

**T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
MAKİNE FAKÜLTESİ
ENDÜSTRİ MÜHENDİSLİĞİ BÖLÜMÜ**



MAKİNE ÖĞRENMESİ İLE KONUT FİYAT TAHMİNİ

**GALİP ŞAHİN
OKAN DEMİRKAYA
ZEYNEP YAZICIOĞLU**

LİSANS BİTİRME TEZİ

**DANIŞMAN
PROF. DR. SELÇUK ÇEBİ**

İSTANBUL¹, 2024

ÖNSÖZ

Bu tezin hazırlanmasında katkıda bulunan, bu çalışma boyunca bize rehberlik eden ve desteğini esirgemeyen danışmanımız Prof. Dr. Selçuk Çebi'ye en içten teşekkürlerimizi sunarız. Kendisinin bilgi ve deneyimi, bu tezin oluşmasında büyük bir rol oynamıştır.

Bu çalışma, İstanbul'daki konut fiyatlarının belirlenmesi ve tahmin edilmesi amacıyla gerçekleştirilmiştir. Ev fiyatlarını etkileyen faktörlerin çeşitliliği ve bu faktörlerin ev sahipleri ve yatırımcılar üzerindeki önemi göz önünde bulundurularak, özellikle büyük şehirlerdeki fiyat dalgalanmalarının olumsuz etkilerini azaltmak hedeflenmiştir. Bu bağlamda, İstanbul'da konut fiyatlarını etkileyen unsurları tespit etmek ve bu fiyatların tahmin edilmesi için çeşitli makine öğrenmesi tekniklerini kullanmak amacıyla kapsamlı bir analiz yapılmıştır.

Bu çalışmanın, ilgili alanda çalışan araştırmacılara ve uygulayıcılara faydalı olmasını diliyorum. Tezimizin ileride yapılacak çalışmalara katkı sağlaması ve ekonomik politikaların geliştirilmesinde yol gösterici olması en büyük temennimizdir.

Haziran, 2024

GALİP ŞAHİN

OKAN DEMİRKAYA

ZEYNEP YAZICIOĞLU

İçindekiler

KISALTMA LİSTESİ	5
ŞEKİL LİSTESİ.....	6
ÇİZELGE LİSTESİ.....	7
ÖZET.....	8
1. GİRİŞ	11
1.1 Problemin Tanımı.....	11
1.2 Problemin Endüstriyel/Akademik Önemi	12
1.3 Konuya Etki Eden Parametreler ve Tanımları	12
1.4 Çözüm Yöntemleri/Araçlar/Teknikleri	12
2. LİTERATÜR ARAŞTIRMASI.....	13
2.1 Makine Öğrenimi ile Konut Fiyatı Tahmini konusunda yapılan çalışmalarda kullanılan teknikler.....	14
2.2 Makine Öğrenimi ile Konut Fiyatı Tahmini konusunda yapılan çalışmalarda kullanılan parametreler	18
3. KULLANILAN YÖNTEMLER	24
3.1 Lineer Regresyon	24
3.2 Karar Ağacı	24
3.3 Rastgele Orman.....	25
3.4 Gradient Boosting	25
3.5 k-En Yakın Komşu (k-EYK) Algoritması	25
3.6 Sinir Ağları	26
4. UYGULAMA.....	27
4.1 Verilerin Toplanması	27
4.2 Verilerin Temizlenmesi	28
4.3 Verilerin Düzenlenmesi.....	28
4.4 Verilerin Normalize Edilmesi.....	29
4.5 Korelasyon Matrisi	30
4.6 Yöntemlerin Uygulanması.....	31

4.6.1	Veri Setinin Hazırlanması	32
4.6.2	Model Listesi ve Başlatılması	33
4.6.3	Modellerin Değerlendirilmesi	34
4.6.4	Sonuçların Görselleştirilmesi	35
5.	SONUÇ VE ÖNERİLER.....	37
5.1	Sonuç.....	37
	KAYNAKLAR.....	39

KISALTMA LİSTESİ

YSA	Yapay sinir ağları
PR	Lineer Regresyon
K-NN	K en yakın komşu
NB	Naive Bayes
KA	Karar ağacı
RO	Rastgele Orman
DVM	Destek Vektör Makinesi
Torbalı A.	Torbalı Ağaçlar Topluluğu
GB	Gradyan Arttırma
Bayesian R.	Bayesian Regresyon

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1	Yıllara Göre Yöntemlerin Kullanım Sayıları.....	17
Şekil 1.2	Kullanılan Parametrelerin Dağılımı.....	23
Şekil 1.3	Lineer Regresyon Grafiği.....	24
Şekil 1.4	Karar Ağacı Düğümleri.....	26
Şekil 1.5	Sinir Ağları.....	27
Şekil 1.6	Ev İlan Detayları.....	28
Şekil 1.7	Temizlenmiş Ev İlan Detayları.....	28
Şekil 1.8	Düzenlenmiş Ev İlan Detayları.....	29
Şekil 1.9	Normalizasyon formülü.....	29
Şekil 1.10	Normalize Edilmiş Ev İlan Detayları.....	30
Şekil 1.11	Korelasyon Matrisi.....	30

ÇİZELGE LİSTESİ

Sayfa

Çizelge 1.1	Yapılan Çalışmalarda Kullanılan Teknikler.....	12
Çizelge 1.2	Yapılan Çalışmalarda Kullanılan Parametreler.....	16
Çizelge 1.3	Kullanılan Tekniklerin Yüzdeleri.....	19
Çizelge 1.4	Yöntemlerin Performans Değerleri.....	33

MAKİNE ÖĞRENMESİ İLE KONUT FİYAT TAHMİNİ

Galip ŞAHİN

Okan DEMİRKAYA

Zeynep YAZICIOĞLU

Endüstri Mühendisliği Bölümü

Lisans Bitirme Tezi

Tez Danışmanı: Prof. Dr. Selçuk ÇEBİ

İstanbul'daki konut fiyatlarının belirlenmesi ve tahmin edilmesi amacıyla gerçekleştirilmiştir. Ev fiyatlarını etkileyen faktörlerin çeşitliliği ve bu faktörlerin ev sahipleri ve yatırımcılar üzerindeki önemi göz önünde bulundurularak, özellikle büyük şehirlerdeki fiyat dalgalanmalarının olumsuz etkilerini azaltmak hedeflenmiştir. Bu çalışma, İstanbul'da konut fiyatlarını etkileyen unsurları tespit ederek, bu fiyatların tahmin edilmesi için çeşitli makine öğrenmesi tekniklerini kullanmayı amaçlamaktadır. Problemin tanımı yapılmış ve evlerin insan yaşamındaki temel ihtiyaçlardan biri olduğu, nüfus artışıyla birlikte artan talebin fiyatların sürekli olarak değişmesine neden olduğu belirtilmiştir. İstanbul gibi büyük şehirlerde bu değişim daha hızlı ve belirgin olduğundan, ev sahipleri ve yatırımcılar için fiyatların öngörülebilir olması büyük önem taşımaktadır. Bu çalışmanın, bu ihtiyaca cevap vermek amacıyla tasarlandığı vurgulanmıştır.

Problemin endüstriyel ve akademik önemi vurgulanmıştır. Ev fiyatlarını etkileyen faktörlerin belirlenmesi, bir ülkenin ekonomik durumu hakkında önemli bilgiler sağlamakta ve ekonomik politikaların geliştirilmesine katkı sunmaktadır. Ayrıca, ev fiyatlarının tahmin edilmesi, yatırımcılar için risklerin azaltılması ve portföy yönetiminin etkin bir şekilde yapılabilmesi açısından da önemlidir. Kentsel planlama kararlarının doğruluğu da bu analizler sayesinde artırılabilir.

Konuya etki eden parametreler ve tanımları ele alınmıştır. Ev fiyatlarını etkileyen faktörler genellikle coğrafi, ekonomik, sosyal ve demografik faktörlerden oluşmaktadır. Konum, altyapı, ulaşım gibi faktörlere bağlı olarak büyük önem taşıyırken, evin özellikleri de fiyatları etkileyen unsurlar arasındadır. Altyapı, ulaşım, eğitim ve sağlık gibi faktörler, ev fiyatlarını etkileyen önemli unsurlar arasında yer almakta ve bu koşulların

bölgede sağlanması fiyatları etkilemektedir. Ekonomik koşullar da ev fiyatlarını etkileyen önemli faktörlerdendir; istihdam seviyeleri, faiz oranları ve enflasyon oranları gibi ekonomik göstergeler, fiyatları etkilemektedir. Talep ve arz dengesi de fiyatları etkileyen önemli faktörler arasında yer almakta olup, talebin artması fiyatların yükselmesine neden olabilmektedir. Çözüm yöntemleri, araçlar ve teknikler incelenmiştir. Ev fiyatlarını tahmin etmek için çeşitli makine öğrenmesi yöntemleri kullanılmıştır. Destek vektör regresyonu ve yapay sinir ağları gibi yöntemler, bu alanda yaygın olarak kullanılan tekniklerdir ve ev fiyatlarını etkileyen faktörleri belirlemeye yardımcı olabilmektedir. Ayrıca, en yüksek doğruluk oranını sağlayan modellerin belirlenmesi amacıyla çeşitli değerlendirme metrikleri kullanılarak modellerin performansları karşılaştırılmıştır. Sonuçlar ve öneriler sunulmuştur. Rastgele Orman (Random Forest) modelinin diğer modellere göre daha iyi performans gösterdiği ve ev fiyatlarını tahmin etmede en güvenilir model olarak öne çıktığı tespit edilmiştir. Bu çalışma, makine öğrenmesi yöntemlerinin ev fiyatlarını tahmin etme konusunda etkili araçlar olduğunu göstermektedir. Gelecek çalışmalarda daha büyük ve çeşitli veri setlerinin kullanılması ve model optimizasyon tekniklerinin uygulanması ile performansın daha da artırılması hedeflenmektedir.

HOUSE PRICE PREDICTION WITH MACHINE LEARNING

Galip SAHIN

Okan DEMIRKAYA

Zeynep YAZICIOGLU

Department of Industrial Engineering

BSc. Thesis

Adviser: Prof. Dr. Selcuk CEBI

This thesis aims to determine and predict housing prices in Istanbul. Considering the diversity of factors affecting house prices and the importance of these factors for homeowners and investors, the goal is to mitigate the adverse effects of price fluctuations, especially in large cities. This study aims to identify the factors influencing housing prices in Istanbul and use various machine learning techniques to predict these prices. The problem is defined by stating that houses are one of the fundamental needs in human life, and the increasing demand due to population growth causes prices to change continuously. Since this change is faster and more pronounced in large cities like Istanbul, it is crucial for homeowners and investors to have predictable prices. It is emphasized that this study is designed to meet this need. The industrial and academic importance of the problem is highlighted. Identifying the factors affecting house prices provides important information about a country's economic situation and contributes to the development of economic policies. Additionally, predicting house prices is important for investors as it reduces risks and allows for effective portfolio management. The accuracy of urban planning decisions can also be improved through these analyses.

The parameters and their definitions affecting the topic are examined. Factors affecting house prices generally consist of geographical, economic, social, and demographic factors. While location, infrastructure, and transportation are of great importance, the characteristics of the house itself are also among the influencing factors. Infrastructure, transportation, education, and health are significant factors affecting house prices, and providing these conditions in the region can influence prices. Economic conditions also play an important role; employment levels, interest rates, and inflation rates are economic indicators that affect prices. The balance of

supply and demand is another crucial factor influencing prices, with increased demand potentially causing prices to rise.

Solution methods, tools, and techniques are reviewed. Various machine learning methods are used to predict house prices. Techniques such as support vector regression and artificial neural networks are commonly used in this field and can help determine the factors affecting house prices. Additionally, to identify the models that provide the highest accuracy, the performances of different models are compared using various evaluation metrics. The results and recommendations are presented. The Random Forest model was found to perform better than other models, emerging as the most reliable model for predicting house prices. This study demonstrates that machine learning methods are effective tools for predicting house prices. Future studies aim to further improve performance by using larger and more diverse datasets and applying model optimization techniques.

BÖLÜM 1

1. GİRİŞ

1.1 Problemin Tanımı

Evler, insan yaşamının temel ihtiyaçlarından biridir ve nüfus artışıyla birlikte talep de artmaktadır. Ev fiyatlarını etkileyen faktörlerin çeşitliliği, ev sahipleri ve yatırımcılar tarafından bu faktörlerin bilinmesini önemli kılmaktadır. Özellikle İstanbul gibi büyük şehirlerde ev fiyatlarının hızlı değişimi, ev sahipleri ve yatırımcılar üzerinde olumsuz

etkilere sebep olmaktadır. Bu çalışma, İstanbul'daki ev fiyatlarının artışını belirlemek ve tahmin etmek amacıyla gerçekleştirilmektedir.

1.2 Problemin Endüstriyel/Akademik Önemi

Ev fiyatlarını etkileyen faktörlerin belirlenmesi ve fiyatların tahmin edilmesi, hem endüstriyel hem de akademik açıdan büyük önem taşımaktadır. Bu faktörlerin belirlenmesi, bir ülkenin ekonomik durumu hakkında önemli bilgiler sağlar. Bu analizler, ekonomik büyüme, gelir dağılımı ve işsizlik gibi değişkenlerin anlaşılmasına yardımcı olarak ekonomik politikaların geliştirilmesine katkı sağlar. Ev fiyatlarının tahmin edilmesi, yatırımcılar için önemlidir çünkü risklerini azaltır ve portföylerinin etkin bir şekilde yönetilmesine yardımcı olur. Ayrıca, fiyatları etkileyen faktörlerin analizi, kentsel planlama kararlarının doğru bir şekilde alınmasına katkı sağlar.

1.3 Konuya Etki Eden Parametreler ve Tanımları

Ev fiyatlarını etkileyen faktörler genellikle coğrafi, ekonomik, sosyal ve demografik faktörlerden oluşur. Konum, altyapı, ulaşım gibi faktörlere bağlı olarak büyük önem taşırken, evin özellikleri de fiyatları etkileyen unsurlardır. Altyapı, ulaşım, eğitim ve sağlık gibi faktörler, ev fiyatlarını etkileyen önemli unsurlardır ve bu koşulların bölgede sağlanması fiyatları etkileyebilir. Ekonomik koşullar da ev fiyatlarını etkileyen önemli faktörler arasındadır; istihdam seviyeleri, faiz oranları, enflasyon oranları gibi ekonomik göstergeler, fiyatları etkiler. Talep ve arz dengesi de fiyatları etkileyen önemli faktörlerdendir; talebin artması fiyatların yükselmesine neden olabilir.

1.4 Çözüm Yöntemleri/Araçlar/Teknikleri

Ev fiyatlarını tahmin etmek için çeşitli çözüm yöntemleri, araçlar ve teknikler bulunmaktadır. Destek vektör regresyonu ve yapay sinir ağları gibi yöntemler, bu alanda yaygın olarak kullanılan tekniklerdir ve ev fiyatlarını etkileyen faktörleri belirlemeye yardımcı olabilir. Ayrıca, en yakın komşu algoritması (KNN), karar ağaçları ve destek vektör makineleri de ev fiyatlarının tahmin edilmesinde kullanılan teknikler arasındadır. Hangi yöntemin kullanılacağı, mevcut veri setinin yapısına, hedefe ve diğer faktörlere bağlı olarak değişebilir.

BÖLÜM 2

2. LİTERATÜR ARAŞTIRMASI

Hamizah vd. (2020) yılında yaptıkları çalışmada makine öğrenmesi ile ev fiyat tahmini için bir yaklaşım geliştirmişlerdir. Evin insan yaşamının en temel ihtiyaçlarından biri olduğunu ve insanların yaşam standartlarının artmasıyla ev talebinin de hızla arttığından söz etmektedirler. Ev piyasasının bir ülkenin ekonomi ölçeğinde önemli bir etken olduğunu belirtmişlerdir. Ev talebindeki artışlar ev fiyatlarının da artmasına neden olmaktadır. Ev fiyatlarını etkileyen birçok faktör olduğundan ev konusundaki bütün paydaşların fiyatları etkileyen faktörleri bilmek istediklerini belirtmişlerdir. Bu doğrultuda ev fiyat tahminlerinin destek vektör regresyonu ve yapay sinir ağları gibi makine öğrenmesi metodları ile yapılabileceğinden bahsetmektedirler. Bu modellerle ev paydaşlarına mevcut pazarın ev fiyatlarını değerlendirme ve belirleme konusunda yardımcı olmayı hedeflemektedirler. Bu çalışmayı Endonezya' nın Jakarta kenti için yapmışlardır. Çalışmayı, evlerin karakteristik özellikleri belirlemek ve ev fiyat tahmini için kullanılacak model olarak ikiye ayırmışlardır. Ev fiyatını etkileyen faktörleri üç gruba ayırmışlar ve bunları konumsal faktörler, yapısal faktörler ve mahalle faktörleri olarak belirlemişlerdir. Her grubun altında ev fiyatını etkileyen faktörleri tespit etmişlerdir. Çalışmada bu faktörlerin sayısı 19' dur. Bu çalışmada makine öğrenmesi metodu olarak; bir mülkün değerinin sahip olduğu tüm özelliklerin toplamından oluştuğunu ve mülkün bütün özelliklerinin saptanıp pazar fiyatının belirlenmesini sağlayan Hedonik Fiyatlandırma Modeli, değişkenler arasındaki ilişkiyi belirleme için kullanılan Çoklu Regresyon Modeli, Destek Vektör Regresyonu, Yapay Sinir Ağları ve Gradyan Artırma metodları kullanılmıştır(Hamizah vd. 2020).

Thamarai ve Malarvizhi (2019) yılında makine öğrenmesi ile ev fiyat tahmini çalışmasında veri madenciliğinin büyük veri tabanlarından bilgi veya yararlı desen çıkarma işlemi olduğundan, sınıflandırmanın ise sınıf özneliği modelini bulmak için kullanılan veri madenciliği işlevi olduğundan bahsetmektedirler. Karar ağacının sınıflandırma ve tahmin amacıyla kullanılan bir metod olduğundan söz edilmiştir. Özellik Seçimi Ölçütlerinin, veri madenciliği için gereksiz ve tekrarlayan öznelilikleri ortadan kaldırmak için kullanılan bir süreç olduğundan söz etmektedirler. Bu süreçte Bilgi Kazanımı, Kazanç Oranı ve Gini Endeksi algoritmalarını kullanmaktadırlar(Thamarai

ve Malarvizhi 2019). Bu çalışmanın amacı, evlerin özelliklerine ve çevresindeki olanaklara dayanarak fiyatlarını tahmin etmeyi amaçlamaktadır. Evlerin fiyatlarını tahmin etmek için Karar Ağacı ve Çoklu Doğrusal Regresyon yöntemleri kullanılmıştır. Hindistan'ın West Godavari ilçesindeki Tadepalligudem adlı yerin analiz edilmesiyle bir veri seti hazırlanmıştır. Veri seti evlerin yatak odası sayısı, yaşı, ulaşım olanakları, çevredeki okullara uzaklıkları ve alışveriş olanakları gibi özellikler içermektedir. Çalışmada makine öğrenmesi için bir araç olan Scikit-Learn kütüphanesi kullanılmışlardır. Scikit-learn, Kümeleme algoritmaları, Regresyon, Sınıflandırma gibi algoritmalar ile tasarlanmıştır.

2.1 Makine Öğrenimi ile Konut Fiyatı Tahmini konusunda yapılan çalışmalarda kullanılan teknikler

Çalışma	Kullanılan Modeller											
	YSA	LR	K-NN	NB	KA	RO	Bayesian R.	Loj. R	Çoklu Regresyon	Torbalı A.	Gradyan Boosting	
1.Barut, vd. 2022	X	X										
2.Burhan, 2023			X	X	X	X				X		
3.Oral, vd. 2021	X				X							
4.Adetunjia, vd. 2021						X						
5.Ravikumar, vd. 2017	X					X			X			
6. Beimer, vd. 2019		X			X	X			X			

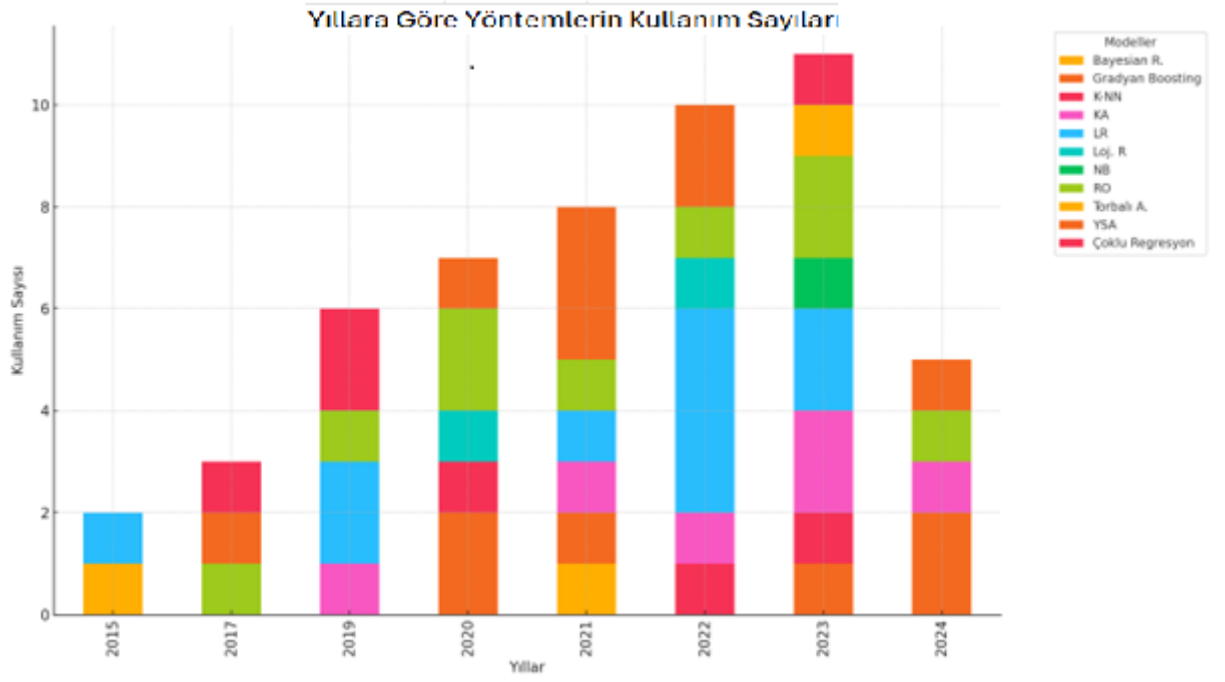
7. Mostofi, vd. 2021	X										
8. Kumar, 2019		X						X			
9. Dabreo, vd. 2021		X									X
10.Begu m, vd. 2022			X		X			X			
11. Yugesh, vd. 2023		X			X						
12. Basysyar , vd. 2022		X									
13. Modi, vd. 2020			X					X			
14. Chowha an, vd. 2023		X				X					X
15. Ng, 2015		X					X				
16. Foryś, 2022	X	X									
17. Hjort, vd. 2024											X
18. Zhang,	X						X				

vd. 2021											
19. Potrawa , vd. 2020	X					X					X
20. Soltani, vd. 2024	X				X	X					X
21. Truong, vd. 2020						X					X
22. Foryś, vd. 2022		X									
23. Lahmiri, vd. 2023									X		
24. Adetunji , vd.2022						X					

Çizelge1.1

Yapılan Çalışmalarda Kullanılan Teknikler

Bu tablo, konut fiyatlarının tahmin edilmesi için çeşitli makine öğrenimi modellerinin kullanıldığı çalışmaları özetlemektedir. Bu çalışmaların amacı, farklı algoritmaların ve yaklaşımların konut fiyatlarının daha doğru tahmin edilmesine nasıl katkıda bulunabileceğini incelemektir. Tablodaki çeşitli araştırmalar, farklı modellerin ve tekniklerin kullanıldığını göstermektedir.



Şekil 1.1

Yıllara Göre Yöntemlerin Kullanım Sayıları

Konut fiyatlarının tahmin edilmesi üzerine yapılan bazı çalışmalarda yalnızca birkaç model kullanılırken, diğerlerinde birçok model bir arada kullanılmıştır. Bu, farklı modellerin avantajlarını ve dezavantajlarını değerlendirmek ve en etkili tahmin yöntemini belirlemek için yapılan bir yaklaşımdır. Araştırmalarda çeşitli yaklaşımların ve algoritmaların incelendiğini göstermektedir.

2.2 Makine Öğrenimi ile Konut Fiyatı Tahmini konusunda yapılan çalışmalarda kullanılan parametreler

Para metr eler	M2	Bina Yaşı	Dem ogra fik Fakt örler	Ev Tipi	Oda Sayı sı	Asan sör	Eml akVe rgisi Oran ı	Bahç e	Kat Sayı sı-Num arası	Balk on	Kon um	Ulaşım
1. Barut, vd. 2022	X	X			X				X			
2. Burhan, 2023	X	X						X	X			
3. Oral, vd. 2021	X	X		X	X	X		X	X	X		
4. Adet unjia , vd. 2021			X		X							
5. Ravi kumar, vd. 2017			X		X						X	
6. Beim er, vd. 2019	X	X	X	X	X						X	X
7. Most ofi, vd. 2021	X	X			X				X	X	X	
8. Kum ar,		X	X	X	X							

2019												
9. Dabr eo, vd. 2021	X	X	X	X	X		X				X	X
10. Begu m, vd. 2022	X				X						X	X
11. Yuge sh, vd. 2023	X				X						X	
12. Basy syar, vd. 2022			X		X						X	X
13. Mod i, vd. 2020	X			X	X						X	X
14. Cho wha an, vd. 2023	X				X					X	X	
15. Ng, 2015	X	X			X						X	
16. Fory ś,	X										X	X

2022												
17. Hjort , vd. 2024	X	X	X		X						X	
18. Zhan g, vd. 2021			X								X	
19. Truo ng, vd. 2020	X		X		X						X	
20. Solta ni, vd. 2024	X	X			X						X	
21. Truo ng, vd. 2020	X	X	X		X	X	X				X	X
22. Fory ş, vd. 2022	X										X	X
23. Lah miri, vd. 2023		X									X	X

Çizelge 1.2

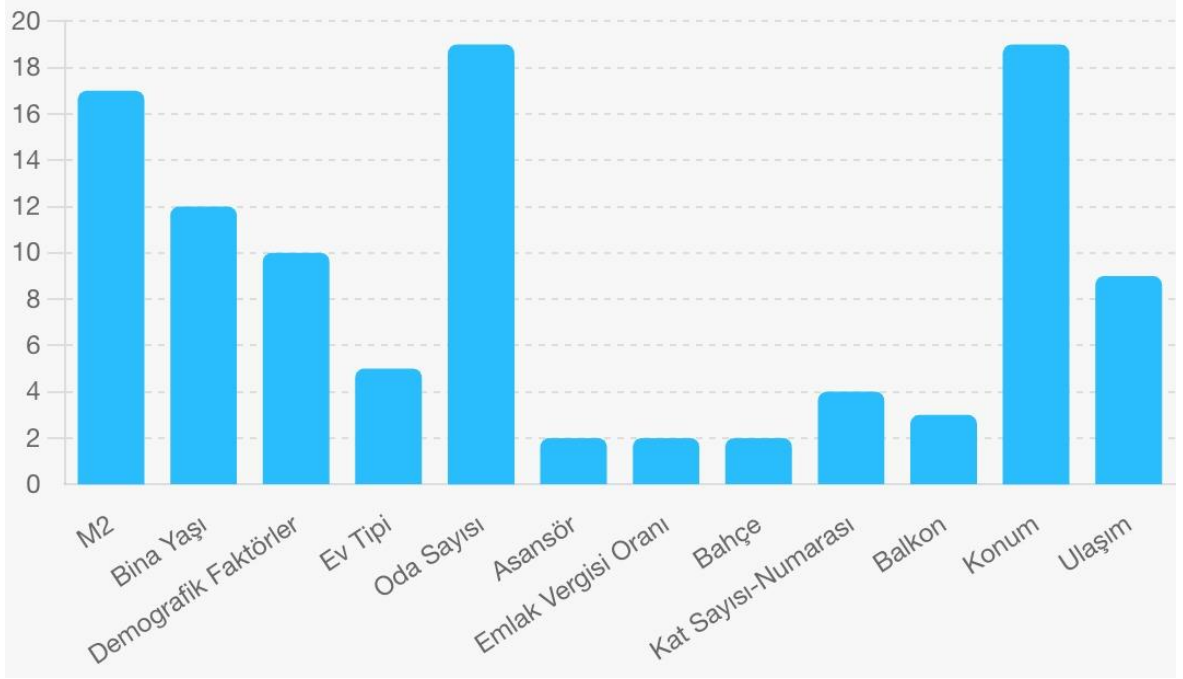
Yapılan Çalışmalarda Kullanılan Parametreler

Tablodaki parametrelerin çeşitli araştırmalarda kullanılması, konut fiyatlarının tahmin edilmesinde hangi faktörlerin önemli olduğunu anlamamıza yardımcı olabilir. Konutun alanı genellikle fiyatı belirlemede kritik bir faktördür. Daha büyük bir alana sahip evler genellikle daha yüksek fiyatlarla ilişkilendirilir. Bu parametre, konutun fiziksel boyutunu temsil eder ve genellikle bir konutun fiyatı üzerinde belirleyici bir etkiye sahiptir. Bir konutun yaşı, fiyatını belirlemede önemli bir rol oynar. Yeni yapılmış binalar genellikle daha yüksek fiyatlarla satılırken, eski binalar genellikle daha düşük fiyatlarla ilişkilendirilir. Bu, genellikle konutun durumu ve kalitesi ile ilişkilendirilen bir faktördür. Bir bölgedeki demografik özellikler, konut fiyatlarını etkileyebilir. Örneğin, gelir seviyesi, eğitim düzeyi ve yaş ortalaması gibi faktörler, bir bölgedeki konut talebini ve dolayısıyla fiyatlarını etkileyebilir. Bu parametreler, konut pazarının tüketici taleplerini ve tercihlerini yansıtabilir.

MODEL	KULLANILDIĞI ÇALIŞMA SAYISI	YÜZDE (%)
Lineer Regresyon (LR)	12	%50.0
Rastgele Orman (RO)	9	%37.5
Yapay Sinir Ağları (YSA)	8	%33.3
Karar Ağaçları (KA)	6	%25.0
K-En Yakın Komşu (K-NN)	5	%20.8
Gradyan Boosting	5	%20.8
Çoklu Regresyon	4	%16.7
Naive Bayes (NB)	3	%12.5
Bayesian Regresyon (Bayesian R.)	3	%12.5
Torbalı Ağaçlar (Bagging)	3	%12.5
Lojistik Regresyon (Loj. R)	2	%8.3

Çizelge 1.3

Kullanılan Tekniklerin Yüzdeleri



Şekil 1.2

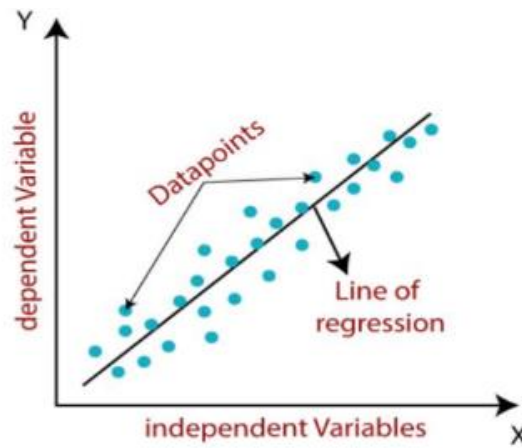
Kullanılan Parametrelerin Dağılımı

Konut tipi, fiyatı belirlemede önemli bir rol oynar. Örneğin, daireler genellikle daha uygun fiyatlıdır, villa veya müstakil evler ise daha yüksek fiyatlarla satılır. Bu parametre, konutun tasarımı, konumu ve özellikleriyle ilgili genel bir kategorizasyon sağlar. Konutun oda sayısı, genellikle fiyatı belirlemede önemli bir faktördür. Daha fazla odaya sahip konutlar genellikle daha yüksek fiyatlarla satılırken, daha az odaya sahip olanlar daha uygun fiyatlı olabilir. Bu, bir konutun genel kullanılabilirliği ve alanı ile ilişkilendirilir. Asansör varlığı, genellikle konutun fiyatını artırır, özellikle de yüksek katlı binalarda. Bu, konutun yaşam kalitesini ve konforunu artırabilecek bir özellik olarak kabul edilir.

3. KULLANILAN YÖNTEMLER

3.1 Lineer Regresyon

Doğrusal Regresyon, bir regresyon görevi gerçekleştiren bir makine öğrenimi algoritmasıdır. Genellikle değişkenler arasındaki ilişkiyi bulmak ve tahmin yapmak için kullanılır. Önerilen Doğrusal Regresyon yönteminin iş akışını göstermektedir. Doğrusal regresyon, Şekil'de gösterildiği gibi, verilen bir bağımsız değişkenin (x) bağımlı değişkenin (y) değerini tahmin etmek için kullanıldığı görevi gerçekleştirir. Bu nedenle, Doğrusal Regresyon kullanarak x (girdi) ve y (çıkı) arasında doğrusal bir ilişki bulur(Kheya vd. 2022).

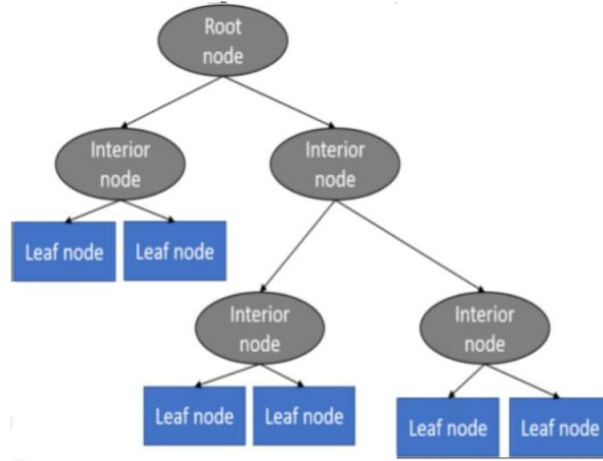


Şekil 1.3

Lineer Regresyon Grafiği (Kaynak: Housing Price Prediction with Machine Learning, 2022)

3.2 Karar Ağacı

Karar Ağacı, en yaygın kullanılan uygulamalardan biridir. Hem Regresyon hem de Sınıflandırma görevlerini çözmek için kullanılabilir. Belirli bir veri noktasıyla ağacın tamamında dolaşarak, ağacın yaprağına ulaşana kadar Doğru/Yanlış sorularına cevap verir. Yaprak düğümdeki bağımlı değişkenin nihai tahmini, ortalama değerdir. Veri noktası için birden fazla yineleme ile, Ağaç doğru bir ölçüm tahmin edebilir. Şekil 4, önerilen Karar Ağacı yönteminin iş akışını temsil etmektedir(Kheya vd. 2022).



Şekil 1.4

Karar Ağacı Düğümleri (Kaynak: Housing Price Prediction with Machine Learning, 2022)

3.3 Rastgele Orman

Rastgele orman algoritması hem sınıflandırma hem de regresyon tahmini yapmak için kullanılabilir, bu nedenle regresyon ormanları olarak da adlandırılır. Ana süreç, verilerin rastgele seçimi ve değişkenlerin rastgele seçimi temelinde birçok karar ağacı geliştirilmesi ve bu ağaçlara dayanarak bağımlı değişkenin sınıfını sağlamaktır. Bu algoritmanın veri setinde kullanılmasının başlıca avantajı, eksik değerleri ele alabilmesi ve eksik verilerin doğruluğunu koruyabilmesidir. Modelin aşırı uyum yapma olasılığı düşüktür ve büyük ölçekli veri setlerine uygulandığında yüksek boyutluluğu işleyebilir. Regresyon ağaçlarında, sonuç sürekli olacaktır(Akande vd. 2022).

3.4 Gradient Boosting

Gradient boosting hem regresyon hem de sınıflandırma için kullanılabilir. Gradient boosting, regresörlerden oluşan regresyon modelleri üretmek için bir tekniktir. Bu fikir, regresyon için tekrar tekrar izlenen bir prosedürün uygulanmasıdır. Burada, verilerin basit regresyon tahminicisini öğreniyoruz, ardından hata kalıntısını hesaplıyoruz. Her bir veri noktası başına hata miktarını belirliyoruz ve hata kalıntısını tahmin etmek için yeni bir model öğreniyoruz. Ana konsept, bir dizi tahmin yapmak, hataları bulmak ve bu hataları azaltmaktır(Ganjisaffar vd. 2011).

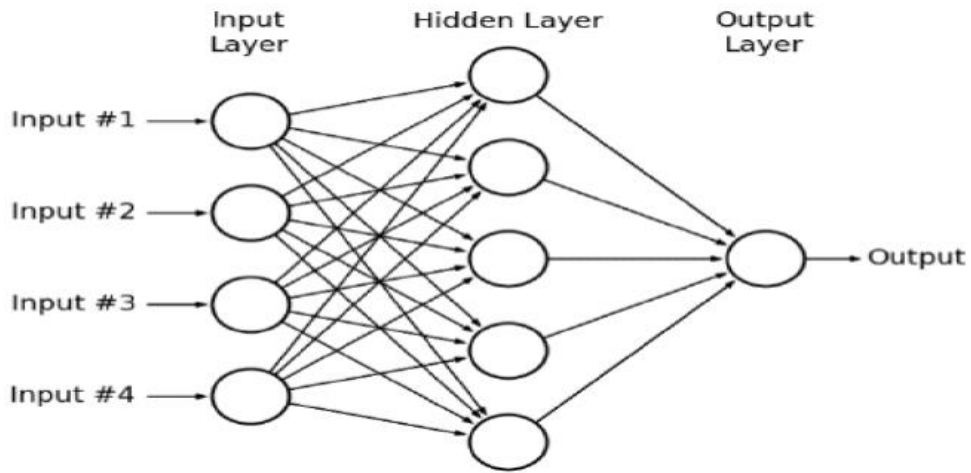
3.5 k-En Yakın Komşu (k-EYK) Algoritması

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 aldığını 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).

3.6 Sinir Ağları

MLP, çok katmanlı algılayıcı (multilayer perceptron) anlamına gelir ve yapay sinir ağlarının bir parçasıdır. Tek katmanlı algılayıcının yapısına sahiptir, ancak bir veya daha fazla gizli katmana sahiptir. Bu yapıda, gizli katman doğrudan giriş katmanına bağlanır. Giriş değerleri algılayıcılara sunulur ve algılayıcılar, herhangi bir doğrusal olarak ayrılabilir girdi kümesini sınıflandırır. Algılayıcıya sunulan giriş değerleri, tahmin edilen çıktı istenen çıkışla aynı ise performans tatmin edici kabul edilir ve ağırlıklarda değişiklik yapılmaz. Eğer eşleşme sağlanmazsa hatayı azaltmak için ağırlıkların değiştirilmesi gerekir (Koskela ve diğerleri; 1996).



Şekil 1.5

Sinir Ağları(Kaynak: House price Prediction,1996)

4. UYGULAMA

4.1 Verilerin Toplanması

Bu çalışmada, İstanbul'daki ev fiyatlarını analiz etmek ve tahmin etmek amacıyla çeşitli veriler toplanmıştır. Verilerin toplanma süreci şu adımlardan oluşmaktadır:

Veri Kaynaklarının Belirlenmesi: İlk olarak, ev ilanlarına ait verilerin toplandığı kaynaklar belirlenmiştir. Bu kaynaklar, popüler emlak siteleri ve bu sitelerde yer alan ilanlardır. Bu siteler, geniş veri tabanları ve düzenli olarak güncellenen içerikleri ile güvenilir veri sağlayıcıları olarak seçilmiştir.

Veri Toplama Aracı ve Yöntemi: Verilerin toplanması için Python programlama dili kullanılarak bir web scraping (web kazıma) scripti yazılmıştır. BeautifulSoup kütüphanesi kullanılarak, ilan sayfalarındaki gerekli bilgileri çekmek için bir otomasyon oluşturulmuştur. Script, belirli bir tarih aralığında yer alan ilanları hedef almış ve ilanların detay sayfalarındaki bilgileri sistematik olarak toplamıştır.

Verilerin Filtrelenmesi ve İşlenmesi: Web scraping sürecinde elde edilen ham veriler, veri temizleme ve ön işleme adımları ile işlenmiştir. İlan başlıkları, fiyat bilgileri, konum koordinatları ve ilan detayları gibi önemli bilgiler JSON formatında toplanmıştır. Fiyat bilgileri ve ilan detayları, ilan sayfalarındaki HTML etiketlerinden ve JSON verilerinden çekilmiştir.

Veri Çerçevesinin Oluşturulması: Elde edilen ham veriler, pandas kütüphanesi kullanılarak bir veri çerçevesine (DataFrame) dönüştürülmüştür. Bu veri çerçevesi, ilan başlığı, fiyat, koordinatlar ve diğer ilan detaylarını içermektedir. Verilerin düzenlenmesi ve analiz edilebilmesi için uygun yapıya getirilmiştir.

Veri Depolama ve Güncelleme: Toplanan ve işlenen veriler, Excel dosyaları olarak saklanmıştır. Her bir ilan detayının eksiksiz ve doğru bir şekilde kaydedilmesi için düzenli olarak güncellemeler yapılmış ve ilan detayları Excel dosyasına yazılmıştır.

Eksik Verilerin Tamamlanması ve Doğrulama: Toplanan verilerde eksiklik veya hata olması durumunda, eksik veriler manuel olarak kontrol edilmiş ve tamamlanmıştır. Doğrulama adımları ile verilerin doğruluğu ve tutarlılığı sağlanmıştır.

Sonuç olarak, İstanbul'daki ev fiyatlarını tahmin etmek amacıyla kullanılan veri seti, yukarıda belirtilen adımlar izlenerek oluşturulmuştur. Verilerin toplanması sürecinde kullanılan yöntemler, veri kalitesini artırmak ve analizlerin doğruluğunu sağlamak amacıyla titizlikle uygulanmıştır. Bu veri seti, çalışmanın temelini oluşturarak, ev fiyatlarını etkileyen faktörlerin belirlenmesi ve fiyat tahmin modellerinin geliştirilmesi için kullanılmıştır.

4.2 Verilerin Temizlenmesi

Excel üzerinde topladığımız veriler web sitelerinden aldığımız ev ilan detayları ve bu ev ilanlarına ait market, ulaşım, eczane, okul gibi bölgelere uzaklığını göstermektedir. Şekil 1.6’ da gösterildiği gibi topladığımız ev ilanlarının 25’ er adet paratmereye sahiptir. Bu parametrelere ait veri girişleri veri aldığımız web sitelerinin kullanıcıları tarafından yapılmaktadır. Dolayısıyla bazı verilerin girişleri, zorunlu giriş olmadığı için yapılmamaktadır. Veri girişi yapılmayan parametreler dosyamızdan silinmiştir.

İlan Başlığı	Fiyat	Koordinatların Numarası	İncelleme	Kategorisi	Metrekare	Oda Sayısı	Bulunduğu Kat	Kısıtma Tipi	Diye Uygun	Durum	İçerisinde	Aldat	Banyo Sayısı	Balkon Sayısı	WC Sayısı	Ülke	Gezilebilir	Yatırım
ESA Gayri 3050000 TL	41.189614921615	30 Nisan 2 Satılık	80 M2	3+1	1.Kat	Kombi Doğ Krediye Uy Sıfır	Hayır	100 TL	2	1	2	Hayır	27 Mart 2021					
ANADOLU 3300000 TL	41.183214921936	27 Mart 2 Satılık	126 M2	3+1	2.Kat	Kombi Doğ Krediye Uygun	Hayır	1					27 Mart 2021					
Avçılarda 4600000 TL	41.002614922570	30 Nisan 2 Satılık	150 M2	3+2	3.Kat	Kombi Doğ Krediye Uygun	Hayır	2					27 Mart 2021					
ESA Gayri 5500000 TL	41.190114922597	30 Nisan 2 Satılık	180 M2	6+2	Çatı Duple	Kombi Doğ Krediye Uy Sıfır	Hayır	100 TL	4		4		27 Mart 2021					
AVCILAR C900000 TL	40.990414923704	04 Nisan 2 Satılık	95 M2	2+1	2.Kat	Kombi Doğ Krediye Uy Sıfır	Hayır	1			1	Hayır	27 Mart 2021					
AVCILAR A950000 TL	40.998514923713	04 Nisan 2 Satılık	135 M2	3+1	2.Kat	Merkezi D Krediye Uy Sıfır	Hayır	1	1	1	Hayır	27 Mart 2021						
METROBÜ 5750000 TL	40.990014923943	27 Mart 2 Satılık	135 M2	4+2	3.Kat	Kombi Doğ Krediye Uygun	Hayır	2					27 Mart 2021					
BAĞCILAR 7000000 TL	41.036114925080	28 Mart 2 Satılık	120 M2	3+1	1.Kat	Kombi Doğ Krediye Uygun	Hayır	2					28 Mart 2021					
BAĞCILAR Y:3500000 TL	41.039514925109	28 Mart 2 Satılık	100 M2	3+1	1.Kat	Kombi Doğ Krediye Uygun	Hayır	1					28 Mart 2021					
CUMHURİ 3300000 TL	41.017214925142	01 Nisan 2 Satılık	65 M2	2+1	3.Kat	Kombi Doğ Krediye Uygun Değil	Hayır	1					28 Mart 2021					

Şekil 1.6

Ev İlan Detayları

Literatür araştırmalarımız ve pazar araştırmalarımız sonucunda ev fiyatına etki eden parametreler tek tek incelenmiş olup ev fiyatına etki etmeyen parametreler ilgili dosyamızdan çıkarılmıştır. Bunun sonucunda Şekil 1.7’ de görülen dosya oluşturulmuştur. Bu dosyada da görüldüğü gibi toplanan verilerden kullanılmak üzere; fiyat, koordinatlar, net metrekare, oda sayısı, bulunduğu kat, banyo sayısı, binanın yaşı, site içerisinde olup olmadığı parametreleri seçilmiştir.

İlan Numarası	İlan Başlığı	Fiyat	Koordinatlar	Net Metrekare	Oda Sayısı	Bulunduğu Kat	Banyo Sayısı	Binanın Yaşı	Site İçerisinde
14125193	%70 PEŞİN %30 ELDEN TAKSİTLE	1100000 TL	('lat': 41.08203979465168, 'lon': 28.6629675 M2	1+1	5.Kat	1	0 (Yeni)	Hayır	
14125883	GoldHouse dan Satılık Yeşilköy d	4100000 TL	('lat': 40.959152199762556, 'lon': 28.820445 M2	1+1	Bahçe Katı	1	4	Hayır	
14134709	Başakşehir huzzak Tworde 1+1 s:	3600000 TL	('lat': 41.0696003081825, 'lon': 28.77449968 M2	1+1	6.Kat	1	4	Evet	
14138756	Bayrampaşa Yıldırım mahallesinc	6500000 TL	('lat': 41.06553288091083, 'lon': 28.8870387 M2	2+1	4.Kat	1	0 (Yeni)	Hayır	
14138886	Bayrampaşa Yıldırım mahallesinc	8700000 TL	('lat': 41.06582611011002, 'lon': 28.88730165 M2	3+2	Çatı Dupleks	2	0 (Yeni)	Hayır	
14142304	BAHÇEŞEHİRDE EBEVEYN BANYO	2549000 TL	('lat': 41.073261565017, 'lon': 28.6640157120 M2	3+1	1.Kat	2	4	Hayır	
14143334	BAHÇEŞEHİR LOKASYONUNDA EE	4690000 TL	('lat': 41.082733936299356, 'lon': 28.6684130 M2	3+1	3.kat	2	0 (Yeni)	Hayır	
14147143	DG HOMEART'TAN ACARKENT'TE	12500000 TL	('lat': 41.11592576777344, 'lon': 29.13405750 M2	9+ Oda		6+	0 (Yeni)	Evet	

Şekil 1.7

Temizlenmiş Ev İlan Detayları

4.3 Verilerin Düzenlenmesi

Kullanılmayan parametreler ve gereksiz verilerin çıkarılması sonrasında verilerin normalizasyonu gerekmektedir. Bu normalizasyon öncesinde sütunlarda bulunan

verilerin sayısal değerlere çevirilmesi gerekmektedir. Bu değerlere çevrilmeden önce sütunlarda bulunan yazı karakterleri ortadan kaldırılmıştır. Fiyat sütunundan ‘TL’ net metrekare sütunundan ‘M2’, bulunduğu kat sütunundan ‘.kat’ ve diğer yazılar, binanın yaşı sütunundan ‘(Yeni)’ yazıları kaldırılmıştır.

Koordinatlar sütunundan sadece enlem ve boylam verilerinin sayısal kısımları alınmıştır. Bu veriler ayrı sütunlara yazdırılmıştır. Oda sayısı sütununda bulunan oda sayısı ve ‘+’ ile gösterilen ek oda değerleri toplanmıştır ve yeni bir değer oluşturulmuştur. Site içerisinde sütununda bulunan ‘Evet’ ve ‘Hayır’ değerleri 1 ve 0 olacak şekilde değiştirilmiştir. Bu verilerin düzenlenmesi sonucunda Şekil 1.8’ de görülen dosya oluşturulmuştur.

İlan Numarası	lat	lon	Net Metrekare	Oda Sayısı	Bulunduğu Kat	Banyo Sayısı	Binanın Yaşı	Site İçerisinde	Fiyat
14226393	40.997688519481	28.840638175606728	230	6	2	3	0	0	6000000
14231845	40.97693607172903	28.713810592889786	120	4	2	2	0	0	3600000
14236935	41.01003712484566	28.85212776142089	120	5	2	2	0	0	4000000
14237585	41.0538660222639	28.83165458724975	100	4	2	2	3	1	8200000
14239081	41.00693636351042	28.860680980625386	185	6	2	2	0	0	20000000
14239808	40.97581623887031	28.718247301876545	110	3	2	2	2	0	4500000
14242516	41.011567573160306	28.833284148245294	145	5	2	2	0	0	4000000
14242612	41.010402563791594	28.834928787427494	75	3	2	1	0	0	3000000

Şekil 1.8

Düzenlenmiş Ev İlan Detayları

İlan detaylarına ait market, eczane, okul, ulaşım gibi ihtiyaç noktalarına uzaklık farklı bir dosyada tutulmaktadır. Her bir ilanının bu ihtiyaç noktalarına uzaklıkları toplanmış ve ortalaması alınmıştır. Bu değerler ilan detayları dosyasında bir sütun oluşturularak, ilan numaraları ile eşleştirilip dosyaya aktarılmıştır.

4.4 Verilerin Normalize Edilmesi

Verileri normalize edebilmek için temizleme ve düzenleme işlemleri yapılmıştır. Sonrasında normalize etme işlemi için Şekil 1.9’ da bulunan formül kullanılmıştır.

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Şekil 1.9

Normalizasyon formülü

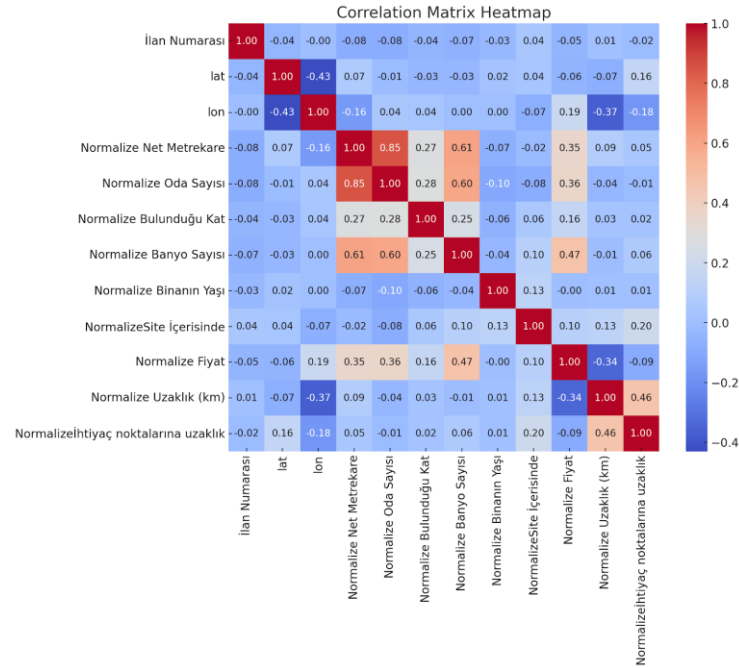
Sütunlarda bulunan verilerin normalizasyonu sonucunda Şekil 1.10' da bulunan dosya oluşturulmuştur. Bu dosya korelasyon matrisini oluşturmak ve diğer makine öğrenmesi metodlarında veriyi kullanabilmek için son haline ulaşmıştır. Bütün veriler 1 ile 0 arasında değerlerden oluşmaktadır.

İlan Numar	lat	lon	Net Metrekare	Oda Sayısı	Bulunduğu Ka	Banyo Sayısı	Binanın Yaşı	Site İçerisin	Fiyat	Uzaklık (km)	İhtiyaç noktalarına uzaklık
14226393	40.99769	28.84064	0.9111111111	0.714285714	0.666666667	0.666666667	1	0	0.105691057	0.139144301	0.088154536
14231845	40.97694	28.71381	0.4222222222	0.428571429	0.666666667	0.666666667	0	0	0.056910569	0.270286214	0.151658664
14236935	41.01004	28.85213	0.4222222222	0.571428571	0.666666667	0.666666667	0	0	0.06504065	0.126802529	0.110744521
14237585	41.05387	28.83165	0.3333333333	0.428571429	0.666666667	0.666666667	0.75	1	0.150406504	0.159547232	0.071253784
14239081	41.00694	28.86068	0.7111111111	0.714285714	0.666666667	0.666666667	0	0	0.390243902	0.118124179	0.055583135
14239808	40.97582	28.71825	0.3777777778	0.285714286	0.666666667	0.666666667	0.5	0	0.075203252	0.266083697	0.058425555
14242516	41.01157	28.83328	0.5333333333	0.571428571	0.666666667	0.666666667	0	0	0.06504065	0.145946068	0.106157989

Şekil 1.10

Normalize Edilmiş Ev İlan Detayları

4.5 Korelasyon Matrisi



Şekil 1.11

Korelasyon Matrisi

Korelasyon Matrisi Analizi

Bu çalışmada, İstanbul'daki ev fiyatlarını etkileyen faktörler üzerinde bir korelasyon analizi gerçekleştirilmiştir. Korelasyon matrisi, her bir değişkenin diğer değişkenlerle olan doğrusal ilişkisini ortaya koymaktadır. Bu bölümde, korelasyon matrisinden elde edilen bulgular detaylı bir şekilde yorumlanacaktır.

Genel Gözlemler: Korelasyon matrisi, değişkenler arasındaki ilişkileri -1 ile +1 arasında bir değerle ifade eder. Pozitif değerler, iki değişken arasında doğrudan bir ilişki olduğunu gösterirken, negatif değerler ters bir ilişki olduğunu belirtir. 0'a yakın değerler ise iki değişken arasında anlamlı bir ilişki olmadığını ifade eder.

Ev Fiyatı ile İlişkili Değişkenler:

Net Metrekare (0.35): Ev fiyatları ile net metrekare arasında pozitif bir korelasyon vardır. Bu, evin büyüklüğünün artmasıyla fiyatın da arttığını göstermektedir. Bu ilişki beklenen bir durumdur, çünkü daha büyük evler genellikle daha yüksek fiyatlara satılmaktadır.

Oda Sayısı (0.36): Oda sayısı ile ev fiyatları arasında da pozitif bir korelasyon bulunmaktadır. Oda sayısının artması, evin fonksiyonelliğini ve konforunu artırdığı için fiyat üzerinde olumlu bir etki yapmaktadır.

Banyo Sayısı (0.47): Banyo sayısının ev fiyatları üzerindeki etkisi de oldukça belirgindir. Birden fazla banyoya sahip evler, özellikle büyük aileler için daha çekici olabilmekte ve bu da fiyatı artırmaktadır.

Diğer Değişkenler Arasındaki İlişkiler:

Net Metrekare ve Oda Sayısı (0.85): Evin net metrekare alanı ile oda sayısı arasında güçlü bir pozitif korelasyon bulunmaktadır. Bu, daha büyük evlerin genellikle daha fazla odaya sahip olduğunu göstermektedir.

Banyo Sayısı ve Oda Sayısı (0.61): Banyo sayısı ile oda sayısı arasında da pozitif bir ilişki bulunmaktadır. Bu, daha fazla odalı evlerin genellikle daha fazla banyoya sahip olduğunu göstermektedir.

Uygulamalı Yorumlar: Bu korelasyon analizi, ev fiyatlarını etkileyen temel faktörleri belirlemede önemli bir araçtır. Özellikle net metrekare, oda sayısı, banyo sayısı gibi fiziksel özellikler ile fiyat arasında güçlü pozitif ilişkiler görülmektedir. Bu bulgular, yatırımcılar ve gayrimenkul geliştiricileri için stratejik kararlar alırken dikkate alınmalıdır. Ayrıca, kentsel planlama ve altyapı yatırımlarında bu faktörlerin göz önünde bulundurulması, daha doğru ve etkin politikaların geliştirilmesine yardımcı olacaktır.

Sonuç olarak, korelasyon matrisi ev fiyatlarını tahmin etmede ve bu fiyatları etkileyen faktörleri anlamada değerli bir araçtır. Bu analiz, gelecekte yapılacak olan derinlemesine çalışmalar için sağlam bir temel oluşturmaktadır.

4.6 Yöntemlerin Uygulanması

Bu bölümde, ev fiyatlarını tahmin etmek için kullanılan makine öğrenmesi yöntemlerinin uygulanması ve elde edilen sonuçlar detaylı bir şekilde açıklanacaktır. Çalışmamızda çeşitli makine öğrenmesi algoritmaları kullanılmıştır. Kullanılan kütüphaneler ve yöntemler aşağıda sırasıyla açıklanmıştır.

Kullanılan Kütüphaneler

Pandas:

Veri manipülasyonu ve analizi için kullanılır.

CSV, Excel, SQL gibi çeşitli dosya formatlarını okuma ve yazma işlemlerini destekler.

Veri temizleme, yeniden şekillendirme ve gruplama gibi işlemleri kolaylaştırır.

Kaynak: Pandas Documentation

(https://pandas.pydata.org/docs/user_guide/index.html)

NumPy:

Büyük, çok boyutlu diziler ve matrislerle çalışmayı sağlar.

Yüksek performanslı matematiksel ve mantıksal işlemler sunar.

Diğer veri bilimi kütüphanelerinin temelini oluşturur.

Kaynak: NumPy Documentation (<https://numpy.org/doc/stable/index.html>)

Scikit-learn:

Sınıflandırma, regresyon, kümeleme gibi makine öğrenmesi algoritmaları sunar.

Model seçimi, çapraz doğrulama ve hiperparametre optimizasyonu için araçlar sağlar.

Kullanıcı dostu API ve geniş topluluk desteği ile öğrenme ve uygulama süreçlerini destekler.

Kaynak: Scikit-learn Documentation. (<https://scikit-learn.org/stable/>)

4.6.1 Veri Setinin Hazırlanması

Veri seti, ev fiyatlarını içeren bir Excel dosyasından yüklenmiştir. Bu veri seti üzerinde analiz ve modelleme yapabilmek için Pandas kütüphanesi kullanılmıştır. İlk olarak, gerekli kütüphane import edilmiştir ve ardından Excel dosyası okunarak bir DataFrame oluşturulmuştur. Özellikler (X) ve hedef değişken (y) olarak iki ayrı değişken seti oluşturulmuştur. Özellikler, ilan numarası ve fiyat dışında kalan tüm sütunları içerirken, hedef değişken sadece fiyat sütununu içerir.

```
import pandas as pd

# Excel dosyasını yükle
```



```
file_path = 'ev_fiyatlari_uzaklikli_v2_normalize.xlsx'
df = pd.read_excel(file_path)

# Özellikler ve hedef değişken
X = df.drop(columns=['İlan Numarası', 'Fiyat'])
y = df['Fiyat']
```

Makine öğrenmesi modellerinin performansını değerlendirmek için veri setini eğitim ve test olarak ikiye ayırmak gerekir. Scikit-learn kütüphanesinin `train_test_split` fonksiyonu kullanılarak veri seti %80 eğitim ve %20 test olarak ayrılmıştır. Bu şekilde model, eğitim verisi üzerinde öğrenme işlemi gerçekleştirirken, test verisi üzerinde modelin genel performansı değerlendirilir.

```
from sklearn.model_selection import train_test_split

# Eğitim ve test verisi ayırma
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                                    random_state=42)
```

4.6.2 Model Listesi ve Başlatılması

Bu bölümde, Scikit-learn kütüphanesinden çeşitli regresyon modelleri kullanılmıştır. Bu modeller arasında Doğrusal Regresyon, Destek Vektör Makineleri, Karar Ağaçları, Rastgele Ormanlar, Gradient Boosting, K-En Yakın Komşu ve Yapay Sinir Ağları bulunmaktadır. Her bir model, ilgili sınıfın bir örneği olarak başlatılmış ve bir sözlük yapısında saklanmıştır.

```
from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.neural_network import MLPRegressor

# Modellerin listesi ve başlatılması
models = {
```

```
"Linear Regression": LinearRegression(),
"Support Vector Machine": SVR(),
"Decision Tree": DecisionTreeRegressor(random_state=42),
"Random Forest": RandomForestRegressor(random_state=42),
"Gradient Boosting": GradientBoostingRegressor(random_state=42),
"K-Nearest Neighbors": KNeighborsRegressor(),
"Neural Network": MLPRegressor(random_state=42, max_iter=500)
}
```

4.6.3 Modellerin Değerlendirilmesi

Her bir model, eğitim verisi ile eğitilmiş ve test verisi ile değerlendirilmiştir. Model performansını değerlendirmek için dört farklı metrik kullanılmıştır:

Ortalama Mutlak Hata (MAE): Tahmin edilen değerler ile gerçek değerler arasındaki mutlak hataların ortalamasıdır. Düşük MAE, modelin tahminlerinde daha doğru olduğunu gösterir.

Ortalama Kare Hatası (MSE): Tahmin edilen değerler ile gerçek değerler arasındaki hataların karelerinin ortalamasıdır. MSE, hataları daha büyük değerlerle cezalandırdığı için, büyük hataların etkisini daha belirgin hale getirir.

Ortalama Mutlak Yüzde Hatası (MAPE): Tahmin edilen değerler ile gerçek değerler arasındaki mutlak yüzdesel hataların ortalamasıdır. MAPE, tahminlerin doğruluğunu yüzdesel olarak gösterir ve genellikle yüzdelik hata olarak ifade edilir.

R-Kare (R²) Skoru: Modelin açıklayabildiği toplam varyasyonun yüzdesidir. R² skoru 1'e ne kadar yakınsa, modelin tahmin gücü o kadar yüksektir.

Model performansını değerlendirmek için gerekli olan metrikler `sklearn.metrics` modülünden import edilmiştir.

Çalışmamızda kullanılan farklı yöntemleri değerlendirmek için yazılmış olan fonksiyon, verilen bir makine öğrenmesi modelini değerlendirir ve modelin performansını dört farklı metrik kullanarak ölçer. Fonksiyon dört parametre alır:

model: Değerlendirilecek makine öğrenmesi modeli

X_train: Eğitim verisinin özellikleri

X_test: Test verisinin özellikleri

y_train: Eğitim verisinin hedef değişkeni

y_test: Test verisinin hedef değişkeni

```
from sklearn.metrics import mean_absolute_error, mean_squared_error,
mean_absolute_percentage_error, r2_score

# Modelleri değerlendirme fonksiyonu
def evaluate_model(model, X_train, X_test, y_train, y_test):
    model.fit(X_train, y_train) # Modeli eğitim verisi ile eğitir
    y_pred = model.predict(X_test) # Test verisi üzerinde tahminler yapar
    mae = mean_absolute_error(y_test, y_pred) # Ortalama Mutlak Hata (MAE)
    hesaplar
    mse = mean_squared_error(y_test, y_pred) # Ortalama Kare Hatası (MSE)
    hesaplar
    mape = mean_absolute_percentage_error(y_test, y_pred) # Ortalama Mutlak
    Yüzde Hatası (MAPE) hesaplar
    r2 = r2_score(y_test, y_pred) # R-Kare (R2) Skoru hesaplar
    return mae, mse, mape, r2 # Bu metrikleri geri döner

# Her modeli değerlendir ve sonuçları sakla
results = {}
for name, model in models.items():
    mae, mse, mape, r2 = evaluate_model(model, X_train, X_test, y_train,
    y_test)
    results[name] = [mae, mse, mape, r2]
```

4.6.4 Sonuçların Görselleştirilmesi

Model performans sonuçları, bir veri çerçevesine dönüştürülmüş ve görselleştirilmiştir. Aşağıda her model için elde edilen performans metrikleri sunulmuştur.

```
# Sonuçları daha iyi görselleştirmek için DataFrame'e dönüştür
results_df = pd.DataFrame(results, index=['MAE', 'MSE', 'MAPE', 'R2']).T
```

```
import ace_tools as tools; tools.display_dataframe_to_user(name='Model
Evaluation Results', dataframe=results_df)

print(results_df)
```

Sonuçların DataFrame'e Dönüştürülmesi: results sözlüğünde saklanan performans metrikleri, pandas kütüphanesi kullanılarak bir DataFrame'e dönüştürülmüştür. Bu işlem, sonuçların daha kolay analiz edilebilmesi ve görselleştirilebilmesi için yapılır. index parametresi, DataFrame'in satır başlıklarını belirlerken, .T metodu, DataFrame'in transpozunu alarak satır ve sütunların yer değiştirmesini sağlar. Veri Çerçevesinin Görselleştirilmesi: ace_tools kütüphanesi kullanılarak oluşturulan DataFrame'in görselleştirilmesi sağlanır. tools.display_dataframe_to_user fonksiyonu, results_df DataFrame'ini kullanıcıya görsel olarak sunar. Bu sayede, her model için elde edilen performans metrikleri açık bir şekilde görüntülenebilir. DataFrame'in Konsolda Yazdırılması: print(results_df) ifadesi, results_df DataFrame'ini konsola yazdırarak sonuçların metin formatında da görüntülenmesini sağlar.

Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Mean Absolute Percentage Error (MAPE)	R-squared (R2)
Linear Regression	0.049225	0.006890	0.524071	0.368713
Support Vector Machine	0.075986	0.009985	0.545471	0.085157
Decision Tree	0.028033	0.003709	0.265983	0.660166
Random Forest	0.022727	0.002179	0.221836	0.800310
Gradient Boosting	0.029257	0.002972	0.284613	0.727719
K-Nearest Neighbors	0.033456	0.004561	0.556632	0.582132
Neural Network	0.046120	0.006394	0.712045	0.414182

Çizelge 1.4

Yöntemlerin Performans Değerleri

Yapılan analizler sonucunda, Rastgele Orman (Random Forest) modelinin diğer modellere göre daha iyi performans gösterdiği tespit edilmiştir. Rastgele Orman modeli, en düşük Ortalama Kare Hatası (MSE) ve en yüksek R2 skoru ile en doğru tahminleri sağlamıştır. Diğer modeller de makul performans sergilemiş olsa da,

Rastgele Orman modeli ev fiyatlarını tahmin etmede en güvenilir model olarak öne çıkmaktadır.

Bu çalışmanın sonuçları, makine öğrenmesi yöntemlerinin ev fiyatlarını tahmin etme konusunda etkili araçlar olduğunu göstermektedir. Gelecek çalışmalarda, daha büyük ve çeşitli veri setlerinin kullanılması ve model optimizasyon tekniklerinin uygulanması ile performansın daha da artırılması hedeflenmektedir.

BÖLÜM 5

6. SONUÇ VE ÖNERİLER

6.1 Sonuç

Bu çalışmada, normalize edilmiş veri setlerini kullanarak çeşitli makine öğrenmesi modellerinin ev fiyatlarını tahmin etme performanslarını analiz ettik. Değerlendirilen modeller arasında Linear Regression, Support Vector Machine, Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors ve Neural Network modelleri bulunmaktadır. Amacımız, farklı değerlendirme metriklerine dayanarak en doğru tahminleri hangi modelin sağladığını belirlemektir.

Analizimiz, Random Forest modelinin tüm diğer modellerden daha iyi performans gösterdiğini ortaya koymuştur. Random Forest modelinin Ortalama Kare Hatası (MSE) 0.002179 ve R-kare değeri 0.800310 olarak bulunmuştur. Bu sonuç, Random Forest modelinin ev fiyatlarındaki varyansın yaklaşık %80.03'unu açıklayabildiğini göstermektedir ve bu da onu test edilen modeller arasında en güvenilir model yapmaktadır. Decision Tree ve Gradient Boosting modelleri de sırasıyla 0.660166 ve 0.727719 R-kare değerleri ile güçlü performans sergilemiştir.

Linear Regression ve Support Vector Machine (SVM) modelleri ise diğer modellere göre daha düşük performans sergilemişlerdir. Linear Regression modelinin MSE değeri 0.006890 ve R-kare değeri 0.368713 iken, SVM modelinin MSE değeri 0.009985 ve R-kare değeri yalnızca 0.085157 olarak bulunmuştur. Bu modellerin düşük performansı, ev fiyatları gibi karmaşık ve çok değişkenli bir problemin doğrusal olmayan ilişkilerini yakalamakta yetersiz kalabileceğini göstermektedir. K-Nearest Neighbors (KNN) modeli

de, 0.004561 MSE ve 0.582132 R-kare deęerleri ile makul bir performans sergilemiř, ancak Random Forest, Decision Tree ve Gradient Boosting modelleri kadar yksek bir performans gsterememiřtir. KNN modelinin yksek MAPE (Mean Absolute Percentage Error) deęeri (0.556632), bu modelin bazı tahminlerde nemli sapmalar gsterebileceęine iřaret etmektedir.

Neural Network modeli ise 0.006394 MSE ve 0.414182 R-kare deęerleri ile orta dzeyde bir performans sergilemiřtir. Bu model, doęrusal olmayan iliřkileri yakalayabilme potansiyeline sahip olmasına raęmen, veri setimizin byklę ve modelin karmařıklıęı nedeniyle dięer gçl modellerin gerisinde kalmıřtır.

Her bir model iin ayrıntılı performans metrikleri ařaęıdaki tabloda sunulmuřtur:

Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Mean Absolute Percentage Error (MAPE)	R-squared (R2)
Linear Regression	0.049225	0.006890	0.524071	0.368713
Support Vector Machine	0.075986	0.009985	0.545471	0.085157
Decision Tree	0.028033	0.003709	0.265983	0.660166
Random Forest	0.022727	0.002179	0.221836	0.800310
Gradient Boosting	0.029257	0.002972	0.284613	0.727719
K-Nearest Neighbors	0.033456	0.004561	0.556632	0.582132
Neural Network	0.046120	0.006394	0.712045	0.414182

Genel olarak, bu çalışmanın sonuçları, karmaşık ve çok değişkenli problemler için ensemble yöntemlerinin (özellikle Random Forest ve Gradient Boosting) üstün performans sergilediğini göstermektedir. Bu modeller, veri setimizdeki çeşitli değişkenlerin etkileşimlerini ve doğrusal olmayan ilişkilerini etkili bir şekilde yakalayabilmiştir. Sonuç olarak, ev fiyatlarının tahmin edilmesinde en güvenilir ve doğruluk oranı yüksek modeller olarak Random Forest ve Gradient Boosting öne çıkmaktadır.

Gelecek çalışmalar için, modellerin performansını daha da artırmak amacıyla daha büyük ve çeşitli veri setlerinin kullanılması, model optimizasyon tekniklerinin (örneğin, hyperparameter tuning) uygulanması ve model interpretability (yorumlanabilirlik) üzerinde daha fazla durulması önerilmektedir. Ayrıca, farklı bölgeler ve piyasa koşulları altında modellerin performansını incelemek, modellerin genellenabilirliğini ve gerçek dünya uygulamalarındaki güvenilirliğini değerlendirmek açısından önemli olacaktır.

KAYNAKLAR

- [1] Aydemir, E., Aktürk, C., & Yalçinkaya, M. A. (2020). Yapay zekâ ile konut fiyatlarının tahmin edilmesi. *Turkish Studies - Applied Sciences*, 15(2), 183-194.
- [2] Burhan, H. A. (2023). Konut fiyatları tahmininde makine öğrenmesi sınıflandırma algoritmalarının kullanılması: Kütahya kent merkezi örneği. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 76, 221-237.
- [3] Naga Satish, G., Raghavendran, Ch. V., Sugnana Rao, M. D., & Srinivasulu, Ch. (2019). House price prediction using machine learning. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(9), 717-721.
- [4] Thamarai, M., & Malarvizhi, S. P. (2020). House price prediction modeling using machine learning. *I.J. Information Engineering and Electronic Business*, 2020(2), 15-20.
- [5] Oral, M., Okatan, E., & Kırbaş, İ. (2021). Makine öğrenme yöntemleri kullanarak konut fiyat tahmini üzerine bir çalışma: Madrid örneği. 3rd International Young Researchers Student Congress, 09-11 December 2021, Burdur, Turkey.
- [6] Zulkifley, N. H., Abdul Rahman, S., Ubaidullah, N. H., & Ibrahim, I. (2020). House price prediction using a machine learning model: A survey of literature. *I.J. Modern Education and Computer Science*, 2020(6), 46-54.
- [7] Beimer, J., & Francke, M. (2019). Out-of-sample house price prediction by hedonic price models and machine learning algorithms. *ResearchGate*.

- [8] Mostofi, F., Toğan, V., & Başağa, H. B. (2021). House price prediction: A data-centric aspect approach on performance of combined principal component analysis with deep neural network model. *Journal of Construction Engineering, Management & Innovation*, 4(2), 106-116.
- [9] Kumar, A. (2019). House rent price prediction. *International Research Journal of Engineering and Technology (IRJET)*, 6(4), 3188.
- [10] Dabreo, S., Rodrigues, S., Rodrigues, V., & Shah, P. (2020). Real estate price prediction. In *Proceedings of Fr. Conceicao Rodrigues College of Engineering*.
- [11] Begum, A., Kheya, N. J., & Zahid, Z. R. (2022). Housing price prediction with machine learning. *International Journal of Innovative Technology and Exploring Engineering*.
- [12] VN, M., Yugesh, J., NL, G., & Reddy, M. (2023). House price prediction using linear regression. *International Research Journal of Modernization in Engineering Technology and Science*, 5(4).
- [13] Basysyar, F. M., & Dwilestari, G. (2022). House price prediction using exploratory data analysis and machine learning with feature selection. *Acadlore Transactions on Machine Learning*, 1(1), 11-21.
- [14] Modi, M., Sharma, A., & Madhavan, P. (2020). Applied research on house price prediction using diverse machine learning techniques. *International Journal of Scientific & Technology Research*, 9(4), 371.
- [15] Chowhaan, M. J., Nelli, S., Shaik, S., [diğer yazarlar]. (2023). Machine learning approach for house price prediction. *Asian Journal of Research in Computer Science*, 16(2).
- [16] Ng, A. (2015). Machine learning for a London housing price prediction mobile application. Individual Project: Final Report. Imperial College London, Department of Computing.
- [17] Burhan, H. A. (2023). Konut Fiyatları Tahmininde Makine Öğrenmesi Sınıflandırma Algoritmalarının Kullanılması: Kütahya Kent Merkezi Örneği. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 76, 221-237. DOI: 10.51290/dpusbe.1249461
- [18] Adetunjia, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2021). House Price Prediction using Random Forest Machine Learning Technique. *Procedia Computer Science*, 00, 000–000.