

Kira Tahminleme Sistemi

Proje Açıklaması

Bu projede, konut kiralalarının tahmin edilmesi amacıyla bir makine öğrenmesi uygulaması gerçekleştirilmiştir. Kullanılan veri seti, Hindistan'daki konut kiralalarına ilişkin çeşitli özellikleri içermektedir. Amaç, bu özellikler yardımıyla kira tutarını tahmin etmektir. Proje kapsamında aşağıdaki adımlar gerçekleştirilmiştir:

1. Veri Kümesi ve Ön İşleme

* Veri seti: House_Rent_Dataset.csv

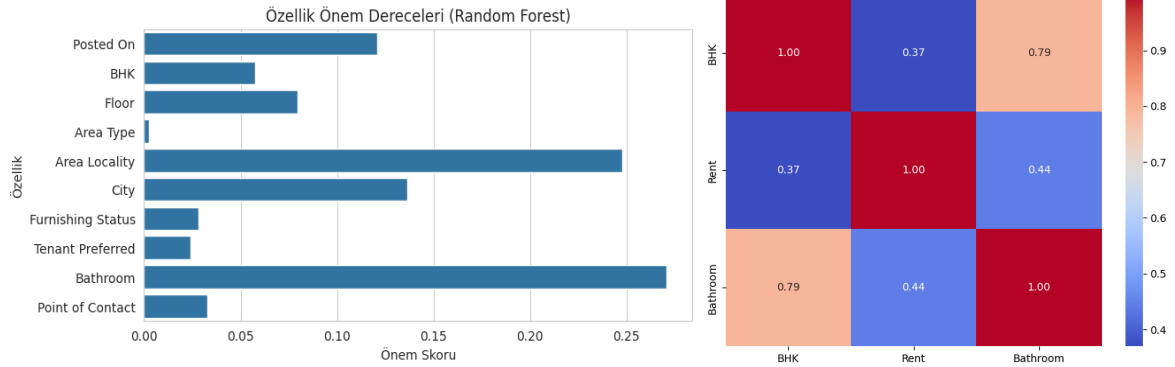
* "Size" sütunu veri setinden çıkarılmıştır.

* Veri seti 4746 konuttan oluşmakta, oda sayısı (BHK), banyo sayısı, kat bilgisi, konum ve ilan tarihi gibi değişkenler içermektedir. Sayısal olmayan veriler sayısallaştırılmış, eksik değerler ortalama ile doldurulmuştur. Özellikle kira değişkeni büyük farklılıklar gösterdiğinden (ortalama \approx 35.000 TL, medyan \approx 16.000 TL), bazı veriler çarpık dağılım sergilemiştir. Tüm kategorik sütunlar LabelEncoder ile modele uygun hale getirilmiştir.

2. Veri Görselleştirme:

* Korelasyon matrisi ve çeşitli dağılım grafikleri ile değişkenler arasındaki ilişkiler analiz edilmiştir.

* Özellik önem dereceleri görselleştirilmiştir.



3. Model Eğitimi:

* Linear Regression, Random Forest Regressor ve XGBoost Regressor modelleri eğitildi.

* Random Forest için GridSearchCV ile hiperparametre optimizasyonu yapıldı.

* Veri seti eğitim ve test olarak bölündü (train_test_split).

4. Performans Değerlendirme:

* MAE (Mean Absolute Error), MSE (Mean Squared Error), R2 Score, MAPE (Mean Absolute Percentage Error) metrikleri kullanılarak değerlendirme yapıldı.

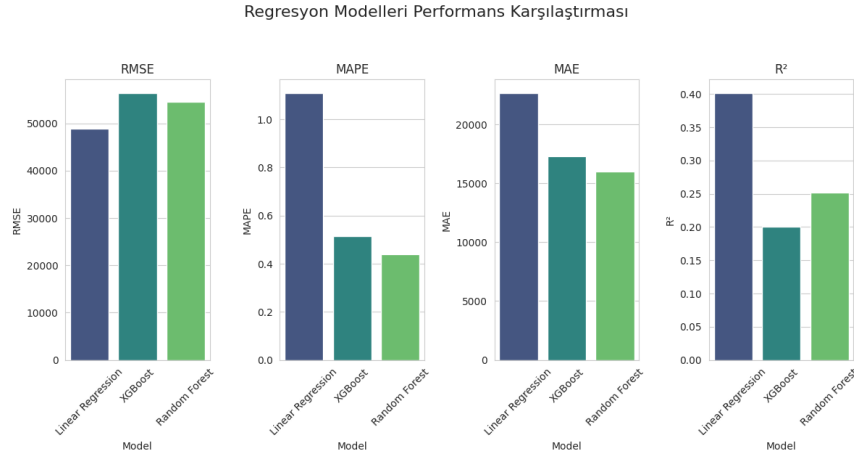
* Linear Regression, düşük RMSE ve yüksek R^2 skoruyla istatistiksel olarak güçlü görünse de MAPE değeri oldukça yüksektir. Bu da modelin özellikle büyük kira değerlerini tahmin ederken ciddi sapmalar yaptığını gösterir.

*XGBoost Regressor ise daha gelişmiş bir model olmasına rağmen bu veri kümesinde beklenen başarıyı gösterememiştir. Yüksek MAPE ve RMSE değerleri, modelin fazla varyanslı verilere karşı duyarlı olduğunu ve belki de aşırı öğrenme (overfitting) riskine sahip olduğunu düşündürmektedir.

*Random Forest Regressor, düşük MAE ve MAPE değerleriyle en dengeli ve başarılı model olmuştur; özellikle yüksek değişkenli kira değerlerine karşı daha dayanıklıdır.

5. Model Karşılaştırması:

Aşağıda üç modelin performansı özetlenmiştir:



6. Yöntem Seçimi:

*Tablo ve metrik değerlendirmesine göre en düşük hatalı model **Random Forest Regressor** olarak belirlenmiştir. Bu model, ortalama mutlak hata (MAE) ve MAPE açısından diğer modellere göre daha iyi sonuç vermiştir. Özellikle MAPE değerinin %43.80 ile çok daha düşük olması, tahminlerin oran olarak daha güvenilir olduğunu gösterir.

7.Sonuç ve Geliştirme Önerileri:

*Bu çalışma, konut kiralарını tahmin etmek amacıyla üç farklı regresyon modelinin karşılaştırmalı analizini içermektedir. Elde edilen sonuçlar, model hatalarının henüz yüksek olduğunu ve veri setindeki aşırı değerlerin (örneğin 3,500,000 TL gibi çok büyük kiralар) performansı etkilediğini göstermektedir. Sınırlılıklar arasında veri kümesinin yalnızca belirli bir şehre ait olması ve kimi özelliklerin yorumlanabilirliğinin düşük olması sayılabilir. İleride yapılabilecek geliştirmeler için veri setinin büyütülmesi, ek özellik mühendisliği (örneğin ulaşım kolaylığı gibi) önerilebilir. Ayrıca farklı modellerin denenmesi ve model optimizasyonunun kapsamının genişletilmesi çalışmanın doğruluğunu artırabilir. Elde edilen model, mevcut haliyle bile kiralарın yaklaşık tahminini sunabilir; ancak gerçek dünya uygulamaları için sürekli güncellenmesi gerekebilir.