



LAPORAN TUGAS

Pembelajaran Mesin

02 – Ananda Ayu Sekar Wiranti



**INFORMATICS ENGINEERING STUDY PROGRAM
DEPARTMENT OF INFORMATION TECHNOLOGY**

1. Import data yang akan digunakan NumPy memiliki kemampuan untuk membentuk objek N-dimensional array, yang mirip dengan list pada Python. Sedangkan Pandas biasa digunakan untuk membuat tabel, mengubah dimensi data, mengecek data, dan lain sebagainya.

```
import numpy as np
import pandas as pd
```

2. Memasukan data file 50_Startups

```
# memasukan data pada file 50_Startups

data = pd.read_csv('50_Startups.csv')
data.head()
```

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94

3. Kemudian merubah dari bentuk nominal menjadi integer dengan one hot encoder, kemudian menambahkan kolom transformer digunakan untuk transformasi stateless seperti mengambil log frekuensi, melakukan penskalaan khusus

```
# merubah ke bentuk one hot encoder dan mengetahui fitur keluaran untuk transformasi column transformer

from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import make_column_transformer

transformer = make_column_transformer((OneHotEncoder(), ['State']), remainder='passthrough')

transformer1 = transformer.fit_transform(data)
transformer1_data = pd.DataFrame(transformer1, columns=transformer.get_feature_names())
print(transformer1_data.head())
```

```
onehotencoder__x0_California  onehotencoder__x0_Florida  \
0                             0.0                      0.0
1                             1.0                      0.0
2                             0.0                      1.0
3                             0.0                      0.0
4                             0.0                      1.0

onehotencoder__x0_New York  R&D Spend  Administration  Marketing Spend  \
0                             1.0    165349.20      136897.80      471784.10
1                             0.0    162597.70      151377.59      443898.53
2                             0.0    153441.51      101145.55      407934.54
3                             1.0    144372.41      118671.85      383199.62
4                             0.0    142107.34       91391.77      366168.42
```

4. Kemudian melanjutkan dengan split data

```
X = transformer1_data.iloc[:, :-1].values
y = transformer1_data.iloc[:, -1].values

y = y.reshape(len(y), 1)
y.shape

(50, 1)

# split
from sklearn.model_selection import train_test_split

Xm_train, Xm_test, ym_train, ym_test = train_test_split(X, y, test_size=0.2, random_state=50)
Xm_test.shape
Xm_train.shape

(40, 6)
```

5. Kemudian menghitung hasil perhitungan dengan multiple regression

```
# multiple Regretion (memprediksi suatu nilai berdasarkan dua variabel atau lebih)

from sklearn.linear_model import LinearRegression

mlr = LinearRegression()
mlr.fit(Xm_train, ym_train)

ym_pred = mlr.predict(Xm_test)

gabung = np.concatenate((ym_test, ym_pred), axis=1)
gabung

array([[ 96712.8      , 84457.16115252],
       [ 90708.19     , 72543.44544734],
       [191792.06     , 186070.97934498],
       [ 81229.06     , 67431.79651296],
       [152211.77     , 150237.4771915 ],
       [126992.93     , 113635.43439889],
       [ 96479.51     , 88703.00224459],
       [134307.35     , 124781.17620402],
       [101004.64     , 101217.77951759],
       [108552.04     , 113863.68825244]])
```

6. Hasil perhitungan dengan multiple regression

```
from sklearn.metrics import r2_score

r2_mlr = r2_score(ym_test, ym_pred)
print(r2_mlr)

0.897134285307714
```

