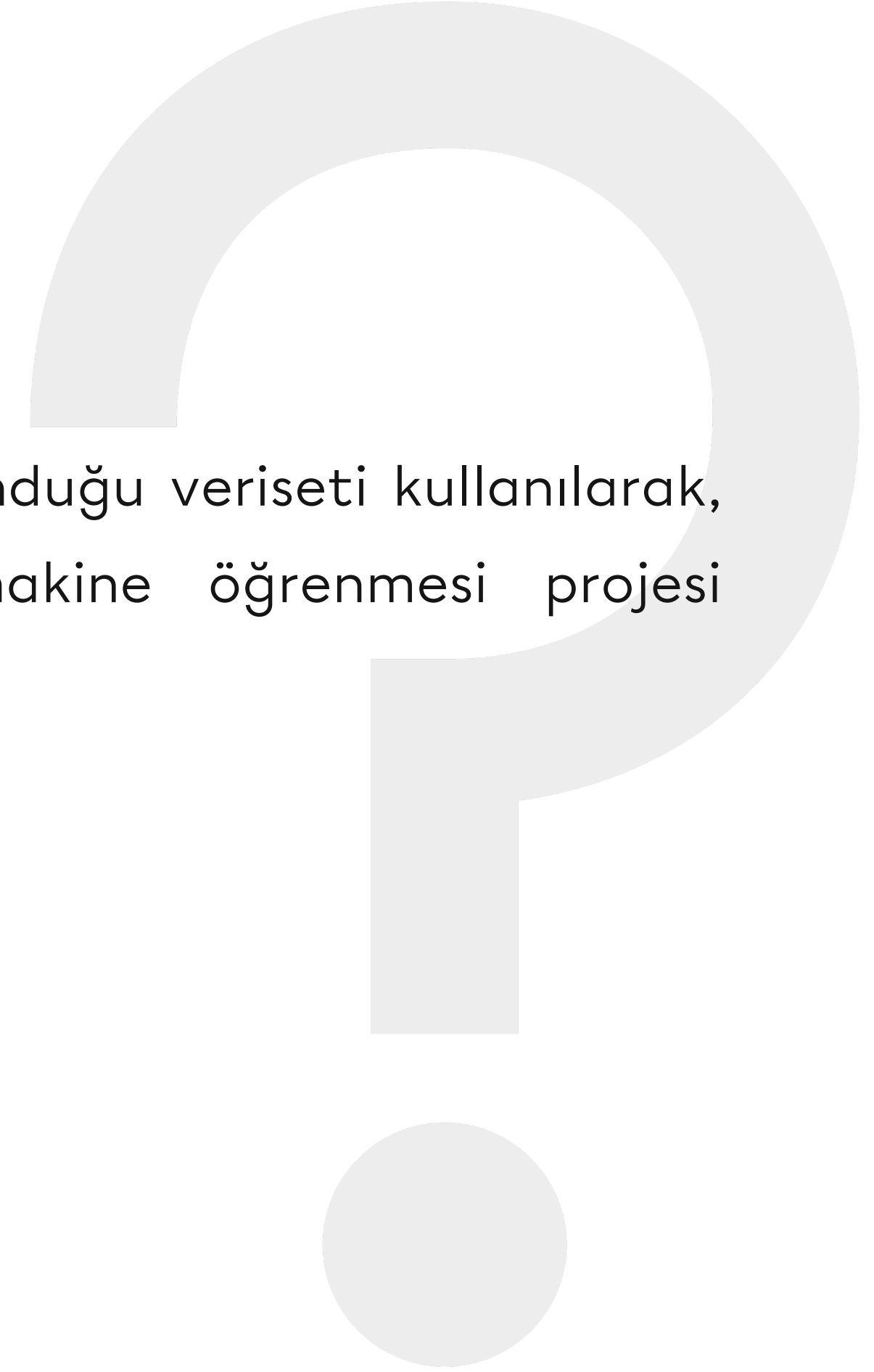




Ev Fiyat Tahmin Modeli

İş Problemi

Her bir eve ait özelliklerin ve ev fiyatlarının bulunduğu veriseti kullanılarak, farklı tipteki evlerin fiyatlarına ilişkin bir makine öğrenmesi projesi gerçekleştirilmek istenmektedir.

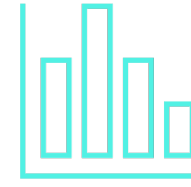


Ames, Iowa'daki konut evlerinden oluşan bu veri seti içerisinde 79 açıklayıcı değişken bulunduruyor. Kaggle üzerinde bir yarışması da bulunan projenin veri seti ve yarışma sayfasına aşağıdaki linkten ulaşabilirsiniz. Veri seti bir kaggle yarışmasına ait olduğundan dolayı train ve test olmak üzere iki farklı csv dosyası vardır. Test veri setinde ev fiyatları boş bırakılmış olup, bu değerleri sizin tahmin etmeniz beklenmektedir.

<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/overview/evaluation>



Toplam Gözlem
1460



Sayısal Değişken
38



Kategorik Değişken
43

Görev 1: Keşifçi Veri Analizi

Adım 1: Train ve Test veri setlerini okutup birleştiriniz. Birleştirdiğiniz veri üzerinden ilerleyiniz.

Adım 2: Numerik ve kategorik değişkenleri yakalayınız.

Adım 3: Gerekli düzenlemeleri yapınız. (Tip hatası olan değişkenler gibi)

Adım 4: Numerik ve kategorik değişkenlerin veri içindeki dağılımını gözlemleyiniz.

Adım 5: Kategorik değişkenler ile hedef değişken incelemesini yapınız.

Adım 6: Aykırı gözlem var mı inceleyiniz.

Adım 7: Eksik gözlem var mı inceleyiniz.

Görev 2: Feature Engineering

Adım 1: Eksik ve aykırı gözlemler için gerekli işlemleri yapınız.

Adım 2: Rare Encoder uygulayınız.

Adım 3: Yeni değişkenler oluşturunuz.

Adım 4: Encoding işlemlerini gerçekleştiriniz.

Görev 3: Model Kurma

Adım 1: Train ve Test verisini ayırınız. (SalePrice değişkeni boş olan değerler test verisidir.)

Adım 2: Train verisi ile model kurup, model başarısını değerlendiriniz.

Bonus: Hedef değişkene log dönüşümü yaparak model kurunuz ve rmse sonuçlarını gözlemleyiniz. Not: Log'un tersini (inverse) almayı unutmayınız.

Adım 3: Hiperparemetre optimizasyonu gerçekleştiriniz.

Adım 4: Değişken önem düzeyini inceleyiniz.

Bonus: Test verisinde boş olan salePrice değişkenlerini tahminleyiniz ve Kaggle sayfasına submit etmeye uygun halde bir dataframe oluşturup sonucunuzu yükleyiniz.

miuul

miuul.com