



Feature Engineering





İş Problemi

Bir makine öğrenmesi pipeline'ı için gerekli olan veri ön işleme ve değişken mühendisliği scriptinin hazırlanması gerekmektedir.

Veri seti bu script'ten geçirildiğinde modellemeye hazır hale gelmesi beklenmektedir.

Veri Seti Hikayesi

Veri seti Titanic gemi kazasında bulunan kişilere ait veri setidir.

768 gözlem ve 12 değişkenden oluşmaktadır.

Hedef değişken "Survived" olarak belirtilmiş olup;

1: kişinin hayatta kalmasını,

0: kişinin hayatta kalamamasını ifade etmektedir.



PROBLEMI DEĞİŞKENLER GÖREVLER

Değişkenler

Survived - Hayatta Kalma 0 Öldü, 1 Hayatta Kaldı

Pclass - Bilet Sınıfı

1 = 1. sınıf, 2 = 2.sınıf, 3 = 3.sınıf

Sex - Cinsiyet

Age - Yaş

Sibsp - Titanicte'ki kardeş / eş sayısı

Parch – Titanicte'ki ebeveyn / çocuk sayısı

Embarked: - Yolcunun gemiye biniş yaptığı liman (C = Cherbourg, Q = Queenstown, S = Southampton

Fare - Bilet ücreti

Cabin: Kabin numarası



Proje Görevleri

DSMLBC | VERI BILIMI OKULU

GÖREV One of the second of th

Görev 1:

Çalışma dizininde helpers adında bir dizin açıp, içerisine data_prep.py adında bir script ekleyiniz.

Feature Engineering bölümünde kendimize ait tüm fonksiyonları, bu script içerisine toplayınız.

Burada olması gereken fonksiyonlar:

- outlier_thresholds
- replace_with_thresholds
- check_outlier
- grab_outliers

- remove_outlier
- missing_values_table
- missing_vs_target
- label_encoder

- one_hot_encoder
- rare_analyser
- rare_encoder



GÖREV 02

Görev 2:

titanic_data_prep adında bir fonksiyon yazınız.

Bu fonksiyon için gerekli olan veri ön işleme ya da EDA fonksiyonları, helpers içerisinde yer alan eda.py ve data_prep.py dosyaları içerisinden alınız.



Görev 3:

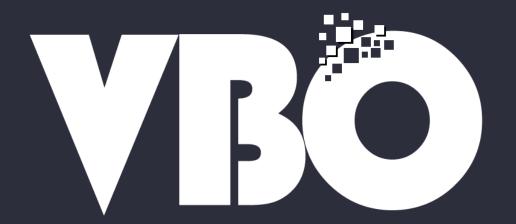
GÖREV
03

Veri ön işleme yaptığınız veri setini pickle ile diske kaydediniz.





Data Science and Machine Learning Bootcamp



İletişim: veribilimi@veribilimiokulu.com