

LAPORAN JOBSHEET 2

1. Melakukan pemaanginaln library yang dibutuhkan

```
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
```

✓ 2.5s

2. Berikut ini adalah fungsi untuk membaca atau import file “50_startups.csv setelah itu ada fungsi head() yang memanggil 5 data teratas

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

[4] ✓ 0.8s

	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. Berikut ini adalah proses One Hotr Encoder untuk mengubah column menjadi kategori

```
transform_list = [
    ('encoded', OneHotEncoder(dtype='int'), [3]),
    ('skip', 'passthrough', ["R&D Spend", "Administration", "Marketing Spend", "Profit"])
]
```

[5] ✓ 0.3s

4. Setelah itu merupakan proses dari column transformer untuk mengubah isi dari data

```
ct = ColumnTransformer(transform_list)
trans = ct.fit_transform(data)
```

[6] ✓ 0.7s

```
data_new = pd.DataFrame(trans, columns=ct.get_feature_names())
data_new.head()
```

[7] ✓ 0.1s

	encoded_x0_California	encoded_x0_Florida	encoded_x0_New York	R&D Spend	Administration	Marketing Spend	Profit
0	0.0	0.0	1.0	165349.20	136897.80	471784.10	192261.83
1	1.0	0.0	0.0	162597.70	151377.59	443898.53	191792.06
2	0.0	1.0	0.0	153441.51	101145.55	407934.54	191050.39
3	0.0	0.0	1.0	144372.41	118671.85	383199.62	182901.99
4	0.0	1.0	0.0	142107.34	91391.77	366168.42	166187.94

5. Setelah itu adalah proses Linear Regrersion dengan membuat variable x dan y dimana variable x untuk mengambil data dari seluruh kolom setelah proses column transformer, sedangkan y digunakan untuk mengambil data column terakhir

```
x = data_new.iloc[:, :-1].values
y = data_new.iloc[:, 6].values
```

[8] ✓ 0.1s

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

[9] ✓ 0.4s

... (50, 1)

6. Lalu proses pemanggilan library Linear Regression dengan membuat 4 variabel lalu dilakukan train_test_split untuk variable x dan y dan diberikan size = 0.2 dengan jumlah data 50

```

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=50)

```

[10] ✓ 0.5s

```

mlr = LinearRegression()
mlr.fit(x_train, y_train)

ym_pred = mlr.predict(x_test)

```

[11] ✓ 0.1s

7. Berikut proses penggabungannya

```

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

```

[12] ✓ 0.5s

```

... 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]])

```

8. Terakhir melakukan prediksi menggunakan r2_score, dengan hasil seperti berikut

```

r2_ml = r2_score(y_test, y_pred)
print(r2_ml)

```

[13] ✓ 0.4s

```

... 0.897134285307714

```