

LAPORAN PRAKTIKUM

BIG DATA

PERTEMUAN KE – 03



Disusun Oleh :

NIM : 195610007

Nama : Ara Widhi Astutik

Kelas : Sistem Informasi-1

UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA

2021/2022

BAB I

EKSPLORASI DATA

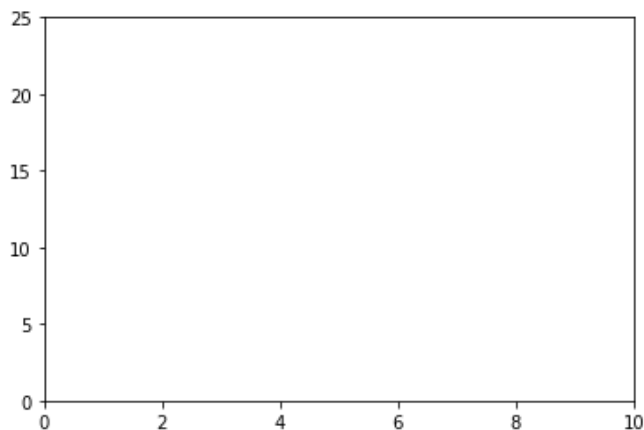
PEMBAHASAN

Praktik

1. Menggunakan fungsi Membuat Ploting data pada sumbu ordinat
 - a. Membuat garis ordinat untuk plotting data menggunakan perintah dibawah ini :

```
In [4]: import matplotlib.pyplot as plt  
plt.axis([0,10,0,25])
```

```
Out[4]: (0.0, 10.0, 0.0, 25.0)
```

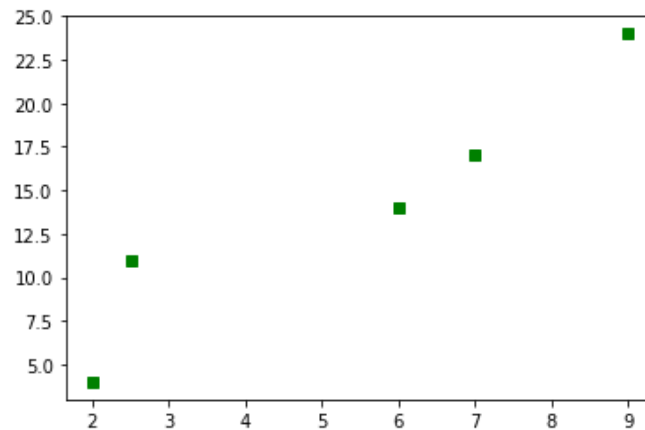


Uraian :

Source code diatas akan menampilkan garis koordinat untuk plotting data, menggunakan `plt.axis()` untuk menetapkan batas x dan y dengan satu panggilan dengan menetapkan `[xmin, xmax, ymin, ymax]`.

- b. Menampilkan data ordinat untuk data $\{(2,4), (2.5,11), (6,14), (7,17), (9,24)\}$ gunakan coding dibawah ini

```
In [5]: plt.plot([2,2.5,6,7,9],[4,11,14,17,24],'gs')
plt.show()
```

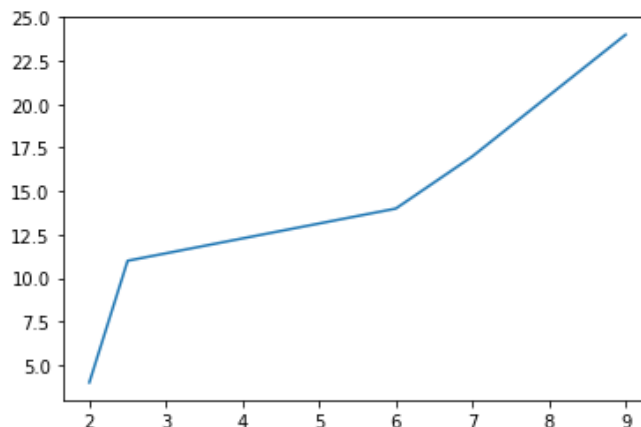


Uraian :

Jika pada langkah pertama membuat garis koordinat sesuai yang ditentukan, maka pada langkah ini akan menentukan titik koordinat dari masing-masing data (data yang ditentukan user), menggunakan `plt.show()` diakhir setiap code untuk visualisasi.

- c. Membuat garis pada titik-titik ordinat diatas dengan menghilangkan “gs” dan tampilan dibawah ini

```
In [6]: plt.plot([2,2.5,6,7,9],[4,11,14,17,24])
plt.show()
```

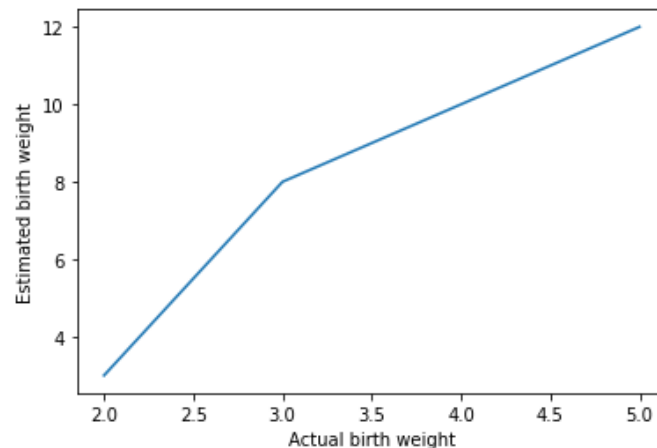


Uraian :

Penggunaan “gs” pada langkah sebelumnya membuat titik koordinat sebelum nantinya akan dibuat grafik dari sebuah data. Pada langkah ini ketika penggunaan “gs” dihilangkan maka akan menampilkan garis sesuai titik koordinat yang telah ditentukan.

- d. Membuat plotting data dengan memberi nama variabel sumbu x dan sumbu y, seperti dibawah ini

```
In [7]: import matplotlib.pyplot as plt
plt.plot([2,3,4,5],[3,8,10,12])
plt.xlabel('Actual birth weight')
plt.ylabel('Estimated birth weight')
plt.show()
```



Uraian :

Sama seperti pada langkah sebelumnya yaitu membuat grafik dari sebuah data yang dituangkan dalam bentuk seperti output diatas, namun yang membedakannya adalah pada langkah ini akan memberikan nama label disetiap sumbunya yaitu pada sumbu X dan sumbu Y. Maka penambahan penulisan source codenya adalah plt. Xlabel dan plt.ylabel.

2. Manajemen data file csv

- a. Menampilkan data → Buatlah data seperti dibawah ini dan simpan dengan nama data3.csv

Mendefiniskan library yang akan kita gunakan. Data disimpan dengan nama file data3.csv dalam satu folder dengan file python yang sedang running. Kita tampilkan menggunakan pandas dengan variabel dataset.

```
In [10]: #mempersiapkan library
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

dataset = pd.read_csv('data3.csv')
dataset
```

Out[10]:

	Nilai UTS	Nilai UAS
0	80	87
1	78	81
2	72	70
3	68	65
4	78	90
5	74	78
6	90	87
7	76	78
8	73	81
9	66	88
10	89	94
11	75	86

Activate Window
Go to Settings to act

Untuk memastikan data tersimpan dan dapat dibaca oleh python gunakan script dibawah ini

```
In [11]: dataset.keys()
```

Out[11]: Index(['Nilai UTS', 'Nilai UAS'], dtype='object')

```
In [12]: dataset.shape
```

Out[12]: (12, 2)

Uraian :

Input [22] digunakan untuk menampilkan nama variabel pada data3.csv

Input [23] digunakan untuk melihat jumlah record dan jumlah atribut dari data3.

b. Scatter Plot Data

```
In [8]: import pandas as pd
pd.plotting.register_matplotlib_converters()
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
In [9]: insurance = 'insurance.csv'
insurance_data = pd.read_csv(insurance)
insurance_data.head()
```

```
Out[9]:
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	yes	southwest	1725.55230
2	28	male	33.000	3	no	southwest	4449.46200
3	33	male	22.705	0	no	southwest	21984.47061
4	32	male	28.880	0	no	southwest	3866.85520

Apakah semua data bisa ditampilkan? Jelaskan!

Uraian :

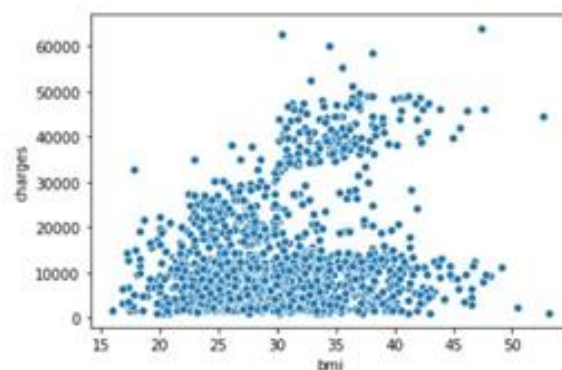
Untuk menampilkan semua data seperti output diatas, maka terlebih dahulu membuat data 'insurance.csv'. jika dilihat dari source code diatas menggunakan matplotlib inline yaitu digunakan untuk meng-embed gambar plot statis didalam notebook. Dan menggunakan import seaborn as sns , dimana seaborn sendiri merupakan library untuk membuat grafik dan statistik dengan menggunakan Python. Library tersebut dibangun berdasarkan library matplotlib serta terintegrasi dengan struktur data pada pandas.

Langkah terakhir adalah membaca dan menampilkan data yang ada pada insurance.csv, menggunakan perintah pd.read_csv untuk membaca dan perintah insurance_data.head() untuk menampilkannya.

Fungsi untuk membuat scatter plot

```
In [5]: sns.scatterplot(x=insurance_data['bmi'], y=insurance_data['charges'])
```

```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2da5a4638>
```



Uraian :

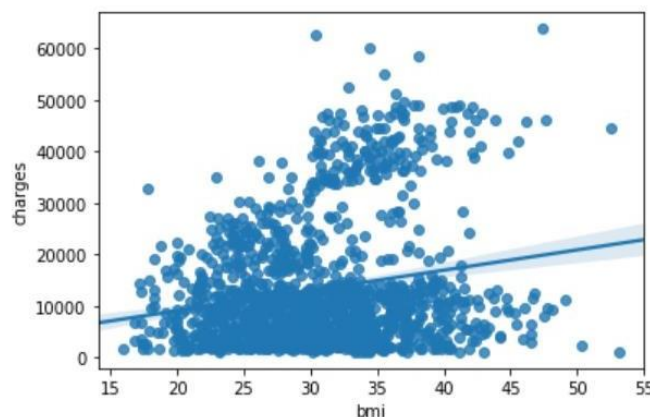
- Install seaborn

```
Command Prompt
C:\Users\Asus>pip install seaborn
Collecting seaborn
  Downloading seaborn-0.11.2-py3-none-any.whl (292 kB)
    | 292 kB 192 kB/s
Requirement already satisfied: numpy>=1.15 in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from seaborn) (1.18.5)
Requirement already satisfied: matplotlib>=2.2 in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from seaborn) (3.4.2)
Collecting scipy>=1.0
  Downloading scipy-1.8.0-cp38-cp38-win32.whl (33.4 MB)
    | 33.4 MB 297 kB/s
Requirement already satisfied: pandas>=0.23 in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from seaborn) (1.2.2)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from matplotlib>=2.2->seaborn) (1.3.1)
Requirement already satisfied: cycler>=0.10 in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from matplotlib>=2.2->seaborn) (0.10.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from matplotlib>=2.2->seaborn) (2.4.7)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from matplotlib>=2.2->seaborn) (2.8.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from matplotlib>=2.2->seaborn) (8.2.0)
Requirement already satisfied: pytz>=2017.3 in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from pandas>=0.23->seaborn) (2021.1)
Requirement already satisfied: six in c:\users\asus\appdata\local\programs\python\python38-32\lib\site-packages (from cycler>=0.10->matplotlib>=2.2->seaborn) (1.15.0)
Installing collected packages: scipy, seaborn
Successfully installed scipy-1.8.0 seaborn-0.11.2
```

Sebelum menampilkan data kedalam bentuk seaborn, maka terlebih dahulu menginstall seaborn pada python seperti pada gambar diatas. Setelah itu mulai memanggil data ke dalam seaborn untuk ditampilkan sesuai ketentuan yang telah ditentukan. Pada source code diatas akan menampilkan data 'bmi' dan 'charges' pada "insurance.scv" untuk kemudian nantinya akan ditampilkan dalam bentuk seaborn. Scatterplot sendiri digunakan untuk mengetahui apakah sebuah variabel memiliki hubungan yang berbanding lurus atau berbanding terbalik.

Untuk membuat plotting regresi dengan variabel x = bmi dan y = charges

```
In [6]: sns.regplot(x=insurance_data['bmi'], y=insurance_data['charges'])
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2d9cef5f8>
```



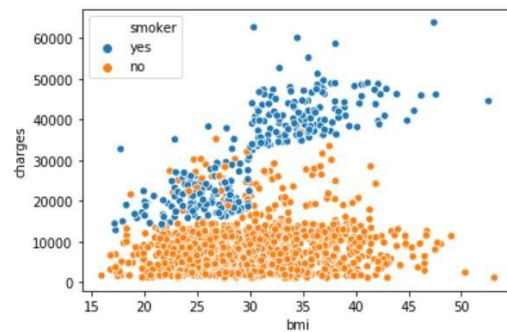
Uraian :

Pada source code diatas menggunakan perintah sns.regplot, ini digunakan untuk memplot data dan model regresi linier fit.

Menampilkan plot data kategori dari salah satu variabel smoker

```
In [13]: sns.scatterplot(x=insurance_data['bmi'], y=insurance_data['charges'],  
                        hue = insurance_data['smoker'])
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x7fa2d9301c18>
```



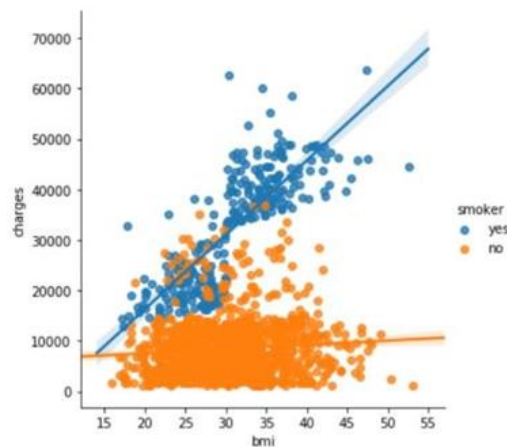
Uraian :

Pada langkah ini, sama seperti langkah membuat scatter plot pada diawal, namun yang membedakannya adalah menambahkan data 'smoker', sehingga tampilan yang dihasilkan akan menampilkan banyak data dari 'bmi', 'charges', dan 'smoker'.

Menampilkan prediksi data tentang smoker melalui garis regresi

```
In [18]: sns.lmplot(x='bmi', y='charges', hue='smoker', data = insurance_data)
```

```
Out[18]: <seaborn.axisgrid.FacetGrid at 0x7fa2d81ffd68>
```



Uraian :

Sedangkan untuk menampilkan prediksi data tentang smoker menggunakan perintah sns.lmplot. Jika regplot() melakukan fit dan plot model regresi linier sederhana. Maka lmplot() menggabungkan regplot() dan FacetGrid.

Latihan

1. Jelaskan fungsi Seaborn.

Uraian :

Seaborn adalah salah satu library open-source dari Python yang berguna untuk menciptakan visualisasi data statistik dengan tampilan berkualitas tinggi. Seaborn dibangun dengan basis library Python lainnya, Matplotlib serta terintegrasi dengan struktur data Pandas.

2. Jelaskan perbedaan scatterplot, regplot, Implot

Uraian :

- a. Scatterplot

Adalah sebuah grafik yang biasa digunakan untuk melihat suatu pola hubungan antara 2 variabel. Untuk bisa menggunakan scatterplot, skala data yang digunakan haruslah skala interval dan rasio.

- b. Regplot

Metode ini digunakan untuk memplot data dan model regresi linier fit.

- c. Implot

Jika `regplot()` melakukan fit dan plot model regresi linier sederhana. Maka `Implot()` menggabungkan `regplot()` dan `FacetGrid`.

BAB II

DAFTAR PUSTAKA

[http://doditsuprianto.blogspot.com/2018/06/visualisasi-dengan-matplotlib.html#:~:text=Metode%20plt.axis\(\)%20memungkinkan,%2C%20xmax%2C%20ymini%2C%20ymax%5D%3A](http://doditsuprianto.blogspot.com/2018/06/visualisasi-dengan-matplotlib.html#:~:text=Metode%20plt.axis()%20memungkinkan,%2C%20xmax%2C%20ymini%2C%20ymax%5D%3A)

<https://segalahal.com/2020/11/03/pengenalan-matplotlib-library-python-untuk-visualisasi-data/#:~:text=%E2%80%9C%25matplotlib%20inline%E2%80%9D%20digunakan%20untuk,plot%20statis%20di%20dalam%20notebook.>

<https://auftechnique.com/berkenalan-dengan-seaborn-python/>

<https://www.gubukpintar.com/2020/12/seaborn-python.html#:~:text=Seaborn%20adalah%20salah%20satu%20library%20open%2Dsource%20dari%20Python%20yang,terintegrasi%20dengan%20struktur%20data%20Pandas.>

<https://www.prasetyowijaya.com/2009/02/scatter-plot.html>

<https://www.geeksforgeeks.org/python-seaborn-regplot-method/>

<https://datascience.stackexchange.com/questions/44192/what-is-the-difference-between-regplot-and-Implot-in-seaborn>