



Araştırma Makalesi / Research Article

KONUT FİYATLARI TAHMİNİNDE MAKİNE ÖĞRENMESİ SINIFLANDIRMA ALGORİTMALARININ KULLANILMASI: KÜTAHYA KENT MERKEZİ ÖRNEĞİ

Hasan Arda BURHAN¹

Öz

Artan şehirleşme ve hızlı ekonomik gelişmeler konut piyasalarının büyümesine neden olmakta; bu bağlamda hem kentlere doğru göçün bir sonucu, hem de konutların yatırım ve servet koruma araçları olarak görülmeleri nedeniyle konut talebi de zaman içinde yükselmektedir. Öte yandan konut fiyatları birçok faktörden etkilenmekte olup, literatürde bunları içeren farklı modeller ve çeşitli yöntemler dâhilinde konut fiyat tahmin çalışmaları yapılmaktadır. Konut piyasalarının Türkiye'nin ekonomik büyüme araçlarından biri olarak taşıdığı öneme paralel olarak bu çalışmada Kütahya ili merkez ilçesinde hanehalkının büyük çoğunlukla ikamet ettiği 4 odalı (3+1) konut fiyatlarının sınıflandırma tahmini, konut fiyatlarını etkileyen çeşitli kriterler ve bunlara ilişkin verilerin yer aldığı bir modelin farklı makine öğrenmesi sınıflandırma algoritmalarında çalıştırılmasıyla gerçekleştirilmiştir. Analiz sonucunda Naive Bayes (NB) yöntemi dışındaki k-En Yakın Komşu (k-EYK), Karar Ağaçları (KA) ve Rastgele Orman (RO) algoritmaların doğruluk oranlarının %60 üzerinde olduğu, temel performans ölçütü olan AUC skorlarına göre ise en başarılı tahmin yönteminin RO ve bunu takiben k-EYK olduğu görülmüştür. Sonuç olarak ele alınan değişkenler ve veri seti bağlamında söz konusu iki yöntemin Kütahya ili kent merkezindeki 4 odalı dairelerin konut fiyatlarının tahmininde başarılı sonuç verdiği görülmüştür.

Anahtar Kelimeler: Konut fiyatları, makine öğrenmesi, karar ağaçları, rastgele orman, Kütahya.

JEL Kodları: C13, C38, R32

USING MACHINE LEARNING CLASSIFICATION ALGORITHMS IN FORECASTING HOUSING PRICES: THE CASE OF KUTAHYA CITY CENTER

Abstract

The increasing urbanization and rapid economic development have led to growing housing markets. This has resulted in a rising demand for housing due to migration towards cities and the acceptance of housing as both an investment and a wealth protection tool. On the other hand, housing prices are affected by many factors, thereby studies on housing price forecasts are carried out within the scope of different models and various methods. In line with the significant importance of housing markets as one of Turkey's economic growth tools, this study includes the forecast of the classification of 4-bedroom (3+1) housing prices in the central district of Kütahya province. The dataset used includes various criteria that affect housing prices, and different machine learning classification algorithms that are utilized. As a result, it is seen that the accuracy rates of the k-Nearest Neighbor (k-NN), Decision Trees (DT) and Random Forest (RF) algorithms were above 60%. As the main performance criterion, the AUC scores yielded that the most successful classification forecast is acquired by the RF algorithm, which is followed by the k-NN. In the context of the variables and data set included in this study, it was seen that successful outcomes in the classification forecasts of 4-bedroom apartments' housing prices in the city center of Kütahya were obtained by these two methods.

Keywords: Housing prices, machine learning, decision trees, random forest, Kutahya.

JEL Codes: C13, C38, R32

¹ Dr. Öğr. Üyesi, Kütahya Dumlupınar Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, arda.burhan@dpu.edu.tr, ORCID: 0000-0003-4043-2652

Başvuru Tarihi (Received): 09.02.2023 **Kabul Tarihi** (Accepted): 24.04.2023

Giriş

Barınma ihtiyacının karşılanmasındaki temel rolünün yanı sıra bir yatırım aracı olarak da değerlendirilen konutlar, bu bağlamda hem günlük hayatın devam ettirildiği güvenli sığınaklar hem de ülke paralarının olası değer kayıplarına karşılık birer servet koruma aracı olarak insan yaşamı ve ekonomiler açısından büyük önem taşımaktadırlar. Öte yandan ülke ekonomileri geliştikçe kentlere doğru göç hareketliliğinin ve konut talebinin arttığı bilinmektedir. Türkiye’de büyüme politikasının temel araçlarından biri olarak kabul edilen konut piyasası, konut fiyatlarını gayrimenkul yatırımı ve ikamet kararlarını etkilemesi açısından tüketiciler, işletmeler, yönetici ve karar alıcılar ile politika yapımcılar için en önemli ilgi odağı haline getirmektedir. Diğer taraftan konut fiyatlarında görülen değişimler, ülke Gayri Safi Yurtiçi Hasıla (GSYH)’sının ve enflasyonunun izlediği seyrin bir anlamda göstergesi olmakta, bu doğrultuda da konut fiyatlarının takip edilmesi gereklilik arz etmektedir (Case vd., 2005). Türkiye İstatistik Kurumu (TÜİK) tarafından açıklanan güncel verilere göre 2022 Ocak-Kasım dönemindeki Türkiye’de konut satışları, 2021 yılı aynı döneme göre %1’lik artış göstermiştir (TÜİK, 2022b). Bununla birlikte Türkiye Cumhuriyeti Merkez Bankası (TCMB)’nin açıkladığı en son verilere göre Konut Fiyat Endeksi (KFE) bir önceki aya kıyasla 2022 Kasım ayında %3,8 oranında, 2021 Kasım ayına göre ise nominal olarak %174,3 ve reel olarak %54 oranında artış göstermiştir (TCMB, 2022a). Bu bağlamda söz konusu endekse ek olarak kredi faiz oranları, GSYH, altın ve kur fiyatları ile tüketici güven endeksi gibi farklı değişkenler de konut talebi ve fiyatlamasında göz önünde bulundurulmuş faktörler olarak belirtilebilmektedir (Akay vd., 2019; Korkmaz, 2019). Buna ek olarak Rahadi vd. (2015)’ne göre konut fiyatları aynı zamanda fiziksel, tarz (konsept) ve lokasyon olmak üzere üç özellikten de etkilenmekte olup oda sayısı, bina yaşı gibi özellikler fiziksel nitelikleri; geleneksel, modern, akıllı ev, vb. nitelikler konsepti; çeşitli önemli tesis ve kurumlara yakınlık ise lokasyon özelliklerini ifade etmektedir. Öte yandan konut fiyatlama örüntülerinin anlaşılması ve bunu etkileyen faktörlerin belirlenebilmesi yalnızca tüketicilerin değil, aynı zamanda gayrimenkul endüstrisindeki mülk sahipleri, analistler ve politika yapımcılarının yanı sıra şehir ve bölge planlama yetkililerinin de ilgilendiği bir konudur. KFE’nin, genel anlamda konut fiyatlarındaki seyrin takip edilmesinde en temel göstergelerden biri olduğu bilinmekle birlikte, konut fiyatlarının konum, şehir, nüfus, fiziksel özellikler gibi birçok kriterle ilişkili olması, bu tahmin sürecinin tek bir konutun fiyatını tahmin etmede kabataslak ve yaklaşık bir tahminci olan ve ortalama fiyat değişikliklerini izleyen bir tekrar satış endeksi olan KFE’den daha fazla bilgiye ihtiyaç duymasına yol açmaktadır (Adetunji vd., 2022). Bu bağlamda konut piyasalarının ülke ekonomileri açısından önemine ve konut fiyatlarını etkileyen faktörlerin çeşitliliğine paralel olarak literatürde farklı yöntemler dâhilinde fiyat tahmin çalışmalarının yapıldığı ifade edilebilmektedir. Özellikle son yıllarda bilhassa yapay zekâ ve makine öğrenmesi tabanlı tahmin sistemlerinin, karar verici ve araştırmacılara konut fiyatlaması sürecinde yardımcı olabileceği açıklıkla görülmüştür. Bir veri kümesindeki kuralları, ilişkileri ve kalıpları belirlemeye yönelik uygun bir yaklaşım sunan makine öğrenmesi algoritmaları, kredi analizi, görüntü tanıma, meteoroloji, tıp, dolandırıcılık tespiti, müşteri ilişkileri, biyoinformatik, tarım gibi birçok alanda kullanılmaktadır (McQueen vd., 1995; Liakos vd., 2018). Buna ek olarak konut piyasasının hali hazırdaki büyüme sürecinde, Destek Vektör Makineleri (DVM), Karar Ağaçları (KA), Yapay Sinir Ağları (YSA), Rastgele Orman (RO), k-En Yakın Komşu (k-EYK), Genetik Algoritmalar (GA), Gradient Boosting (GB), Naive Bayes (NB), Classification and Regression Tree (CART), GRNN, BPNN, RIPPER, AdaBooST, gibi makine öğrenmesi algoritmalarının da konut fiyatı tahmininde sıklıkla kullanılır hale geldiği söylenebilir (Monika vd., 2021).

Diğer taraftan 2022 yılında TÜİK tarafından yayınlanan adrese dayalı nüfus kayıt sistemi sonuçları incelendiğinde Kütahya ili toplam nüfusu 580.701 olup, Merkez ilçe nüfusu 282.243 ve nüfus artış hızı ise binde 17.8 ile binde 7.1 olan Türkiye ortalamasının on puan üzerindedir (TÜİK, 2022a). Yine TÜİK’in gerçekleştirdiği “*İstatistiklerle Aile*” çalışmasına ilişkin veriler incelendiğinde

Türkiye’de ortalama hanehalkı büyüklüğü 2008 yılında 4 iken 2021 yılında bu sayı 3,2’ye gerilemiş; çocuklu çiftlerden oluşan hanehalkının toplam hanehalkına oranı 2008 yılında %54,7 iken bu değer 2021 yılına gelindiğinde %40,8 olmuştur (TÜİK, 2021a). Kütahya ili özelinde son istatistiklere göre toplam hanehalkı sayısı 192.678 ve bunlar içinde tek çekirdek aileden ve en az bir çekirdek aile ve diğer kişilerden oluşan hanehalkı toplamı 146.193 olup, hanehalkının büyük çoğunluğunun tek çekirdek aileden (120.039) oluştuğu ifade edilebilmektedir (TÜİK, 2021a). Öte yandan TÜİK’in 2021 yılı “*Bina ve Konut Nitelikleri Araştırması*”na göre tek çekirdek aileden ve en az bir çekirdek aile ve diğer kişilerden oluşan hanehalklarının çoğunlukla 4 odalı (salon dâhil) konutlarda ikamet ettikleri (sırasıyla %50,1 ve %49,1) görülmüştür (TÜİK, 2021b). Yine aynı araştırmaya göre Kütahya özelinde ise hanehalkının %67,7’sinin ev sahibi olduğu, %39’unun 4 odalı konutlarda ikamet ettiği ve binaların %41,2’sinin 2001 yılı ve sonrasında inşa edildiği belirtilmiştir (TÜİK, 2021b). Yine TÜİK tarafından sağlanan söz konusu istatistiklere göre Kütahya kent merkezinde konut satışlarında 2017 yılına kadar görülen belirgin artış, aynı ivmeyi devam ettirememiş ve 2019 yılında yaklaşık olarak 2015 yılı rakamlarına geri dönmüş olup 2020 yılında az miktarda artış ve 2021 yılında ise küçük miktarda azalış göstermiştir.

Buna uygun olarak bu çalışmada seçili makine öğrenmesi algoritmalarının Kütahya ili kent merkezindeki konut fiyat tahmin sınıflandırma performansları ve yeterlilikleri, ilgili istatistiklere göre Kütahya ilinde hanehalkının büyük çoğunluğunu oluşturan tek çekirdek aileden ve en az bir çekirdek aile ve diğer kişilerden ibaret hanehalklarının büyük oranda ikamet ettikleri 4 odalı (salon dâhil) konutlar için konut piyasası literatürü göz önünde bulundurularak belirlenen seçili kriterleri içeren bir model yardımıyla değerlendirilmiş ve analiz edilmiştir. İlgili literatürde makine öğrenmesi tekniklerinin kullanıldığı ve Kütahya ili konut fiyatlarını odak alan herhangi bir çalışmanın bulunmaması, bu çalışmanın temel motivasyonunu teşkil etmektedir. Çalışmanın takip eden bölümünde ilgili alandaki literatür taraması verilmiş olup, ikinci bölümde analizde kullanılan veri ve yöntemlere ilişkin açıklamalar, uygulama bölümünde ise söz konusu yöntemlerle gerçekleştirilen analizin ayrıntılı açıklamaları yer almaktadır. Sonuç ve tartışmalar bölümünün içeriğini ise elde edilen sonuçlar ve bunlara ilişkin yorum ve değerlendirmeler oluşturmaktadır.

1. Literatür Taraması

Literatürde konut fiyatlarının tahmininde makine öğrenmesi algoritmalarının yanı sıra istatistiksel ve ekonometrik modellerle yapılmış olan çok sayıda çalışma yer almaktadır. Dünya genelinde makine öğrenmesi algoritmaları kullanılarak konut fiyat tahminine ilişkin yapılmış çalışmalardan bazılarına ilişkin detaylar Tablo 1’de verilmektedir:

Tablo 1: Konut Fiyatı Tahmininde Makine Öğrenmesi Algoritmalarının Kullanıldığı Çalışmalar

Referans	Algoritma	Referans	Algoritma
Chen vd. (2017)	DVM	Pai ve Wang (2020)	DVM,CART, GRNN, BPNN
Embaye vd. (2021)	Ridge ,KA, Bagging, RO, GB	Park ve Bae (2015)	C4.5, RIPPER, NB, AdaBoost
Fu (2018)	DVM, YSA, RO	Plakandaras vd. (2015)	EEMD, DVM
Gu vd. (2011)	GA, DVM	Rafiei ve Adeli (2016)	Deep Belief Restricted Boltzmann Machine, GA
Ho vd. (2021)	DVM, RO, GB	Rawool vd. (2021)	KA, RO
Huang (2019)	KA, GB, RO, DVM	Rico-Juan ve de La Paz (2021)	K-EYK, KA, RO, AdaBoost, XGBoost, vb.
Kamal vd. (2020)	KA, RO	Thamarai ve Malarvizhi (2020)	KA
Li vd. (2009)	DVM	Vasquez ve Chellamuthu, (2021)	DVM
Monika vd. (2021)	Extreme-GB, GB, RO, DVM	Vineeth vd. (2018)	YSA

İstatistiksel ve ekonometrik temelli çalışmalara ise Acar (2020), Alfiyatin (2017), Cabrera vd. (2011), Çağlayan Akay vd. (2019), İslamoğlu ve Nazlıoğlu (2019), Kangallı Uyar ve Yayla (2016), Kouwenberg ve Zwinkels (2014), Liu ve Wu (2020), Milunovich (2020), Özen (2022), Özgüler vd. (2022), Sağlam ve Abdioğlu (2020), Sevinç (2022), Webb vd. (2016), Wei ve Cao (2017), Yang vd. (2018); Yayar ve Deniz (2014), Zohrabyan vd. (2008) örnek olarak verilebilir. Öte yandan ilgili literatürde makine öğrenmesi yöntemlerinin Türkiye'deki konut fiyatlarının tahmininde kullanıldığı çalışmaların sınırlı sayıda olduğu ifade edilebilmektedir. Buna göre; Yılmaz vd. (2018), Eskişehir'deki konut fiyatlarını YSA yöntemi ile tahmin ettikleri çalışmalarında 13 farklı kriter dahilinde 2017 yılı Ocak-Mart ayları arası verilerin yer aldığı 19 farklı model geliştirerek model performanslarını karşılaştırmışlardır. Elde edilen sonuçlar, YSA'nın konut fiyatlarının tahmininde etkili bir yöntem olduğunu ortaya koymuştur. Savaş (2019) Ankara ili Gölbaşı ilçesindeki konut fiyatlarının tahmini için 42 kriter içeren dört farklı veri setine DVM ve Çoklu Lineer Regresyon (ÇLR) yöntemlerini uygulamıştır. Sonuçlara göre DVM uygulaması sonuçlarının tutarlı olduğu ve kullanılan yöntemlerin klasik gayrimenkul değerlendirme yöntemleri içinde karlı bir seçenek olarak değerlendirilebileceği ifade edilmiştir. Ulvi ve Özkan (2019), YSA ve Bulanık Mantık yaklaşımlarından faydalanarak Konya, Selçuklu ilçesinde yer alan konutlara ilişkin 8 değişken altındaki verilerin yer aldığı modelle fiyat tahminlemesi yaptıkları çalışmalarında, piyasadaki fiyatlara daha çok yaklaşmış olması nedeniyle YSA yönteminin daha isabetli tahminler veren bir yöntem olduğu belirtilmiştir. Aydemir vd. (2020) geliştirdikleri konut verilerini toplayan ve kaydeden bir sistem vasıtasıyla İstanbul'daki konut fiyatları için 176 farklı özelliğe dayanan 14 farklı algoritma ile tahminleme gerçekleştirmişlerdir. RO ve Derin Öğrenme (DÖ) yöntemlerinin kullanıldığı çalışmada en başarılı yaklaşım, 852 konut için tutarlı tahminlerde bulunan RO algoritması olmuştur. Araştırmacılar, bu bağlamda farklı yapay zeka yöntemlerinin de konut piyasası için fiyat tahmininde kullanılabilir olduğunu belirtmişlerdir. Erkek vd. (2020) Ocak 2015-Aralık 2019 arası 5 yıllık dönem için 15 açıklayıcı değişken ve İstanbul, İzmir, Ankara'ya ait verilerin yer aldığı bir model çerçevesinde yürüttükleri çalışmalarında, DVM ve YSA yöntemleriyle tahminler gerçekleştirmişlerdir. YSA algoritmalarının daha başarılı sonuçlar

verdiği çalışmanın bir diğer sonucuna göre lokasyon ve konut büyüklüğü en önemli değişkenler olarak öne çıkmıştır. Selçi (2021) 2013 Ocak-2019 Aralık dönemi için aylık verilerden ve döviz kurları, yabancılara konut satışı, konut kredisi faizleri, Tüketici Fiyat Endeksi (TÜFE), yeni ve yeni olmayan KFE değişkenlerinden oluşan bir model dahilinde YSA yaklaşımıyla konut fiyatlarını tahmin etmiştir. Elde edilen sonuçlara göre 2020 Ocak dönemi için yapılan tahminlerin gerçeğe oldukça yakın olduğu, ayrıca mevsimsellik tespitinde YSA'nın oldukça etkili sonuç verdiği ortaya çıkmıştır. Altun (2022), Ankara Etimesgut ilçesi için 2020 yılı Haziran-Aralık döneminde kamu konut ihale bedeli verilerini kullanmışlardır. 16 bağımsız değişken içeren YSA modelleriyle yapılan tahmin sonuçlarına göre %93 doğruluk oranı elde edilmiş, bu bağlamda ilgili yöntemin konut değerlemesinde kullanılabilir bir yapı sunduğu ortaya çıkmıştır. Doğan vd. (2022) çalışmalarında Ankara Keçiören ilçesinde yer alan konutlar için geliştirdikleri 11 parametre içeren 20 ayrı YSA modelini kullanarak fiyat tahmininde bulunmuşlardır. Tahmin sonuçlarına göre YSA yaklaşımının konut fiyatlarını öngörmeye yüksek tahmin oranı ile başarılı sonuç verdiği görülmüştür. Kayakuş vd. (2022), RO, YSA ve DVM yöntemlerini kullanarak 9 değişken ve Ocak 2013-Kasım 2020 arası aylık verileri içeren içeren modeller çerçevesinde konut metrekare fiyatlarını tahmin etmişlerdir. Her üç yöntemin de başarılı sonuç verdiği çalışmanın, özellikle Türkiye gibi gelişmekte olan ülkeler için önem taşıyan konut inşa ve fiyat öngörülerini için yol gösterici olacağı belirtilmiştir. Özdemir vd. (2022), Sakarya ili Serdivan, Adapazarı, Arifiye and Karasu ilçeleri için DÖ, Polinomsal Regresyon ve RO algoritmaları kullanarak konut fiyat tahmininde bulunmuşlardır. 7 değişken içeren modeller çerçevesinde yapılan hesaplamalar sonucunda en iyi öngörü sonucunun DÖ yöntemi ile elde edildiği, bunu ise RO algoritmasının takip ettiği ifade edilmiştir.

2. Veriler ve Yöntem

Çalışma verileri Kütahya ili Merkez ilçe için popüler bir e-ticaret sitesine çoklu bağlantılar yoluyla, bu sitede 2022 Aralık ayı itibariyle yer alan merkezi sistem, kombi veya yerden ısıtma özelliğine sahip satılık tüm 3+1 daire ilanlarından derlenmiştir. Veri setinde 5 değişken altında 621 adet gözlem yer almaktadır. Konut fiyatını etkileyen değişkenler Sevinç (2022)'in Lancaster (1966) temelli hedonik fiyatlama modeli yaklaşımına uygun olarak belirlenmiştir. Buna göre modeldeki değişkenler ve kısa açıklamaları Tablo 2'de verilmektedir:

Tablo 2: Model Değişkenleri ve Açıklamaları

Bağımlı Değişken	Düzeyley	Açıklama
Konut Fiyatı (ordinal)	0) 500.000-1.000.000	Dairenin fiyatının TL cinsinden değeri
	1) 1.000.001-1.500.000	
	2) 1.500.001-2.000.000	
	3) 2.000.001-2.500.000	
	4) 2.500.001 ve üzeri	
Bağımsız Değişkenler	Düzeyley	Açıklama
Net m ² (ordinal)	0) 50-100	Dairenin sahip olduğu alanın m ² cinsinden değeri
	1) 101-150	
	2) 151-200	
	3) 201-250	
	4) 250 ve üzeri	
Bina Yaşı (ordinal)	0) 0 (Yeni)	Dairenin yer aldığı binanın yıl cinsinden yaşı
	1) 1	
	2) 2	
	3) 3	
	4) 4	
	5) 5-10	
	6) 11-15	
	7) 16-20	
	8) 21-25	
	9) 26-30	
10) 31 ve üzeri		
Bulunduğu Kat (ordinal)	0) Zemin	Dairenin binada yer aldığı kat
	1) 1. Kat	
	2) 2. Kat	
	3) 3. Kat	
	4) 4. Kat	
	5) 5. Kat	
	6) 6. Kat	
	7) 7. Kat	
	8) 8. Kat	
	9) 9. Kat	
10) 10. Kat ve üzeri		
Site İçinde (nominal)	0) Evet	Dairenin sitede yer alan bir binada olup olmadığı
	1) Hayır	

Tablo 2.'de görüldüğü üzere çalışmada kullanılan model, net m², bina yaşı, bulunduğu kat, site içinde olup olmama durumunu bağımsız değişken; konut fiyatını ise bağımlı değişken olarak içermektedir.

Öte yandan makine öğrenmesi yöntemleri, genel anlamda denetimli/gözetimli (supervised), denetimsiz/gözetimsiz (unsupervised) ve takviyeli (reinforcement) olmak üzere üç farklı kategori altında ele alınmaktadır (Ho vd., 2021). Bir denetimli öğrenme modeli, (x,y) olarak ifade edilebilecek bir gözlem kümesinde, model eğitimi için kullanılacak olması nedeniyle önceden bilinen y değişken değerlerini gerektirmektedir (Vasquez ve Chellamuthu, 2021). Denetimsiz öğrenmede ise algoritmalar, önceden etiketlenmemiş verileri, benzer özellikteki gözlemin gruplandırılması için kullanır (Savaş, 2019). Takviyeli öğrenme yöntemi, verili ödül-ceza bilgisine istinaden nasıl bir davranış şeklinin tercih edileceğinin belirleneceği durumlarda kullanılmaktadır (Murphy, 2012). Denetimli makine öğrenmesi, içeriğinde regresyon ve sınıflandırma olmak üzere iki yaklaşımı barındırmakta olup; regresyonda sürekli nitelikteki veri kümesinden yararlanılarak

çıktı tahmini yapılmakta iken, sınıflandırmada hangi sınıfa ait olduğu bilinmeyen bir örnek için en uygun sınıf belirlenir (Uğuz, 2021). Bu çalışmada kullanılan yöntem olan sınıflandırma, y değişkeni çıktıları ifade etmekte olup $y \in \{1, 2, \dots, C\}$ şeklinde gösterilmekte; C değeri, sınıf sayısını göstermekle birlikte $C=2$ olması halinde ikili (binary), $C>2$ olması halinde ise çoklu (multi-class) sınıflandırma söz konusu olmaktadır (Murphy, 2012). İkili sınıflandırma için Lojistik Regresyon, DVM ve k-DVM yöntemleri; çoklu sınıflandırma için ise k-EYK, NB, KA ve RO algoritmaları kullanılmaktadır (Géron, 2019). Çalışmaya dahil edilen sınıflandırma algoritmalarına ait özet niteliğinde açıklamalar bölümün devamında verilmiştir.

2.1. k-En Yakın Komşu (k-EYK) Algoritması

k-EYK sınıflandırma algoritması, eğitim setinde yer alan ve test verisi x' 'e en yakın k adet noktayı baz alarak bu kümede veri setindeki sınıflardan kaç tanesinin yer aldığı saymakta ve buna göre tahmin edilmek istenen verinin ait olduğu sınıfı belirlemektedir (Brownlee, 2017). k-EYK için en temel hususu, en yakın komşuları belirleyebilmek için test verisi ile eğitim verisi örnekleri arasındaki uzaklığın hesaplanması teşkil etmekte olup, Öklidyen, Manhattan, Chebyshev, Minkowski ve City-block, k-EYK algoritması dahilinde kullanılan uzaklık ölçütlerine örnek olarak verilebilmektedir (Ali vd., 2019). Bu doğrultuda algoritma için önemli olan bir diğer faktör k parametre değerinin belirlenmesi olarak ifade edilmektedir. Bu değer çok büyük olarak belirlenmesi fazla sayıda sınıflanmış veri diğerlerine baskın çıkarak sonuçların yanlı olmasına; çok küçük belirlenmesi ise eğitim setinde yer alan çok sayıda örneğin kullanılabilirlik avantajını yok olmasına neden olduğu için en uygun k değerinin belirlenmesinde genellikle eğitim ve test veri setlerinde denemelerle optimizasyon yaklaşımı tercih edilir (Mucherino vd., 2009).

2.2. Naive Bayes (NB) Algoritması

Girdi değişkenlerinin (öznitelik) koşullu bağımsız olduğu varsayımını temel alan NB algoritması, böylece belli bir çıktıya (sınıf) ilişkin koşullu yoğunluğu tek boyutlu yoğunlukların çarpımı olarak yazma imkânı vermektedir;

$$p(X|y = c, \theta) = \prod_{j=1}^D p(x_j|y = c, \theta_{jc}) \quad (1)$$

ve böylece elde edilen model, NB sınıflandırıcısı olarak adlandırılır (Murphy, 2012). Her bir öznitelige ilişkin değerlerin tek tek hesaplanmasındansa koşullu bağımsızlığı esas olarak olasılık hesaplamalarını, bu sayede algoritmanın işleyişini sadeleştiren algoritma bu özelliği nedeniyle saf (naive) olarak adlandırılmakta ve ikili ya da çoklu sınıflandırma problemleri için kullanılabilir (Brownlee, 2017). Öznitelik sayısının fazla olduğu veri setleri için yapılan uygulamalarda bazı algoritmalarda sorun yaşanabilmekte iken NB algoritması için böyle bir durumun söz konusu olmadığı belirtilmektedir (Uğuz, 2021). Bu bağlamda NB algoritmasının etkin performans göstermesinin, her bir öznitelige tek tek odaklanarak her sınıf için bu özniteliklerden ayrı ayrı istatistikler elde etmesi ile ilgili olduğu ifade edilmektedir (Müller ve Guido, 2017)

2.3. Karar Ağacı (KA) Sınıflandırma Algoritması

Farklı girdi değişkenleriyle yapılan testlerle en benzer eğitim örneğini bulan KA algoritmasındaki ağaç şekli benzetimi, modelin karar düğüm ve yapraklarından oluşması, kökten başlayarak her bir karar düğümünün girdi değişkenine bir ayırma (bölme) testi uygulaması ve çıktıya göre dallardan birinin tercih edilerek bir yaprağa ulaşıldığında arayışın sonlandırılmasını kapsamaktadır (Alpaydın, 2021). Sürecin başarısında saf olmama ölçütlerinden (impurity measure) faydalanılmakta olup, literatürde Entropi ve sürekli veriler için Gini ölçütleri tavsiye edilmektedir (Pajankar ve Joshi, 2022). En uygun bölme işlemi için ağaç öğrenme algoritması kullanılarak optimize edilmek istenen bir amaç fonksiyonu tanımlanır ve f , bölmeyi yapacak olan öznitelik, D_p

ve D_j sırasıyla esas ve j . düğüme ait veri setleri, I , saf olmama ölçütü, N_p toplam eğitim verisi ve N_j ise j . düğümdeki örnek sayısını ifade etmek üzere KA algoritması;

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^m \frac{N_j}{N_p} I(D_j) \quad (2)$$

şeklinde gösterilmektedir (Raschka vd., 2022). Makine öğrenmesi uygulamalarında kullanılan KA temelli algoritmalara, C4.5, C5.0, Hunt, CHi-squared Automatic Interaction Detector (CHID) ve CART algoritmaları örnek olarak verilebilir (Patel ve Rana, 2014).

2.4. Rastgele Orman (RO) Sınıflandırma Algoritması

Bir tahminin varyansını azaltmak, birçok tahminin ortalaması ile mümkün olabilmekte; örneğin M adet farklı ağaç birçok veri altkümesi kullanılarak eğitilmekte, rassal olarak yer değiştirme yoluyla seçilmekte ve tüm birim f_m , m . ağaç olmak üzere;

$$f(x) = \sum_{m=1}^M \frac{1}{M} f_m(x) \quad (3)$$

ile torbalama (bagging) adı verilen işlem yoluyla hesaplanmaktadır (Murphy, 2012). Bu şekilde düzenlenen karar ağaçlarının ileri versiyonunu temsil eden RO algoritması, önyüklemesi yapılmış T adet örnek yaratarak CART sınıflandırıcılarını inşa eder (Breiman, 2001). Öznitelik seçim ölçütü olarak Gini ölçütünün tercih edildiği RO algoritmasında herhangi bir T eğitim seti için rassal olarak seçilen bir örneğin C_i sınıfına ait olasılığı, $f(C_i, T)/|T|$ söz konusu olasılığı temsil etmek üzere Gini ölçütü kullanılarak;

$$\sum \sum_{j \neq i} (f(C_i, T)/|T|)(f(C_j, T)/|T|) \quad (4)$$

şeklinde ifade edilir (Pal, 2005). RO algoritması, sınıflandırma sürecinde farklı öznitelik veya değişkenlerin görel önemlerini değerlendirilmesinde; aynı zamanda özellikle veri boyutlarının çok yüksek olduğu ve her tahmin edici değişkenin sınıflandırma modelini, en iyi değişkeni seçecek şekilde nasıl etkilediğinin bilinmesinin önemli olduğu durumlarda kullanılmaktadır (Rodriguez-Galiano vd., 2012).

3. Uygulama

Çalışma içeriğindeki analiz, Python programlama dili ve makine öğrenmesi uygulamalarında sıklıkla kullanılan uygulama kütüphanelerinden Scikit-learn aracılığıyla gerçekleştirilmiştir. Bir önceki bölümde ifade edildiği üzere denetimli makine öğrenmesi yaklaşımında kurulan modeller, bilinen çıktı değişkenlerini içeren veriler kullanılarak eğitilmektedir. Bu doğrultuda öncelikle veriler, modeli eğitmek üzere eğitim seti ve modeli değerlendirmek üzere test seti olmak üzere iki bölüme ayrılır (Amr, 2020). Literatürde bölümlendirme yüzdesi ile ilgili farklı yaklaşımlar söz konusu olmakla birlikte genel olarak kullanılan oranların eğitim seti için %80, test seti için ise %20'dir (Müller ve Guido, 2017).

Verilerin ilgili oranda bölünmesini takiben model ve algoritma değerlendirme aşamasına geçilmiş olup, çalışmada kullanılan tüm makine öğrenmesi algoritmalarınca izlenen aynı sürecin adımları aşağıda verilmiştir:

Şekil 1: Analizde Kullanılan Algoritmalara Ait Değerlendirme Adımları



Her ne kadar birçok algoritmada zaman içerisinde kategorik değişkenlerin algılanmasına yönelik geliştirmeler yapılmış olsa da, bunların analizde sayısal değerlere dönüştürülerek kullanılmasının

daha iyi performansa ve veri özelliklerinin daha eşit bir şekilde boyutlandırılmasına imkan sağladığı ifade edilmektedir (Saleh, 2020). Bu doğrultuda model dâhilinde yer alan kategorik değişkenlere ait veriler, Scikit-learn kitaplığından faydalanılarak *label encoding* işlemi ile sayısal değerlere dönüştürülmüştür. Sonraki adımda verinin büyük bölümünü içeren eğitim seti, kalan bölümü için tahminler üretecek şekilde eğitilmektedir (Pajankar ve Joshi, 2022). Eğitim sürecinin devamında, test setine ait tahminler çoklu sınıflandırma makine öğrenmesi algoritmaları k-EYK, NB, KA, RO kullanılarak tahmin edilmiştir. Elde edilen tahmin ve gerçekleşen değerlere ait sonuçların bir bölümü Tablo 3'te verilmektedir:

Tablo 3: *Kısmi Tahmin ve Gerçekleşen Değer Karşılaştırmaları*

k-EYK Tahmini	NB Tahmini	KA Tahmini	RO Tahmini	Gerçekleşen Değer
0*	2	0*	0*	0
0*	2	0*	0*	0
1*	2	1*	1*	1
4*	2	4*	4*	4
0*	2	0*	0*	0
4*	2	4*	4*	4
0*	2	0*	0*	0
4*	2	4*	4*	4
4	2	4	4	0
0*	2	0*	0*	0
0*	2	0*	0*	0
0*	2	0*	0*	0
4*	2	4*	4*	4
0*	2	0*	0*	0
0	2	0	4	1
4	2	0*	4	0
4	4	4	4	0
0*	2	0*	0*	0
0*	2	0*	0*	0
0*	2	0*	0*	0
4	2	4	4	0
0	2	0	0	1
0*	2	0*	0*	0
0	4	4	4	1
0*	2	0*	0*	0
4	2	4	4	1
0*	2	0*	0*	0
0*	2	0*	0*	0
4*	4*	4*	4*	4
0*	2	0*	0*	0

* Tutarlı tahminleri ifade etmektedir

Tablo 3, yukarıda değinildiği üzere test set değerlerinin bir bölümünü içermekte olup, son adımda analizde kullanılan makine öğrenmesi yöntemlerinin performanslarına ait bütünlüklü bir değerlendirmenin sunulabilmesi için öncelikle ilgili literatürde hata matrisi (confusion matrix) olarak ifade edilen ölçütten faydalanılmıştır. Bu bağlamda her bir yönteme ilişkin doğru/yanlış sınıflandırma adetlerinin gösterildiği hata matrisleri hesaplanarak Tablo 4'te verilmiştir:

Tablo 4: Konut Satış Fiyatı Tahminlerine İlişkin Hata Matrisleri

Konut Fiyatı (TL)		500.000-1.000.000	1.000.001-1.500.000	1.500.001-2.000.000	2.000.001-2.500.000	2.500.001 ve üzeri
k-EYK Tahmin						
Gerçek	500.000-1.000.000	51	0	0	0	9
	1.000.001-1.500.000	10	1	0	0	2
	1.500.001-2.000.000	1	0	0	0	1
	2.000.001-2.500.000	1	0	0	0	1
	2.500.001-üzeri	13	1	0	0	34
NB Tahmin						
Gerçek	500.000-1.000.000	0	0	55	0	5
	1.000.001-1.500.000	0	0	10	0	3
	1.500.001-2.000.000	0	0	0	0	2
	2.000.001-2.500.000	0	0	1	0	1
	2.500.001-üzeri	0	0	35	0	13
KA Tahmin						
Gerçek	500.000-1.000.000	49	4	0	0	7
	1.000.001-1.500.000	9	2	0	0	2
	1.500.001-2.000.000	0	1	0	0	1
	2.000.001-2.500.000	1	0	0	1	0
	2.500.001-üzeri	19	1	0	0	28
RO Tahmin						
Gerçek	500.000-1.000.000	48	2	0	0	10
	1.000.001-1.500.000	7	2	0	0	4
	1.500.001-2.000.000	0	1	0	0	1
	2.000.001-2.500.000	1	0	0	1	0
	2.500.001-üzeri	17	1	0	0	30

Tablo 4, hata matrisleri aracılığıyla elde edilen doğru ve yanlış sınıflandırma adetlerini vermekte ve doğru sınıflandırma adetleri, köşegen değerlerden görülebilmektedir. Buna göre gerçek satış fiyatı 500.000 ile 1.000.000 TL arasında olan konutlardan k-EYK yöntemi ile 51, KA ile 49 ve RO ile 48'inin fiyat aralığı doğru sınıflandırılmış olup, NB yöntemi ile doğru tahminleme gerçekleştirilememiştir. Benzer şekilde gerçek satış fiyatı 1.000.001 ile 1.500.000 TL arasında olan konutlardan k-EYK yöntemi ile 1, KA ve RO ile 2'sinin fiyat aralığı doğru sınıflandırılmış olup NB yöntemi ile isabetli sınıflandırma gerçekleştirilememiştir. Kalan sınıflandırma tahminleri benzer şekilde yorumlanabilmektedir. Öte yandan hata matrisleri, makine öğrenmesi sınıflandırma problemlerinde performans değerlendirmeleri için tek başlarına yeterli olmayıp model değerlendirmelerinde sıklıkla Receiver Operator Characteristic (ROC) eğrisi ve bu eğrinin altındaki alana ait skoru ifade eden Area Under the Curve (AUC) değerlerinin hesaplanmasına başvurulmaktadır (Uğuz, 2021). Buna göre hata matrislerinden elde edilen doğru pozitif/negatif ve yanlış pozitif/negatif değerleri kullanılarak öncelikle doğruluk (accuracy), kesinlik (precision), bunlara bağlı olarak doğru pozitif oranı (duyarlılık - sensivity/recall), yanlış pozitif oranı ve özgüllük değerleri hesaplanmakta; ROC eğrisi ise duyarlılık ve özgüllüğün birlikte gösterimini içermektedir (Nellore, 2015). ROC eğrisinin köşegeni rassal tahmin olarak ifade edilebilmekte ve bu köşegenin altına düşen tahminlerin rassal tahminlerden daha kötü olduğu, en başarılı sınıflandırma tahmininin ise grafiğin sol üst köşesine denk geldiği; elde edilen ROC eğrisinin altındaki alanın hesaplanması ile de kullanılan modelin nihai sınıflandırma performansını gösteren AUC değerinin elde edildiği belirtilmektedir (Raschka vd., 2022).

AUC değeri, 0 ile 1 arasında değer almakta olup 1'e eşit alan mümkün olan en iyi değeri ifade etmektedir (Amr, 2020). Her ne kadar ROC eğrisi ikili sınıflandırmalar için kullanılıyor olsa da ele alınan çoklu sınıflandırma problemlerini çoklu ikili sınıflandırmalar haline getiren One-vs-Rest (OvR) ve One-vs-One (OvO) yaklaşımları ile de bu eğriler elde edilebilmektedir (Géron, 2019;

Raschka vd., 2022). OvR yönteminde model hedef sınıf ve diğer sınıflara ait veriler kullanılarak eğitilmekte, k adet eğitim verisi sınıfı söz konusu olduğunda, yine k adet sınıflandırıcı yaratılmakta iken; OvO yöntemi ise eğitim setindeki tek bir çift sınıfı seçerek ikili sınıflandırıcıları kullanmakta ve eğitim verisi sınıf sayısı k ise, yaratılan sınıflandırıcı sayısı $k(k-1)/2$ olmaktadır (Jang vd., 2020). Bu yolla elde edilecek AUC değerleri, rassal değer 0.5 olduğu öncülüne göre 0,5-0,6 aralığı için başarısız, 0,6-0,7 aralığı için kötü, 0,7-0,8 aralığı için uygun, 0,8-0,9 aralığı için iyi ve 0,9-1 aralığı için ise mükemmel sınıflandırma sonucu olarak kabul edilmektedir (Safari vd., 2016; Müller ve Guido, 2017). Bu doğrultuda model değerlendirme sürecinin son adımı olan yöntem performanslarının değerlendirilmesi aşamasında her bir modele ait Doğruluk ve AUC (OvO ve OvR) değerleri hesaplanmış olup sonuçlar Tablo 5’te verilmiştir:

Tablo 5: Doğruluk ve AUC (OvO ve OvR) Değerleri

Yöntem	Doğruluk oranı	AUC(OvR)	AUC (OvO)
k-EYK	0.688	0.75485	0.69888
NB	0.104	0.58967	0.53359
KA	0.640	0.69547	0.65371
RO	0.648	0.79031	0.75681

Model değerlendirme sonuçlarının gösterildiği Tablo 5’te yer alan değerler, hata matrislerinden elde edilen modellerin genel doğruluk oranının, NB yöntemi hariç %60 üzerinde olduğunu göstermektedir. Asıl performans değerlendirme ölçütleri olan AUC değerleri OvR ve OvO yaklaşımları bağlamında incelendiğinde ise RO yönteminin sınıflandırma için uygun performans verdiği; k-EYK yönteminin OvR’ye göre uygun, ancak OvO yaklaşımına göre kötü; NB yönteminin her iki ölçekte de başarısız; KA yaklaşımının ise her iki kritere göre de kötü performans gösterdiği ifade edilebilmektedir. Bu bağlamda analizde kullanılan makine öğrenmesi sınıflandırma algoritmalarından özellikle RO ve sonrasında K-EYK yaklaşımlarının, Kütahya ili kent merkezindeki 4 odalı dairelerin konut fiyatlarının tahmininde uygun performans gösterdiği belirtilebilmektedir

4. Sonuç ve Tartışmalar

İkamete mahsus gayrimenkuller, insanlar için korunma, sığınma, sağlığın muhafaza edilmesi gibi yaşamsal işlevleri yerine getirirken genel olarak konut piyasaları, ülke ekonomileri için özellikle inşaat ve finans sektörleri vasıtasıyla itici güç vazifesi görmektedirler. Bu bağlamda konut fiyat eğilimlerinin insanların satın alma, yatırım yapma, çalışıp yaşayacakları şehirleri seçme gibi önemli kararlarını büyük ölçüde etkileyen bir özelliğe sahip olduğu, bu nedenle konut fiyat tahminine ilişkin yaklaşım ve çalışmaların birçok araştırmacı ve karar vericinin dikkatini çektiği söylenebilmektedir. Literatürde otoregresif, vektör otoregresif, vektör hata düzeltme ve bunların varyasyonlarından oluşan farklı ekonometrik yaklaşımların yanı sıra, son yıllarda DVM, KA, RO, NB, torbalama (bagging), boosting, topluluk öğrenmesi (ensemble learning) gibi makine öğrenmesi model ve algoritmalarının da değişen ölçek ve çeşitlilikteki konut fiyatı tahminleme çalışmalarında sıklıkla kullanılır olduğu görülmektedir (Xu ve Zhang, 2021).

Öte yandan temsil içeriği Türkiye geneli olmak üzere, gözlemlenebilir konut niteliklerine bağlı fiyat değişimlerinin hedonik regresyon yöntemiyle elde edilmiş bir endeks ile ifade edildiği KFE’nin TCMB tarafından yayımlanan son istatistiklerine göre ilgili endeksteeki artışın, bir önceki yıla göre nominal olarak %170’ten, reel olarak ise %50’den fazla olduğu belirtilmiştir (TCMB, 2022a). Bu bağlamda Türkiye’de hızla artan konut fiyatlarının halk ve politika yapımcılar açısından önemli bir husus haline geldiği açıklıkla belirtilebilmektedir. Öte yandan aynı endeks istatistikleri, çalışmanın odak ili olan Kütahya’yı içeren TR33 bölgesi (Afyonkarahisar, Kütahya, Manisa, Uşak) açısından incelendiğinde, endeksteeki artışın 2022 yılında bir önceki yıla göre yaklaşık %130, 2020 yılına göre ise %248 olduğu, Kütahya ili özelinde ise yıllık yüzde değişimin %145.06 olduğu gözlenmektedir (TCMB, 2022b; TCMB, 2022c). Diğer taraftan Kütahya ilindeki

hanehalkının yaklaşık %76'sının tek çekirdek aile ile en az bir çekirdek aile ve diğer kişilerden oluştuğu ve söz konusu hanehalklarının büyük çoğunlukla salon dahil 4 odalı evlerde ikamet ettikleri görülmektedir (TÜİK, 2021a; TÜİK, 2021b).

Yukarıda ifade edilen tüm faktörler göz önünde bulundurularak bu çalışmada, Kütahya ili Merkez ilçesinde yer alan 4 odalı (3+1) konut fiyatlarının sınıflandırma tahmini, Sevinç (2022)'in Lancaster (1966) temelli hedonik fiyatlama modeli yaklaşımına uygun olarak konut fiyatını etkileyen net m², bina yaşı, bulunduğu kat, site içinde olup olmama durumu bağımsız değişkenlerini ve konut fiyatı bağımlı değişkenini içeren model ile k-EYK, NB, KA, RO makine öğrenmesi sınıflandırma algoritmaları ile gerçekleştirilmiştir. Makine öğrenmesi tekniklerinin uygulanmasında önemli bir adım olan veri setinin eğitim ve test seti olarak iki bölüme ayrılmasını takiben analizin devamında modelde yer alan kategorik değişkene ait dönüşümler gerçekleştirilmiş, sonrasında eğitim seti ile eğitilen modellerle test seti değerleri tahmin edilmiştir. Elde edilen tahmin sonuçları ve hata matrislerinin tek başlarına, doğru ve yanlış sınıflandırma adetleri ve doğruluk oranlarını vermek dışında sağlıklı tahmin performans değerlendirme ölçütleri olmamaları sebebiyle, analiz sürecinin son aşamasında AUC (OvO ve OvR) skorları hesaplanmıştır. Doğruluk oranlarının, NB yaklaşımı hariç %60 üzerinde olduğu bilgisine ek olarak asıl performans değerlendirme ölçütü olan AUC skorlarına göre k-EYK yönteminin yalnızca OvR'ye göre uygun; RO algoritmasının ise hem OvR hem OvO ölçeğinde uygun sınıflandırma performansı gösterdiği ve bu iki yöntemin Kütahya ili kent merkezindeki 4 odalı dairelerin konut fiyatlarının tahmininde başarılı sonuç verdiği görülmüştür.

Çalışmanın önceki bölümlerinde ifade edildiği üzere bu çalışmanın temel motivasyonunu Türkiye genelinde makine öğrenmesi tekniklerinin konut fiyatı tahmininde kullanıldığı kısıtlı sayıda çalışmanın olması ve bilhassa Kütahya ili özelinde bir çalışmanın yer almaması teşkil etmekle birlikte, bu durum analiz sonucunda elde edilen bulguların önceki çalışmaların bulgularıyla karşılaştırılma imkanından yoksun kalmasına da neden olmaktadır. Öte yandan söz konusu karşılaştırma yöntem perspektifinden yapılırsa; bu çalışmada elde edilen sonuçların, Türkiye'de konut fiyatı tahminlerinin YSA ile yapıldığı Yılmaz vd. (2018), Ulvi ve Özkan (2019), Erkek vd. (2020), Selçi (2021), Altun'a (2022) ait çalışmalar ile DVM, RO gibi diğer yöntemleri içeren Savaş (2019), Aydemir vd. (2020), Kayakuş vd. (2022) ve Özdemir vd. (2022)'in çalışmalarında elde edilen sonuçlara da uygun olarak, makine öğrenmesi algoritmalarının isabetli konut fiyat tahminleri vererek bu bağlamda kullanılabilir olduğunu gösterdiğini söylemek mümkündür. Çalışma sınırlılıkları ölçeğinde değerlendirildiğinde ise ilk olarak analiz dahilindeki algoritmaların, veri setinin söz konusu e-ticaret sitesinin çoklu bağlantılar sürecinde sahip olduğu kadarıyla veriyi içermesi nedeniyle daha büyük veri setiyle çalışma avantajından yoksun kaldığı ifade edilebilir. Diğer taraftan konut fiyatlarına etki eden birçok faktör söz konusu olmakla birlikte bu çalışmada kullanılan modelde sınırlı sayıda değişken yer almıştır. Son olarak, önceki bölümlerde ifade edildiği üzere literatürdeki farklı çalışmalarda faydalanılan birçok makine öğrenmesi algoritması içinden k-EYK, NB, KA, RO çoklu sınıflandırma algoritmalarının kullanılmış olması, bu çalışmaya ilişkin bir diğer sınırlılık olarak vurgulanabilir.

Öte yandan ülkemiz literatüründe konut fiyat tahmininde makine öğrenmesi algoritmalarının kullanıldığı sınırlı sayıda çalışma içinde Kütahya iline odaklanan bir çalışmanın bulunmaması, bilhassa resmi veriler gözetildiğinde Kütahya'da yaşayan hanehalklarının büyük çoğunlukla ikamet ettikleri 4 odalı daireler ölçeğinde konut fiyat tahmininin yapıldığı bir çalışmanın yapılmamış olması, farklı iller temel alınarak yapılmış sınırlı sayıdaki diğer çalışmalarda yer alan yöntemlerden farklı ve güncel makine öğrenmesi sınıflandırma yaklaşımlarının kullanılmış olması ve elde edilen sonuçlar bağlamında çalışmada kullanılan makine öğrenmesi algoritmalarının konut fiyat tahmininde kullanılabilir olduğu sonucuna varılması, bu çalışmanın ilgili literatüre katkısı olarak ifade edilebilir. İlerleyen çalışmalarda farklı ve bütünlük yöntemlerle birlikte daha büyük

veri setleri kullanılarak ve modelde yer alan değişken sayısı artırılarak bu çalışmanın sınırlılıkları aşılabılır.

Kaynakça

- Acar, T. (2020). Determining housing prices using the semiparametric estimation within the hedonic price model framework: Case study of istanbul housing market example. *Ekonomi, Politika & Finans Araştırmaları Dergisi*, 5(3), 561-575.
- Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2022). House price prediction using random forest machine learning technique. *Procedia Computer Science*, 199, 806-813.
- Akay, Ö., Çelik, C., ve Kıral, G. (2019). Konut talebine göre benzer özellik gösteren Türkiye illerinin panel veri kümeleme analizi ile belirlenmesi. *Uluslararası Ekonomi ve Yenilik Dergisi*, 5(2), 231-245.
- Alfiyatin, A. N. (2017). Modeling house price prediction using regression analysis and particle swarm optimization case study: Malang, east java, indonesia. *International Journal of Advanced Computer Science and Applications*, 8(10), 323-326.
- Ali, N., Neagu, D., & Trundle, P. (2019). Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. *SN Applied Sciences*, 1, 1559.
- Alpaydın, E. (2021). *Machine learning - revised and updated edition*. Cambridge, MA: The MIT Press.
- Altun, Ö. (2022). Yapay zekâ yöntemleriyle hazine taşınmazlarının değerlendirilmesi: Yapay sinir ağları ile kamu konutları üzerine bir uygulama. *Türkiye Arazi Yönetimi Dergisi*, 4(2), 62-73.
- Amr, T. (2020). *Hands-on machine learning with scikit-learn and scientific python toolkits*. Birmingham: Packt Publishing Ltd.
- Aydemir, E., Aktürk, C., ve Yalçınkaya, M. A. (2020). Yapay zekâ ile konut fiyatlarının tahmin edilmesi. *Turkish Studies-Information Technologies and Applied Sciences*, 15(2), 183-194.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- Brownlee, J. (2017). *Master machine learning algorithms*. Erişim Adresi: <https://machinelearningmastery.com/master-machine-learning-algorithms/>.
- Cabrera, J., Wang, T., & Yang, J. (2011). Linear and nonlinear predictability of international securitized real estate returns: A reality check. *Journal of Real Estate Research*, 33(4), 565-594.
- Case, K. E., Quigley, J. M., & Shiller, R. J. (2005). Comparing wealth effects: The stock market versus the housing market. *Advances in Macroeconomics*, (5)1, 1-34.
- Chen, J.-H., Ong, C. F., Zheng, L., & Hsu, S.-C. (2017). Forecasting spatial dynamics of the housing market using support vector machine. *International Journal of Strategic Property Management*, 21(3), 273-283.
- Çaglayan Akay, E., Topal, K. H., Kizilarlan, S., & Bulbul, H. (2019). Forecasting of Turkish housing price index: ARIMA, random forest, ARIMA-random forest. *Pressacademia*, 10(10), 7-11.
- Doğan, O., Bande, N., Genç, Y., ve Akyön, F. Ç. (2022). Keçiören/Ankara özelinde konut rayiç değerlerinin yapay sinir ağları metodu kullanılarak tahmini. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 35, 113-128.

- Embaye, W. T., Zereyesus, Y. A., & Chen, B. (2021). Predicting the rental value of houses in household surveys in tanzania, uganda and malawi: Evaluations of hedonic pricing and machine learning approaches. *Plos One*, 16(2), e0244953.
- Erkek, M., Çayırılı, K., & Hepşen, A. (2020). Predicting house prices in turkey by using machine learning algorithms. *Journal of Statistical and Econometric Methods*, 9(4), 31-38.
- Fu, T. (2018). Forecasting second-hand housing price using artificial intelligence and machine learning techniques. *8th international conference on mechatronics, computer and education informationization (MCEI 2018)* (s. 269-273). Shenyang: Atlantis Press.
- Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow*. Sebastopol, CA: O'Reilly Media, Inc.
- Gu, J., Zhu, M., & Jiang, L. (2011). Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications*, 38(4), 3383–3386.
- Ho, W. K., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research*, 38(1), 48–70.
- Huang, Y. (2019). Predicting home value in california, united states via machine learning modeling. *Statistics, Optimization & Information Computing*, 7(1), 66–74.
- İslamoğlu, B., ve Nazlıoğlu, Ş. (2019). Enflasyon ve konut fiyatları: İstanbul, ankara ve izmir için panel veri analizi. *Siyaset, Ekonomi ve Yönetim Araştırmaları Dergisi*, 7(1), 93-99.
- Jang, J., Seo, M., & Kim, C. O. (2020). Support weighted ensemble model for open set recognition of wafer map defects. *IEEE Transactions on Semiconductor Manufacturing*, 33(4), 635-643.
- Kamal, N., Chaturvedi, E., Gautam, S., & Bhalla, S. (2020). House price prediction using machine learning. *Emerging Technologies in Data Mining and Information Security - Proceedings of IEMIS 2020, Volume 3* (s. 799-811). Kolkata: Springer.
- Kangallı Uyar, S. G., ve Yayla, N. (2016). Konut fiyatlarının hedonik fiyatlama yaklaşımına göre mekânsal ekonometrik modeller ile tahmini: İstanbul konut piyasası örneği. *Social Sciences*, 11(4), 26-342.
- Kayakuş, M., Terzioğlu, M., & Yetiz, F. (2022). Forecasting housing prices in Turkey by machine learning methods. *Aestim*, 80, 33-44.
- Korkmaz, Ö. (2019). The relationship between housing prices and inflation rate in Turkey: Evidence from panel Konya causality test. *International Journal of Housing Markets and Analysis*, 13(3), 427-452.
- Kouwenberg, R., & Zwinkels, R. (2014). Forecasting the us housing market. *International Journal of Forecasting*, 30(3), 415–425.
- Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 132-157.
- Li, D.-Y., Xu, W., Zhao, H., & Chen, R.-Q. (2009). A SVR based forecasting approach for real estate price prediction. *International Conference on Machine Learning and Cybernetics* (s. 970–974). Baoding: IEEE.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674-2703.

- Liu, L., & Wu, L. (2020). Predicting housing prices in china based on modified Holt's exponential smoothing incorporating whale optimization algorithm. *Socio-Economic Planning Sciences*, 72, 100916.
- McQueen, R. J., Gamer, S. R., Nevill-Manning, C. G., & Witten, I. H. (1995). Applying machine learning to agricultural data. *Computers and Electronics in Agriculture*, 12, 275-293.
- Milunovich, G. (2020). Forecasting Australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting*, 39, 1098-1118.
- Monika, R., Nithyasree, J., Valarmathi, V., Hemalakshmi, G. R., & Prakash N. B. (2021). House price forecasting using machine learning methods. *Turkish Journal of Computer and Mathematics Education*, 12(11), 3624-3232.
- Mucherino, A., Papajorgji, P. J., & Pardalos, P. M. (2009). K-nearest neighbor classification. A. Mucherino, P. J. Papajorgji, & P. M. (Ed.) *Data mining in agriculture içinde* (ss. 83-106). New York: Springer.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. Cambridge: MIT Press.
- Müller, A. C., & Guido, S. (2017). *Introduction to machine learning with python - a guide for data scientists*. Sebastopol, CA: O'Reilly Media, Inc.
- Nellore, S. B. (2015). Various performance measures in binary classification-an overview of ROC study. *International Journal of Innovative Science*, 2(9), 596-605.
- Özdemir, M., Yıldız, K., & Büyüktanır, B. (2022). Housing price estimation with deep learning: A case study of sakarya turkey. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, 9(1), 138-151.
- Özen, K. (2022). Konut fiyatları üzerinde etkili olan sosyo ekonomik unsurların hedonik fiyat modeliyle araştırılması: Panel veri analizi. *Ağrı İbrahim Çeçen Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 8(2), 503-522.
- Özgüler, İ. C., Büyükkara, G., Z., & Küçüközmen, C. C. (2022). Discovering the fundamentals of turkish housing market: A price convergence framework. *International Journal of Housing Markets and Analysis*, 16(1), 116-145.
- Pai, P.-F., & Wang, W.-C. (2020). Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences*, 10(17), 5832.
- Pajankar, A., & Joshi, A. (2022). *Hands-on machine learning with python - implement neural network solutions with scikit-learn and pytorch*. New York, NY: Apress.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217-222.
- Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data. *Expert Systems with Applications*, 42(6), 2928-2934.
- Patel, B. R., & Rana, K. K. (2014). A survey on decision tree algorithm for classification. *International Journal of Engineering Development and Research*, 2(1), 1-5.
- Plakandaras, V., Gupta, R., Gogas, P., & Papadimitriou, T. (2015). Forecasting the U.S. real house price index. *Economic Modelling*, 45, 259-267.
- Rafiei, M. H., & Adeli, H. (2016). A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management*, 142(2), 04015066.

- Rahadi, R. A., Wiryono, S., Koesrindartoto, D., & Syamwil, I. B. (2015). Factors influencing the price of housing in Indonesia. *International Journal of Housing Markets and Analysis*, 8(2), 169-188.
- Raschka, S., Liu, Y. & Mirjalili, V. (2022). *Machine learning with pytorch and scikit-learn - develop machine learning and deep learning models with python*. Birmingham: Packt Publishing Ltd.
- Rawool, A. G., Rogye, D. V., Rane, S. G., & Bharadi, V. A. (2021). House price prediction using machine learning. *Iconic Research and Engineering Journals*, 4(11), 29-33.
- Rico-Juan, J. R., & de La Paz, P. T. (2021). Machine learning with explainability or spatial hedonics tools? an analysis of the asking prices in the housing market in alicante, spain . *Expert Systems with Applications*, 171, 114590.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.
- Safari, S. A., Elfil, M., & Negida, A. (2016). Evidence based emergency medicine; part 5 receiver operating curve and area under the curve. *Emergency*, 4(2), 111-113.
- Sağlam, C., ve Abdioğlu, Z. (2020). Türkiye’de tüketici fiyatları ile hedonik konut fiyatları arasındaki ilişki: Panel veri analizi. *Yaşar Üniversitesi E-Dergisi*, 15(57), 117-128.
- Saleh, H. (2020). *The machine learning workshop*. Birmingham: Packt Publishing Ltd.
- Savaş, B. (2019). *Makine öğrenme algoritmalarının konut değer tahmininde kullanımı: Ankara gölbaşı uygulaması*. [Yayımlanmamış yüksek lisans tezi]. Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü, Konya.
- Selçi, B. Y. (2021). Türkiye’nin konut satışı değerlerinin yapay sinir ağları ile öngörülmesi. *Ekoist: Journal of Econometrics and Statistics*, (35), 19-32.
- Sevinç, V. (2022). Determining the flat sales prices by fat characteristics using bayesian network models. *Computational Economics*, 59, 549-577.
- TCMB. (2022a, Kasım). *Konut Fiyat Endeksi - Kasım 2022*. Erişim Adresi: <https://www.tcmb.gov.tr/wps/wcm/connect/tr/tcmb+tr/main+menu/istatistikler/reel+sektor+istatistikleri/konut+fiyat+endeksi>.
- TCMB. (2022b). *Konut Fiyat Endeksi - TR33*. Erişim Adresi: https://evds2.tcmb.gov.tr/index.php?/evds/serieMarket/collapse_26/5949/DataGroup/turkish/bie_hkfe/.
- TCMB. (2022c). *Konut Fiyat Endeksi - Yıllık Yüzde Değişim*. Erişim Adresi: <https://evds2.tcmb.gov.tr/index.php?/evds/dashboard/310>.
- Thamarai, M., & Malarvizhi, S. P. (2020). House price prediction modeling using machine learning. *I.J. Information Engineering and Electronic Business*, 2, 15-20.
- TÜİK. (2021a). *İstatistiklerle Aile*. Erişim Adresi: <https://data.tuik.gov.tr/Bulten/Index?p=İstatistiklerle-Aile-2021-45632>.
- TÜİK. (2021b). *Bina ve Konut Nitelikleri Araştırması*. Erişim Adresi: <https://data.tuik.gov.tr/Bulten/Index?p=Bina-ve-Konut-Nitelikleri-Arastirmasi-2021-45870>.

- TÜİK. (2022a). *Adrese Dayalı Nüfus Kayıt Sistemi*. Erişim Adresi: <https://data.tuik.gov.tr/Bulten/Index?p=Adrese-Dayali-Nufus-Kayit-Sistemi-Sonuclari-2022-49685>.
- TÜİK. (2022b). *Konut Satış İstatistikleri, Kasım 2022 (Rapor No. 45683)*. Erişim Adresi: <https://data.tuik.gov.tr/Bulten/Index?p=Konut-Satis-Istatistikleri-Kasim-2022-45683>.
- Uğuz, S. (2021). *Makine öğrenmesi - teorik yönleri ve python uygulamaları ile bir yapay zeka ekolü*. Ankara: Nobel Akademik Yayıncılık.
- Ulvi, C., ve Özkan, G. (2019). Taşınmaz değerlemede yapay zeka tekniklerinin kullanılabilirliği ve yöntemlerin karşılaştırılması. *Geomatik*, 4(2) ,134-140.
- Vasquez, C., & Chellamuthu, V. (2021). House price prediction with statistical analysis in support vector machine learning for regression estimation. *Curiosity: Interdisciplinary Journal of Research and Innovation*, 1(2), 22311.
- Vineeth, N., Ayyappa, M., & Bharathi, B. (2018). House price prediction using machine learning algorithms. *International Conference on Soft Computing Systems* (s. 425-433). Singapore: Springer.
- Webb, R. I., Yang, J., & Zhang, J. (2016). Price jump risk in the us housing market. *The Journal of Real Estate Finance and Economics*, 53(1), 29–49.
- Wei, Y., & Cao, Y. (2017). Forecasting house prices using dynamic model averaging approach: Evidence from china. *Economic Modelling*, 61, 147–155.
- Xu, X., & Zhang, Y. (2021). House price forecasting with neural networks. *Intelligent Systems with Applications*, 12, 200052.
- Yang, J., Yu, Z., & Deng, Y. (2018). Housing price spillovers in China: A high-dimensional generalized VAR approach. *Regional Science and Urban Economics*, 68, 98-114.
- Yayar, R., ve Deniz, G. (2014). Mersin kent merkezinde konut piyasası fiyatlarının hedonik tahmini. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 14(3), 87-100.
- Yılmazel, Ö., Afşar, A., ve Yılmazel, S. (2018). Konut fiyat tahmininde yapay sinir ağları yönteminin kullanılması. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (20), 285-300.
- Zohrabyan, T., Leatham, D. J., & Bessler, D. A. (2008). *Cointegration analysis of regional house prices in US*. (Report No. 1292-2016-102324). Erişim Adresi: <https://ageconsearch.umn.edu/record/48138>.