**Proje Ön Gösterimi**

Bir veya daha fazla sorunu geleneksel Makine Öğrenimi yöntemleriyle değerlendireceğiniz bir projede çözmek istiyor musunuz? Sizleri, geleneksel Makine Öğrenimi yöntemleriyle elde ettiğiniz sonuçları değerlendireceğiniz bir proje bekliyor. Projedeki sorunları regresyon ve ağaç temelli yöntemler kullanarak çözmeniz gerekiyor. Bu projeyi tamamlamalısınız, burada veriyi inceleme, model için veriyi hazırlama ve keşifsel veri analizi yöntemlerini öğrenebileceksiniz. Portföyünüzde böyle bir projeye sahip olmalısınız!

**Veri Seti**

Melbourne, Avustralya'nın Victoria eyaletinin başkenti ve en büyük şehri olup, hem Avustralya hem de Okyanusya'da ikinci en kalabalık şehirdir. Veri kümesi, Melbourne'deki evlerin birçok özelliğini ve fiyatlarını içermektedir.

**Data Setindeki Veriler**

**●** Suburb

● Address

● Rooms: Number of rooms

● Price: Price in Australian dollars, target variable

● Method: S - property sold; SP - property sold prior; PI - property passed in; PN - sold prior not disclosed;

SN - sold not disclosed; NB - no bid; VB - vendor bid; W - withdrawn prior to auction; SA - sold after

auction; SS - sold after auction price not disclosed. N/A - price or highest bid not available.

● Type: br - bedroom(s); h - house,cottage,villa, semi,terrace; u - unit, duplex; t - townhouse; dev site -

development site; o res - other residential.

● SellerG: Real Estate Agent

● Date: Date sold

● Distance: Distance from CBD in Kilometres

● Regionname: General Region (West, North West, North, North east ...etc)

● Propertycount: Number of properties that exist in the suburb.

● Bedroom2 : Scraped # of Bedrooms (from different source)

● Bathroom: Number of Bathrooms

● Car: Number of carspots

● Landsize: Land Size in Metres

● BuildingArea: Building Size in Metres

● YearBuilt: Year the house was built

● CouncilArea: Governing council for the area

● Lattitude

● Longtitude

**Proje’nin Adımları**

**1.Bir Google Colaborratory Dosyası Oluşturun**

● Dosyanın .ipynb uzantılı olduğundan emin olun

● Projede detayları açıklayan yorum satırlarının bulunduğundan emin olun.

● Projeyi sunarken, bu .ipynb dosyasının hücrelerini gönderin, böylece hücreler çalıştırılır ve sonuçlar görünür hale gelir.

**2.Gerekli Kütüphanelerin İçeri Aktarılması**

● Projede kullanılacak olan kütüphaneleri Colab ortamına içe aktarın.

● Veri analizi için NumPy, Pandas, Seaborn ve Matplotlib kütüphanelerini içe aktarın.

● Model oluşturma ve modelin performansını değerlendirme için sklearn.model\_selection, sklearn.metrics, sklearn.ensemble, sklearn.linear\_model, sklearn.tree ve sklearn.neighbours kütüphaneleri ve modüllerini içe aktarın.

**3.Proje Tanımı**

Bu proje için Melbourne Konut veri setini projemize yüklememiz gerekiyor. Topladığımız verilerin kalitesi ve miktarı, tahmin modelimizin ne kadar iyi olabileceğini belirleyecektir. Bu nedenle veri setini çok dikkatli bir şekilde incelememiz gerekiyor. Melbourne Konut veri setini kullanarak bir evin fiyatını tahmin edeceğiz, ki bu gerçek bir yaşam örneğidir. Herhangi bir maliyeti değerlendirmeden önce, veriyi ön işleme tekniklerini kullanarak analiz etmeye başlayacağız. Ardından modellerimizi oluşturacağız ve projeyi tamamlamak için performanslarını ölçeceğiz.

**4.Veri Toplama ve Gözlemleme**

● read\_csv() ile veri kümesini projeye yükleyin ve ilk 5 sütunu gözlemleyin.

● Veri kümesinin şeklini, sütun sayısını ve boyutunu bulun.

● Veri kümesinin bilgilerini gösterin, bu bilgiler sütun sayısını, sütun etiketlerini, sütun veri tiplerini, bellek kullanımını, aralık indeksini ve her sütundaki hücre sayısını (null olmayan değerler) içerir.

**5.Keşifsel Veri Analizi**

● Veri kümesinin betimsel istatistiklerini inceleyin.

● Bazı değişkenlerin değerleri nesne olarak verilmiştir. Aynı zamanda kategorik değerlerin de olduğunu gözlemliyoruz. Bu, veri kümesini incelediğimizde bize sorun çıkarabilir. Bu nedenle, bu tür durumlarda değişkenleri kategorik olarak tanımlamamız gerekmektedir.

● Yinelemeli verileri kontrol edin. Eğer yinelemeli veriler varsa, veri kümesinden temizleyin.

● Veri kümesindeki aykırı verileri temizleyin. Veri kümesini incelediğinizde, aykırı verilerin genellikle "Landsize" ve "Buildingarea" değişkenlerinde olduğunu gözlemleyeceksiniz.

○ Veri kümesindeki aykırıları tespit ederken z-skor yöntemini kullanmanızı bekliyoruz!

● Veri kümesindeki eksik verileri bulun ve temizleyin.

○ Veri kümesinden eksik verilerin banyo ve araba değişkenlerinde olduğunu gözleyebilirsiniz. Veri kümesindeki eksik değerlerin mod yöntemi kullanılarak doldurulmasını bekliyoruz. Bunun için aşağıdaki kodu kullanabilirsiniz.

HINT:

for column in categorical\_columns:

data[column] =

data[column].fillna(data[column].mode().iloc[0])

**6.Veri Görselleştirme:**

○ Fiyat dağılımını görselleştirmek için bir histogram oluşturun.

○ Tüm sayısal değişkenler ile fiyat değişkeni arasındaki ilişkiyi görmek için bir çift grafik çizin.

○ Seaborn üzerinde bir ısı haritası kullanarak bir korelasyon matrisi çizin.

○ Kategorik değişkenler için Etiket Kodlayıcı ve Tekil Kodlayıcı (One Hot Encoder) uygulayın.

**7.Model Seçimi**

● Fiyat tahmini yapacağımızdan, x ve y değişkenlerimizi doğru bir şekilde belirlememiz gerekmektedir.

● Model eğitiminin performansını artırmak için verilerimizi eğitim ve test verilerine bölmek.

● Aşağıda belirtilen modelleri ön işlenmiş veriler kullanarak eğitin.

HINT: Please use this dictionary

models = {

'Lasso': {

'model': Lasso()

},

'LinearRegression': {

'model': LinearRegression()

},

'Ridge': {

'model': Ridge()

},

'ElasticNet': {

'model': ElasticNet()

},

'KNeighborsRegressor': {

'model': KNeighborsRegressor()

},

'RandomForestRegressor': {

'model': RandomForestRegressor()

},

'GradientBoostingRegressor': {

'model': GradientBoostingRegressor()

},

'AdaBoostRegressor': {

'model': AdaBoostRegressor(n\_estimators = 5, learning\_rate = 1.2, loss = 'exponential', random\_state = 2)

},

**8.Model Değerlendirmesi**

● Modelleri birbirleriyle karşılaştırma.

● Değerlendirme metrikleri (MAE, MSE, RMSE, R2) kullanarak en iyi performans gösteren modeli seçme.

Example:

MAE: 245102.450828939

MSE: 111946245183.27266

RMSE: 334583.6893562994

R2: 0.49244357398844063

**9.Proje Teslimi**

● Projeyi, .ipynb uzantılı bir kod dosyası olarak hazırlamanız ve tüm hücreleri çalıştırmanız gerekmektedir.

● Hazırladığınız bu dosyaları bir GitHub deposuna eklemeniz ve bu depo bağlantısını aşağıda verilen formun içine eklemeniz gerekmektedir.

● Proje bir ekip olarak yapılacaktır. Oluşturulan ekipler en fazla 5 kişiden oluşmalıdır.

● Proje ekibiniz hakkında bilgi gönderebilirsiniz bu form aracılığıyla.

● Form Bağlantısı: [Form Bağlantısı]