



HOUSE PRICE PREDICTION

ELIF BUSE PURUT

ZEKI ONUR YÜKSEK

TALHA BURAK GÜRSEL

İş Problemi

- Bir ev alırken yada satışa sunarken fiyatının ne olduğunu merak ettiğimizde kullanabileceğimiz bir makine öğrenmesi modeli.
- Model sayesinde aradığımız evin özelliklerini girerek aradığımız şartlardaki bir evin fiyatlarına erişebiliyoruz.



Veri Seti Hikayesi

- Ames, Iowa'daki konut evlerinden oluşan bu veri seti içerisinde 80 veri seti değişkeni, 44 kendi ürettiğimiz değişkenden oluşuyor. İçerisinde farklı parametreler barındırarak gelişmiş bir tahmin modeli kurmaya olanak tanıyor. 'SalePrice' özelliği hedef özelliktir ve regresyonla sonucu tahmin edilecek değerleri tutmaktadır.

Sayısal değişken : 81

Kategorik Değişken: 43

21.05.2023
Toplam Gözlem: 1460



Yol Haritası ve Kullanılan Kütüphaneler



Keşifçi veri analizi (EDA) işlemlerimizi yaparak genel resme bakılması

Verideki değerleri inceleyip aykırılıkları ve boş değerleri tespit edip gerekli işlemlerin yapılması

Modelimizin daha iyi öğrenebilmesi için yeni değişkenler türetilmesi ve encode edilmeleri

Modelin kurlumu

Hiperparametre optimizasyonu

Performans iyileştirme

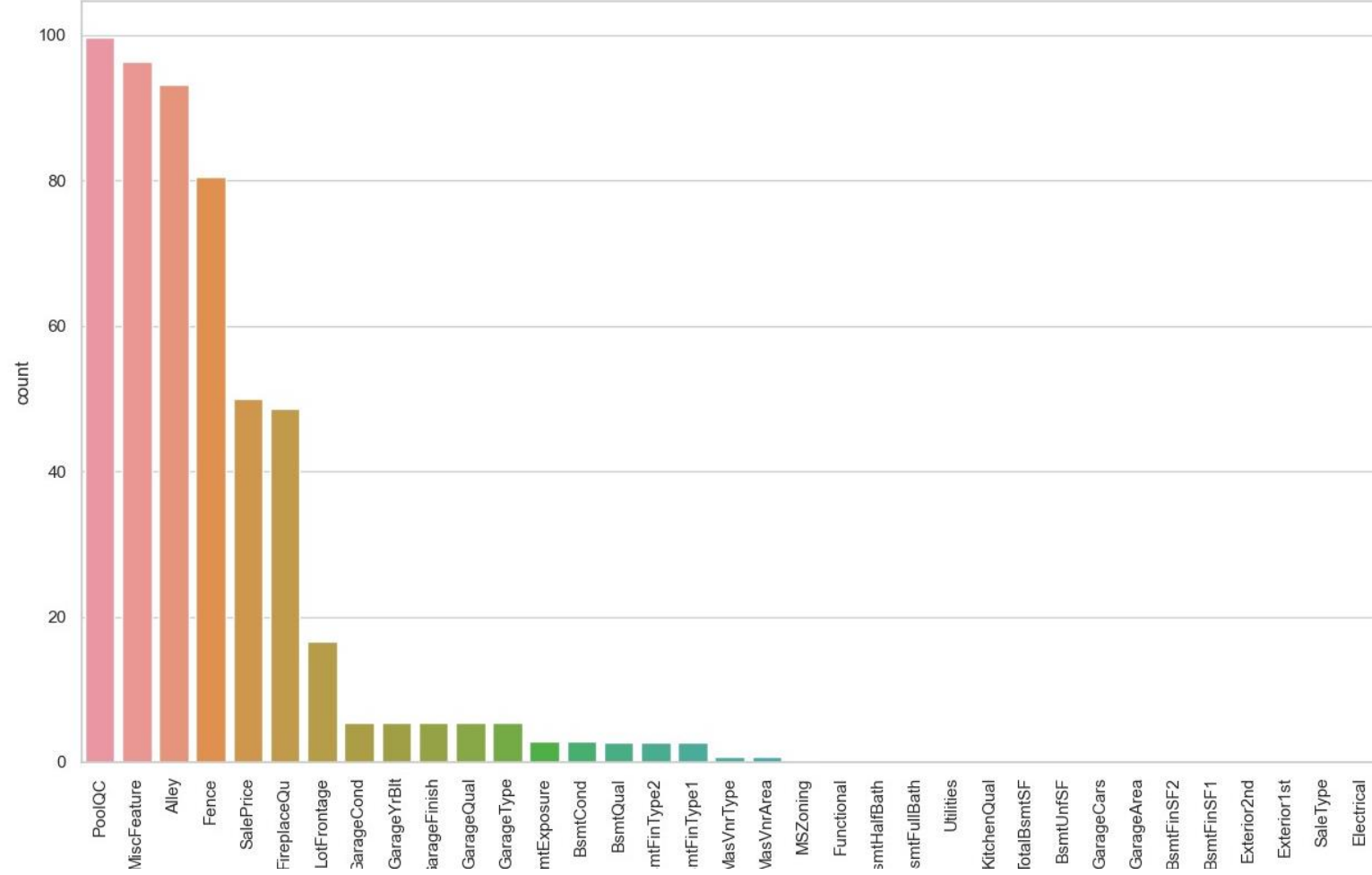
Genel değerlendirme

Veri Ön İşleme

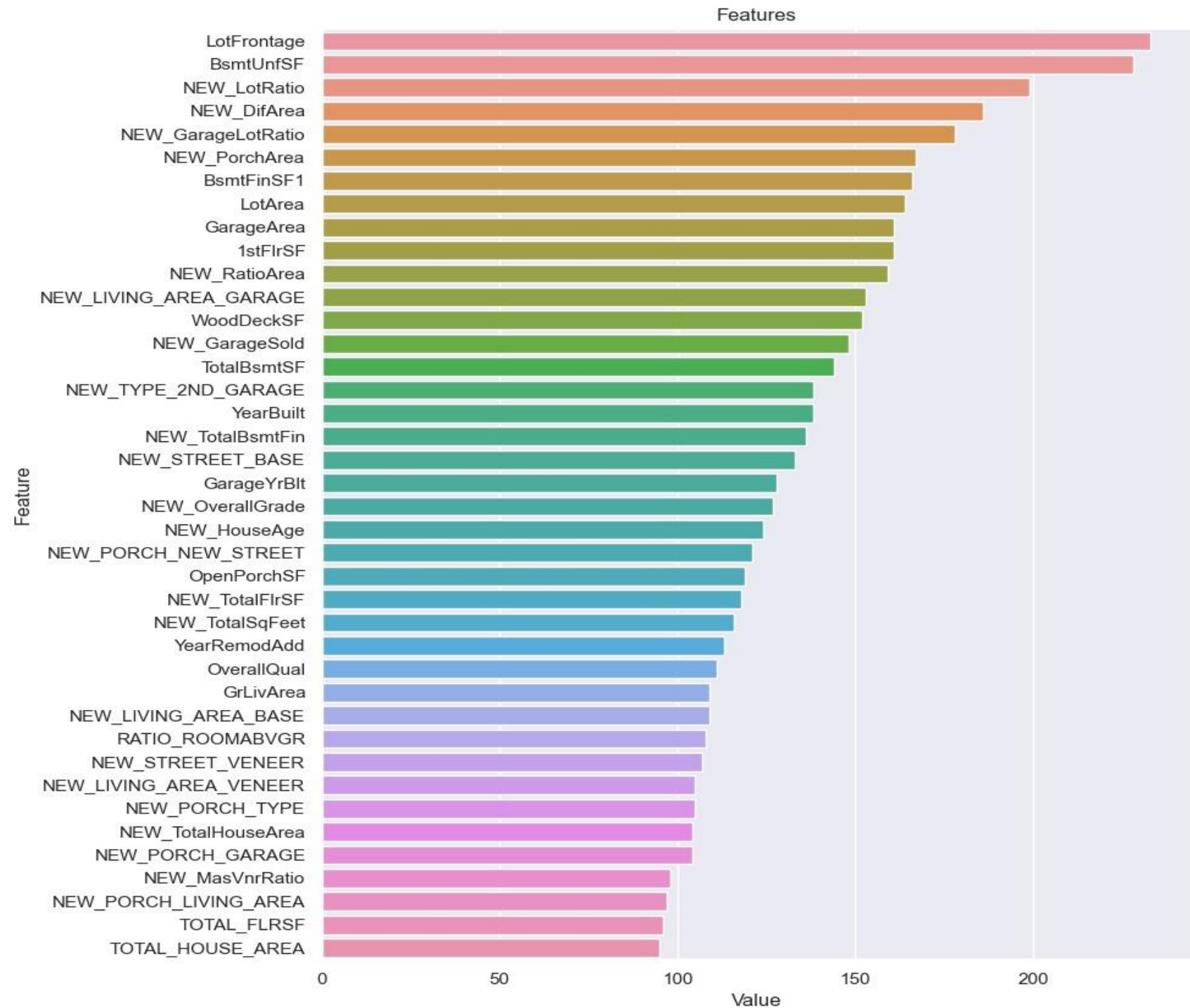
- Eksik Değerler:

Object değişkenlerde eksik değerlerin çok fazla olduğu keşfedildi. Bu boş değerlerin aslında o evin içerisinde aranan özelliğin barınmamasından kaynaklı olduğu anlaşıldı.

Ve bulunan değerlere flag işlemi uygulandı.



Değişken Önemi



Yeni Elde Edilen Önemli Değişkenler

NEW_LotRatio

Total arazinin yaşam alanına oranı

NEW_DiffArea

Kullanılabilir yaşam alanı

NEW_GarageLotRatio

Total arazinin garaja oranı

NEW_PorchArea

Kapalı-Açık yürüyüş alanlarının toplamı

NEW_RatioArea

Evin bulunduğu alanın total araziye oranı

NEW_Living_Area_Garage
Yaşam alanının garaja oranı

Model Kurulumu ve Başarı Değerlendirmesi

- LightGBM modelini tercih etme sebebimiz hep hız hem de başarı oranına baktığımızda elimizde en optimum sonucu üreten model oldu. Modeli direk kullanarak doğal olarak başarılı bir sonuç üretmede ve hiper parametre optimizasyonu ile başarı değerimizi yükselttik.
- Başarı değerimizi daha iyi görebilmemiz için ise rmse nin ardından mape değerini kullandık.
- MAPE (Mean Absolute Percentage Error) Hata değerlerinin birimlerinde farklılık olduğunda kullanılan bir fonksiyondur. Yüzde olarak ifade edilir. Fonksiyonun %10'un altında çıkması yüksek doğruluğa sahip model, %10 ile %20 arasında çıkması doğru tahmin modeli olduğunu göstermektedir.
- Ortalama karekök hatası (RMSE), bir model veya tahminci tarafından tahmin edilen değerler (örnek veya popülasyon değerleri) ile gözlemlenen değerler arasındaki farkların karekökünü hesaplayarak sonuç döndüren bir ölçüttür.

RMSE	MAPE
0.0339	0.15

DİNLEDİĞİNİZ İÇİN TEŞEKKÜRLER

ELİF BUSE PURUT

ZEKİ ONUR YÜKSEK

TALHA BURAK GÜRSEL