

# VERİ MADENCİLİĞİ

## İçindekiler

Verilerin Temizlenmesi .....	2
1-Train Verilerindeki Missing Değerler İçin İzlenen Yol .....	2
2-Test Verilerindeki Missing Değerler İçin İzlenen Yol .....	2
3-Kategorik Değişkenler .....	2
Kullanılan Kütüphaneler ve Versiyonlar .....	2
Test Sonucu Elde edilen Değerler .....	3
KAYNAK SİTELER .....	3

!!!! Eğer Editör kodları çalıştırmazsa colab linkinden çalıştırabilirsiniz. !!!!

[https://colab.research.google.com/drive/1DjmHNYNrSOZYEAox0Q8IOlcwBR\\_Cpp7s?usp=share\\_link](https://colab.research.google.com/drive/1DjmHNYNrSOZYEAox0Q8IOlcwBR_Cpp7s?usp=share_link)

# Verilerin Temizlenmesi

## 1-Train Verilerindeki Missing Değerler İçin İzlenen Yol

Sayısal değişkenlerde credit\_amount ve age kolunda missing değerler az olduğu için missing değerler yerine ortalamayı atadım.

- Credit\_amount = 3 missing değer
- Age = 9 missing değer

Kategorik değerlerde ise missing değerler yerine verilerin modlarını atadım.

- Credit history = 3 missing değer
- Property\_magnitude = 3 missing değer
- Employment de missing yok

## 2-Test Verilerindeki Missing Değerler İçin İzlenen Yol

Sayısal verilerde ve kategorik verilerde missing değerleri sildim. 250 değerden 228 değer kaldı.

- Credit\_amount = 2 missing değer
- Age = 6 missing değer
- Credit\_history = 6 missing değer
- Employment = 3 missing değer
- Property\_magnitude = 5 missing değer

## 3-Kategorik Değişkenler

Train ve Test verilerindeki bütün Kategorik değişkenler (Credit\_history, Employment, Property\_magnitude, Class) one-hot encoding ile sayısal değişkene dönüştürüldü.

## Kullanılan Kütüphaneler ve Versiyonlar

Versiyon: pip 22.3.1 (python 3.7)

**import** math

**import** numpy **as** np

**import** pandas **as** pd

**from** sklearn.naive\_bayes **import** GaussianNB

**from** sklearn.preprocessing **import** LabelEncoder

## Test Sonucu Elde edilen Değerler

### TEST SONUCU

Accuracy (Doğruluk): 0.675

TP rate (Gerçek Pozitif oranı): 0.712

TN rate (Gerçek Negatif oranı): 0.433

TP (Gerçek Pozitif) adedi: 141

TN (Gerçek Negatif) adedi: 13

FP Adedi: 57

FN Adedi: 17

## KAYNAK SİTELER

<https://towardsdatascience.com/data-cleaning-with-python-and-pandas-detecting-missing-values-3e9c6ebcf78b>

<https://stackoverflow.com/questions/39173813/pandas-convert-dtype-object-to-int>

<https://realpython.com/python-data-cleaning-numpy-pandas/>

<https://tirendazakademi.medium.com/pandas-kütüphanesi-nedir-905012e654b8>

<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>

<https://stackoverflow.com/questions/66056695/what-does-labelencoder-fit-do>

[https://pandas.pydata.org/docs/reference/api/pandas.get\\_dummies.html](https://pandas.pydata.org/docs/reference/api/pandas.get_dummies.html)

<https://pandas.pydata.org/docs/reference/api/pandas.concat.html>

<https://jakevdp.github.io/PythonDataScienceHandbook/05.05-naive-bayes.html>

<https://stackoverflow.com/questions/70187512/x-train-y-train-from-transformed-data>

<https://www.youtube.com/watch?v=jXNkKKR5b80>

[https://www.w3schools.com/python/pandas/ref\\_df\\_dropna.asp](https://www.w3schools.com/python/pandas/ref_df_dropna.asp)

<https://note.nkmk.me/en/python-pandas-nan-fillna/>

<https://www.youtube.com/watch?v=woHxvbBLarQ&t=243s>