

Revangga Adytya - 26

PROGRAM STUDI TEKNIK INFORMATIKA JURUSAN TEKNOLOGI INFORMASI



LAPORAN TUGAS MINGGU 3

1. Gambar di bawah ini merupakan pemanggilan library di python

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
✓ 0.1s
```

2. Gambar yang di bawah ini merupakan library dari panda untuk membaca file "50_startups.csv"lalu ada kode program "data_awal.head()" digunakan untuk menampilkan 5 file teratas

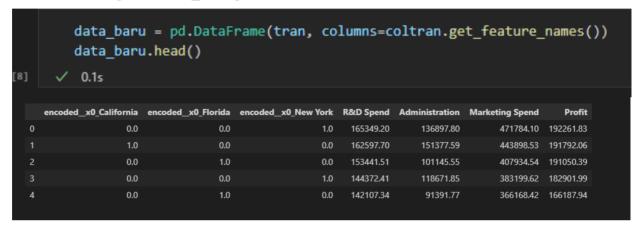
```
data_awal = pd.read_csv('50_Startups.csv')
data_awal.head()
0.2s
R&D Spend Administration Marketing Spend
                                                State
                                                           Profit
 165349.20
                 136897.80
                                  471784.10 New York 192261.83
  162597.70
                                  443898.53 California 191792.06
                 101145.55
  153441.51
                                  407934.54
                                               Florida 191050.39
  144372.41
                                  383199.62 New York 182901.99
                 118671.85
  142107.34
                  91391.77
                                  366168.42
                                               Florida 166187.94
```

3. dibawah ini merupakan proses dari OneHotEncoder



4. Setelah dilakukan proses OneHotEncoder maka dilakukan proses column transformer untuk mengubah isi dari column state

5. selanjutnya adalah menggabungkan isian dari column state yang sudah dipecah menadi satu dengan file "50 startups.csv



6. Lalu proses selanjutnya adalah proses Linear Regeresion, pada bagian tahap pertama dibuat variable X dan y, X digunakan untuk mengambil semua column dari table yang sudahdi lakukan proses Column Transformer



7. proses selanjutnya yaitu proses pemanggilan library LinearRegression, pada proses ini dibuat 4 variable dan dilakukan proses train_test_split yang isinya terdapat variable X, dan y

8. selanjutnya yaitu penggabungan dari Linear Regression

```
gabung = np.concatenate((ym_test, ym_pred), axis=1)
   gabung
array([[ 96712.8 , 84457.16115252],
      [ 90708.19
                    , 72543.44544734],
      [191792.06
                   , 186070.97934498],
                   , 67431.79651296],
      [ 81229.06
                   , 150237.4771915 ],
      [152211.77
      [126992.93
                    , 113635.43439889],
                   , 88703.00224459],
      [ 96479.51
                   , 124781.17620402],
      [134307.35
      [101004.64
                   , 101217.77951759],
      [108552.04
                     , 113863.68825244]])
```

9. Selanjutnya dilakukan prediksi menggunakan r2_score, nilai yang sempurna adalah nilai yang mendekati 1





10.