

TUGAS MACHINE LEARNING
“Regression”



Disusun oleh:

R.Muhammad Azmi Herdi Shofiyullah / 2041720079

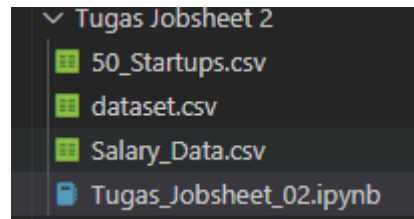
TI-3A / 21

D4 TEKNIK INFORMATIKA
TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
2022

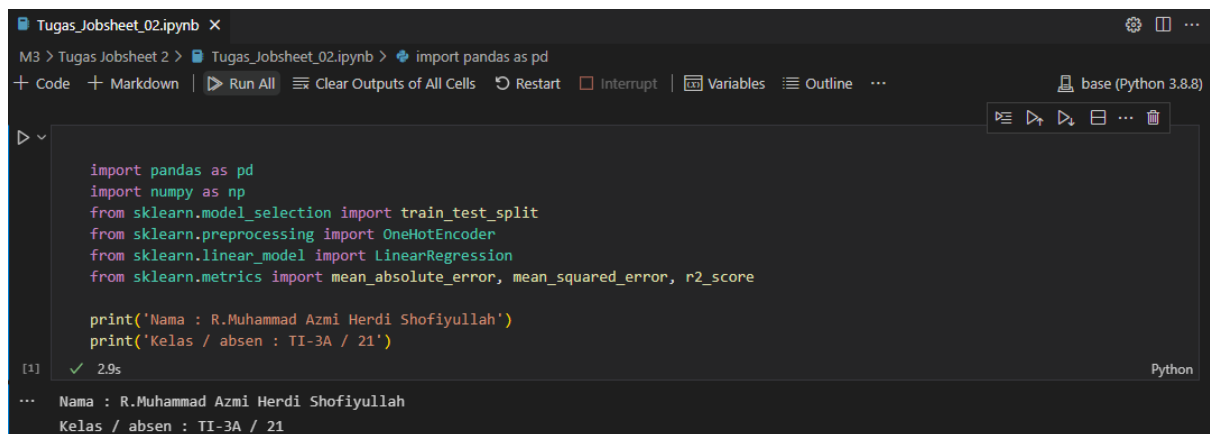
Laporan

Berikut adalah hasil dari praktikum yang telah saya jalankan :

1. Menyiapkan file yang dibutuhkan dalam satu folder yang sama.



2. Melakukan import library yang nantinya akan dibutuhkan pada saat melakukan praktikum.



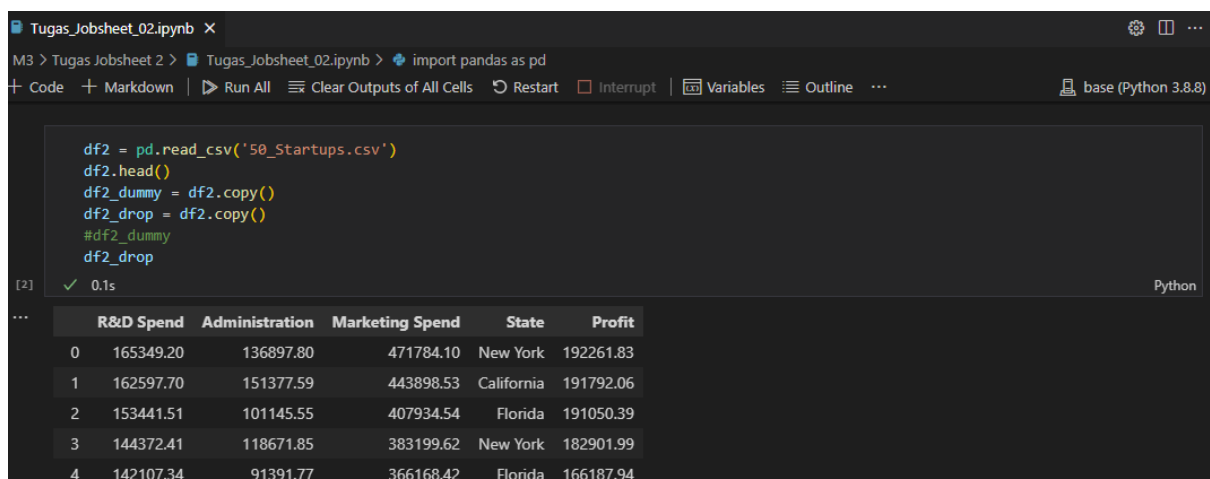
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OneHotEncoder
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

print('Nama : R.Muhammad Azmi Herdi Shofiyullah')
print('Kelas / absen : TI-3A / 21')
```

[1] ✓ 2.9s Python

... Nama : R.Muhammad Azmi Herdi Shofiyullah
Kelas / absen : TI-3A / 21

3. Menampilkan data yang sudah disiapkan sebelumnya serta melakukan copy data (dummy) agar jika terjadi kesalahan maka data utama tidak akan berubah.



```
df2 = pd.read_csv('50_Startups.csv')
df2.head()
df2_dummy = df2.copy()
df2_drop = df2.copy()
#df2_dummy
df2_drop
```

[2] ✓ 0.1s Python

...

	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

4. Membuat One Hot Encoding pada data 'State'.

```
Tugas_Jobsheet_02.ipynb X
M3 > Tugas Jobsheet 2 > Tugas_Jobsheet_02.ipynb > import pandas as pd
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Interrupt | Variables | Outline ... base (Python 3.8.8)

data_oh = pd.get_dummies(df2_drop['State'])
data_oh

[3] ✓ 0.1s Python

...
  California  Florida  New York
0           0         0         1
1           1         0         0
2           0         1         0
3           0         0         1
4           0         1         0
```

5. Melakukan penggabungan data antara data hasil One Hot Encoding dengan data awal.

```
Tugas_Jobsheet_02.ipynb X js02_regression.ipynb
M3 > Tugas Jobsheet 2 > Tugas_Jobsheet_02.ipynb > gabung = np.concatenate((ym_test, ym_pred), axis=1)
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Interrupt | Variables | Outline ... base (Python 3.8.8)

df2_dummy = df2_dummy.drop('State', axis=1)
df2_new = pd.concat([data_oh, df2_dummy], axis='columns')
df2_new.head()

[4] ✓ 0.1s Python

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

6. Membuat variable bebas yaitu X dan y.

```
M3 > Tugas Jobsheet 2 > Tugas_Jobsheet_02.ipynb > gabung = np.concatenate((ym_test, ym_pred), axis=1)
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Interrupt | Variables | Outline ... base (Python 3.8.8)

X = df2_new.iloc[:, :-1].values
y = df2_new.iloc[:, -1].values

[5] ✓ 0.4s Python
```

7. Mengubah baris kolom dari variable bebas yang telah dibuat sebelumnya.

```
Tugas_Jobsheet_02.ipynb X js02_regression.ipynb
M3 > Tugas Jobsheet 2 > Tugas_Jobsheet_02.ipynb > gabung = np.concatenate((ym_test, ym_pred), axis=1)
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Interrupt | Variables | Outline ... base (Python 3.8.8)

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

[6] ✓ 0.6s Python

...
(50, 1)
```

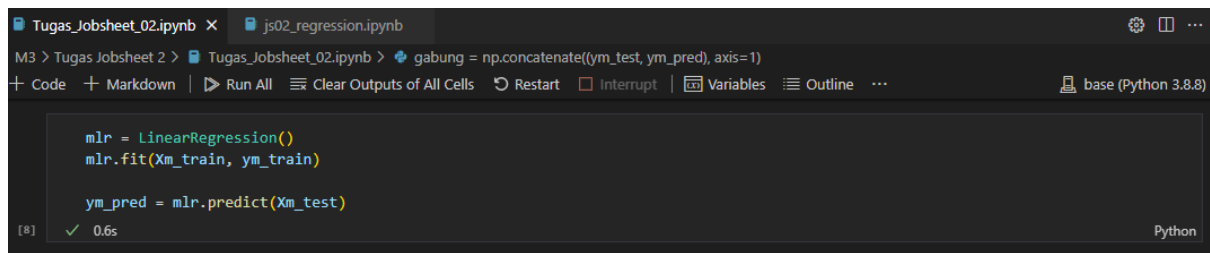
8. Membuat pemisahan data uji dan data latih.

```
Tugas_Jobsheet_02.ipynb X js02_regression.ipynb
M3 > Tugas Jobsheet 2 > Tugas_Jobsheet_02.ipynb > gabung = np.concatenate((ym_test, ym_pred), axis=1)
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Interrupt | Variables | Outline ... base (Python 3.8.8)

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

[7] ✓ 0.6s Python
```

9. Melakukan multiple Linear Regression pada data latih.



A screenshot of a Jupyter Notebook interface. The top bar shows two tabs: 'Tugas_Jobsheet_02.ipynb' and 'js02_regression.ipynb'. The notebook is running on 'base (Python 3.8.8)'. The code cell contains the following Python code:

```
M3 > Tugas Jobsheet 2 > Tugas_Jobsheet_02.ipynb > gabung = np.concatenate((ym_test, ym_pred), axis=1)

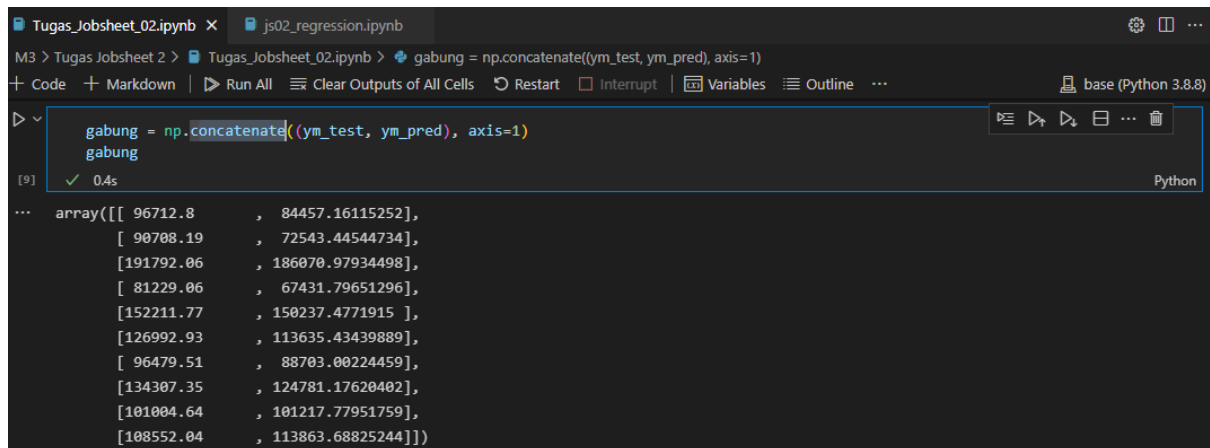
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Interrupt | Variables | Outline ...

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

ym_pred = mlr.predict(Xm_test)
```

The output of the cell is a green checkmark and the text '[8] ✓ 0.6s'.

10. Menggabungkan array data latih dengan data prediksi.



A screenshot of a Jupyter Notebook interface. The top bar shows two tabs: 'Tugas_Jobsheet_02.ipynb' and 'js02_regression.ipynb'. The notebook is running on 'base (Python 3.8.8)'. The code cell contains the following Python code:

```
M3 > Tugas Jobsheet 2 > Tugas_Jobsheet_02.ipynb > gabung = np.concatenate((ym_test, ym_pred), axis=1)

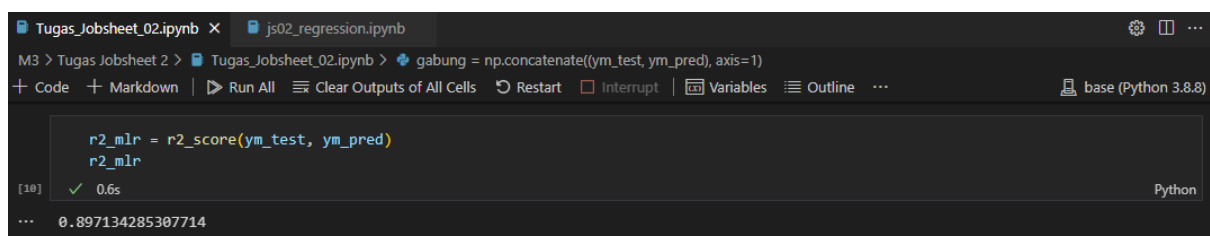
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Interrupt | Variables | Outline ...

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

The output of the cell is a green checkmark and the text '[9] ✓ 0.4s'. Below the output, a large array is displayed:

```
... 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]])
```

11. Melakukan proses r2 score.



A screenshot of a Jupyter Notebook interface. The top bar shows two tabs: 'Tugas_Jobsheet_02.ipynb' and 'js02_regression.ipynb'. The notebook is running on 'base (Python 3.8.8)'. The code cell contains the following Python code:

```
M3 > Tugas Jobsheet 2 > Tugas_Jobsheet_02.ipynb > gabung = np.concatenate((ym_test, ym_pred), axis=1)

+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Interrupt | Variables | Outline ...

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

The output of the cell is a green checkmark and the text '[10] ✓ 0.6s'. Below the output, the R-squared score is displayed:

```
... 0.897134285307714
```