Hazırlayan: İzzet Esener

**Teknik Dokümantasyon: Ev Kira Tahmin Modeli**

**Proje Özeti**

Bu proje, bir ev kira tahmin modeli geliştirmeyi amaçlamaktadır. Farklı makine öğrenimi algoritmaları (Linear Regression, Decision Tree, Random Forest, XGBoost) kullanılarak, ev özelliklerine göre kira fiyatı tahmini yapılmıştır. Veri ön işleme, model eğitimi ve değerlendirme adımları detaylı bir şekilde uygulanmıştır.

Github : <https://github.com/theIzzet/House-Rent-Prediction>

**Kullanılan Veri Seti**

Veri seti olarak [House\_Sent\_Prediction](https://www.kaggle.com/datasets/iamsouravbanerjee/house-rent-prediction-dataset) kullanıldı. Veri setinin özellikleri şu şekilde:

**Özellikler:**

* + Rent (Hedef Değişken)
  + BHK (Oda Sayısı)
  + Size (m²)
  + Bathroom (Banyo Sayısı)
  + Area Type (Alan Türü)
  + City (Şehir)
  + Furnishing Status (Mobilya Durumu)
  + Tenant Preferred (Kiracı Tercihi)
  + Point of Contact (İletişim Noktası)
* **Boyut:** Orijinal veri seti 4746 satır ve 12 sütundan oluşmaktadır.

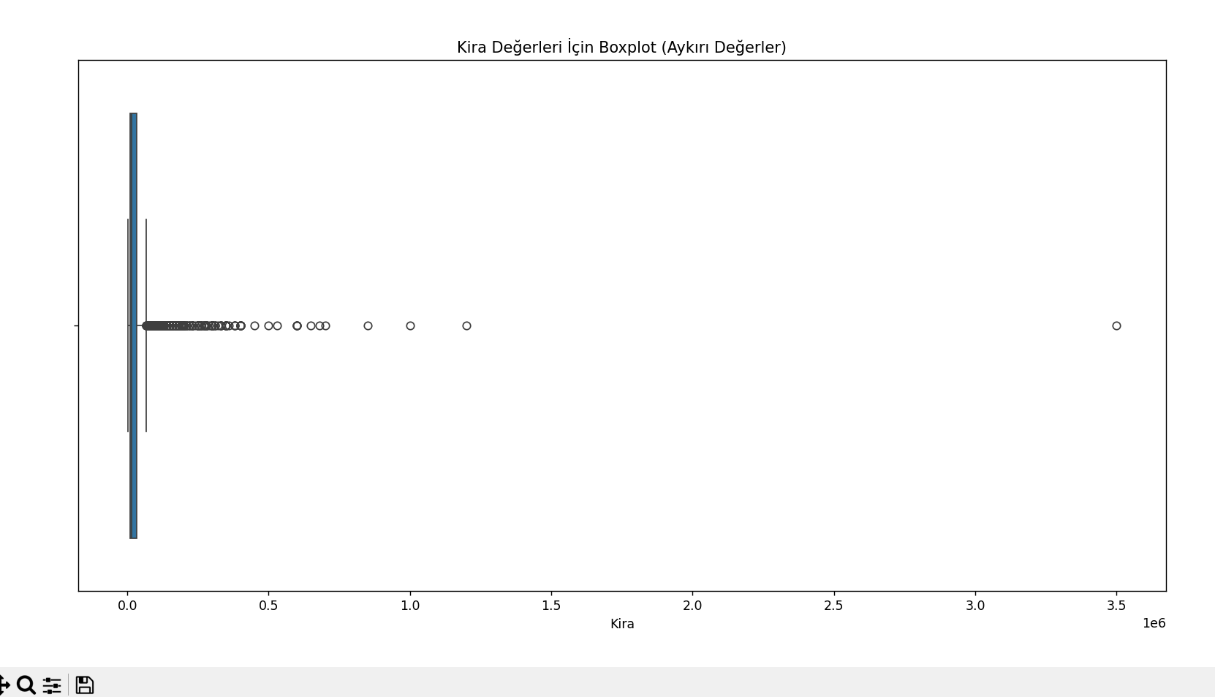
**Veri Ön İşleme Adımları**

**1.Gereksiz Sütunların Çıkarılması**

* Posted On, Area Locality, Floor gibi model için anlamsız olan sütunlar çıkarıldı.

**2.Aykırı Değerlerin Tespiti ve Temizlenmesi**

* **Boxplot analizi** ile Rent sütunundaki aykırı değerler tespit edildi.
* **500.000'den yüksek** kira değerleri aykırı kabul edilerek veri setinden çıkarıldı.

****

**3.Kategorik Değişkenlerin Dönüştürülmesi**

* Area Type, City, Furnishing Status, Tenant Preferred, Point of Contact gibi kategorik değişkenler **One-Hot Encoding** ile sayısallaştırıldı.

**4.Verinin Eğitim ve Test Kümelerine Ayrılması**

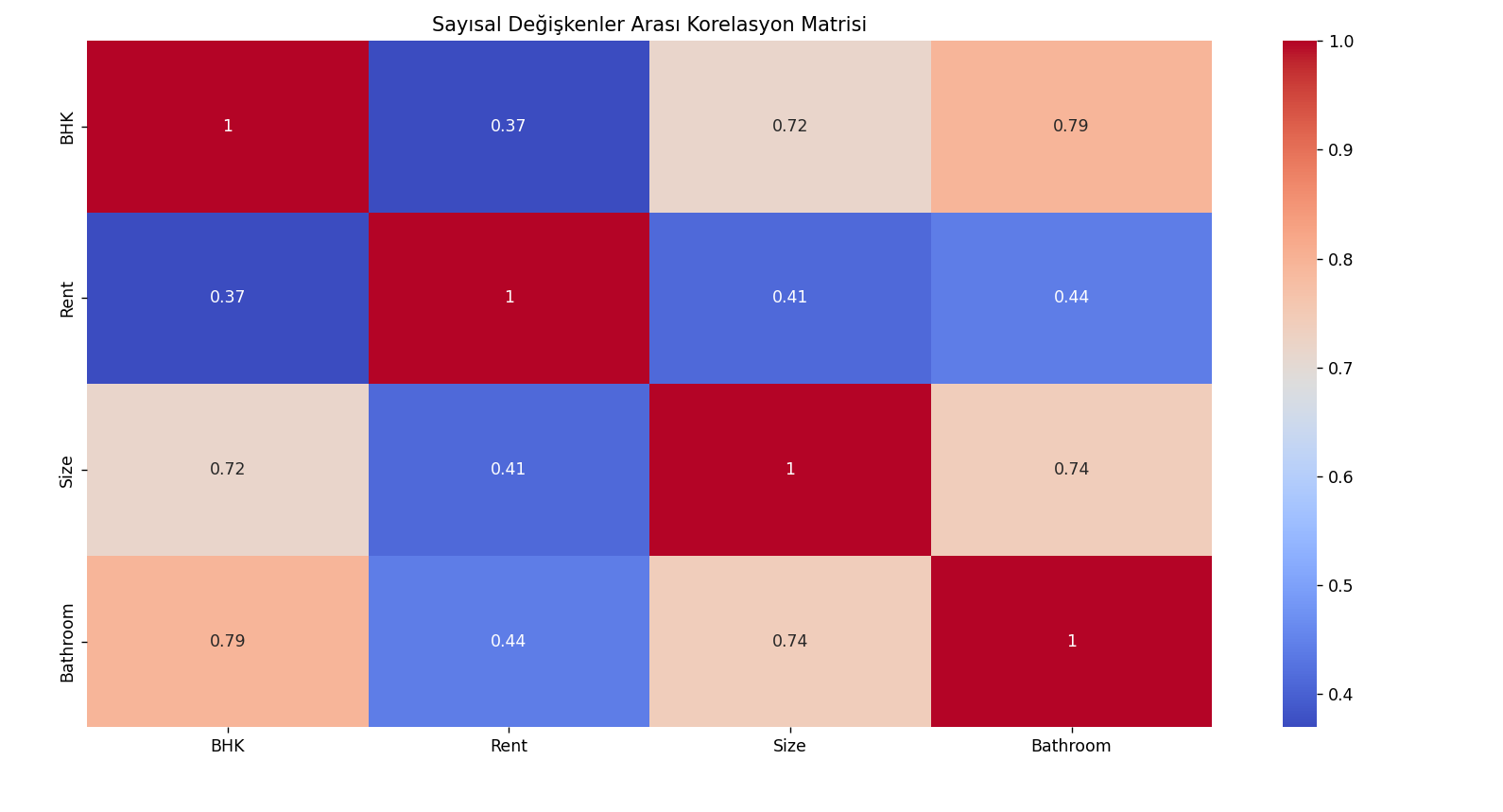
* Veri, **%70 eğitim** ve **%30 test** olarak ayrıldı.

**5.Verinin Standart Ölçeklenmesi**

* StandardScaler kullanılarak sayısal değerler normalize edildi.

**Korelasyon Analizi**

Sayısal değerlerin korelasyon analizi sonucu şu şekildedir:



**Kullanılan Makine Öğrenimi Modelleri**

**Linear Regression**

**Performans: MAE:** 0.352, **RMSE:** 0.540, **R²:** 0.530

**Decision Tree Regressor**

**Performans:** **MAE:** 0.229, **RMSE:** 0.528, **R²:** 0.550

**Random Forest Regressor**

**Performans: MAE:** 0.197, **RMSE:** 0.433, **R²:** 0.698

**XGBoost Regressor**

**Performans: MAE:** 0.190, **RMSE:** 0.429, **R²:** 0.703

**Model Performans Karşılaştırması**

Ölçeklenmiş Veriler Üzerinde model performans karşılaştırması

| **Model** | **MAE** | **RMSE** | **R²** |
| --- | --- | --- | --- |
| Linear Regression | 0.352 | 0.540 | 0.530 |
| Decision Tree | 0.229 | 0.528 | 0.550 |
| Random Forest | 0.197 | 0.433 | 0.698 |
| **XGBoost** | **0.190** | **0.429** | **0.703** |

**Ölçeklenmemiş Veriler Üzerinde** model performans karşılaştırması

| **Model** | **MAE** | **RMSE** | **R²** |
| --- | --- | --- | --- |
| Decision Tree | 11519.99 | 26454.64 | 0.577 |
| Random Forest | 10123.03 | 22253.36 | 0.701 |
| **XGBoost** | **9835.36** | **22152.75** | **0.703** |

**Sonuç:**

En iyi performans gösteren model **XGBoost** oldu (R² = 0.703).

**Ölçekleme**, bazı modellerin performansını artırdı (Decision Tree ve Random Forest).

