç duymaz. Bazı problemlerde etiketleri veya sınıfları bilinmeyen verilerden faydalı bilgiler elde edilebilir. Denetimli öğrenme modelinde etiketi bilinenlerden tümevarım yöntemi ile etiketi bilinmeyenler sınıflandırılır iken denetimsiz öğrenme mevcut etiketsiz verilerin özniteliklerine odaklanır; özniteliklerden örüntüler çıkarır. Denetimsiz öğrenme modelleri kümeleme, anomali tespiti, yoğunluk tahmini gibi görevlerde kullanılır. En yaygın denetimsiz öğrenme algoritmaları şunlardır:

- K-Ortalamalar Kümeleme
- Hiyerarşik Küme Analizi
- Temel Bileşen Analizi

4.2.3 Yarı Denetimli Öğrenme

Yarı denetimli öğrenme sınıflandırma ve regresyon görevleri için modeli eğitmek üzere hem etiketli hem de etiketsiz verileri kullanan denetimli ve denetimsiz öğrenme modellerini tek bir çatı altında toplayan makine öğrenmesi türüdür. Bu öğrenme türü genellikle veri etiketlemenin zaman alıcı ve maliyetli olduğu durumlarda kullanılır. Bu öğrenme türünde verinin çok az miktarının etiketlenmiş verilerden oluştuğu görülür. Nesne tanıma, yüz tanıma, hastalık tanısı, doğal dil işleme gibi uygulamalarda kullanılır.

4.2.4 Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme türünde eğitim ve test aşamaları aynı anda gerçekleşmektedir. Sistem bilgi toplamak ve topladığı bilgilerle kendisini eğitmek için sürekli çevresiyle etkileşime girer ve deneyler yapar. Bu bir dizi deneyin sonucunda

sistem ödöl veya ceza almaktadır. Sistem bu ödöl veya cezalara göre davranışlarını/ öğrenimini optimize eder. Pekiştirmeli öğrenme türünde amaç sistemin ödöllelerini maksimize ve cezalarını minimize etmektir. Birçok akıllı robot yürümeyi öğrenmek için pekiştirmeli öğrenme algoritmaları ile öğrenimini tamamlar. Pekiştirmeli öğrenme sınıflandırma, borsa tahmini, robotik operasyon gibi görevlerde yaygın olarak kullanılır.

4.3 Metodoloji

Veri bilimi ve makine öğrenmesi çalışmaları son yıllarda hızla büyüyen ve gelişen bir disiplin haline gelmiştir. İş dünyasında karar verme süreçlerini geliştirmek, operasyonları takip etmek ve müşteri deneyimlerini iyileştirmek için veri analizi ve makine öğrenmesi hayati bir rol oynamaktadır. Bu süreçleri başarı ile yönetebilmek, takip edebilmek ve uygulamak için yapılandırılmış bir yaklaşıma gerek duyulmaktadır.

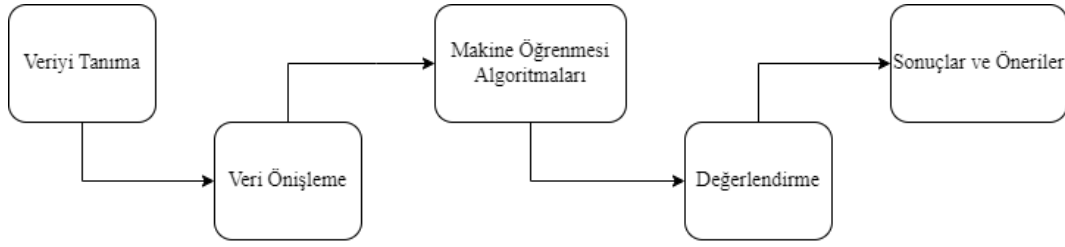
CRISP-DM (Cross-Industry Standard Process for Data Mining) modeli veri madenciliği ve veri bilimi için sektörden bağımsız bir süreç modelidir. İşin anlaşılmasında uygulamaya kadar altı iteratif aşamadan oluşur [50].

CRISP-DM:

- İşin Anlaşılması (Business Understanding)
- Verinin Anlaşılması (Data Understanding)
- Veri Hazırlığı (Data Preparation)
- Modelleme (Modelling)
- Değerlendirme (Evaluation)
- Dağıtım/Uygulama (Deployment)

aşamalarından oluşmaktadır [51].

Bu çalışmada da CRISP-DM süreç modeli uygulanarak uluslararası kabul görmüş bir metodoloji sırası belirlenmiştir. Bu çalışmaya ait iş akış şeması Şekil 4.5'te verilmiştir.



Şekil 4.5 İş akış şeması

4.4 Veri Önileme

Makine öğrenmesinin belirli bir görevdeki başarısını etkileyen faktörlerden biri veri önilemedir. Modele girecek verilerin temsili ve kalitesi modelin performansını doğrudan etkiler. Çok fazla alakasız ve gereksiz bilgi varsa veya gürültülü ve güvenilir veriler veri setinde mevcutsa, eğitim aşamasında bilginin keşfi o kadar zor olmaktadır. Makine öğrenmesi problemlerinde veri hazırlama ve filtreleme adımları önemli miktarda işlem süresi almaktadır. Veri önileme; veri temizleme, veri normalizasyonu, veri dönüşümü, veri entegrasyonu, eksik verilerin çıkarılması veya tamamlanması, aykırı veya gürültülü veri tespiti gibi işlemleri içerir [52]. Başarılı bir veri önileme aşamasından sonra elde edilen nihai veri seti, daha sonra uygulanacak makine öğrenmesi algoritması için güvenilir ve uygun bir kaynak olarak kabul edilebilir.

Ham veri seti eksik, hatalı, gürültülü, tutarsız veya aykırı veriler içerebilir. Bu tür verilerin ham veri setinden ayıklanması veri temizleme olarak adlandırılır. Bir sayının yanlışlıkla metin olarak kaydedilmesi veya ondalıklı sayılara virgül yerine nokta konulması gibi hatalı veriler, fiyat bazındaki bazı verilerde Türk Lirası bazı verilerde Amerikan Doları kullanılması veya bir bireyin yaşı ile tutarlı olmayan doğum tarihi gibi tutarsız veriler, sensör arızası nedeniyle hava sıcaklığı değerinin 150°C ölçülmesi veya bir öğrencinin notunun 250 girilmesi gibi gürültülü verilerin tespit edilip düzeltilmesi ya da veri setinden temizlenmesi gerekmektedir.

Veri bilimi ve makine öğrenmesi projelerinde en çok karşılaşılan ve çözümü en kolay problemlerden biri veri setinde eksik verilerin bulunmasıdır (Bkz. Tablo 5.1). Eksik verinin varlığı, veri setinde belirli bir değişkenin gözlemlenmemesi ya da değişkenin gözlem değerinin işlenmemesi ile ortaya çıkar. Model performansını etkilememesi açısından eksik veri sorunu, eksik veri içeren gözlemler ya da sütunların veriden tamamen çıkarılması ya da eksik verilerin ortalama, medyan, mod, ileri veya geri değer vb. yöntemlerle tamamlanması ile çözümlenmelidir.

Aykırı veriler, veri setindeki gözlemlerin büyük bir bölümünden farklı değerlere sahip olan ve genellikle veri setinin genel dağılımına uymayan verilerdir (Bkz.

Tablo 5.3). Aykırı değerler, veri setinin çoğunluğunun uyum gösterdiği genel istatistiki dağılımından sapan ve genel sonuçlar çıkarmak için kullanılmayacak biçimde diğer verilerden belirgin şekilde uzakta bulunur [53]. Aykırı değer analizi için kullanılan tekniklerden biri Yerel Aykırı Faktör (Local Outlier Factor, LOF) analizidir. Yerel aykırı faktör tekniği [54] tarafından geliştirilen yoğunluk tabanlı bir aykırı değer hesaplama yöntemidir. Bu yöntem bir gözlemin çevresindeki komşulara göre ne kadar izole olduğunun sayısal bir ifade ile tespit edilmesine dayanır. Bu teknikte ilk olarak bir a nesnesine en yakın k . nesnenin öklid mesafesi hesaplanır. Bu mesafeye k -mesafe adı verilir. Bu mesafe içinde kalan nesneler tarafından bir $kNN(a)$ kümesi oluşturulur. Bu kümedeki bir b nesnesine erişilebilirlik mesafesi (4.1)' deki eşitlik ile hesaplanır.

$$reach - dist_{a,b} = \max((k - mesafe(b), d(a, b))) \quad (4.1)$$

Bu eşitlikte $d(a, b)$, a 'dan b 'ye olan öklid mesafesidir. Bir sonraki adımda a 'nın yerel erişilebilirlik yoğunluğu (4.2)'deki gibi hesaplanır.

$$lrd_a = \frac{k}{\sum_{b \in kNN(a)} reach - dist_{a,b}} \quad (4.2)$$

Yerel erişilebilirlik yoğunluğu hesabından sonra yerel aykırı faktör değeri (4.3) eşitliği ile hesaplanır.

$$LOF_a = \frac{\frac{1}{k} \sum_{b \in kNN(a)} lrd_b}{lrd_a} \quad (4.3)$$

Yerel aykırı faktör değeri 1'e yakın bir değerse yoğunluklar aynı olacağından a gözlemi aykırı olmayacaktır. Eğer bu değer 1'den büyükse a gözlemi aykırı gözlem olacaktır.

Veri ön işleme işlemlerinden bir diğeri de veri dönüşümüdür. Bu işlemde verilerin yapısı ve/veya formatı değiştirilerek makine öğrenmesi modellere uygun hale getirilir. Veri dönüşümüne normalizasyon, encoding gibi örnekler verilebilir.

Makine öğrenmesi algoritmaları genellikle sayısal verilerle birlikte çalışırlar. Kategorik veriler sayısal verilere dönüştürülerek yani encoding (karakter kodlaması) yapılarak makine öğrenmesi modellerinin verileri anlaması sağlanmalıdır. Bilinen karakter kodlaması türlerine label encoding ve one-hot encoding örnek verilebilir.

Label encoding, kategorik verilerin sınıflarını sayısal bir değer ile ifade eden bir dönüşüm işlemidir. Bu yöntem özellikle ordinal kategorik değişkenli bir veride en düşük sınıfa 0'dan başlanarak sınıf önemine göre bütün sınıf türlerine sayısal değerlerin atandığı bir veri dönüştürme yöntemidir.

One-hot encoding, kategorik verileri her kategori için bir ikilik sütun oluşturarak sayısal verilere dönüştürme yöntemidir. Her bir kategori kendi sütununda 1, diğer sütunlarda 0 değerini alarak bir dizi ikili değişken ile temsil edilir. Böylece her bir kategorik değer makine öğrenmesi modellerinin anlayabileceği bir formata dönüşmüş olur. One-hot encoding dönüşümüne bir örnek Tablo 4.1'de verilmiştir.

Tablo 4.1 One-hot encoding örneği

Tapu Türü	Hisseli	Kat İrtifaklı	Kat Mülkiyetli
Hisseli	1	0	0
Kat İrtifaklı	0	1	0
Kat Mülkiyetli	0	0	1

One-hot encoding işlemlerinde karşılaşılan en büyük problem Dummy değişken tuzağıdır. Tablo 4.1'de görüldüğü gibi tüm kategorilere bir encoding işlemi uygulandığında çoklu doğrusal bağlanıma neden olabilir. Bu durum regresyon analizinde modeli olumsuz etkiler. Bu üç sütundan ikisinin değerini bildiğimizde üçüncüsü tahmin edilebilir. [1,0] Hisseli, [0,1] Kat İrtifaklı kategorisini temsil ediyorsa [0,0] vektörünün Kat Mülkiyetli kategorisini temsil ettiği anlaşılabılır.

Veri indirgeme, veri ön işleme işlemlerinden bir diğeridir. Makine öğrenmesi modellerinde veri işleme süresini azaltarak performansı arttırmak, veri depolama maliyetlerini düşürmek amacıyla veri indirgeme işlemi yapılır. Veri indirme hem gözlem hem de değişkenlerin boyutu azaltarak bilgi kaybı yaşamadan ya da minimum kayıpla modelin çalışmasını sağlar. Öznitelik seçimi, özellik çıkarımı, gözlem seçimi gibi işlemler veri indirgemeye örnek verilebilir.

Bir veri kümesindeki özniteliklerin tümü makine öğrenmesi modeli için gerekli olmayabilir; hatta bazı öznitelikler modeli olumsuz etkileyebilir. Bu yüzden veri kümesindeki gereksiz ve önemsiz özniteliklerin modelden çıkarılması gerekebilir. Bu gereksiz ya da önemsiz özniteliklerin veri kümesinden çıkarılması işlemine öznitelik seçimi denir. Öznitelik seçimi yöntemlerinden biri korelasyon analizidir. Korelasyon analizi iki değişkenin birbiriyle ne kadar ilişkili olduğunu gösterir [55]. Bir değişkendeki değişim diğer değişkende de aynı yönde değişime neden oluyorsa aralarında pozitif, ters yönde değişime neden oluyorsa aralarında negatif korelasyon var demektir. Bu değişimin gücünü gösteren değer olan korelasyon katsayısı değeri -1 ile +1 arasında değer alır. -1 değeri iki değişken arasında negatif tam korelasyon,

+1 değeri pozitif tam korelasyon olduğu anlamına gelmektedir. 0 değeri ise iki değişken arasında bir ilişkinin bulunmadığını ifade eder. Korelasyon katsayısının hesaplanması (4.4) eşitliğinde gösterilmiştir.

$$r = \frac{1}{n} \sum_{i=1}^n \frac{(a_i - \bar{a})}{\sigma_a} \frac{(b_i - \bar{b})}{\sigma_b} \quad (4.4)$$

Bu eşitlikte n gözlem sayısını, $\bar{a}$ ve $\bar{b}$ değişkenlerin ortalama değerlerini, σ_a ve σ_b değişkenlerin standart sapmalarını ifade etmektedir.

Bağımsız değişkenler arasında güçlü ilişkilerin olması çoklu bağlanım (multicollinearity) sorununu da beraberinde getirir [56]. Bu sorun regresyon modellerindeki bağımsız değişkenler arasında güçlü bir ilişki olduğunda modellerin tahminlerinin doğruluğunu ve değişkenlerin modele olan etkilerinin yorumlanmasını etkiler. (4.5) eşitliğindeki gibi bir y_i modelini açıklarken x_{i1} ve x_{i2} arasında yüksek bir korelasyon varsa bu iki bağımsız değişkenin etkilerinin ayrı ayrı belirlenmesi zorlaşmaktadır. Korelasyonu yüksek bağımsız değişkenlerden birini veri setinden çıkarmak çoklu bağlanım problemini giderme yöntemlerinden biridir.

$$y_i = \alpha_{i0} + \alpha_{i1}x_{i1} + \alpha_{i2}x_{i2} + \dots + \alpha_{in}x_{in} + \varepsilon_i \quad (4.5)$$

Bunun dışında çoklu bağlanım probleminin diğer etkileri şunlardır:

- Katsayı tahminlerinin standart hatalarını artırır, bağımsız değişkenlerin önemsiz görünmesine sebep olabilir [57].
- Bir bağımsız değişkendeki değişimin bağımlı değişken üzerindeki etkisi doğru ölçülemez.
- Model performansının anlamlılığını düşürür. Yüksek çoklu bağlanım durumlarında modelin belirleme katsayısı (R^2) değeri yüksek olsa bile bu durum yanıltıcıdır.
- Modelde aşırı uyum (overfitting) sorunlarına yol açabilir.

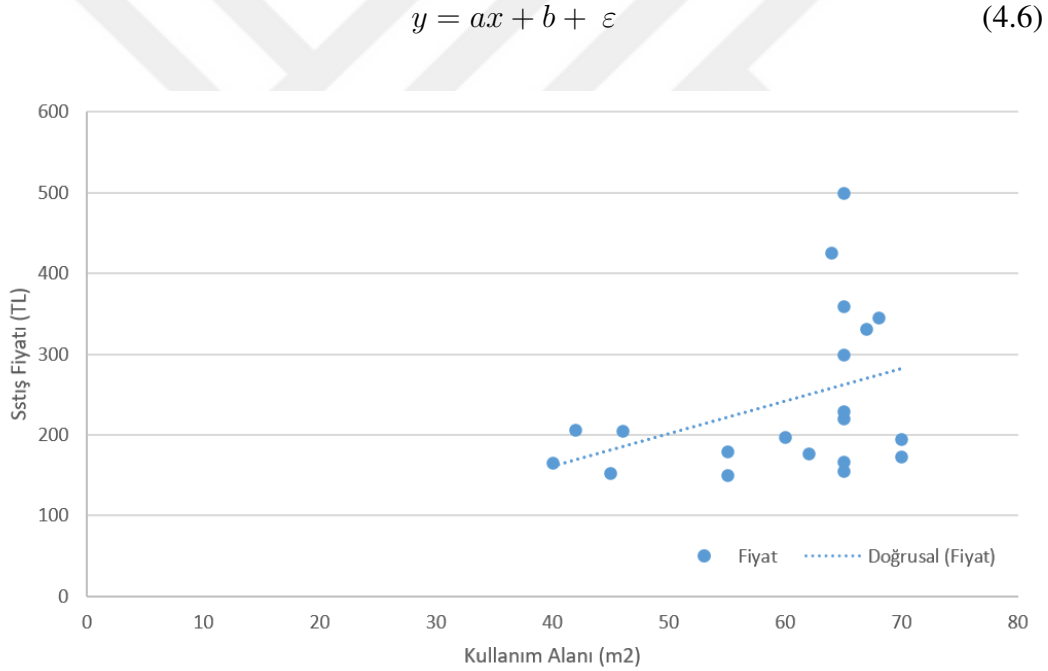
4.5 Makine Öğrenmesi Algoritmaları

Makine öğrenmesi algoritmaları belirli bir problemin çözümünde kullanılan, veri setlerine dayalı ve güncellenmesi kolay algoritmalar. Kullanılacak algoritma

problemin ve veri setinin türüne göre farklılık gösterir. Bu çalışmada kullanılan problem bir tahmin problemi olduğundan bu kısımda tahmin problemlerinde kullanılan Çoklu Doğrusal Regresyon, Coğrafi Ağırlıklı Regresyon, Rastgele Orman, AdaBoost, XGBoost ve CatBoost makine öğrenmesi algoritmalarına değinilmiştir.

4.5.1 Çoklu Doğrusal Regresyon

Regresyon, belirli bir rastgele değişkenin nasıl davranış gösterdiğini açıklamak için kullanılan matematiksel modeldir [58]. Bu değişken satılık bir evin satış fiyatı, bir firmanın mali büyüme hızı veya ekimi yapılan bir mahsulün verimi olabilir. Basit bir doğrusal regresyon modeli (4.6) eşitliğinde ifade edilen iki boyutlu uzaydaki verinin doğrusunu çizmeye çalışır (Şekil 4.6).

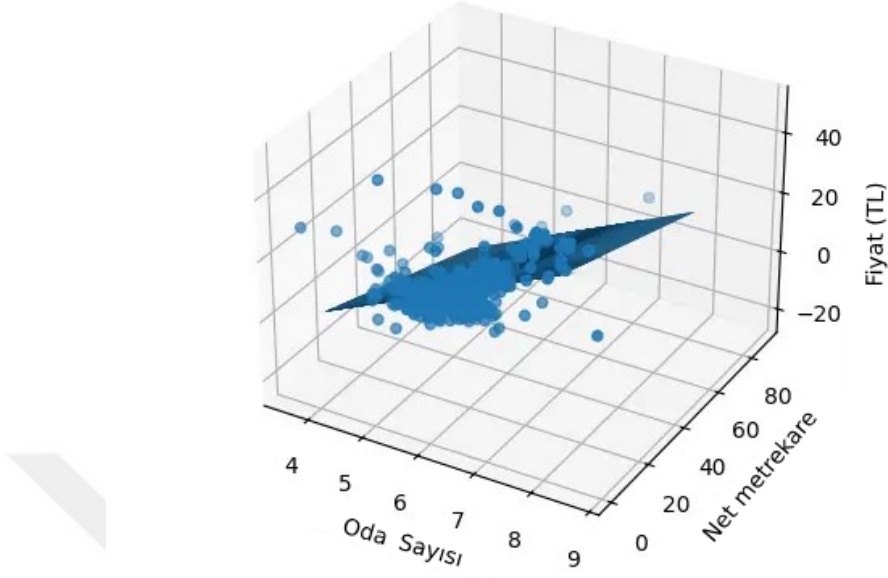


Şekil 4.6 Basit doğrusal regresyon grafiği

Bu eşitlikte davranışı açıklanacak olan değişken y bağımlı değişkeni, x ise açıklayıcı yani bağımsız değişkeni ifade eder. Basit bir doğrusal regresyonda bağımlı değişkeni açıklayacak tek bir bağımsız değişken bulunur. Çoklu doğrusal regresyonda ise (4.7) eşitliğinde belirtildiği gibi y bağımlı değişkenini açıklayan iki veya daha fazla bağımsız değişken bulunmaktadır (Şekil 4.7).

$$y = \alpha_0 x_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n + \beta + \varepsilon \quad (4.7)$$

Bu eşitliklerde y ve x dışında; α değişken katsayısını, β sabit değeri ve ε ise hata payını temsil etmektedir.



Şekil 4.7 Çoklu doğrusal regresyon grafiği

Bağımsız değişken sayısı ne kadar fazla olursa, bağımsız değişken ile bağımlı değişken arasındaki ilişki daha doğru ve kolay açıklanabilir. Ancak bağımsız değişken sayısının artırılması ek gözlemler gerektirdiğinden zaman alıcı ve masraflı olmaktadır. Bu nedenle toplam varyansı en az sayıda değişkenle açıklamak bu modelin esas amacıdır [59].

(4.2) eşitliğinde görüldüğü gibi Çoklu Doğrusal Regresyon modeli birden fazla açıklayıcı/bağımsız değişkenin bulunduğu bir model olması sebebiyle her bir açıklayıcı/bağımsız değişkenin modele etkisi farklılık gösterebilir. Değişkenlerin modeldeki ağırlıklarını değerlendirmek için birkaç yöntem bulunmaktadır:

- **Mümkün olan tüm regresyonlar yöntemi (All possible regression):** Bu yöntemde değişkenlerin 1'den n 'e kadar ki tüm kombinasyonlarını içeren regresyon modelleri oluşturulur. R^2 skoru en yüksek modeller seçilerek bu modellerdeki değişkenlerin her birinin sırası ile modelde olmaması durumundaki katkıları kontrol edilir [60].
- **Değişken ekleme yöntemi (Forward selection):** Bu yöntemde en iyi tek değişkenli modele en yüksek katkıyı sağlayacak değişken ilave edilir. Değişken ilavesi ile regresyon modelinin hata kareler toplamındaki değişimin

önemine bakılır. Bu yöntem eklenen değişkenle hesaplanan hata kareler toplamında önemli bir değişim olmayıncaya kadar devam ettirilir. [60].

- Değişken eleme yöntemi (Backward selection): Değişken eleme yöntemi tüm bağımsız değişkenlerle çoklu doğrusal regresyon modelinin kurulması ile başlar. En yüksek p-değerine sahip değişken modelden çıkarılır ve yeni bir model oluşturulur. Bu işlem modeldeki tüm değişkenler için 0,05 p-değerinin altında bir değere sahip olana kadar tekrar edilir.
- Değişken ekleme ve eleme yöntemi (Stepwise selection): Bir modelde kullanılacak bağımsız değişkenlerin iteratif olarak seçilmesini içeren bir regresyon modeli oluşturma yöntemidir. Değişken ekleme ve eleme yöntemlerinin bir kombinasyonudur.

4.5.2 Coğrafi Ağırlıklı Regresyon

[61], uzayın farklı noktalarında farklı ilişkilerin var olmasına izin veren çoklu doğrusal regresyon modelini kalibre ederek coğrafi ağırlıklı regresyon adını verdikleri bir teknik geliştirmişlerdir. Bir apartman dairesine ek bir yatak odası eklendiğinde marjinal bir fiyat artışı olduğu düşünülebilir fakat bu fiyat değişimi apartman dairesinin bulunduğu konumdaki sosyal, kültürel ve ekonomik duruma göre değişiklik gösterebilir. Ek yatak odasının apartman dairesi üzerindeki fiyat değişimi bekar veya yaşlı çiftlerin ağırlıklı yaşadığı mahalleler de olumsuz olabilirken, çocuklu ve kalabalık ailelerin ağırlıklı olduğu mahallelerde olumlu olabilir. İşte bu gibi uzay üzerindeki konumsal varyasyonlar Coğrafi Ağırlıklı Regresyon ile açıklanabilmektedir. Bu teknik karmaşık mekânsal varyasyonların tanımlanmasına, haritalanmasına ve modellenmesine olanak sağlamaktadır [61].

Coğrafi ağırlıklı regresyon, geleneksel çoklu doğrusal regresyona dayanan parametrik bir modeldir. Çoklu doğrusal regresyona ait (4.7) eşitliğinde her bir bağımsız değişkenin katsayısı sabitken, Coğrafi ağırlıklı regresyonun (4.8) ve (4.9) eşitliklerinde belirtildiği gibi bağımsız değişkenlerin katsayısı veri noktalarının konuma bağlı olarak mekânsal değişiklikler göstermektedir [12].

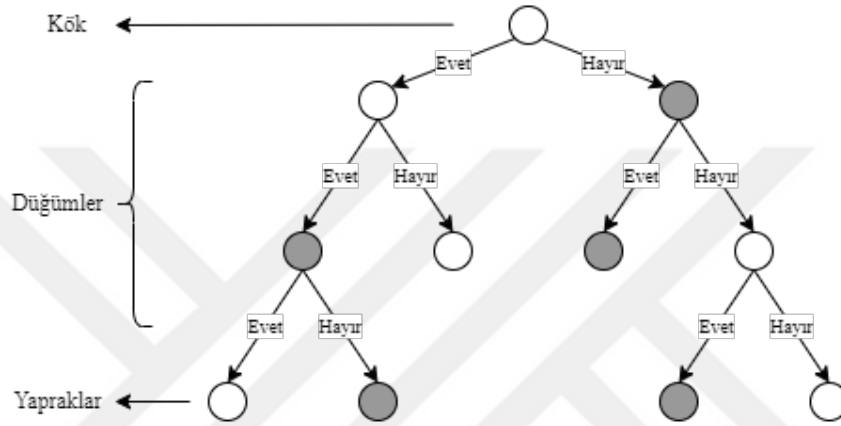
$$y_i = \alpha_{i0} + \alpha_{i1}x_{i1} + \alpha_{i2}x_{i2} + \dots + \alpha_{in}x_{in} + \varepsilon_i \quad (4.8)$$

$$y_i(\mu_i, \nu_i) = \alpha_{i0}(\mu_i, \nu_i) + \sum_{r=1}^n \alpha_{ir}(\mu_i, \nu_i)x_{ir} + \varepsilon_i(\mu_i, \nu_i) \quad (4.9)$$

Eşitlikte (μ_i, ν_i) i. noktanın koordinatı, y_i noktanın bağımlı değişkeni, x_{ir} bağımsız değişkeni, α_{ir} değişken katsayısını temsil ederken ε_i ise hata payıdır.

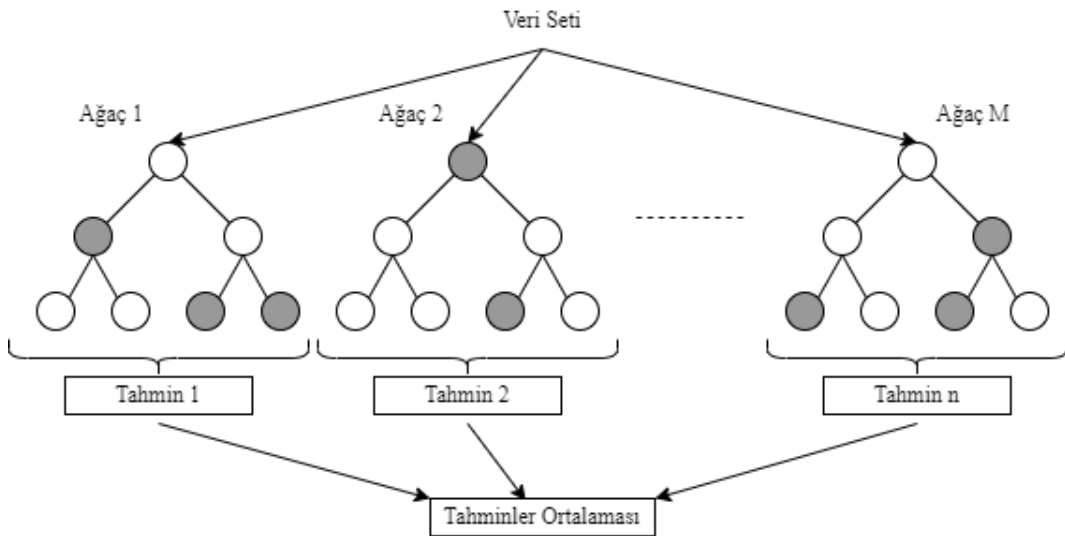
4.5.3 Rastgele Orman

Rastgele Orman, [62] tarafından ortaya çıkarılan sınıflandırma ve regresyon problemlerinin çözümü için kullanılan karar ağacı temelli denetimli bir makine öğrenmesi algoritmasıdır. Karar ağaçları Şekil 4.8’de gösterildiği gibi kök, yaprak ve düğümlerden oluşan ağaca benzer bir yapıdır. Karar ağacının başlangıç noktası, verilerin sınıflandırılması veya tahmini amacıyla sorular sorularak kök düğümde oluşturulmaktadır. Bu süreç dallanmamış düğümler ya da yapraklara ulaşılan kadar yani kesin sınıflama veya en yakın tahmin sonucu elde edilene kadar sürdürülmektedir.



Şekil 4.8 Karar ağacı yapısı [63]

Şekil 4.9’de çalışma prensibi sunulan Rastgele Orman aslında basit karar ağaçları topluluğunun özel bir türüdür ve bazı girdi verileri kullanılarak topluluktaki her bir ağaç tarafından yapılan sınıflandırma problemlerinde çoğunluk oylamasına veya regresyon problemlerinde ortalamaya dayalı bir tahmin verir [10].



Şekil 4.9 Rastgele orman algoritması [64]

Rastgele Orman M adet rastgele karar ağacından oluşur. Bir ağaç kümesindeki j. Ağaç göz önüne alındığında her x sorgu noktasında tahmin edilen değer (4.10)'te gösterilmiştir:

$$t_j(x) = m_n(x; \emptyset_j, \partial_n) \quad (4.10)$$

Burada x sorgu noktası veya girdi verisi olarak adlandırılır. $\emptyset_j$ j. Ağacın rastgele değişkenleri, ∂_n ise eğitim veri setini temsil eder. Rastgele ormanın genel tahmin değeri ise M adet rastgele karar ağacının tahminlerinin ortalamasıdır (4.11):

$$T(x) = \frac{1}{M} \sum_{j=1}^M m_n(x; \emptyset_j, \partial_n) \quad (4.11)$$

Rastgele orman algoritmasına kullanıcı iki parametre sunar, bunlardan birincisi ağaç sayısı (n) ve ağaç yapısındaki karar düğümleri için kullanılacak olan değişken sayısı (m)'dir. Bu parametreler belirlendikten sonra eğitim veri setinin 2/3'ü öğrenme verisi (inBag) ve 1/3'ü test verisi (Out of-Bag, OOB) olarak belirlenir [65].

Rastgele Orman algoritması çeşitli veri kümeleri üzerinde yüksek doğrulukla tahmin yapma ve sınıflandırma yeteneği sayesinde yaygın olarak kullanılmaktadır. Bu algoritma yüksek boyutlu verileri işleyebilmesi, karmaşık ilişkileri ortaya çıkarması ve aşırı uyum riskini minimumda tutma yetenekleri ile öne çıkar. Zayıf özniteliklere düşük ağırlık vererek verimli çalışır ve diğer yöntemlere göre daha hızlı sonuçlar verir. Eksik ya da gürültülü veriler ile aykırı değerlere başa çıkmada başarılıdır. Hem kategorik hem de sürekli veri tipleri ile çalışabilmektedir [19].

4.5.4 Adaptive Boosting (AdaBoost)

Boosting düşük doğrulukta olan birçok basit zayıf öğreniciyi bir araya getirerek daha isabetli bir tahmin modeli elde etmeyi amaçlayan bir makine öğrenmesi yöntemidir [66]. [67]'in çalışmasındaki AdaBoost algoritması ilk boosting algoritmasıdır. AdaBoost algoritması hatalı tahminler yapan zayıf bir sınıflandırıcıyı yüksek doğruluk oranına sahip güçlü bir sınıflandırıcıya dönüştürür [68]. Algoritma belirli bir zayıf öğrenme algoritmasını eğitim verileri üzerindeki çeşitli dağılımlar üzerinden iteratif şekilde çalıştırır ve ardından zayıf öğrenici tarafından üretilen sınıflandırıcıları tek bir bileşik sınıflandırıcıda birleştirir [69]. Her bir iterasyonda zayıf öğreniciyi yanlış sınıflandırılan örneklerle odaklanarak güçlendirir ve zayıf model eğitilir. Eğitilen zayıf modelin performansına göre örneklerin ağırlıkları güncellenir. Son aşamada tüm zayıf modellerin tahminleri belirlenen ağırlıklarla birleştirilerek nihai tahmin elde edilir.

AdaBoost algoritmasında girdi olarak (4.12)'da gösterilen m adet S eğitim seti verilir.

$$S = (a_1, b_1), (a_2, b_2), \dots, (a_m, b_m) \quad (4.12)$$

Bu eşitlikte a_m gözlem, b_m sınıf etiketidir. Ardından bir zayıf öğrenici algoritması seçilerek toplam T iterasyon sayısı belirlenir. Eğitim setindeki her bir örnek eşit olarak (4.13)'teki gibi ağırlıklandırılır.

$$\omega_i = \frac{1}{m} \quad (4.13)$$

Her bir $t = 1, \dots, T$ iterasyonu için zayıf öğrenici seçilen algoritma ile mevcut ω_i ağırlığı ile eğitilir ve sınıflandırma yapılır. Bu sınıflandırmanın sonunda (4.14) eşitliği ile zayıf öğrenicinin hata payı hesaplanır.

$$\varepsilon_t = \sum_{i=1}^m \omega_i I(b_i \neq h_t(a_i)) \quad (4.14)$$

(4.8) eşitliğinde $I()$ ifadesi indikatör fonksiyonunu, $h_t(a_i)$ ifadesi t . iterasyondaki öğrenicinin tahminini, b_i ise gerçek sınıf etiketi değerini ifade eder. Hata oranı hesaplandıktan sonra zayıf öğrenicinin performans değeri (4.15) eşitliğiyle hesaplanır.

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \varepsilon_t}{\varepsilon_t} \quad (4.15)$$

Bu performans ölçütüne göre yanlış sınıflandırılan örneklerin ağırlıkları (4.16) eşitliğiyle güncellenir.

$$\omega_i = \omega_i e^{(\alpha_t I(b_i \neq h_t(a_i)))} \quad (4.16)$$

İterasyonlar tamamlandığında nihai tahmin (4.17) şeklinde hesaplanır.

$$F(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right) \quad (4.17)$$

4.5.5 Extreme Gradient Boosting (XGBoost)

Klasik gradyan arttırma modeli karar ağaçları gibi zayıf tahmin modellerini ardışık şekilde birleştirerek daha güçlü bir model elde etmeyi amaçlar. XGBoost, gradyan arttırma algoritmasının optimize edilmiş, yüksek performans gösteren versiyonudur. [70] tarafından geliştirilen XGBoost, ağaç güçlendirme için yüksek etkili, ölçeklenebilir bir makine öğrenmesi algoritmasıdır. Geleneksel karar ağaçları yönteminin aksine XGBoost sadece önceki ağaçlardan öğrenmekle kalmaz aynı zamanda önceki ağacın hatalarını azaltmak için ilgili yapraklardaki puanları toplar ve sonunda daha doğru sonuç üretir. XGBoost, öğrenmeyi hızlandırabilen

ve yüksek hassasiyette tahmin yapabilen paralel ve dağıtık hesaplama gibi birçok başka özelliğe sahip bir modeldir [71]. Ayrıca bu model çeşitli kayıp fonksiyonlarını kullanarak aşırı uyuma dirençli, büyük verilere rahatlıkla uygulanabilen bir makine öğrenmesi algoritmasıdır [17].

XGBoost verilere bir ilk karar ağacı uydurarak başlar. Ardından ikinci bir model, ilk modelin düşük performans gösterdiği durumları doğru bir şekilde tahmin etmeye odaklanır. Bu güçlendirme işlemi birçok kez tekrarlanır ve böylelikle birbirini izleyen her model, önceki tüm modellerin eksikliğini gidermeye çalışır [72].

XGBoost algoritmasına girdi olarak (4.18)'de gösterilen etiketli eğitim seti verilir. Bu eğitim setinde a_n özellik, b_n etikettir. Amaç a_n özelliklerinden yola çıkarak b'_n tahmini ile b_n 'i bulmaktır.

$$S = (a_1, b_1), (a_2, b_2), \dots, (a_n, b_n) \quad (4.18)$$

Modeli eğitmek a_n özellikli, b_n etiketli eğitim verilerine en iyi uyan θ parametrelerini bulmak demektir. Modeli eğitmek için modelin eğitim verilerine ne kadar uyumlu olduğunu ölçmek üzere amaç fonksiyonu tanımlanması gerekir (4.19).

$$obj(\theta) = L(\theta) + \Omega(\theta) \quad (4.19)$$

Amaç fonksiyonu iki bileşenden oluşur bunlardan biri $L(\theta)$ eğitim kaybı fonksiyonu diğeri ise $\Omega(\theta)$ düzenleme terimidir. Ortalama karesel hata en yaygın eğitim kaybı fonksiyonlarından biridir (4.20).

$$L(\theta) = \sum_{n=1} (b_n - b'_n)^2 \quad (4.20)$$

Karar ağaçlarında tahminin matematiksel modeli her bir karar ağacından gelen tahmin skorlarının toplamıdır (4.21).

$$b'_n = \sum_{k=1}^K F_k(a_n) \quad (4.21)$$

Bu eşitlikte K karar ağacı sayısı, F_k tahmin fonksiyonudur. XGBoost iteratif bir güçlendirme algoritmasıdır. İterasyon bir başlangıç (etiketlerin ortalaması gibi) tahmin değeri ile başlar (4.22) ve her t . iterasyonda modele yeni bir ağaç eklenir.

$$b'_0 = \frac{1}{n} \sum_{i=0}^n b_i \quad (4.22)$$

XGBoost modelin her iterasyonunda hataları iyileştirmek için gradyanları kullanır. Bunu 2. Dereceden Taylor Açılımı ile sağlar (4.23).

$$Obj^{(t)} = \sum_{i=1}^n \left[g_i F_t + \frac{1}{2} h_i b_i^2 \right] + \Omega(F_t) \quad (4.23)$$

Bu eşitlikte g_i kayıp fonksiyonunun 1. Türevi olan gradyanı, h_i ise kayıp fonksiyonunun 2. Türevi olan hessian'ı ifade eder.

4.5.6 CatBoost

CatBoost, Yandex şirketi tarafından geliştirilmiş açık kaynaklı, gradyan arttırma tabanlı bir makine öğrenmesi algoritmasıdır. Gradyan arttırma algoritmasının performansını arttırmak için XGBoost, LigthGBM gibi diğer gradyan arttırma algoritmalarına alternatif olarak olarak sunulmuştur. [73], çalışmasıyla tanıtılmış olan bu algoritma, Category (Kategori) ve Boosting (Arttırma) kelimelerinin birleştirilmesiyle adlandırılmıştır.

Bu algoritmayı diğer gradyan arttırma algoritmalarından ayıran iki önemli özellik bulunmaktadır. Bunlardan birincisi veri setinde bulunan kategorik verileri başarılı bir şekilde ele almasıdır. Bu algoritma kategorik verilerle veri ön işlemeye gerek olmadan eğitim sırasında ilgilenir. Algoritmanın bir diğer avantajı da ağaç yapısını seçerken yaprak değerlerini hesaplamak için kendine has bir şema kullanmasıdır, bu da aşırı uyumu azaltmaya yardımcı olur [74].

Yüksek öğrenme hızı, kategorik, sayısal ve metinsel verilerle bir arada çalışabilmesi ve görselleştirme seçenekleri CatBoost'u farklı gradyan arttırma algoritmalarından ayıran diğer özellikleridir.

CatBoost, denetimli makine öğrenmesi için 'Sıralı hedef istatistikleri' ve 'Sıralı artırım' adları verilen yenilikleri getirmiş açık kaynaklı bir gradyan destekli karar ağacı uygulamasıdır [75]. CatBoost one-hot kodlama kullanmadığında kategorik değişkenleri kodlamak için 'Sıralı hedef istatistiği' adı verilen bir teknik kullanmaktadır. Hedef istatistik, bir veri kümesindeki kategorik bir özneliğin belirli değerleriyle ilişkili temel gerçek çıktı değerlerinden hesaplanan değeridir. Bu algoritma kategorik değişkenlerle başa çıkmak için bir özelliğin kategorik değerlerini bir hedef istatistiği ile değiştirir.

Sıralı hedef istatistiği hesaplamasındaki en önemli husus veriyi eğitim ve test veri seti olarak ayırmaktır. Elimizde $S = (a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)$ gözlemlerinden oluşan bir veri kümesi olduğunu varsayalım; burada $a_n = a_{n1}, a_{n2}, \dots, a_{nm}$

gibi sayısal ve kategorik olmak üzere m özellikten oluşan bir vektör ve b_n etiket değerleridir. Bu veri kümesindeki bazı veriler $L(\theta)$ kayıp fonksiyonunu en aza indirmek amacıyla eğitim için, bazı veriler de test için kullanılır. CatBoost'un eğitim verileri S öğelerine rastgele bir permütasyon σ ile keyfi bir sıralama ile seçilir.

CatBoost'un kategorik değişkenleri nasıl kodladığına dair bir diğer önemli husus Iverson parantezleridir. $[x = y]$ parantezi, $x = y$ olduğunda 1, aksi halde 0 değerine sahip olur. Bu fonksiyon CatBoost'un kategorik bir özelliği sayısal bir değere dönüştürmek için uyguladığı eşitlikte önemli bir rol oynar.

CatBoost basit olarak kategorik veriyi (4.24) eşitliği ile sayısal değere dönüştürür:

$$\hat{a}_k^i = \frac{\sum_{a_j \in S_k} [a_j^i = a_k^i] \cdot b_j + \alpha p}{\sum_{a_j \in S_k} [a_j^i = a_k^i] + \alpha} \quad (4.24)$$

Bu eşitlikte:

- a_j^i ve a_k^i sırasıyla j. ve k. gözlemlerdeki i. kategorik özelliğin değerleri,
- S_k k. gözlem için permütasyona göre sıralanmış gözlemler kümesi,
- b_j j. gözlemin hedef değişkeni,
- α sıfırdan büyük düzgünleştirme parametresi,
- p veri kümesindeki etiketin ortalama değeridir.

4.6 Model Değerleme Kriterleri

Makine öğrenmesi algoritmalarının tahmin problemlerinde kullanılmasının yaygınlaşmasıyla birlikte gayrimenkul değerleme alanında makine öğrenmesi modellerinin kullanılmasına ilişkin literatürde çok sayıda örnek bulunmaktadır. Bu kulvardaki makine öğrenmesi yaklaşımları değerlendirme sürecini kısaltmış, değerlemenin güvenilirliğini arttırmıştır. Modellerin gelişmesiyle birlikte gayrimenkul değerlerini tahmin etmek için gittikçe artan sayıda algoritma ortaya çıkarken, açık kaynak kodlu yazılımlar bu algoritmaların kullanımını kolaylaştırmaktadır. Kullanılan bu modellerin performansını doğru bir şekilde değerlendirmek, en uygun modeli seçmeye ve bu modeller sayesinde doğru, güvenilir bir gayrimenkul değerlemesine yardımcı olur. Birden fazla model arasından en uygun tahmin modelinin seçilmesi ve performansının ölçülmesi gayrimenkul değerlemesinin bu alandaki önemli unsurlarından biridir.

Makine öğrenmesi modelleri üzerinde yapılan çalışmalarda kullanılan en yaygın metrikler şu şekildedir:

- Ortalama Kök Karesel Hata (Root Mean Square Error (RMSE))
- Ortalama Mutlak Hata (Mean Absolute Error (MAE))
- Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error (MAPE))
- Belirleme Katsayısı (Coefficient of Determination – R-square (R^2)) [76]; [77]).

Ortalama kök karesel hata, tahmin edilen değerler ile gerçek değerler arasındaki farkın karelerinin ortalamasının kareköküyle hesaplanan bir hata metriğidir. Ortalama kök karesel hatanın gözlem sayısı, a_i gerçek değer ve $\hat{a}_i$ tahmin edilen değer olmak üzere (4.25)'da eşitliği verilmiştir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \hat{a}_i)^2} \quad (4.25)$$

Ortalama mutlak hata, tahmin edilen değer ile gerçek değer arasındaki mutlak farkların ortalamasıdır. Ortalama mutlak hata n gözlem sayısı, a_i gerçek değer ve $\hat{a}_i$ tahmin edilen değer olmak üzere (4.26) eşitliğinde gösterilmiştir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |a_i - \hat{a}_i| \quad (4.26)$$

Ortalama mutlak yüzde hata, mutlak hataların gerçek değerlerin mutlak değerine oranlarının yüzdesidir. Ortalama mutlak yüzde hata, n gözlem sayısı, a_i gerçek değer ve $\hat{a}_i$ tahmin edilen değer olmak üzere (4.27) eşitliğinde gösterilmiştir.

$$MAPE = 100 \frac{1}{n} \sum_{i=1}^n \left| \frac{a_i - \hat{a}_i}{a_i} \right| \quad (4.27)$$

Belirleme katsayısı, istatistik ve regresyon analizlerinde bir modelin bağımsız değişkenlerinin bağımlı değişkendeki varyansı ne kadar iyi açıkladığını gösteren bir ölçüdür. Belirleme katsayısının hesabı n gözlem sayısı, a_i gerçek değer, $\hat{a}_i$ tahmin edilen değer ve $\bar{a}$ gerçek değerlerin ortalaması olmak üzere (4.28) eşitliğinde gösterilmiştir.

$$R^2 = 1 - \frac{\sum_{i=1}^n (a_i - \hat{a}_i)^2}{\sum_{i=1}^n (a_i - \bar{a})^2} \quad (4.28)$$

5

UYGULAMA

Uluslararası Değerlendirme Görevlileri Birliği'ne göre toplu değerlendirme, ortak veriler, standartlaştırılmış yöntemler ve istatistiksel testler kullanarak belirli bir tarihten itibaren belirli sayıdaki gayrimenkulün değerinin belirlenmesi sürecidir [78]. Bu çalışmada çalışma alanı olan İstanbul'un Bayrampaşa ilçesindeki 2023 yılının Mayıs ve Haziran aylarına ait 1267 adet satılık konut verisine bazı istatistik hesaplamalar ve veri ön işleme adımları uygulanarak 4. Bölümde açıklanan Makine Öğrenmesi algoritmaları uygulanmıştır. Ayrıca verilerden coğrafi referanslı öznitelikler çıkarılarak coğrafi referanslı özniteliklerin tahmin doğruluğundaki önemine değinilmiştir. Böylece gayrimenkullerin toplu değerlemesi üzerine yapılan bu çalışmada makine öğrenmesi tekniklerinin kullanılabilirliği tartışılmıştır.

5.1 Çalışma Alanı

Çalışma alanımız İstanbul ilinin Avrupa yakasında yer alan bir ilçesi olan Bayrampaşa'dır. Bayrampaşa 11 mahallesi (Altıntepe, Cevatpaşa, İsmetpaşa, Kartaltepe, Kocatepe, Muratpaşa, Orta, Terazidere, Vatan, Yıldırım, Yenidoğan) ve yaklaşık 280.000 yerleşik nüfusuyla İstanbul'un merkezinde bulunan en prestijli ilçelerinden biridir.

15 Temmuz Şehitler İstanbul Otogarı, Forum İstanbul Alışveriş Merkezi, İstanbul Büyükşehir Belediyesi Meyve-Sebze Hali, M1 metro hattı ve T4 tramvay hattı gibi iş, eğitim, sağlık, ticaret, kültür, alışveriş, spor ve önemli ulaşım noktalarını barındırmasından ve ilçe sınırının kuzeyinden TEM (E-80) ile ortasından O-3 bağlantı yollarının geçmesinden dolayı günlük yaşamda nüfusu 1.000.000'u aşan bu ilçe, doğusunda Eyüpsultan, batısında Esenler, güneyinde Zeytinburnu ve kuzeydoğusunda Gaziosmanpaşa ile komşudur. Eyüpsultan'ın bir semti olan Bayrampaşa 1990 yılında ilçe statüsü kazanmıştır [79]. 45° 05 Enlem, 28° 90 Boylamında yer alan Bayrampaşa, yaklaşık 9,62 km² yüzölçüme sahiptir (Bkz. Harita 5.1).

Birçok önemli noktayı bünyesinde bulundurması, hem TEM (E-80) hem de O-3 bağlantı yollarının sınırları içinden geçmesi, hem de konum olarak İstanbul'un merkezine yakın olması Bayrampaşa'yı bir çekim merkezi haline getirmiş ve konuta olan talebi arttırmıştır, bu sebeple çalışma alanı olarak Bayrampaşa seçilmiştir. Çalışmamızdaki veri setinde Bayrampaşa'nın 11 mahallesinden de satılık konut ilanları bulunmaktadır. Satılık konut ilanlarının konumsal dağılımı (Bkz. Harita 5.2) ArcGIS Pro 3.1.0 yazılımı kullanılarak görselleştirilmiştir.

5.2 Veri Seti

5.2.1 Veri Seti Özellikleri

Veri setimiz Bayrampaşa ilçesine ait 2023 yılının Mayıs ve Haziran aylarındaki konut satış ilanlarından 1267 örnekleme içermektedir. Veri setinde satılık gayrimenkullerin yapısal, mülki ve coğrafi referanslı özelliklerini içeren 17 adet bağımsız değişken, 1 adet bağımlı değişken bulunmaktadır. Bağımsız değişkenler; satılan dairenin brüt metrekaresini temsil eden 'm² (Brüt)', net metrekaresini temsil eden 'm² (Net)', 'Oda Sayısı', 'Bina Yaşı', bulunduğu katı temsil eden 'Kat', binadaki toplam kat sayısını veren 'Kat Sayısı', 'Isıtma Türü', 'Banyo Sayısı', 'Balkon', 'Eşya', 'Kullanım Durumu', 'Site', 'Aidat (TL)', krediye uygunluk bilgisini içeren 'Kredi', tapu türünü belirten 'Tapu' değişkenleri ile coğrafi referanslar içeren 'Enlem' ve 'Boylam' değişkenleridir. Bağımlı değişken ise gayrimenkulün satış değerini Türk Lirası cinsinden belirten 'Fiyat (TL)' değişkenidir. Çalışmadan kullanılan veri setinin ilk 5 satırı Tablo 5.1'de, veri setine ait değişkenlerin özellikleri Tablo 5.2'de gösterilmiştir. Bağımlı değişken Fiyat değişkenine ait özet istatistik ise Tablo 5.3'te verilmiştir.

Tablo 5.1 Veri setinin ilk 5 satırı

m ² (Brüt)	m ² (Net)	Oda Sayısı	Bina Yaşı	Kat	Kat Sayısı	Isıtma Türü	Banyo Sayısı	Balkon	Eşya	Kullanım Durumu	Site	Aidat (TL)	Kredi	Tapu	Enlem	Boylam	Fiyat (TL)
81	75	2+1	5-10 Arası	Bodrum Kat	4	Kombi (Doğalgaz)	1	Yok	Hayır	Kiracılı	Hayır	-	Evet	Kat Mülkiyeti	41.0418356	28.9057535	1.800.000
200	180	4+2	5-10 Arası	3	4	Kombi (Doğalgaz)	2	Var	Hayır	Boş	Hayır	-	Evet	Kat Mülkiyeti	41.0569174	28.8993531	4.600.000
115	100	2+2	0	4	5	Yerden Isıtma	2	Var	Hayır	Boş	Hayır	-	Evet	Kat Mülkiyeti	41.0386109	28.9033629	3.700.000
120	115	3+1	4	3	5	Kombi (Doğalgaz)	2	Var	Hayır	Mülk Sahibi	Hayır	-	Evet	Kat Mülkiyeti	41.0506417	28.9033724	3.650.000
100	90	2+1	5-10 Arası	2	5	Kombi (Doğalgaz)	2	Var	-	Mülk Sahibi	Hayır	150	-	Kat İrtifakı	41.0538852	28.9013625	4.000.000

Tablo 5.1 ve Tablo 5.2'ye bakıldığında veri setimizde her türden ölçekte veri olduğu görülmektedir. İlk etapta Türkçe karakterlerin İngilizce karakterlere ve kategorik değişkenlerin makine öğrenmesi algoritmalarına hazır hale getirilmesi için sayısal değişkenlere çevrilmesi gerektiği göze çarpmaktadır. Ayrıca Isıtma Türü, Eşya, Aidat (TL), Kredi ve Tapu değişkenlerinde eksik veriler bulunmaktadır.

Tablo 5.3'teki veri setine ait özet istatistik incelendiğinde satış fiyatlarının ortalaması ve medyanı arasındaki farkın fazla olması veri setinde aykırı değerlerin olduğunu göstermektedir. Aynı şekilde Tablo 5.4'deki fiyat değişkenine ait frekans tablosu da incelendiğinde aykırı değerlerin olduğu görülmektedir.

Tablo 5.2 Değişken özellikleri

Değişken Adı	Gözlem Sayısı	Veri Tipi
Brüt metrekare	1267	Sayısal
Net metrekare	1267	Sayısal
Oda Sayısı	1267	Kategorik
Bina Yaşı	1267	Kategorik
Kat	1267	Kategorik
Kat Sayısı	1267	Sayısal
Isıtma Türü	1264	Kategorik
Banyo Sayısı	1267	Sayısal
Balkon	1267	Kategorik
Eşya	1218	Kategorik
Kullanım Durumu	1267	Kategorik
Site	1267	Kategorik
Aidat (TL)	408	Sayısal
Kredi	1230	Kategorik
Tapu	1247	Kategorik
Enlem	1267	Sayısal
Boylam	1267	Sayısal
Fiyat (TL)	1267	Sayısal

Tablo 5.3 Fiyat değişkeninin özet istatistiği

Gözlem Sayısı	1267
Maksimum Fiyat (TL)	25.750.000
Minimum Fiyat (TL)	870.000
Değişim Aralığı (TL)	24.880.000
Ortalama Fiyat (TL)	3.861.310
Standart Sapma (TL)	2.300.761
Medyan Fiyat (TL)	3.300.000
Ortalama Net metrekare Fiyat (TL)	35.551

5.2.2 Veri Önileme Uygulaması

Çalışmamızın bu bölümünde Jupyter Notebook açık kaynak kodlu yazılımında Python programlama dilinin 3.11.5 versiyonunun Pandas, Numpy, Matplotlib ve Seaborn kütüphaneleri kullanılmıştır.

Tablo 5.4 incelendiğinde ‘Aidat (TL)’ değişkeninin yaklaşık %68’i eksik veri içermektedir. Bu sebeple ‘Aidat (TL)’ değişkeni veri setinden çıkarılmıştır. ‘Isıtma Türü’ değişkenindeki 3, ‘Kredi’ değişkenindeki 37 ve ‘Tapu’ değişkenindeki 20 eksik verinin gözlemleri veri setinden silinmiştir. ‘Eşya’ değişkenindeki eksik 49 gözlem ise değişkenin en yüksek frekanslı seçeneği olan ‘Hayır’ seçeneği ile doldurulmuştur. Bu veri temizleme işlemlerinden sonra 1 değişken veri setinden çıkarılmış, 1267 gözlemden 56’sı silinmiş geriye eksiksiz 1211 gözlem kalmıştır.

Tablo 5.4 Fiyat değişkeninin frekans tablosu

Fiyat Aralığı (TL)	Frekans
0-5.000.000	990
5.000.000-10.000.000	253
10.000.000-15.000.000	18
15.000.000 ve üzeri	6

Tablo 5.5 de ki ‘Kat’ kategorik değişkenin frekans bilgileri incelendiğinde fazla sayıda kategoriye sahip olduğu ancak buradaki kategori sayısının azaltılabileceği görülmektedir. Bayrampaşa ilçesi genelinde ‘Giriş Altı Kot 1’ ve ‘Bahçe Katı’ seviyesindeki daireler ‘Bodrum Kat’ seviyesinde, ‘Giriş Kat’ ve ‘Yüksek Giriş’ seviyesindeki dairenin ‘Zemin Kat’ seviyesinde olduğu bilinmektedir. Bu sebeple bu değişkenleri birleştirmek kategori sayısını azaltarak veriyi daha kaliteli hale getirecektir. Yukarıda belirtilen kategoriler kendi aralarında birleştirilmiş ve kategori sayısı azaltılmıştır. ‘Giriş Altı Kot 2’ kategorisinin ismi de ‘2. Bodrum Kat’ olarak güncellenmiştir.

Tablo 5.5 Kat değişkeninin frekans tablosu

Kat	Frekans	Kat	Frekans
Giriş Altı Kot 1	1	8	4
Giriş Altı Kot 2	1	9	1
Bodrum Kat	37	10	3
Bahçe Katı	125	11	5
Giriş Kat	19	12	7
Zemin Kat	14	13	5
Yüksek Giriş	109	14	4
1	192	15	5
2	191	16	1
3	246	17	1
4	136	18	3
5	68	19	1
Çatı Katı	20	22	1
6	5	23	1
7	5		

Tablo 5.1 ve Tablo 5.2’ye bakıldığında ‘Oda Sayısı’ değişkeninin kategorik bir veri olduğu görülmektedir. Bu değişkeni oluşturan kategorilerin aslında sayısal bilgilerden oluşması sebebiyle ‘Oda Sayısı’ değişkeni (örneğin ‘2+1’ kategorisi 3, ‘4+2’ kategorisi 6 olacak şekilde) güncellenmiştir. Ancak burada dikkat edilmesi gereken bir husus bulunmaktadır. CatBoost algoritması için veri setinden bir kopya oluşturulmuştur. CatBoost kategorik verilerle çalışabildiğinden bu algoritma için hazırlanan veri setinde bu dönüşüm uygulanmamıştır. Diğer algoritmalar için

hazırlanan veri setinde ise değişken sayısal veri tipine çevrilmiştir. Güncellenen ‘Oda Sayısı’ değişkenine ait frekans bilgileri Tablo 5.6 da verilmiştir.

Tablo 5.6 Oda Sayısı değişkenin güncel frekans tablosu

Oda Sayısı	Frekans	Oda Sayısı	Frekans
1	1	4.5	1
1.5	3	5	90
2	76	6	132
2.5	4	7	35
3	574	8	17
3.5	3	9	1
4	274		

Veri setinde bulunan Oda Sayısı, Bina Yaşı, Kat, Isıtma Türü, Balkon, Eşya, Site, Kredi ve Tapu değişkenleri ordinal kategorik değişkenlerdir. CatBoost için hazırlanan veri setinde ordinal kategorik değişkenlerin tamamına ordinal tanımlamaları yapılmıştır. Diğer modeller için hazırlanan veri setinde Oda Sayısı değişkeni hariç diğer değişkenlerin ordinal tanımlamaları, ordinal kategorik olarak tanımlanan değişkenlere label encoding dönüşümü, Kullanım Durumu değişkeni nominal kategorik değişken olduğundan sadece bu değişkene one-hot encoding dönüşümü yapılmıştır.

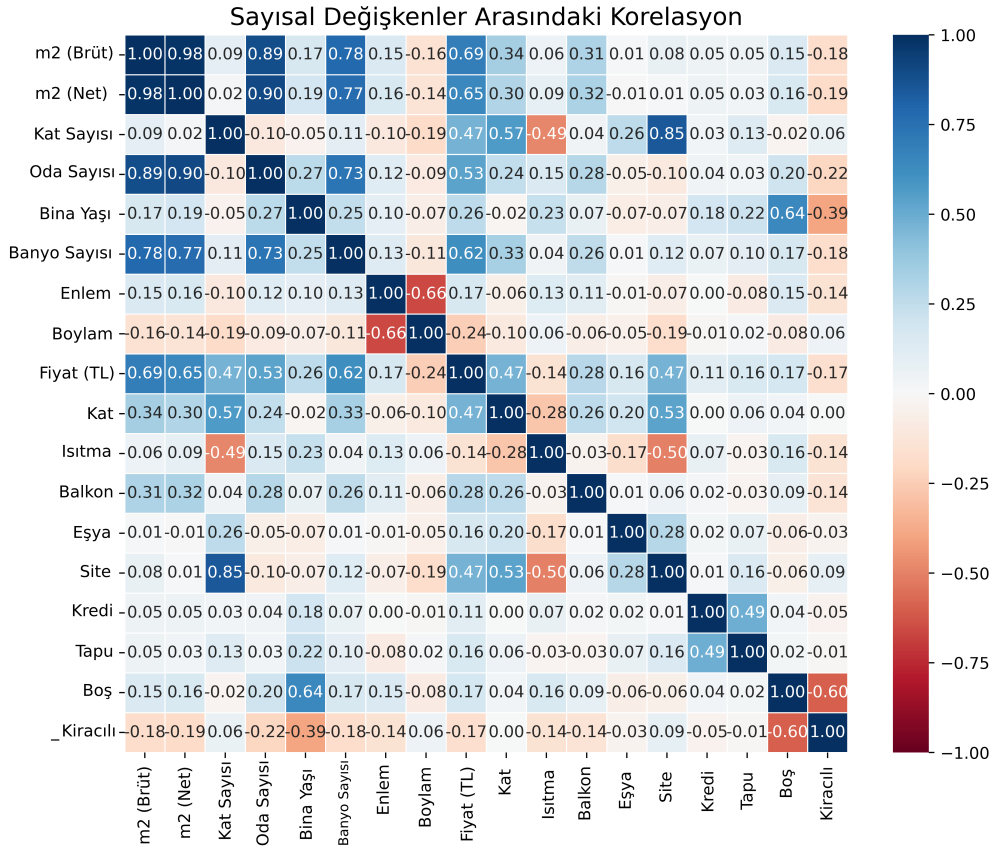
Bir sonraki aşamada yerel aykırı faktör algoritması ile aykırı değer analizi yapılmıştır. Aykırı değerlere sahip 75 gözlem her iki veri setinden de çıkarılmış ve veri setlerinde 1136 adet gözlem kalmıştır.

Sayısal değişkenler arasındaki ilişki Pearson Korelasyon Analizi ile matrise dönüştürülmüştür (Şekil 5.1). Matrise bakıldığında ‘m² (Brüt)’ ve ‘m² (Net)’ değişkenleri arasında %98 gibi bir korelasyon değeri göze çarpmaktadır. Aralarında çok yüksek korelasyonlu değişkenlerin olması veri setinin kalitesini düşüreceğinden ‘m² (Brüt)’ değişkeni veri setinden çıkarılmıştır. Eksik veri analizi ve değişken çıkarma işlemleri sonucu veri seti Tablo 5.7’deki son halini almıştır.

Son olarak her iki veri setinin birer kopyası yapılmış ve kopya veri setlerinden coğrafi referanslar olan Enlem ve Boylam değişkenleri çıkarılmıştır.

5.3 Bulgular

Çalışmamızın bu bölümünde Spyder 5.5.1 yazılımında Python programlama dilinin 3.11.5 versiyonunun Pandas, Numpy ve Scikit-learn kütüphaneleri ile ArcGIS Pro 3.1.0 yazılımı kullanılmıştır.



Şekil 5.1 Korelasyon Matrisi

Veri ön işleme aşamasından geçen veri setleri %33'ü test verisi olacak şekilde AdaBoost, CatBoost, Rastgele Orman, Çoklu Doğrusal Regresyon ve XGBoost algoritmaları ile gerekli hiperparametre ayarlamaları yapılarak eğitilmiştir. Coğrafi referanslı ve coğrafi referanssız veri setleri ile kurulan modellerin, RMSE, MAE, MAPE ve R^2 metrikleri hesaplanmıştır. Sonuçlar Tablo 5.8'de gösterilmiştir.

Sonuçlara bakıldığında genel olarak modellerde coğrafi referanslar kullanıldığında hata oranları (RMSE, MAE ve MAPE) düşmekte, R^2 değeri artmaktadır. Bu durum coğrafi referansların model performansını arttırdığını göstermektedir. CatBoost ve Rastgele Orman, coğrafi referanslı verilerle en düşük hata (RMSE, MAE ve MAPE), en yüksek R^2 değerlerine sahip modeller olmuştur. CatBoost coğrafi referanslı veride 719.002,18 TL RMSE ve 0,86 R^2 değerine sahiptir. Rastgele Orman modeli ise coğrafi referanslı veride 719.986,96 TL RMSE ve 0,86 R^2 değerleri ile CatBoost modeline yakın bir performans göstermiştir. AdaBoost ve Çoklu Doğrusal Regresyon modelleri diğer modellere göre daha yüksek hata metriklerine (RMSE, MAE ve MAPE) ve daha düşük R^2 değerlerine sahip modeller olmuştur. XGBoost modelinin performansı 745.944,49 TL RMSE ve

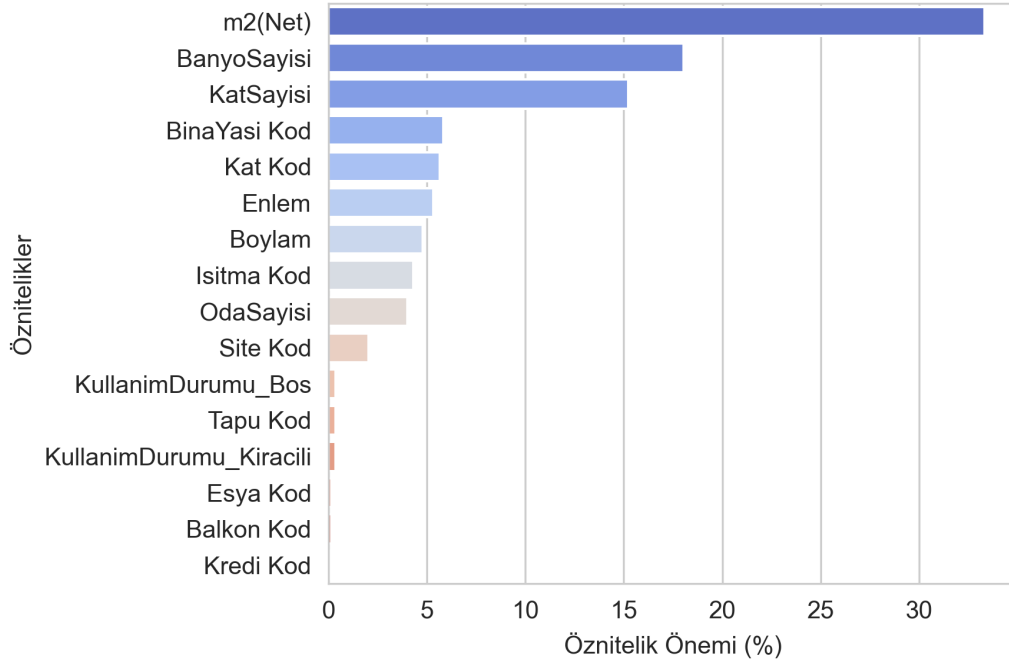
Tablo 5.7 Veri önışleme sonrası güncel veri seti

Değişken Adı	Gözlem Sayısı	Veri Tipi
Net metrekare	1136	Sayısal
Oda Sayısı	1136	Sayısal
Bina Yaşı	1136	Sayısal
Kat	1136	Sayısal
Kat Sayısı	1136	Sayısal
Isıtma Türü	1136	Sayısal
Banyo Sayısı	1136	Sayısal
Balkon	1136	İkilik
Eşya	1136	İkilik
Kullanım Durumu (Boş)	1136	İkilik
Kullanım Durumu (Kiracılı)	1136	İkilik
Site	1136	İkilik
Kredi	1136	İkilik
Tapu	1136	Sayısal
Enlem	1136	Sayısal
Boylam	1136	Sayısal
Fiyat (TL)	1136	Sayısal

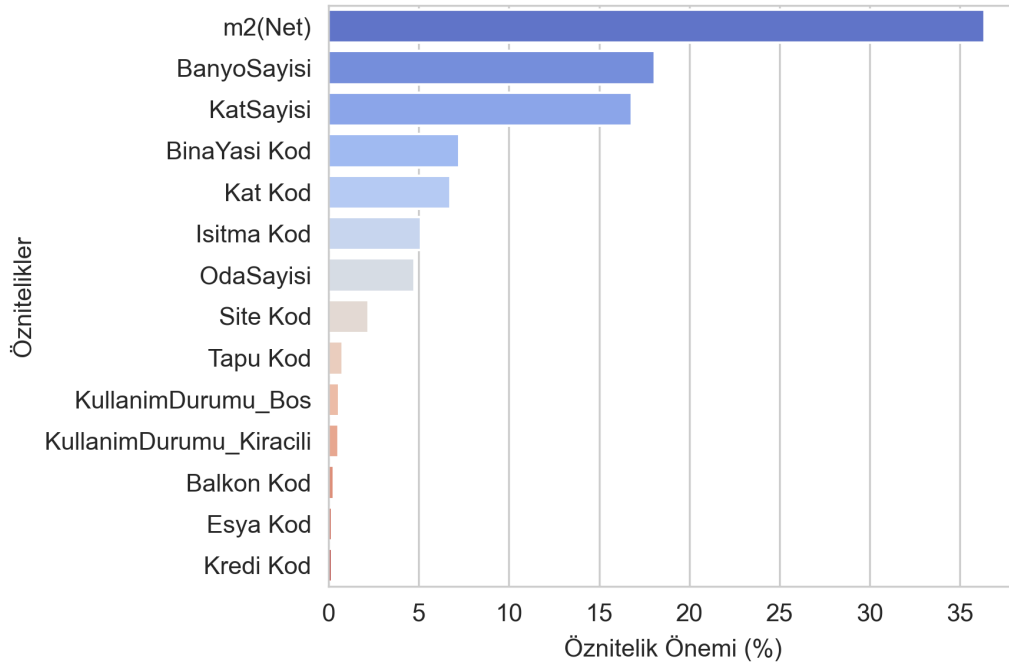
0,85 R^2 değerleri ile CatBoost ve Rastgele Orman modellerine yakın sonuçlar vermiştir. Coğrafi referanslı değişkenler olan Enlem ve Boylam değerlerinin bütün algoritmalarda tahmin doğruluğunu artırdığı görülmektedir. Ancak bu iyileşme her modelde aynı seviyede olmamıştır. Coğrafi referanslar özellikle XGBoost başta olmak üzere CatBoost ve Rastgele Orman modellerinde tahmin doğruluğunu önemli ölçüde arttırmıştır. XGBoost, CatBoost ve Rastgele Orman algoritmalarına ait fiyat tahmin hatalarının standart sapmalarının harita üzerindeki dağılımı sırasıyla Harita 5.3, Harita 5.4 ve Harita 5.5'te gösterilmiştir.

Ayrıca Rastgele Orman ve CatBoost algoritmalarıyla yapılan modellerde özniteliklerin önemi hesaplanmıştır. Coğrafi referanslı ve coğrafi referanssız Rastgele Orman ve CatBoost modellerindeki hangi özniteliklerin model üzerinde ne kadar etkisi olduğunu gösterir grafikler Şekil 5.2, Şekil 5.3, Şekil 5.4 ve Şekil 5.5 'de verilmiştir. Grafiklere bakıldığında coğrafi referanslı özniteliklerin bina yaşı, satılık konutun bulunduğu kat, ısıtma türü, toplam oda sayısı gibi diğer öznitelikler kadar önemli olduğu görülmektedir. Ayrıca coğrafi referanslı özniteliklerin modele eklenmesi satılık konutun net alanının modele olan etkisini azaltmıştır.

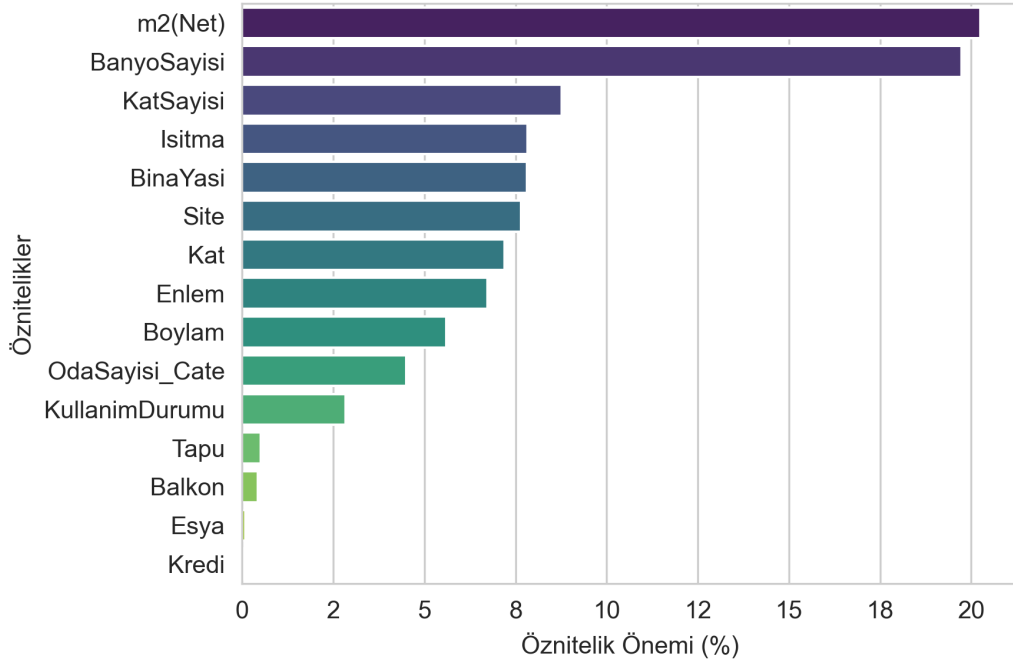
Bu modeller dışında Tablo 5.7'de ki veri seti ile Coğrafi Ağırlık Regresyon algoritması da uygulanmıştır. Uygulanan bu Coğrafi Ağırlık Regresyon modeli 0,87 R^2 skoru vermiştir. Tahmin hatalarının standart sapmalarının harita üzerindeki dağılımı Harita 5.6'te verilmiştir.



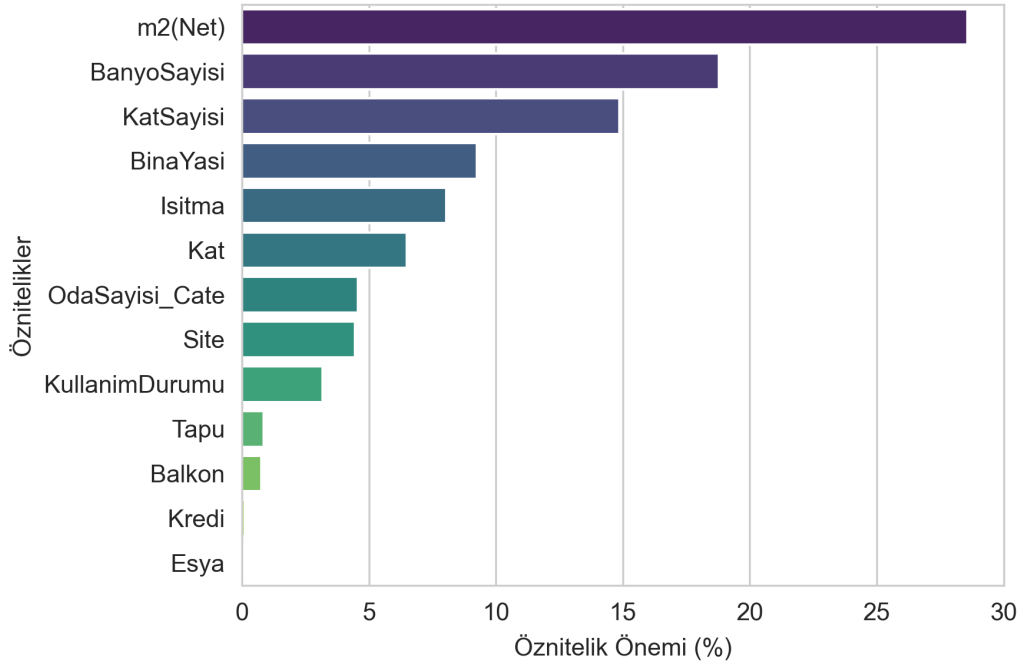
Şekil 5.2 Coğrafi referanslı Rastgele Orman modelindeki özniteliklerin önem grafiği



Şekil 5.3 Coğrafi referanssız Rastgele Orman modelindeki özniteliklerin önem grafiği



Şekil 5.4 Coğrafi referanslı CatBoost modelindeki özniteliklerin önem grafiği



Şekil 5.5 Coğrafi referanssız CatBoost modelindeki özniteliklerin önem grafiği

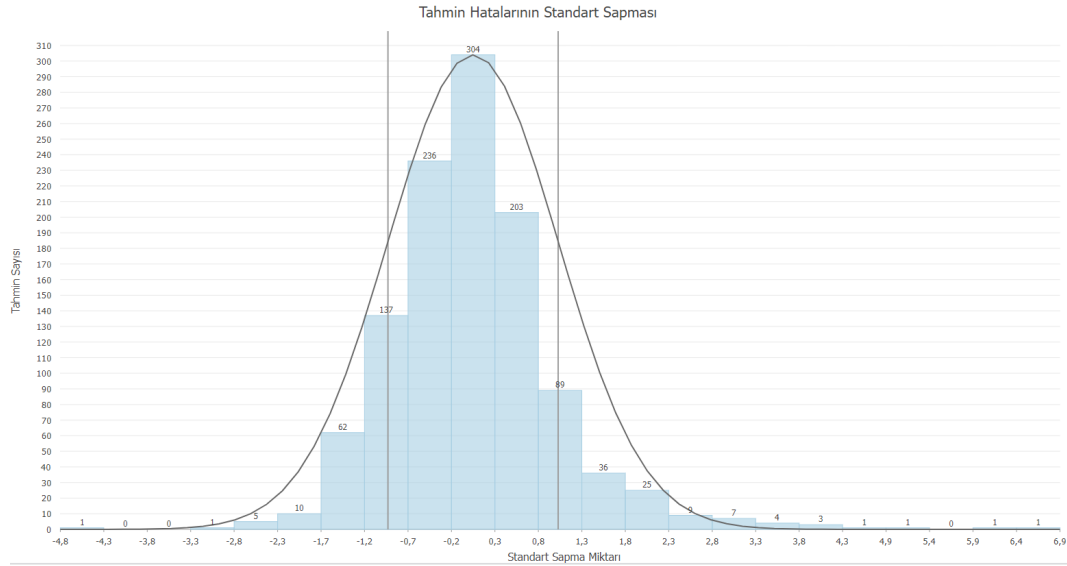
Şekil 5.6’de ki grafikte Coğrafi Ağırlıklı Regresyon modeli sonucu tahmin edilen fiyatlar ile gerçek fiyatlar arasındaki farkın standart sapma dağılımı gösterilmiştir. Bu dağılım normal dağılıma uymaktadır. Hataların büyük çoğunluğu -1 ile 1 standart hata arasında kalmaktadır. Bu sonuçlarla birlikte Coğrafi Ağırlıklı Regresyon algoritmasının da fiyat tahmin modelinde yüksek doğrulukla tahminler yapabildiği anlaşılmıştır.

Harita 5.3, Harita 5.4, Harita 5.5 ve Harita 5.6 gerçek fiyatlar ile modellerle tahmin edilen fiyatlar arasındaki farkların yani hataların standart sapmalarının dağılımını göstermektedir. Haritada renkler kırmızı tonlarında koyulaştıkça sapmalar negatif yönde artmaktadır. Negatif sapmaların olduğu ilanlarda modeller gerçek fiyatların üzerinde tahmin yapmışlardır. Renkler mavi tonlarında koyulaştıkça sapmalar pozitif yönde artmaktadır. Bu durumda da modeller gerçek fiyatlarında altında düşük tahminlerde bulunmuşlardır. Haritalar incelendiğinde sapmaların fazla olduğu bölgeler genelde aynıdır. Ayrıca haritalarda beyaz ve beyaza yakın açık renkler ağırlıktadır, bu durum modellerin fiyat tahminleri ile gerçek fiyatlar arasındaki sapmalarının sıfıra yakın veya düşük olduğunu göstermektedir.

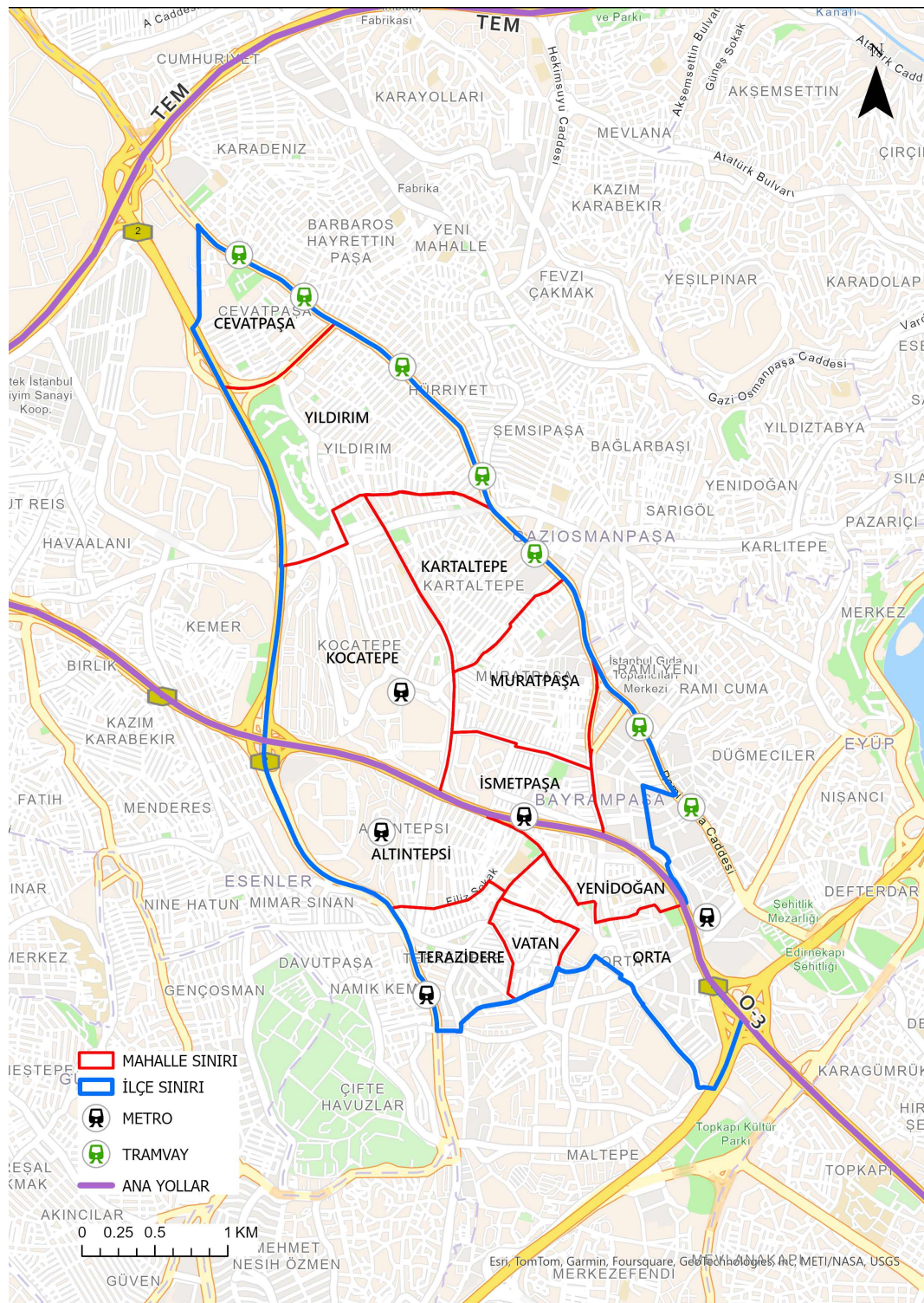
Standart sapma dağılım haritalarına bakıldığında ortak olarak Cevatpaşa Mahallesi’nin belli bir bölgesindeki ilanların $(-2,5)$ standart sapmanın altında olduğu görülmektedir. Bu durumun ortaya çıkmasında ki sebebin Cevatpaşa Mahallesi’nin o bölgesindeki sosyo-kültürel özelliğinden kaynaklandığı düşünülmektedir. Yine haritalara bakıldığında Kartaltepe, Kocatepe ve Muratpaşa Mahallelerinde de $(-2,5)$ standart sapma altında ilanların olduğu görülmektedir. Bu mahalleler O-3 ve TEM yollarına en uzak mahalleler olduğundan gerçek fiyatların düşük kalmış olabileceği tahmin edilmektedir. Haritaların ortak bir özelliği de Yıldırım Mahallesi’nde göze çarpmaktadır. Haritaların tümünde mavi tonların sayı olarak en fazla olduğu mahalle Yıldırım Mahallesi’dir. Bu mahalledeki ilanların gerçek fiyatları modellerin tahmin ettiği fiyatlardan yüksektir, bu durum yine Yıldırım Mahallesi’nin sosyo-kültürel özelliğinden kaynaklanmaktadır. Metro ve tramvay duraklarına yakın ilanlar büyük çoğunlukla beyaz ve beyaza yakın açık renklidir. Bu durum bu bölgelerde fiyatlandırma yapılırken ulaşımın göz önünde bulundurulduğu anlamına gelmektedir.

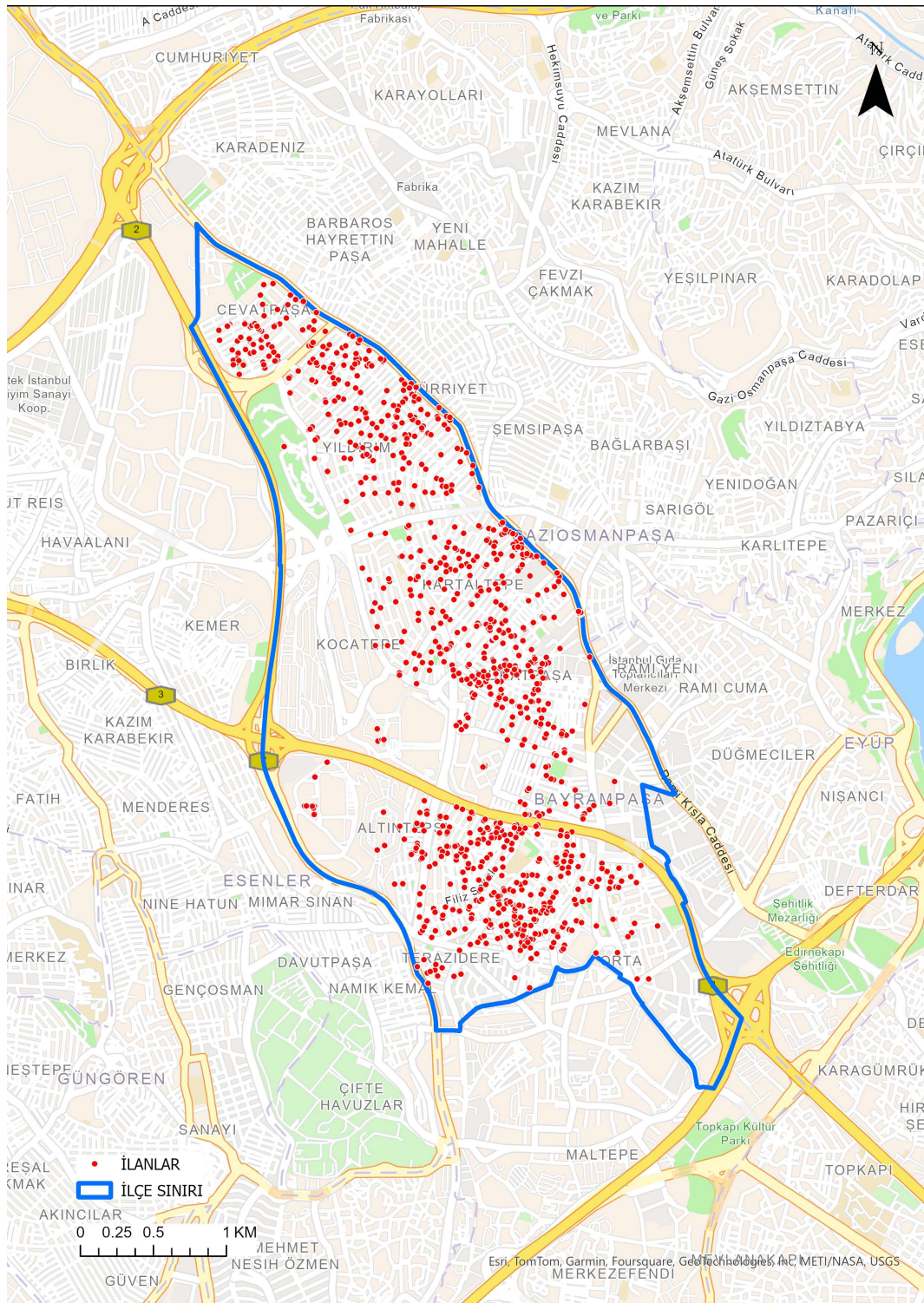
Tablo 5.8 Modellerin tahmin hataları

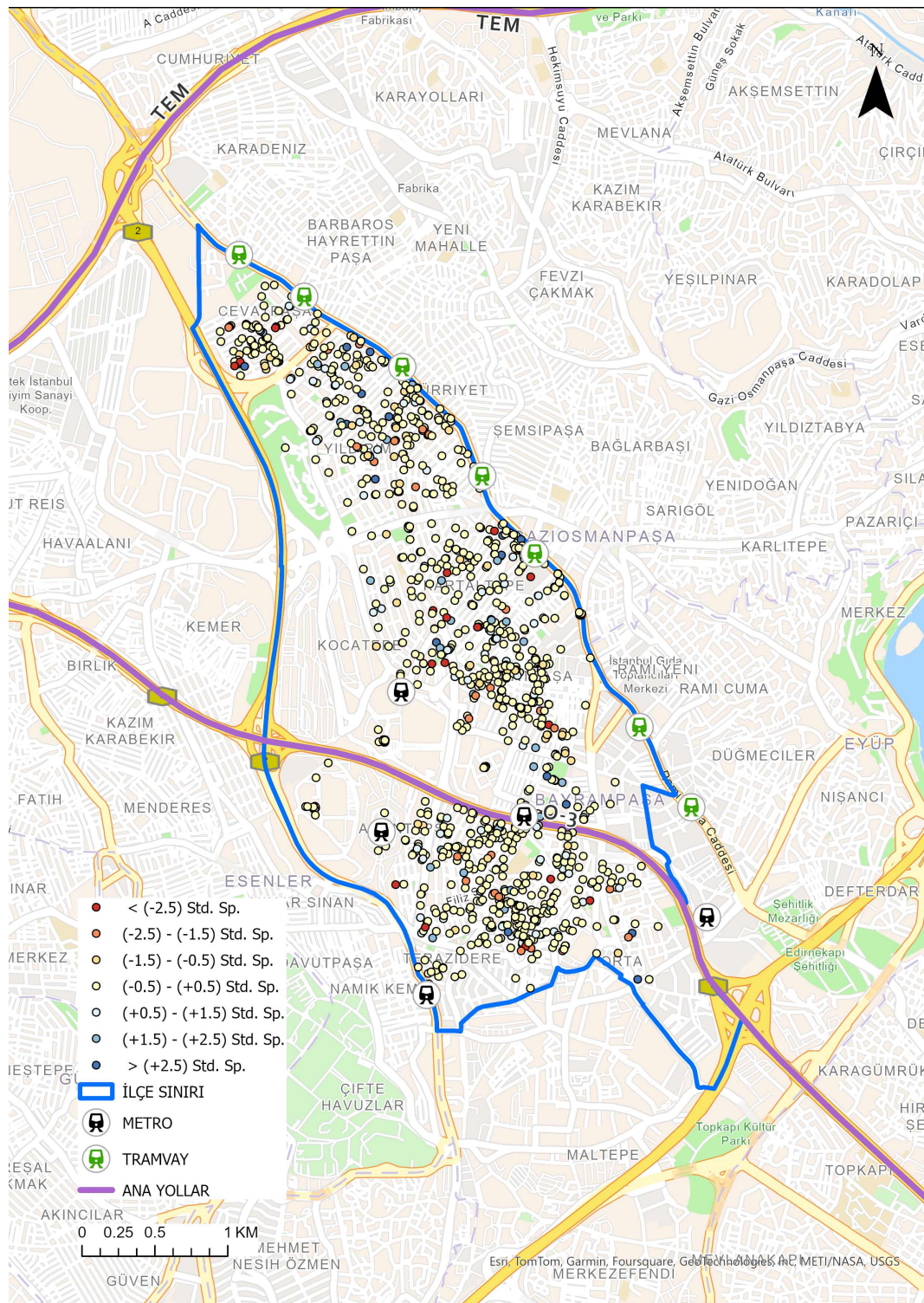
		RMSE	MAE	MAPE	R ²
AdaBoost	Koordinatlı	968.812,48	763.254,12	0,27	0,74
	Koordinatsız	1.001.430,97	795.033,75	0,27	0,73
CatBoost	Koordinatlı	719.002,18	504.953,45	0,14	0,86
	Koordinatsız	810.243,55	586.123,95	0,17	0,82
Ç. Doğrusal Reg.	Koordinatlı	939.658,28	688.014,92	0,20	0,76
	Koordinatsız	963.706,83	704.417,34	0,20	0,75
Rastgele Orman	Koordinatlı	719.986,96	510.986,57	0,14	0,86
	Koordinatsız	837.760,07	609.602,41	0,17	0,81
XGBoost	Koordinatlı	745.944,49	530.098,89	0,16	0,85
	Koordinatsız	903.641,74	652.604,23	0,18	0,78

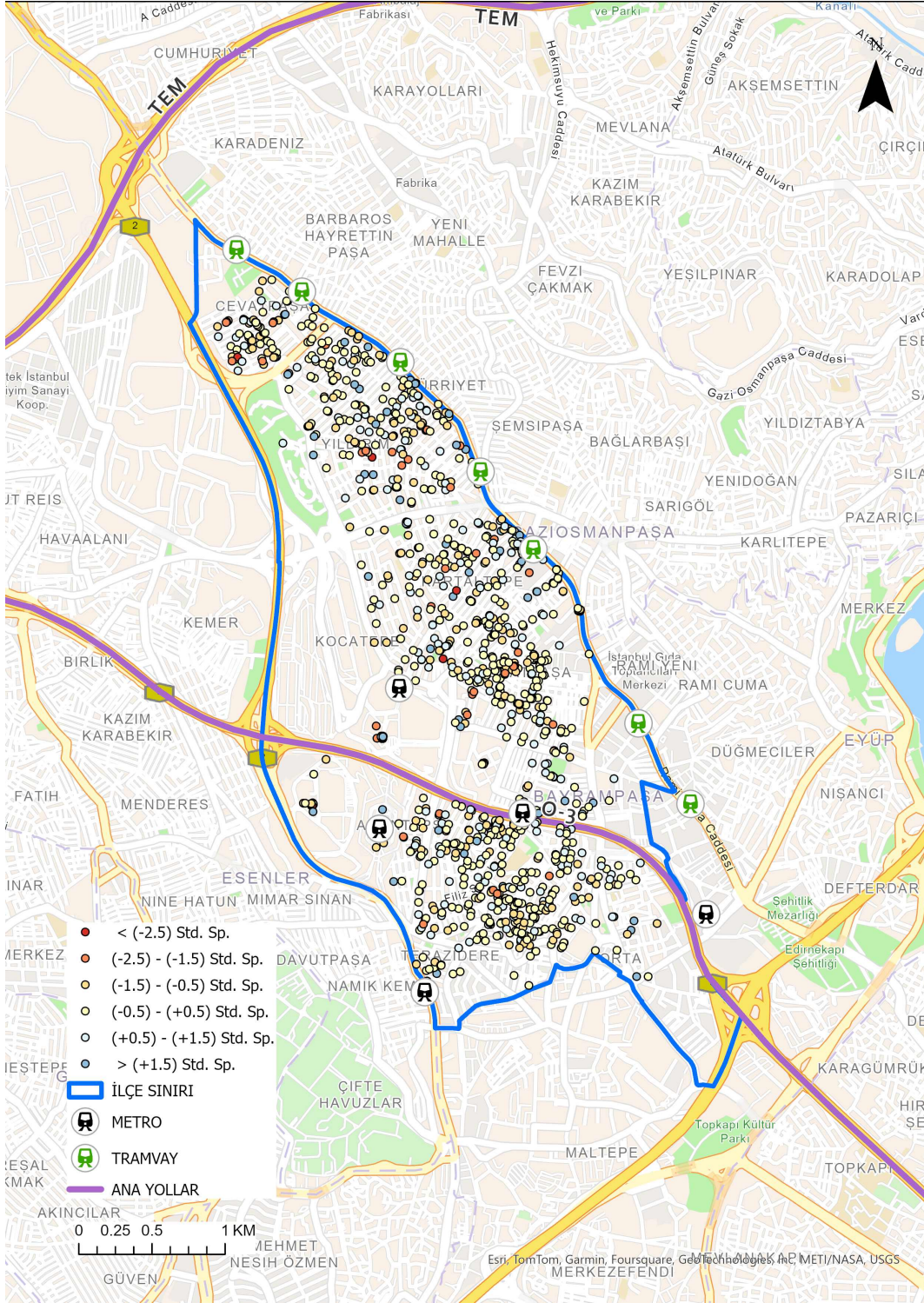


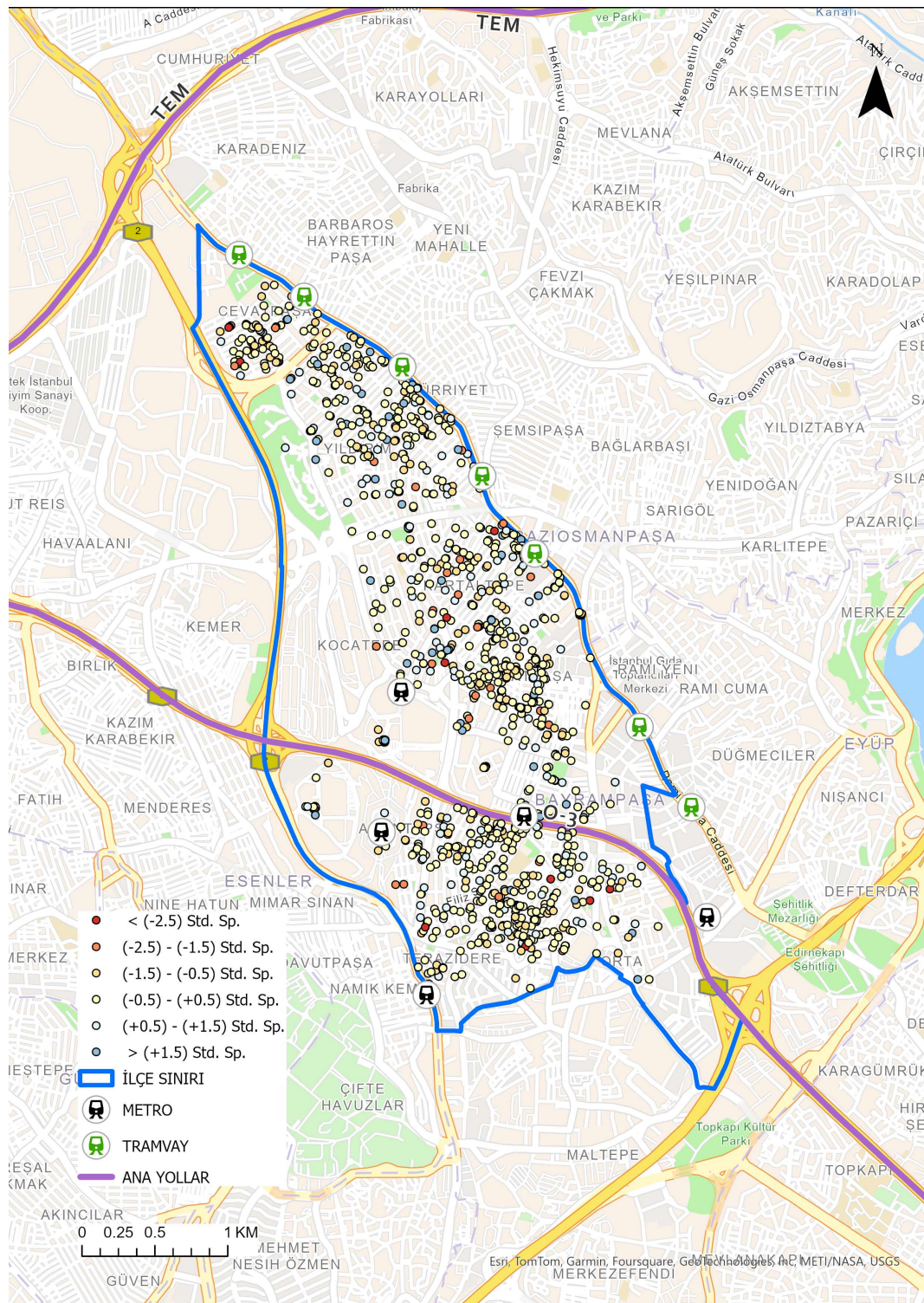
Şekil 5.6 Coğrafi ağırlıklı regresyon modelinin tahminleri ile gerçek fiyatlar arasındaki farkların standart sapma dağılım grafiği

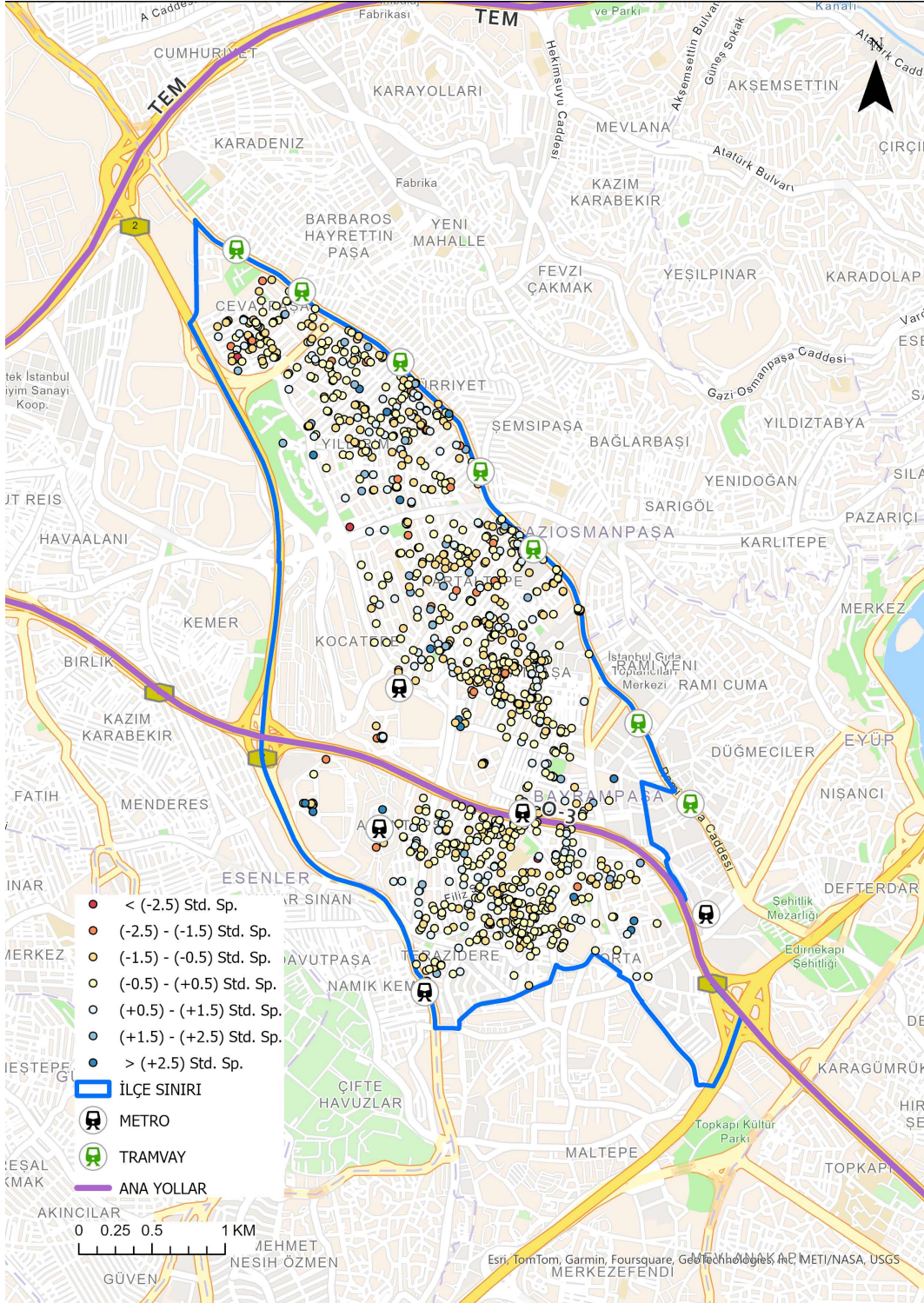












Harita 5.6 Coğrafi ağırlıklı regresyon modelinin tahminleri ile gerçek fiyatlar arasındaki farkların standart sapma dağılım haritası

Uluslararası Değerlendirme Görevlileri Birliği'ne göre toplu değerlendirme, ortak veriler, standartlaştırılmış yöntemler ve istatistiksel testler kullanarak belirli bir tarihten itibaren belirli sayıdaki gayrimenkulün değerinin belirlenmesi sürecidir [78]. Bu kapsamda çalışmamızda literatürde yer alan AdaBoost, CatBoost, Çoklu Doğrusal Regresyon, Rastgele Orman ve Coğrafi Ağırlık Regresyon algoritmalarının toplu değerlendirme problemlerindeki tahmin başarıları ve coğrafi referanslı değişkenler olan Enlem ve Boylam değerlerinin tahmin başarısına etkisi araştırılmak istenerek mevcut literatüre katkı sağlamak amaçlanmıştır.

Bu amaçla yapılan çalışmada İstanbul'un Bayrampaşa ilçesindeki 2023 yılının Mayıs ve Haziran aylarına ait 1267 adet satılık konut verisine bazı istatistik hesaplamalar ve veri ön işleme adımları uygulanarak AdaBoost, CatBoost, Çoklu Doğrusal Regresyon, Rastgele Orman, XGBoost ve Coğrafi Ağırlık Regresyon modellerine hazır hale getirilmiş ve bu algoritmalar kullanılarak tahmin modelleri oluşturulmuştur. Coğrafi Ağırlık Regresyon haricindeki diğer algoritmalarda coğrafi referanslı değişkenlerin etkileri gözlenmiştir. Çalışma sonunda CatBoost, Rastgele Orman, XGBoost ve Coğrafi Ağırlıklı Regresyon algoritmaları ile oluşturulan modellerin tahmin doğruluğunun başarılı olduğu görülmüş, coğrafi referanslı değişkenlerin tahmin modellerinin başarısını arttırdığı gözlenmiştir. Çalışmanın sınırlaması verilerin yalnızca belirli bir bölgeden temsil edilen veriler olmasıdır. Bu sınırlama elde edilen sonuçların genellenmesinin önüne geçmektedir.

Uygulama gayrimenkullerin makine öğrenmesi yöntemleri ile değerlemesinin yukarıdaki sonuçları göz önünde bulundurularak gayrimenkullerin doğru, objektif, güvenli ve hızlı bir şekilde belirlenmesinin mümkün olduğunu göstermektedir. Belediyeler ve Tapu Müdürlüklerinin değerlemeler sonucu elde edilen emlak vergisi, tapu harcı gibi ülke ekonomisine katkı sağlayan gelirlerini doğru, adaletli ve kayıpsız toplaması için bu algoritmaların bir yazılım ile kullanımı kolay bir arayüzünü yaparak kullanması önerilmektedir. Böylelikle ülke genelinde açık kaynak gayrimenkul değer haritalarının bu modeller ile üretilebileceği, bu

haritaların paylaşılması ve sigortacılık, bankacılık gibi gayrimenkul piyasasına yönelik güvenli bir veri tabanının oluşturulabileceği de düşünülmektedir. Çeşitli modellerin oluşturulabilmesi ve emlak fiyatlarının denetlenebilmesi açısından gayrimenkul satış ilan verilerinin emlak sitelerinde açık kaynak olarak paylaşılması da önerilmektedir.

Gelecek araştırmalar için daha geniş bir bölgeye ait çok daha fazla sayıda örnekleme sahip bir veri seti ile bu çalışmanın yapılması sonuçların genellenebilirliğini arttırabilir. Veri setine farklı veri önışleme yöntemleri uygulanarak modellerin başarıları test edilebilir. Ayrıca tahmin modelleri literatürde yer alan diğer regresyon algoritmaları ile kullanılarak, farklı metodolojilerin tahmin başarısı araştırılabilir.



- [1] TÜİK, *Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları, 2007-2023*. erişim adresi: <https://data.tuik.gov.tr/Bulten/Index?p=Adrese-Dayali-Nufus-Kayit-Sistemi-Sonuclari-2023-49684> (erişim tarihi 04/11/2024).
- [2] TÜİK, *Konut Satış İstatistikleri, Aralık 2023*. erişim adresi: <https://data.tuik.gov.tr/Bulten/Index?p=Konut-Sat%C4%B1%C5%9F-%C4%B0statistikleri-Aral%C4%B1k-2023-49516&dil=1> (erişim tarihi 04/11/2024).
- [3] N. Erdem, “Türkiye için bir taşınmaz değerlendirme sistemi yaklaşımı,” *Geomatik*, c. 2, no. 1, ss. 18–39, 2017.
- [4] Y. B. Önal K. Tezcan, “Gayrimenkul değerlendirme ve önemi: vergisel açıdan bir değerlendirme,” *Muhasebe ve Denetime Bakış*, no. 24, ss. 15–36, 2008.
- [5] M. Çete ve T. Yomralıoğlu, “Türkiye için bir arazi idare sistemi yaklaşımı,” *Jeodezi ve Jeoinformasyon Dergisi*, no. 100, ss. 1–43, 2008.
- [6] K. Çelik, “Gayrimenkul değerlemesinde geleneksel yöntemler,” *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, c. 12, no. 1, ss. 246–259, 2022.
- [7] T. Alkan, Y. Dokuz, A. Ecemiş, A. Bozdağ S. S. Durduran, “Using machine learning algorithms for predicting real estate values in tourism centers,” *Soft Computing*, c. 27, no. 5, ss. 2601–2613, 2023.
- [8] S. S. Bilgilioglu H. M. Yılmaz, “Comparison of different machine learning models for mass appraisal of real estate,” *Survey review*, c. 55, no. 388, ss. 32–43, 2023.
- [9] D. Tchunte S. Nyawa, “Real estate price estimation in French cities using geocoding and machine learning,” *Annals of operations research*, c. 308, no. 1, ss. 571–608, 2022.
- [10] E. A. Antipov E. B. Pokryshevskaya, “Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics,” *Expert systems with applications*, c. 39, no. 2, ss. 1772–1778, 2012.
- [11] L. Lemeš A. Akagic, “Prediction of Real Estate Market Prices with Regression Algorithms,” *International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies*, Springer, 2022, ss. 401–411.
- [12] E. Guliker, E. Folmer M. van Sinderen, “Spatial determinants of real estate appraisals in the Netherlands: A machine learning approach,” *ISPRS international journal of geo-information*, c. 11, no. 2, s. 125, 2022.

- [13] B. Park J. K. Bae, "Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data," *Expert systems with applications*, c. 42, no. 6, ss. 2928–2934, 2015.
- [14] M. Tekin I. U. Sari, "Real estate market price prediction model of Istanbul," *Real Estate Management and Valuation*, c. 30, no. 4, ss. 1–16, 2022.
- [15] J. Hong, H. Choi W.-s. Kim, "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*, c. 24, no. 3, ss. 140–152, 2020.
- [16] M. Ozdemir, K. Yıldız B. Büyüktanır, "Housing price estimation with deep learning: a case study of sakarya Turkey," *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, c. 9, no. 1, ss. 138–151, 2022.
- [17] R. Sibindi, R. W. Mwangi A. G. Waititu, "A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices," *Engineering Reports*, c. 5, no. 4, e12599, 2023.
- [18] L. El Mouna, H. Silkan, Y. Haynf, M. F. Nann S. C. Tekouabou, "A comparative study of urban house price prediction using machine learning algorithms," *E3S Web of Conferences*, EDP Sciences, c. 418, 2023, s. 03 001.
- [19] H. Sharma, H. Harsora B. Ogunleye, "An Optimal House Price Prediction Algorithm: XGBoost," *Analytics*, c. 3, no. 1, ss. 30–45, 2024.
- [20] R. M. Almejrb, O. M. Sallabi A. A. Mohamed, "Applying C atboost Regression Model for Prediction of House Prices," *2022 International Conference on Engineering & MIS (ICEMIS)*, IEEE, 2022, ss. 1–7.
- [21] Y. Wang Q. Zhao, "House Price Prediction Based on Machine Learning: A Case of King County," *2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*, Atlantis Press, 2022, ss. 1547–1555.
- [22] E. Jahanshiri, T. Buyong A. R. M. Shariff, "A review of property mass valuation models," *Pertanika Journal of Science & Technology*, c. 19, no. 1, ss. 23–30, 2011.
- [23] T. Dimopoulos A. Moulas, "A proposal of a mass appraisal system in Greece with CAMA system: Evaluating GWR and MRA techniques in Thessaloniki Municipality," *Open geosciences*, c. 8, no. 1, ss. 675–693, 2016.
- [24] D. Wang, V. J. Li H. Yu, "Mass appraisal modeling of real estate in urban centers by geographically and temporally weighted regression: A case study of Beijing's core area," *Land*, c. 9, no. 5, s. 143, 2020.
- [25] S. Sisman A. C. Aydinoglu, "A modelling approach with geographically weighted regression methods for determining geographic variation and influencing factors in housing price: A case in Istanbul," *Land use policy*, c. 119, s. 106 183, 2022.
- [26] A. Açlar V. Çağdaş, *Taşınmaz (Gayrimenkul) Değerleme*. Harita Kadastro Mühendisleri Odası, 2008.

- [27] T. Yomralıoğlu, R. Nişancı, M. Çete E. Candaş, “Dünya’da ve Türkiye’de Taşınmaz Değerlemesi,” *Türkiye’de Sürdürülebilir Arazi Yönetimi Çalıştayı*, c. 26, s. 27, 2011.
- [28] E. Güngör, *Gayrimenkul değerlendirme ve Türkiye’de sermaye piyasalarında gayrimenkul ekspertiz şirketlerine yönelik düzenlemeler yapılmasına ilişkin öneriler*. Sermaye Piyasası Kurulu, 1999.
- [29] E. Köktürk E. Köktürk, *Taşınmaz değerlendirme: Taşınmaz hukuku-İmar hukuku-Değerleme yöntemleri*. Seçkin Yayıncılık, 2016.
- [30] TUDB, *Türkiye Değerleme Standartları Çalışması*. erişim adresi: https://tdub.org.tr/Images/Uploads/tugdes_taslak_metni.pdf (erişim tarihi 04/11/2024).
- [31] E. Pagourtzi, V. Assimakopoulos, T. Hatzichristos N. French, “Real estate appraisal: a review of valuation methods,” *Journal of property investment & finance*, c. 21, no. 4, ss. 383–401, 2003.
- [32] N. Büyükkaracıoğlu, İ. Altınışik H. Uzun, “Türkiye’de Gayrimenkul Değerleme Çalışmalarında Karşılaşılan Sorunlar ve Çözüm Önerileri,” *Selçuk Üniversitesi Sosyal ve Teknik Araştırmalar Dergisi*, no. 14, ss. 77–91, 2017.
- [33] T. Yomralıoğlu, “Taşınmazların Değerlendirilmesi,” *Tapu ve Kadastro Dergisi*, no. 25, ss. 69–76, 2001.
- [34] *Mevzuat Bilgi Sistemi*. erişim adresi: <https://www.mevzuat.gov.tr/> (erişim tarihi 01/12/2024).
- [35] N. Erdem, “Türkiye taşınmaz değerlendirme sisteminin etkinliğinin araştırılması,” *Geomatik*, c. 4, no. 1, ss. 1–13, 2019.
- [36] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of research and development*, c. 3, no. 3, ss. 210–229, 1959.
- [37] Microsoft, *Azure Machine Learning’de derin öğrenme ve makine öğrenmesi karşılaştırması*. erişim adresi: <https://learn.microsoft.com/tr-tr/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2> (erişim tarihi 03/12/2024).
- [38] B. Çobanoğlu, *Veri Bilimi İçin Python*. KODLAB Yayın Dağıtım Yazılım Ltd. Şti., 2021.
- [39] W. S. McCulloch W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The bulletin of mathematical biophysics*, c. 5, ss. 115–133, 1943.
- [40] D. O. Hebb, *The organization of behavior: A neuropsychological theory*. Wiley ve Sons, 1949.
- [41] A. M. Turing, “Computing machinery and intelligence,” *Mind*, c. 59(236), ss. 433–460, 1950.
- [42] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O’Reilly Media, Inc., 2022.

- [43] T. Cover P. Hart, “Nearest neighbor pattern classification,” *IEEE transactions on information theory*, c. 13, no. 1, ss. 21–27, 1967.
- [44] S. University, *Stanford’s robotics legacy*. erişim adresi: <https://news.stanford.edu/stories/2019/01/stanfords-robotics-legacy> (erişim tarihi 04/11/2024).
- [45] G. DeJong, “Generalizations Based on Explanations.,” *IJCAI*, c. 81, 1981, ss. 67–69.
- [46] B.-H. Juang L. R. Rabiner, “Automatic speech recognition—a brief history of the technology development,” *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara*, c. 1, no. 67, s. 1, 2005.
- [47] B. E. Boser, I. M. Guyon V. N. Vapnik, “A training algorithm for optimal margin classifiers,” *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, ss. 144–152.
- [48] R. Memisevic G. Hinton, “Unsupervised learning of image transformations,” *2007 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2007, ss. 1–8.
- [49] R. Sengupta, D. Sengupta, A. K. Kamra D. Pandey, “Artificial Intelligence and Quantum Computing for a Smarter Wireless Network,” *Artificial Intelligence*, c. 7, no. 19, ss. 200–208, 2020.
- [50] C. Schröer, F. Kruse J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Computer Science*, c. 181, ss. 526–534, 2021.
- [51] R. Wirth J. Hipp, “CRISP-DM: Towards a standard process model for data mining,” *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, Manchester, c. 1, 2000, ss. 29–39.
- [52] S. Garcia, S. Ramirez-Gallego, J. Luengo, J. M. Benitez F. Herrera, “Big data preprocessing: methods and prospects,” *Big data analytics*, c. 1, ss. 1–22, 2016.
- [53] S. Şişman A. Ç. Aydınoglu, “Toplu Taşınmaz Değerlemesi için Coğrafi Verinin İşlenmesi ve Rastgele Orman Makine Öğrenmesi Tekniğiyle Tahmin Modelinin Geliştirilmesi,” *Harita Dergisi*, c. 169, ss. 13–27, 2023.
- [54] M. M. Breunig, H.-P. Kriegel, R. T. Ng J. Sander, “LOF: identifying density-based local outliers,” *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, ss. 93–104.
- [55] C. Wheelan, *Naked statistics: Stripping the dread from the data*. WW Norton & Company, 2013.
- [56] A. S. Albayrak, “Çoklu Doğrusal Bağlantı Halinde Enküçük Kareler Tekniğinin Alternatifi Yanlı Tahmin Teknikleri ve Bir Uygulama,” *Uluslararası Yönetim İktisat Ve İşletme Dergisi*, c. 1, no. 1, ss. 105–126, 2005.
- [57] N. Shrestha, “Detecting multicollinearity in regression analysis.,” *American Journal of Applied Mathematics and Statistics*, c. 8, no. 2, ss. 39–42, 2020.

- [58] Q.-H. Luu, M. F. Lau, S. P. Ng T. Y. Chen, “Testing multiple linear regression systems with metamorphic testing,” *Journal of Systems and Software*, c. 182, s. 111 062, 2021.
- [59] S. Narin, O. Doğan, N. Bande Y. Genç, “Keçiören/Ankara özelinde konut rayiç değerlerinin tahmininde çoklu regresyon analizi ve yapay sinir ağları yöntemlerinin karşılaştırılması,” *International Journal of Engineering Research and Development*, c. 15, no. 2, ss. 828–839, 2023.
- [60] Y. Terzi, *SPSS ile İstatiksel Veri Analizi*. erişim adresi: <https://avys.omu.edu.tr/storage/app/public/yukselt/129644/sos208-8%20%C3%A7oklu%20regresyon.pdf> (erişim tarihi 04/11/2024).
- [61] C. Brunsdon, A. S. Fotheringham M. E. Charlton, “Geographically weighted regression: a method for exploring spatial nonstationarity,” *Geographical analysis*, c. 28, no. 4, ss. 281–298, 1996.
- [62] L. Breiman, “Random forests,” *Machine learning*, c. 45, ss. 5–32, 2001.
- [63] T. Kavzoğlu İ. Çölkesen, “Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği,” *Teknolojik Araştırmalar*, c. 2, no. 1, ss. 36–45, 2010.
- [64] IBM, *What Is Random Forest?* Erişim adresi: <https://www.ibm.com/topics/random-forest> (erişim tarihi 03/12/2024).
- [65] F. Erdem, M. A. Derinpınar, R. Nasırzadehdızaı, S. Oy, D. Z. Şeker B. Bayram, “Rastgele orman yöntemi kullanılarak kıyı çizgisi çıkarımı İstanbul örneği,” *Geomatik*, c. 3, no. 2, ss. 100–107, 2018.
- [66] R. E. Schapire, “Explaining adaboost,” içinde *Empirical inference: festschrift in honor of vladimir N. Vapnik*, Springer, 2013, ss. 37–52.
- [67] Y. Freund R. E. Schapire, “A decision-theoretic generalizaion of online learning and an application to boosting,” *Journal of Computer and System Sciences*, c. 1, no. 55, ss. 119–139, 1997.
- [68] C. Ying, M. Qi-Guang, L. Jia-Chen G. Lin, “Advance and prospects of AdaBoost algorithm,” *Acta Automatica Sinica*, c. 39, no. 6, ss. 745–758, 2013.
- [69] Y. Freund, R. E. Schapire ve diğ., “Experiments with a new boosting algorithm,” *icml*, Citeseer, c. 96, 1996, ss. 148–156.
- [70] T. Chen C. Guestrin, “Xgboost: A scalable tree boosting system,” *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, ss. 785–794.
- [71] Y. Zhao, G. Chetty D. Tran, “Deep learning with XGBoost for real estate appraisal,” *2019 IEEE symposium series on computational intelligence (SSCI)*, IEEE, 2019, ss. 1396–1401.
- [72] L. Rampini F. Re Cecconi, “Artificial intelligence algorithms to predict Italian real estate market prices,” *Journal of Property Investment & Finance*, c. 40, no. 6, ss. 588–611, 2022.

- [73] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush A. Gulin, “CatBoost: unbiased boosting with categorical features,” *Advances in neural information processing systems*, c. 31, 2018.
- [74] A. V. Dorogush, V. Ershov A. Gulin, “CatBoost: gradient boosting with categorical features support,” *arXiv preprint arXiv:1810.11363*, 2018.
- [75] J. T. Hancock T. M. Khoshgoftaar, “CatBoost for big data: an interdisciplinary review,” *Journal of big data*, c. 7, no. 1, s. 94, 2020.
- [76] A. Botchkarev, “A new typology design of performance metrics to measure errors in machine learning regression algorithms,” *Interdisciplinary Journal of Information, Knowledge, and Management*, c. 14, ss. 045–076, 2019.
- [77] Z. Lian, Y. Ma, M. Li, W. Lu W. Zhou, “Discovery Precision: An effective metric for evaluating performance of machine learning model for explorative materials discovery,” *Computational Materials Science*, c. 233, s. 112 738, 2024.
- [78] J. K. Eckert, R. J. Gloudemans R. R. Almy, *Property appraisal and assessment administration*. International Association of Assessing Officers, 1990.
- [79] B. B. Başkanlığı, *Bayrampaşa’yı Tanıyalım*. erişim adresi: [https : / / bayrampasa.bel.tr/bayrampasa-yi-taniyalim-167](https://bayrampasa.bel.tr/bayrampasa-yi-taniyalim-167) (erişim tarihi 15/09/2024).

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. Ayar, M. D. ve Şen, A., (2023). "Makine Öğrenimi Yöntemlerini Kullanarak Gayrimenkul Değer Tahmini". 19. Türkiye Harita Bilimsel ve Teknik Kurultayı. 04-06 Ekim. Ankara, Türkiye.

