

Research Article

Received: date:22.10.2024

Accepted: date:24.12.2024

Published: date:31.12.2024

Makine Öğrenmesi Algoritmalarıyla Konut Fiyatlarının Tahmini: Giresun Örneği

Mehmet Akif KARA^{1*}

¹Giresun University, Department of Business Administration, Giresun Türkiye; akifkara28@gmail.com

Orcid: 0000-0003-4308-9933¹

*Correspondence: akifkara28@gmail.com

Özet: İnşaat sektörü son 20 yılın en önemli sermaye birikim alanları arasında yer almaktadır. Konut satışları da bu alanın önemli bir parçasıdır. Konut fiyatlarının seyrinin tahmin edilmesi hem üretici hem de tüketiciler açısından önemlidir. Makine öğrenmesi algoritmaları tahmin çalışmalarında sıklıkla kullanılmakta ve iyi performans göstermektedir. Bu çalışmada Giresun örneği üzerinde 365 veri kullanılarak konut değerlendirme tahmini MATLAB 2023b yazılım programı kullanılarak çözümlenmektedir. Çalışmada RMSE, MAE, R^2 , MSE metrik değerleri kullanılarak makine öğrenmesi algoritmaları tahmin başarılarına göre incelenmektedir. Çalışmada konut değerlendirmek için kullanılan "brüt m², net m², bina yaşı, bulunduğu kat, toplam kat, site içinde olup olmama, asansör, ısıtma, açık ve kapalı otopark," değişkenleri kullanılmıştır. Veriler Türkiye'de yaygın olarak kullanılan ve en çok ziyaret edilen bir e ticaret web sitesinden elde edilmiştir. Çalışma sonucunda Gaussian Process Regression algoritmasının daha iyi bir performans gösterdiği ortaya koyulmuştur. Bu algoritmanın bu veri seti üzerinde R^2 değeri %89 olarak çok yüksek bulunmuştur.

Keywords: makine öğrenmesi, konut değerlendirme, tahmin.

Prediction of Housing Prices Using Machine Learning Algorithms: The Case of Giresun

Abstract: The construction sector has been one of the most significant areas for capital accumulation over the past 20 years. Housing sales are also an important part of this sector. Predicting the trend of housing prices is crucial for both producers and consumers. Machine learning algorithms are frequently used in predictive studies and have shown good performance. This study focuses on housing valuation predictions using 365 data points from Giresun, analyzed with MATLAB 2023b software. The study examines the predictive success of machine learning algorithms using metrics such as RMSE, MAE, R^2 , and MSE. The variables used for housing valuation include "gross m², net m², building age, floor location, total floors, whether it's within a site, presence of an elevator, heating, open and closed parking." The data was obtained from a popular e-commerce website widely used in Turkey. The results indicate that the Gaussian Process Regression algorithm performs better, achieving a high R^2 value of 89% on this dataset.

Keywords: machine learning, residential property valuation, forecasting.

1. Giriş

Yapay Zeka'nın gelişimi ile birlikte geliştirilen birçok algoritma tahmin çalışmalarında da kullanılmaktadır. Makine Öğrenmesi algoritmaları bu bağlamda YZ (AI) yöntemi olarak tercih edilmektedir [1]. Makine öğrenmesi, giriş verileri ile çıkış verileri arasında doğrusal bir formül ile ifade edilemeyen problemleri

çözebilmek için geliştirilmiş olan güncel esnek hesaplama sistemlerine verilen genel bir adlandırmadır [2]. Makine öğrenmesi 3 alt sınıflandırma içermektedir. Bunlar; denetimli, denetimsiz ve yarım denetimli algoritmalarıdır. Denetimli öğrenme modelleri genel olarak sınıflandırma ve regresyon problemlerinde yoğun olarak kullanılırken, denetimsiz öğrenme modelleri ise kümeleme alanında daha çok uygulanmaktadır [3]. Makine öğrenmesi, bilgisayarların verilerden öğrenmesini ve belirli görevleri gerçekleştirmesini sağlamaktadır. Geleneksel programlama yöntemlerinin aksine, bilgisayara belirli kural ya da algoritma yerine, verilerle beslenerek kendi kendine öğrenme yeteneği kazandırılmaktadır.

Bu çalışmanın motivasyonu Türkiye başta olmak üzere gelişmekte olan ülkeler için önemli bir sermaye birikim alanı olan inşaat sektörünün dinamosu olan konut piyasalarının durumunu makine öğrenmesi algoritmalarıyla ortaya koymaktır. Klasik yöntemler genel itibariyle piyasa trendlerini ve tarihsel verileri göz önünde bulundurmaktadır. Makine öğrenmesi algoritmaları ise daha derinlemesine ve veri odaklı bir yaklaşım sunmaktadır. Bu bağlamda büyük veri setlerinden ilişkileri öğrenerek daha doğru tahmin yapma yeteneğine sahip olmaktadır. Ayrıca fiyat tahminlerinin doğruluğunu artırmak için farklı faktörleri de analiz edebilmektedir. Konut değerlendirme bir yanıyla son yılların en önemli sermaye birikim alanı olan inşaat sektörü bir yanıyla da barınma hakkı kavramsallaştırması bağlamında bireyleri doğrudan ilgilendiren bir konudur [4]. Barınma hakkı bağlamında tartışıldığında konutların fiyatlarının ve bu fiyatlara göre sağlıklı, insani ve sosyal olarak yaşanabilir durumda olmaları gerekmektedir [5]. Sermaye birikim açısından özellikle neoliberal politikaların kullanıldığı günümüzde Türkiye’de büyüme politikasının temel araçlarından birisi inşaat piyasasıdır [6]. Bu açıdan birbiriyle bağlantılı çok sayıda sektörü ilgilendirmektedir. İşletmeler, yöneticiler, karar alıcılar ve politika yapıcılar arasında da ilgi odağı haline gelmektedir. Dolayısı ile konut fiyatlarının takip edilmesi önemlilik arz etmektedir.

Çalışmanın amacı makine öğrenmesi algoritmalarıyla konut fiyat tahmini gerçekleştirmektedir. Algoritmaların belirli metrikler dâhilinde bu problem için nasıl performans gösterecekleri de ikincil amaçtır. Çalışmanın örneklemini Giresun ili Merkez ilçesi oluşturmaktadır. Giresun ili gibi yerel piyasalar çeşitli ekonomik, sosyal ve çevresel faktörlerin etkisi altında şekillenmektedir. Dolayısı ile Giresun ilindeki konut fiyatlarını etkileyen çeşitli unsurları belirlemek ve bu unsurları makine öğrenmesi algoritmaları kullanarak fiyat tahmininde kullanmak amaçlanmaktadır. Çalışma verilerinin Giresun ili Merkez ilçede 3+1 daireler ile kısıtlanması ise çalışmanın kısıtlılığını oluşturmaktadır. Çalışmada öncelikli olarak makine öğrenmesi algoritmaları ele alınmış, ardından literatür incelemesi ile alandaki boşluk ortaya koyulmuş ve uygulama gerçekleştirilmiştir. Bu sayede yalnızca yerel konut piyasasına ilişkin bilgiler elde edilmekle kalınmayacak aynı zamanda makine öğrenme algoritmalarının pratik uygulaması da ortaya konulacaktır. Çalışma sonucunda elde edilmesi muhtemel bulguların yerel yönetim ve yatırımcılar başta olmak üzere ilgili paydaşlar açısından stratejik kararlar alınırken yol göstereceği düşünülmektedir.

2. Makine Öğrenmesi

Günümüzde büyük verilerin giderek artış göstermesi ile birlikte bu verilerden yeni çıkarımlar elde edebilmek için makine öğrenmesi algoritmaları kullanılmaktadır. Bu alanda kullanılan algoritmalar, büyük veri kümelerindeki ilişki ve kalıpları tespit edebilmek için eğitilmektedir. Böylece en iyi karar ve tahminleri yapabilmektedir. Makine öğrenmesi uygulamaları, sahip oldukları veri miktarı ile orantılı olarak gelişmekte ve daha fazla veri ile daha iyi sonuçlar üretmeleri sağlanmaktadır. Alana ilişkin literatür incelendiğinde; makine öğrenmesinin üç ana kategoride ele alındığı görülmektedir [7],[8]. Bunlar; denetimli, denetimsiz ve yarı denetimli öğrenme olarak adlandırılmaktadır.

Denetimli öğrenme, makine öğreniminde en yaygın ve popüler yöntemlerden biridir. Bu yöntemde, her eğitim verisi bir çıktı etiketi üretmektedir. Sistem, eğitim süreci aracılığıyla sorunları öğrenir ve belirli bir hedef değişkenin tahmin edilmesi amacıyla etiketli veriler kullanarak model oluşturur. Bu süreç, girdiler (özellikler) ile çıktılar (hedef değişken) arasındaki ilişkileri anlamak için çeşitli algoritmalar kullanır. Denetimli öğrenme, genellikle sınıflandırma ve regresyon problemleri için tercih edilir [9].

Denetimsiz makine öğrenmesi ise, etiketlenmemiş veriler üzerinde içgörü ve ilişkilerin keşfedilmesine odaklanır. Bu yöntemde eğitim sırasında etiketlenmemiş veriler kullanılır ve sistemin, bu verilerden kalıpları keşfetmesi beklenir. Sonuçlar belirsizdir ve sistemden istenen çıktılar önceden bilinmemektedir. Dolayısıyla, bu model eğitim sırasında doğru cevaplarla yönlendirilmemiştir [10].

Yarı denetimli makine öğrenmesi ise, denetimli ve denetimsiz öğrenmenin bir kombinasyonunu temsil eder. Bu yöntemde bazı veriler etiketlenirken, bazıları etiketlenmez. Amaç, etiketlenmemiş küçük bir veri kümesinin etiketli veri ile bir arada kullanılmasıyla öğrenmede belirgin bir iyileşme sağlamaktır. Bu yaklaşım, her iki yöntemin avantajlarını birleştirerek daha etkili sonuçlar elde etmeyi hedefler [3].

Bu çalışmada denetimli makine öğrenmesi algoritmaları kullanılmıştır. Bu algoritmalar;

Destek Vektör Makinaları (SVM): SVM genellikle sınıflandırma problemlerinin çözümünde kullanılan gözetimli öğrenme yöntemlerinden birisidir. Bu yöntemde, bir düzlem üzerine yerleştirilmiş noktaları ayırmak için bir doğru çizer. Bu doğrunun, iki sınıfının noktaları için de maksimum uzaklıkta olmasını amaçlar. Karmaşık ama küçük ve orta ölçekteki veri setleri için uygundur. SVM, el yazısı tanıma, çevre bilimleri, genelleştirilmiş tahmine dayalı kontrol uygulamaları, görüntü sınıflandırma vb. konularda uygulanmaktadır. SVM, genellikle sınıflandırma görevleri için kullanılsa da, regresyon problemlerinde de uygulanabilir. Destek Vektör Regresyonu (SVR), verileri en iyi ayıran hiper düzlemi bulmayı amaçlar ve bu süreçte hata marjını minimize etmeye odaklanır [10].

Rastgele Orman (Random Forest): Rastgele ormanlar veya rastgele karar ormanları, eğitim zamanında çok sayıda karar ağacı oluşturularak ve sınıfların modu (sınıflandırma) veya ortalama/ortalama tahmini olan sınıfın çıktısını alarak çalışan sınıflandırma, regresyon ve diğer görevler için bir topluluk öğrenme yöntemidir. Rastgele ormanlar genellikle karar ağaçlarından daha iyi performans gösterir, ancak doğrulukları gradyan destekli ağaçlardan daha düşüktür [11].

Doğrusal Regresyon (Linear Regression): Doğrusal regresyon, titizlikle çalışılan ve pratik uygulamalarda yaygın olarak kullanılan ilk tip regresyon analizidir. Bunun nedeni, bilinmeyen parametrelerine doğrusal olarak bağlı modellerin, parametreleriyle doğrusal olmayan ilişkili modellere göre daha kolay uydurulması ve elde edilen tahmin edicilerin istatistiksel özelliklerinin belirlenmesinin daha kolay olmasıdır [12].

Karar Ağaçları (Decision Trees): Karar ağaçları, regresyon problemlerinde sürekli hedef değişkenleri tahmin etmek için kullanılır. Veriyi ağaç yapısına bölerek, her düğümde belirli kriterlere dayalı olarak kararlar alır. Bu yaklaşım, veri setindeki karmaşıklıkları yakalamada özellikle etkilidir [13].

Gaussian Process Regression: Bu istatistiksel model, verilerin sürekli bir dağılımını modellemek için kullanılır. Tahminlerindeki belirsizlikleri dikkate alarak sonuçları bir dağılım şeklinde sunar. Tahminler için güven aralıkları belirlemede özellikle etkilidir [14].

Ensemble Bagged Tree: Bu yöntem, birden fazla karar ağacını bir araya getirerek tahmin performansını artırmaya dayanır. Ağaçlar, verinin farklı alt kümeleri üzerinde eğitilir; bu da genel model performansını artırır ve aşırı uyum riskini azaltır [15].

Bu algoritmaların her biri, veri setinin özelliklerine ve ele alınan probleme bağlı olarak farklı avantajlar sunar. Uygun algoritmanın seçimi, genellikle bu faktörlere bağlı olarak değişir ve en iyi sonuçları veren algoritmayı belirlemek için deneme-yanılma süreci ile karşılaştırma gerektirir.

3. Literatür İncelemesi

Literatürde konut fiyatlarının tahmini üzerine özellikle ekonometrik yöntemlerle yapılmış çalışmalar fazlalık göstermektedir. Buna karşı makine öğrenmesi algoritmalarıyla yapılan çalışmalar sınırlı sayıdadır. Literatür incelemesinde Web of Science ve TRDizin veri tabanlarına bakılmıştır. Hem bu veri tabanlarında etki değeri yüksek dergilerin yayımlanması ve ilgili çalışmaların atıf sayılarının yüksek olması tercih sebebidir.

Burhan, çalışmasında bağımlı değişken olarak konut fiyatını, bağımsız değişkenler olarak net m², bina yaşı, bulunduğu kat, site içinde olup olmaması gibi faktörleri almıştır. Çalışmada Naive Bayes yöntemi dışındaki k-En yakın komşu, karar ağaçları ve rastgele orman algoritmalarının doğruluk oranlarının %60 üzerinde olduğu ortaya koyulmuştur. En başarılı tahmin yönteminin ise Rastgele Orman algoritması olduğu görülmüştür [16]. Oral vd. çalışmalarında 11 kriter kullanarak konut fiyatı tahmini için makine öğrenmesi yöntemlerini kullanmıştır. Çalışmada model değerlendirmesi için RMSE, MSE ve R² değerleri kullanılmıştır. Çalışma sonucunda en iyi performans değerine sahip yöntem Bagged Tree Ensemble olmuştur [17]. Doğan vd. çalışmalarında Ankara Keçiören ilçesi örnekleme dahilinde 11 kriter içeren 20 farklı yapay sinir ağları modelini kullanarak fiyat tahmininde bulunmuştur. Çalışma sonucunda YSA yaklaşımının fiyat tahmininde başarılı sonuç verdiği görülmektedir [18]. Kayakuş vd. çalışmalarında makine öğrenmesi algoritmalarından Rastgele Orman, Yapay Sinir Ağları ve DVM yöntemlerini kullanmıştır. 9 değişken ile Ocak 2013-Kasım 2020 arası aylık verileri içeren modeller çerçevesinde konut metrekare fiyatlarını tahmin etmiştir. Üç yöntemde başarılar

sonuçlar vermiştir [19]. Özdemir vd. Sakarya ilinde yer alan 4 ilçe için DÖ, Polinomsal Regresyon ve Rastgele Orman algoritmalarını kullandığı çalışmalarında 7 değişken içeren modeller çerçevesinde konut fiyat tahmininde bulunmuştur. Çalışma sonucunda en başarılı tahmini DÖ yöntemi ile elde etmişlerdir [20]. Garcia vd. Bu araştırma, konut fiyatlarını tahmin etmek için en iyi makine öğrenmesi algoritmalarını belirlemeyi ve COVID-19 pandemisinin bir İspanyol şehrindeki konut fiyatları üzerindeki etkisini nicel olarak ölçmeyi amaçlamaktadır. Metodoloji, veri hazırlama, özellik mühendisliği, hiperparametre eğitimi ve optimizasyonu, model değerlendirme ve seçimi ile nihayetinde model yorumlama aşamalarını içermektedir. Boosting tabanlı topluluk öğrenme algoritmaları (Gradient Boosting Regressor, Extreme Gradient Boosting ve Light Gradient Boosting Machine) ile bagging tabanlı algoritmalar (rastgele orman ve extra-trees regressor) kullanılarak bir doğrusal regresyon modeli ile karşılaştırılmaktadır. Araştırma, COVID-19 pandemisi ilan edilmeden önce ve sonra Alicante (İspanya) emlak pazarına ait coğrafi referanslı mikro verilerle birlikte, kadastro, sosyo-demografik ve ekonomik göstergeler ve uydu görüntüleri gibi diğer tamamlayıcı kaynaklardan elde edilen bilgilerle bir vaka çalışması geliştirmektedir. Sonuçlar, makine öğrenmesi algoritmalarının, emlak pazarı verileri gibi karmaşık verilerin doğrusal olmayan özelliklerine daha iyi uyum sağladıkları için geleneksel doğrusal modellerden daha iyi performans gösterdiğini ortaya koymaktadır. Bagging tabanlı algoritmalar (rastgele orman ve extra-trees regressor) aşırı öğrenme sorunları gösterirken, boosting tabanlı algoritmalar daha iyi performans ve daha düşük aşırı öğrenme göstermektedir [21]. Soltani vd. çalışmalarında çeşitli değişkenlerin konut fiyatı varyasyonları üzerindeki etkilerini keşfetmek için makine öğrenmesi algoritmalarını kullanmıştır. Avustralya'nın Metropolitan Adelaide şehrine ait 32 yıllık konut fiyatı veri seti ile çalışmada 428000 satış işlem kaydı ve 38 açıklayıcı değişken kullanmıştır. Gradient-Boosting ve Rastgele Orman gibi topluluk makine öğrenmesi tekniklerinin daha iyi sonuç verdiğini ortaya koymuştur. Modellerin tahmin başarılarını artırmak için mekansal-zamansal gecikme değişkeni eklenmiş ve başarılı sonuç vermiştir. Çalışma sonucunda bu tür değişkenlerin Makine öğrenmesi uygulamalarında yararlı bir moderatör olacağı vurgulanmıştır [22]. Park ve Bae çalışmalarında ABD'de bulunan Virginia Fairfax şehrindeki 5359 apartman dairesinin konu verilerini analiz etmiştir. Yazarlar, makine öğrenmesi algoritmalarına dayalı bir konut tahmin modeli geliştirmektedir. RIPPER algoritmasının konut fiyat tahmininde diğer modellere göre daha başarılı performans gösterdiğini ortaya koymuştur [23]. Truong vd. Çalışmalarında konut fiyat tahmini için farklı modelleri incelemektedir. Rastgele Orman, XGBoost ve Light GBM gibi üç farklı makine öğrenmesi algoritması ile Hibrit Regresyon ve Yığın Genelizasyon Regresyonu (Stacked Generalization Regression) gibi iki tekniği karşılaştırmıştır. Çalışma sonucunda yazarlar, kullanılan yöntemlerden arzu edilen sonuçlar elde ettiğini vurgulamıştır. Ancak her modelin kendi avantaj ve dezavantajlarını ortaya koymuşlardır. Özellikle zaman karmaşıklığı vurgusu yapılmaktadır [24]. Akay vd. Çalışmalarında ARIMA modeli ile Rassal Orman ve Hibrit Rassal Orman yöntemlerini kullanarak Türkiye Konut Fiyat Endeksi serisi için öngörü performansları karşılaştırılmıştır. Çalışma bulgularına göre hibrit modelin daha iyi performans gösterdiği görülmektedir [25].

4. Materyal ve Yöntem

Çalışmada kullanılan veri seti, Türkiye'de en fazla ziyaret edilen bir satış web sitesi/uygulamasından elde edilmiştir. Veri toplama süreci 13.10.2024-16.10.2024 tarihlerini kapsamaktadır. Bu tarih aralığında web sitesinde bulunan ilanlar değerlendirmeye dâhil edilmiştir. Çalışmada örneklem olarak Giresun Merkez ilçesi seçilmiştir. Giresun'un Doğu Karadeniz Bölgesi'nde en fazla göç veren illerden birisi olması, konut fiyatlarının diğer bölge illerine göre yüksek seviyede olması tercih nedenidir. Aynı zamanda çalışmada 3+1 daireler tercih edilmiş ve sınırlandırma yapılmıştır. Bu bağlamda Giresun İli Merkez ilçesinde 440 ilana ulaşılmıştır. Daha sonra veriler ön işlem sürecinden geçirilmiştir. Farklı ilanlarda farklı emlak şirketleri tarafından paylaşıldığı tespit edilmiş ve veri setinde sadece bir tanesi kullanılmıştır. Analizde kullanılan veri sayısı 365 olarak tamamlanmıştır. Çalışmada kullanılan ve bağımsız değişken olarak ele alınan kriterler literatürde yapılan benzer çalışmalardan elde edilmiştir. Bu çalışmalardan farklı olarak Giresun ili özelinde yaşanan "araç parkı sorunu" bağlamında açık otopark, kapalı otopark gibi değişkenler de kriter olarak alınmış ve toplam kriter sayısı 11 olarak belirlenmiştir. Bağımlı değişken olarak konut fiyatları, bağımsız değişkenler olarak ise brüt m², net m², bina yaşı, bulunduğu kat, toplam kat sayısı, site içinde olup olmama, ısıtma, asansör, açık otopark, kapalı otopark olup olmaması olarak belirlenmiştir. Çalışmada MATLAB2023b yazılım programında yer alan Regression Learner Toolbox aracı kullanılmıştır. Öncelikle; veriler ön işleminden geçirilmiştir. Kategorik veriler

Tablo 2’de gösterildiği biçimde kodlanmış ve programa uygun hale getirilmiştir. Regression Learner Toolbox kendi içerisinde normalizasyon işlemini yaptığı için ayrıca normalizasyona ihtiyaç duyulmamıştır.

4.1. Model Performansını Değerlendirme

Makine öğrenmesi çalışmalarında model performansları çeşitli metriklerle bakılarak değerlendirilmektedir. Bu metrikler arasında en sık kullanılanları, açıklamaları ve matematiksel eşitlikleri Tablo 1’de gösterilmektedir.

Tablo 1. Çalışmada kullanılan metrikler ve açıklamaları

Metrik	Açıklama	Eşitlik
R^2	Modeldeki bağımsız değişkenlere göre bağımlı değişkenin varyasyon oranını ölçer. Regresyon modelinin çok fazla bağımsız değişkeni varsa test verilerinde istenilen başarıyı göstermez. Bu durumlarda Düzeltilmiş R^2 kullanılır.	$1 - \frac{\sum_{t=1}^{ntest} (x_t - \hat{x}_t)^2}{\sum_{t=1}^{ntest} (x_t)^2}$
MSE	Gerçek değerler ile tahmin edilen değerler arasındaki hataların karesinin ortalaması ile tanımlanır	$\frac{1}{ntest} \sum_{t=1}^{ntest} (x_t - \hat{x}_t)^2$
RMSE	MSE’nin kareköküdür.	$\sqrt{\frac{1}{ntest} \sum_{t=1}^{ntest} (x_t - \hat{x}_t)^2}$
MAE	Gerçek değerler ile tahmin edilen değerler arasındaki hataların mutlak değerlerinin ortalaması ile tanımlanır.	$\sum_{i=1}^{ntest} \frac{ x_t - \hat{x}_t }{ntest}$

R^2 için 1’e en yakın değeri alması beklenirken MSE, RMSE ve MAE için en düşük değeri veren tahmin başarılı sayılacaktır.

5. Bulgular

Çalışmada kullanılan veri setinin ilk 10 değeri Tablo 2’de gösterilmektedir. Tablo 2’de yer alan değerler MATLAB programına uygun hale getirilmiştir. Değişkenler arasında yer alan “bulunduğu kat”, “bina yaşı”, “site içinde bulunup bulunmama”, “ısıtma”, “asansör”, “açık otopark” ve “kapalı otopark” değişkenlerine sayısal karşılık atanmıştır ve analize uygun hale getirilmiştir.

Tablo 2. Veri seti (ilk 10 değer)

Brüt m2	Net m2	Bulunduğu Kat	Toplam Kat Sayısı	Bina Yaşı	Site İçinde	Isıtma	Asansör	Açık Otopark	Fiyat	Kapalı Otopark
170	160	bahçe katı	7	2	evet	dgaz	var	var	5950000	yok
165	150	3	7	0	hayır	dgaz	var	var	3350000	yok
140	135	5	9	1	hayır	dgaz	var	var	3850000	yok
140	110	1	10	0	hayır	dgaz	var	var	2900000	var
158	150	2	10	0	evet	dgaz	var	var	3500000	yok
145	128	3	3	1	evet	dgaz	var	var	3975000	yok
165	156	7	7	5-10 arası	evet	merkezi	var	var	3500000	yok
130	125	2	5	5-10 arası	evet	dgaz	var	var	2550000	yok
160	130	7	10	5-10 arası	evet	merkezi	var	var	4000000	var
135	125	9	10	5-10 arası	hayır	dgaz	var	var	4000000	yok

İlk olarak veri seti MATLAB yazılım programında analize uygun hale getirilmiştir. Örneğin; “site içinde olma” kriterine 1 atanırken “site içinde olmama kriterine” 0 değerleri atanmıştır. Daha sonra program içerisinde bulunan Regression Learner Toolbox ile analiz gerçekleştirilmiştir.

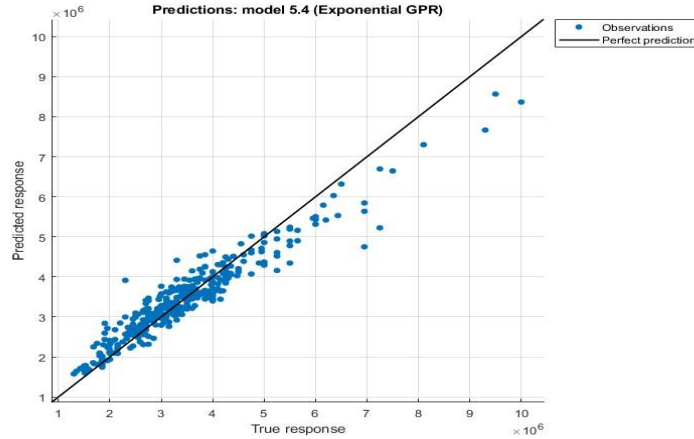
Regression Learner Toolbox ile yapılan analiz sonucunda algoritmaların RMSE, R^2 , MSE ve MAE sonuçları

Tablo 3. Algoritmaların sonuçları

Methods	RMSE	R^2	MSE	MAE
Linear Regression	8.57	0.54	1.73	6.08
Regression Tree	6.68	0.72	4.46	4.81
SVM	7.52	0.65	5.66	3.27
Ensemble Bagged Tree (RO)	7.39	0.66	5.47	5.06
Gaussian Process R.	4.17	0.89	1.73	2.98
Neural Network	7.45	0.65	5.55	5.21

Tablo 3’de gösterildiği üzere en iyi performansı Gaussian Process Regression algoritması göstermiştir. Bu algoritmanın RMSE değeri diğer yöntemlere göre daha düşük olduğu için en iyi performansı gösterdiği ortaya koyulmaktadır. Bu yöntemin yanı sıra Regression Tree yöntemi de en iyi ikinci performansı göstermiştir. R^2 değerleri incelendiğinde ise Linear Regression algoritması hariç tüm algoritmalarının modelin %60 ından fazlasını açıkladığı görülmektedir. R^2 değeri modelin açıklama gücünü ifade etmektedir. Bu bağlamda GPR yönteminin modeli çok iyi açıkladığı görülmektedir.

Grafik 1’de GPR yönteminin tahmin grafiği gösterilmektedir.



Grafik 1. GPR ile elde edilen tahmin grafiği

Grafik 1 incelendiğinde Giresun ilinde konut fiyatlarının belirli bir nokta aralığında yoğunlaştığı görülmektedir.

6. Sonuç ve Tartışma

Bu çalışmada Giresun ili Merkez ilçesinde bir satış sitesinde ilana koyulan 3+1 daireler için konut fiyat tahmininde bulunulmuştur. Makine öğrenmesi algoritmalarının tahmin performansları karşılaştırılmıştır. Çalışmada MATLAB yazılım programı içerisinde bulunan Regression Learner Toolbox aracı kullanılmıştır. Yapılan analiz sonucunda en iyi performansı gösteren yöntem Gaussian Process Regression algoritması olmuştur. Yöntemin MAE, MSE ve RMSE değerleri sırasıyla 0.29, 0.17, 0.41, R^2 değeri ise 0.89 olarak bulunmuştur. Bu yöntemin yanı sıra Neural Network, Support Vector Machine, Regression Trees gibi yöntemlerin de modeli açıklama dereceleri yüksek olarak bulunmuştur. Tüm bu sonuçlar makine öğrenmesi algoritmalarının konut değerlendirme çalışmalarında kullanılabileceğini ve iyi performans gösterdiklerini ortaya koymaktadır.

Konut fiyatlarının tahmin edilmesi birçok açıdan önemlidir. Benzer çalışmalar yapmayı planlayan muhtemel araştırmacıların;

- Farklı veri setleri ve yöntemler kullanarak tahmin performanslarını karşılaştırılması önerilmektedir.
- Aynı veri seti üzerinden çalışacak araştırmacıların ise kriter sayılarını (değişken) artırması ya da azaltması ile tahmin performansı gerçekleştirmeleri önerilebilir.

Yazar Katkıları: Çalışma tek yazarlıdır.

Finansman: "Bu araştırma dışarıdan fon almadı"

Çıkar çatışmaları: "Yazar çıkar çatışması beyan etmemektedir"

Kaynaklar

- [1] E.Egrioglu, U.Yolcu ve E.Baş, Yapay Sinir Ağları, Ankara: Nobel, 2019.
- [2] M.Atalay ve E.Çelik, "Büyük veri analizinde yapay zeka ve makine öğrenmesi uygulamaları", Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Dergisi, C.9, s.22, ss.255-172, 2017.
- [3] M.Mohri, A.Rostamzadeh ve A.Talwalkar, Foundations of Machine Learning, USA: The MIT Press, 2012.

- [4] B.Caliskan, H.A.Özkan Yazar ve A.Keskin, "Understanding sociodemographic characteristics shaping the choice of large homes: lessons from Turkish metropolises", *International Journal of Housing Markets and Analysis*, <https://doi.org/10.1108/IJHMA-07-2024-0090>, 2024.
- [5] R. Rolnik, "Place, inhabitation and citizenship: the right to housing and the right to the city in the comtemporary urban world", *International Journal of Housing Policy*, c.14, s.13, ss.293-300, 2014.
- [6] S. Purkis, "İstanbul'da inşaat odaklı birikimin durdurulamayan yükselişi: konut fazlasına karşın artan konut açığı", *Mülkiye Dergisi*, c. 40, sy. 4, ss. 91-112, 2016.
- [7] O.A. Taiwo, "Types of machine learning algorithms", *New Advances in Machine Learning*, Y.Zhang (Ed.), InTech, University of Portsmouth United Kingdom, pp.3-31, 2010.
- [8] F.Y. Osisanwo et al., "Supervised machine learning algorithms: classification and comparison", *International Journal of Computer Trends and Technology (IJCTT)*, v.48, n.3, pp.128-138, 2017.
- [9] B.Mahesh, "Machine learning algorithms-a reiew", *International Journal of Science and Research*, v.9,n.1, pp.381-386, 2020.
- [10] D.A. Pisner and D.M. Schneyer, "Support vector machine", in *Machine Learning*, Academic Press, pp.101-121, 2020.
- [11] Ö. Akar and O. Güngör, "Rastgele Orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması", *hkmojd*, no. 106, pp. 139-146, December 2012, doi: 10.9733/jgg.241212.1t.
- [12] E.Namlı, R.Ünlü and E.Gül, "Fiyat tahminlemede makine öğrenmesi teknikleri ve doğrusal regresyon yöntemlerinin kıyaslanması: Türkiye'de satılan ikinci el araç fiyatlarının tahminlenmesine yönelik bir vaka çalışması", *KONJES*, vol.7, n.4, pp.806-821, 2019.
- [13] W.Y. Loh, "Classification and regression trees", *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, v.1,n.1, pp.14-23, 2011.
- [14] M.Seeger, "Gaussian processes for machine learning", *International Journal of Neural Systems*, v.14,n.02, pp.69-106, 2004.
- [15] T.G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision tress: Bagging, boosting, and randomization", *Machine Learning*, 40, pp.139-157, 2000.
- [16] B.H. Arda, "Konut fiyatları tahmininde makine öğrenmesi sınıflandırma algoritmalarının kullanılması: Kütahya kent merkezi örneği", *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, s.76, ss.221-237, 2023.
- [17] M.Oral, E.Okatan, İ.Kırbaş, "Makine öğrenme yöntemleri kullanarak konut fiyat tahmini üzerine bir çalışma: Madrid örneği", 3 rd *International Young Researchers Student Congress*, pp. 263-272, 2021.
- [18] O.Doğan et al., "Konut değerlendirme tahmininde yapay sinir ağları ve çok regresyon analizi yöntemlerinin kıyaslanması: Yenimahalle/Ankara örneği", *Türkiye Arazi Yönetimi Dergisi*, v.6, n. 1, pp.21-31, 2024.
- [19] M.Kayakuş, M.Terzioğlu, F.Yetiş, "Forecasting housing prices in Turkey by machine learning methods", *Aestimum*, 80, pp.33-44, 2022.
- [20] M.Özdemir, K.Yıldız ve B.Büyüktanır, "Housing price estimation with deep learning: A case study of sakarya Turkey", *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, c.9,s.1,ss.138-151, 2022.
- [21] R.T. Mora-Garcia, M.F. Cespedes-Lopez, V.R. Perez-Sanchez, "Housing price prediction using machine learning algorithms in COVID-19 times", *Land*, v.11, n.11, pp.1-32, 2022.
- [22] A. Soltani, M.Heydari, F.Aghaei, C.J.Pettit, "Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms", *Cities*, n.131, 2022.
- [23] B. Park and J.K. Bae, "Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data". *Expert systems with applications*, v.42, n. 6, pp.2928-2934, 2015.
- [24] Q.Truong, M.Nguyen, H.Dang, B.Mei, "Housing price prediction via improved machine learning techniques". *Procedia Computer Science*, 174, pp. 433-442, 2020.
- [25] E.Ç. Akay et al., "Türkiye konut fiyat endeksi öngörüsü: ARIMA, rassal orman ve arima-rassal orman". *PressAcademia Procedia*, c.10, s.1, ss.7-11, 2019.