

Nama : Armanita Aning

NIM : 20.01.013.002

Kelas : python C

1. Python 9

```
# Import library pandas
import pandas as pd

# Import library numpy
import numpy as np

# Import library matplotlib dan seaborn untuk visualisasi
import matplotlib.pyplot as plt
import seaborn as sns

# Import modul LinearRegression digunakan untuk memanggil algoritma Linear Regression
from sklearn.linear_model import LinearRegression

# Import modul train_test_split digunakan untuk membagi data kita menjadi training dan testing set
from sklearn.model_selection import train_test_split

# Import Math agar program dapat menggunakan semua fungsi yang ada pada modul math (ex:sqrt)
import math

# Me-non aktifkan peringatan pada Python
import warnings
warnings.filterwarnings('ignore')
```

3. Cek Dataset

```
data = 'CarPrice_Assignment.csv'
dataset = pd.read_csv(data)
dataset.head(10)
```

	car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase	...	enginesize	fu
0	1	3	alfa-romero giulia	gas	std	two	convertible	rwd	front	88.6	...	130	
1	2	3	alfa-romero stelvio	gas	std	two	convertible	rwd	front	88.6	...	130	
2	3	1	alfa-romero Quadrifoglio	gas	std	two	hatchback	rwd	front	94.5	...	152	

```
[ ] dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 26 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   car_ID                 205 non-null   int64   
1   symboling              205 non-null   int64   
2   CarName                205 non-null   object  
3   fueltype               205 non-null   object  
4   aspiration              205 non-null   object  
5   doornumber             205 non-null   object  
6   carbody                205 non-null   object  
7   drivewheel             205 non-null   object  
8   enginelocation         205 non-null   object  
9   wheelbase              205 non-null   float64  
10  carlength              205 non-null   float64  
11  carwidth               205 non-null   float64  
12  carheight              205 non-null   float64  
13  curbweight             205 non-null   int64   
14  enginetype             205 non-null   object  
15  cylindernumber         205 non-null   object  
16  enginesize              205 non-null   int64   
17  fuelsystem             205 non-null   object  
18  boreratio              205 non-null   float64  
19  stroke                 205 non-null   float64  
20  compressionratio       205 non-null   float64  
21  horsepower              205 non-null   int64   
22  peakrpm                205 non-null   int64   
23  citympg                205 non-null   int64   
24  highwaympg             205 non-null   int64   
25  price                  205 non-null   float64  
dtypes: float64(8), int64(8), object(10)
memory usage: 41.8+ KB
```

```
[ ] dataset.isna().sum()
```

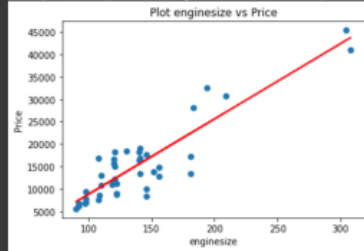
car_ID	0
symboling	0
CarName	0
fueltype	0
aspiration	0
doornumber	0
carbody	0
drivewheel	0

```
[121]
[146]
[ 97]
[122]
[304]]
```

```
[ ] # Prediksi data test
y_prediksi = regressor.predict(X_test)
```

```
plt.scatter(X_test,y_test)
plt.plot(X_test, y_prediksi, c='r')
plt.xlabel('engine size')
plt.ylabel('Price')
plt.title('Plot engine size vs Price')
```

```
Text(0.5, 1.0, 'Plot engine size vs Price')
```



```
# Prediksi harga mobil dengan nilai engine size tertentu
```

```
print('nilai prediksi harga dengan engine size 100 : ',regressor.predict([[90]]))
print('nilai prediksi harga dengan engine size 150 : ',regressor.predict([[150]]))
print('nilai prediksi harga dengan engine size 200 : ',regressor.predict([[200]]))
```

```
nilai prediksi harga dengan engine size 100 : [[7098.5663135]]
nilai prediksi harga dengan engine size 150 : [[17188.98418658]]
nilai prediksi harga dengan engine size 200 : [[25597.66574748]]
```

2. Python 10

```
# IMPORT LIBRARY
import pandas as pd
from matplotlib import pyplot as plt
from matplotlib import pyplot as plt

# READ CSV
df = pd.read_csv("dataasuransi.csv")

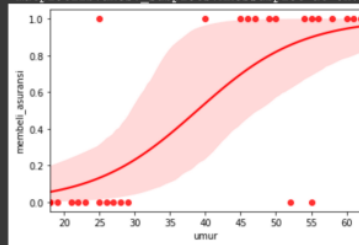
# TAMPIL DATAFRAME
df.head(5)
```

```
[4] ✓ 0.2s
```

	umur	membeli_asuransi
0	22	0
1	25	0
2	47	1
3	52	0
4	46	1

```
[7] # TAMPIL KE DALAM PLOT
sb.regplot(x='umur', y='membeli_asuransi', data=df, logistic=True, color='red')

[7] /usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning:
import pandas.util.testing as tm
/usr/local/lib/python3.7/dist-packages/statsmodels/genmod/families/family.py:890: RuntimeWarning:
n_endog_mu = self._clean((1. - endog) / (1. - mu))
<matplotlib.axes._subplots.AxesSubplot at 0x7f857fc87ad0>
```



```
[8] from sklearn.model_selection import train_test_split

[9] # PISAHKAN DATA TRAINING DAN DATA TEST
X_train, X_test, y_train, y_test = train_test_split(df[['umur']], df.membeli_asuransi,

[10] # TAMPIL NILAI VARIABEL TEST
# (BARIS KODE INI BOLEH DISKIP, TUJUANNYA HANYA UNTUK CEK NILAI)
X_test
```

```
umur
[10] 18  19
      2  47
      13 29
```

```
[11] X_train
```

```
umur
14  49
8   62
1   25
16  25
12  27
11  28
3   52
15  55
7   60
5   56
21  26
26  23
17  58
20  21
[11] 6   55
      25 54
      10 18
      19 18
      0  22
      24 50
      23 45
      22 40
      9  61
      4  46
```

```
[13] # IMPORT LIBRARY REGRESI LOGISTIK
      from sklearn.linear_model import LogisticRegression

[14] # MEMBUAT MODEL
      model = LogisticRegression()

[15] # MODEL MELAKUKAN TRAINING
      model.fit(X_train, y_train)

      LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                          intercept_scaling=1, l1_ratio=None, max_iter=100,
                          multi_class='auto', n_jobs=None, penalty='l2',
                          random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                          warm_start=False)

[16] # PREDIKSI
      model.predict(X_test)

      array([0, 1, 0])

[18] # CEK KEAKURATAN MODEL
      model.score(X_test, y_test)

      1.0

[19] # PREDIKSI JIKA UMUR = 60 TAHUN
      model.predict([[60]])

      array([1])

[27] # PREDIKSI JIKA UMUR = 1 TAHUN
      model.predict([[1]])

      array([0])
```

3. Python 11

```
In [12]: import pandas as pd
          from matplotlib import pyplot as plt
          import seaborn as sb

          df = pd.read_csv("datatitanic.csv")
          df
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.28
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.10
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.750

891 rows x 12 columns



```
In [15]: df.isnull()
```

[illegible]

```

886 False      False      False      False      False      False      False      False      False      False      True      False
887 False      False      False      False      False      False      False      False      False      False      False      False
888 False      False      False      False      False      True       False      False      False      False      True       False
889 False      False      False      False      False      False      False      False      False      False      False      False
890 False      False      False      False      False      False      False      False      False      False      True       False

```

891 rows x 12 columns

```

In [16]: # Jumlah null pada kolom
df.isnull().sum()

```

```

PassengerId    0
Survived        0
Pclass         0
Name           0
Sex            0
Age           177
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin        687
Embarked       2
dtype: int64

```

```

Embarked       2
dtype: int64

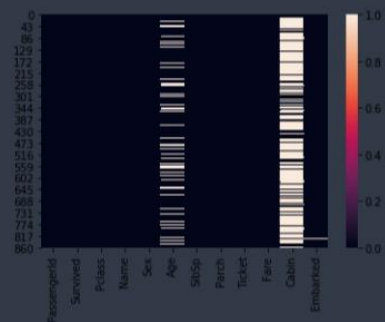
```

```

In [17]: # Melihat data-data yang kosong pada kolom
sb.heatmap(df.isnull())

```

<matplotlib.axes._subplots.AxesSubplot at 0x27cf9467898>



```

In [18]: # Hapus kolom Cabin
# axis=1 atau 'kolom': menerapkan fungsi ke tiap baris.
df.drop("Cabin", axis=1, inplace=True)

```



```

In [29]: model.fit(X_train, y_train)

C:\Users\kevin\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:432: FutureWarning: Default solver
o 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)

LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                    intercept_scaling=1, l1_ratio=None, max_iter=100,
                    multi_class='warn', n_jobs=None, penalty='l2',
                    random_state=None, solver='warn', tol=0.0001, verbose=0,
                    warm_start=False)

In [30]: model.predict(X_test)

array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0], dtype=int64)

In [31]: model.score(X_test, y_test)

0.6111111111111112

In [32]: model.predict([[43]])

array([0], dtype=int64)

In [33]: # KNN

```