

PYTHON ASSIGNMENT 2

Machine Learning

Python adalah sebuah bahasa pemrograman interpretatif multiguna yang dirancang oleh *Python Software Foundation* yang fokusnya pada tingkat keterbacaan kode. Python juga diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan dengan sintaksis kode yang jelas, serta dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Dalam Python sendiri terdapat sebuah pemrograman menggunakan *Machine Learning* (ML).

Machine Learning (ML) adalah cabang aplikasi kecerdasan buatan/*Artificial Intelegant* (AI) yang berfokus pada pengembangan sebuah sistem yang mampu belajar ‘sendiri’ tanpa harus berulang kali di program oleh manusia. Aplikasi ini membutuhkan data sebagai bahan ajar (*training*) sebelum mengeluarkan output dan biasanya aplikasi ini berada dalam domain spesifik atau tidak bisa diterapkan secara general untuk semua permasalahan. Dalam Python sendiri, bahan ajar untuk *Machine Learning* yang paling terkenal adalah Iris.

Python Iris adalah sebuah rangkaian latihan *Machine Learning* pada Python yang paling terkenal, dimana bunga Iris menjadi subjeknya. Dalam *dataset* itu, ada 3 jenis bunga Iris (Iris Setosa, Iris Versicolor, dan Iris Virginica) dimana setiap bunganya memiliki banyak data yang sama. Dataset ini juga berisi panjang serta lebar setiap petal dan sepal yang dimiliki oleh ketiga bunga Iris tersebut. Dalam latihan ini, data haruslah dalam bentuk CSV (Comma Separated Values). CSV sendiri adalah suatu format data dalam basis data dimana setiap record dipisahkan dengan tanda koma (,) atau titik koma.

1. Penjelasan Script

Dalam tugas kedua, data yang digunakan berasal dari situs Statistik Selandia Baru (<https://www.stats.govt.nz/large-datasets/csv-files-for-download/>). Situs ini menyediakan beragam data CSV yang dapat digunakan sebagai subjek dalam latihan *machine learning*

pada Python. Data CSV yang diambil adalah data sensus penduduk nasional Selandia Baru pada 30 Juni 2018. Dari sini kita masuk ke Python 3 dan memasukkan perintah

```
import pandas as pd
```

Setelah memasukkan perintah pada script pertama dan dijalankan, masukkan lagi perintah kedua yaitu input data CSV yang diinginkan.

```
variabel_data = pd.read_csv ('website dataset csv')
```

Data kemudian diberi variabel bernama *population*.

```
population = pd.read_csv  
( 'https://www.stats.govt.nz/assets/Uploads/National-  
population-estimates/National-population-estimates-At-30-  
June-2018/Download-data/national-population-estimates-at-30-  
june-2018-components-of-population-change-csv.csv' )
```

Perintah ini kemudian dijalankan, namun hasilnya tidak dideklarasikan secara langsung. Perintah selanjutnya adalah serangkaian perintah yang jika di deklarasikan akan menampilkan informasi mengenai dataset tersebut.

```
population.info()
```

Jika perintah ini dijalankan, output yang keluar berisi tentang informasi mengenai dataset, berisi jumlah kolom, tipe data, dan banyak penyimpanan yang digunakan.

```
population.describe()
```

Jika perintah ini dijalankan, output yang keluar berupa data table yang sudah ditampilkan pada output perintah info diatas, table berisi *count*, *mean*, *std*, *min*, 3 persenan dengan jeda 25%, dan *max*.

```
population.head()  
population.tail()
```

Perintah head dan tail menunjukkan data tabel. Head sendiri menunjukkan 5 baris teratas dalam tabel, sementara tail menunjukkan 5 baris terakhir dalam tabel.

Setelah mendapatkan informasi yang diinginkan (informasi dari perintah *population.info()*), langkah selanjutnya adalah menginputkan perintah untuk membuat plot atau grafik dari dataset yang sudah dimasukkan. Input grafik pada Python sedikit berbeda dengan input grafik pada MATLAB dimana kita harus meng-import plot sebelum plot itu ditampilkan. Adapun perintah plot adalah sebagai berikut.

```
import matplotlib.pyplot as plt
```

Setelah meng-import plot, langkah selanjutnya adalah memasukkan data-data yang diperlukan untuk membuat plot, seperti ukuran grafik, variabel-variabel, nama (label) dari sumbu x dan y grafik, judul grafik, legenda atau keterangan grafik. Data variabel grafik didapatkan dari hasil *population.info()*.

```
plt.figure(figsize=(20,10))  
plt.plot(population['period'], label = 'Period')  
plt.plot(population['status'], label = 'Status')  
plt.plot(population['natural_increase'], label = 'Natural  
Increase')  
plt.plot(population['net_migration'], label = 'Net Migration')  
plt.plot(population['population_change'], label = 'Population  
Change')
```

```
plt.plot(population['percent_population_change'], label =  
'Percent Population Change')  
plt.plot(population['population'], label = 'Population')  
plt.xlabel('Data Count')  
plt.ylabel('Value')  
plt.title('Plot of each parameter')  
plt.legend();  
plt.show()
```

Perintah `plt.figure(figsize=(20,10))` memiliki maksud bahwa ukuran grafik yang kita inginkan berukuran panjang 20 dan lebar 10. Perlu diingat jika menulis variabel untuk tidak memberi spasi dan sebaiknya ditulis dengan *underscore* (`_`) untuk menjelaskan spasi. Setelah perintah diketik dan dijalankan, grafik yang diinginkan akan muncul.

2. Screenshot Script dan Grafik

```
In [1]: import pandas as pd  
  
In [2]: population = pd.read_csv('https://www.stats.govt.nz/assets/Uploads/National-population-estimates/National-population-estimates-A')  
<----->  
  
In [3]: population.info()  
  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 41 entries, 0 to 40  
Data columns (total 7 columns):  
period                41 non-null object  
status                41 non-null object  
natural_increase      41 non-null float64  
net_migration         41 non-null float64  
population_change     41 non-null float64  
percent_population_change 41 non-null float64  
population            41 non-null float64  
dtypes: float64(5), object(2)  
memory usage: 2.4+ KB
```

Gambar 2.1 Script untuk Grafik

THOHARINNISA PAMBAYUN
03411940000001
DEPARTEMEN TEKNIK GEOFISIKA – ITS

```
In [8]: population.describe()
```

Out[8]:

	natural_increase	net_migration	population_change	percent_population_change	population
count	41.000000	41.000000	41.000000	41.000000	4.100000e+01
mean	7726.829268	8226.829268	15360.975610	0.337073	4.509144e+06
std	985.399510	8082.481809	8098.236785	0.168971	1.851270e+05
min	5500.000000	-5300.000000	1500.000000	0.030000	4.259800e+06
25%	7100.000000	1900.000000	9100.000000	0.210000	4.373900e+06
50%	7600.000000	6900.000000	13800.000000	0.310000	4.442100e+06
75%	8500.000000	16000.000000	23000.000000	0.500000	4.647300e+06
max	9900.000000	21900.000000	29400.000000	0.620000	4.885300e+06

Gambar 2.2 Script untuk Grafik

```
In [9]: population.head()
```

Out[9]:

	period	status	natural_increase	net_migration	population_change	percent_population_change	population
0	2008-06-30	F	8500.0	-3200.0	4200.0	0.10	4259800.0
1	2008-09-30	F	8100.0	2900.0	9800.0	0.23	4269600.0
2	2008-12-31	F	9200.0	2700.0	10700.0	0.25	4280300.0
3	2009-03-31	F	9300.0	5100.0	13200.0	0.31	4293500.0
4	2009-06-30	F	8400.0	1900.0	9100.0	0.21	4302600.0

```
In [11]: population.tail()
```

Out[11]:

	period	status	natural_increase	net_migration	population_change	percent_population_change	population
36	2017-06-30	F	6800.0	10500.0	17300.0	0.36	4793900.0
37	2017-09-30	F	5500.0	18500.0	24000.0	0.50	4817800.0
38	2017-12-31	F	7500.0	19100.0	26600.0	0.55	4844400.0
39	2018-03-31	F	7300.0	19900.0	27200.0	0.56	4871600.0
40	2018-06-30	P	6200.0	7500.0	13700.0	0.28	4885300.0

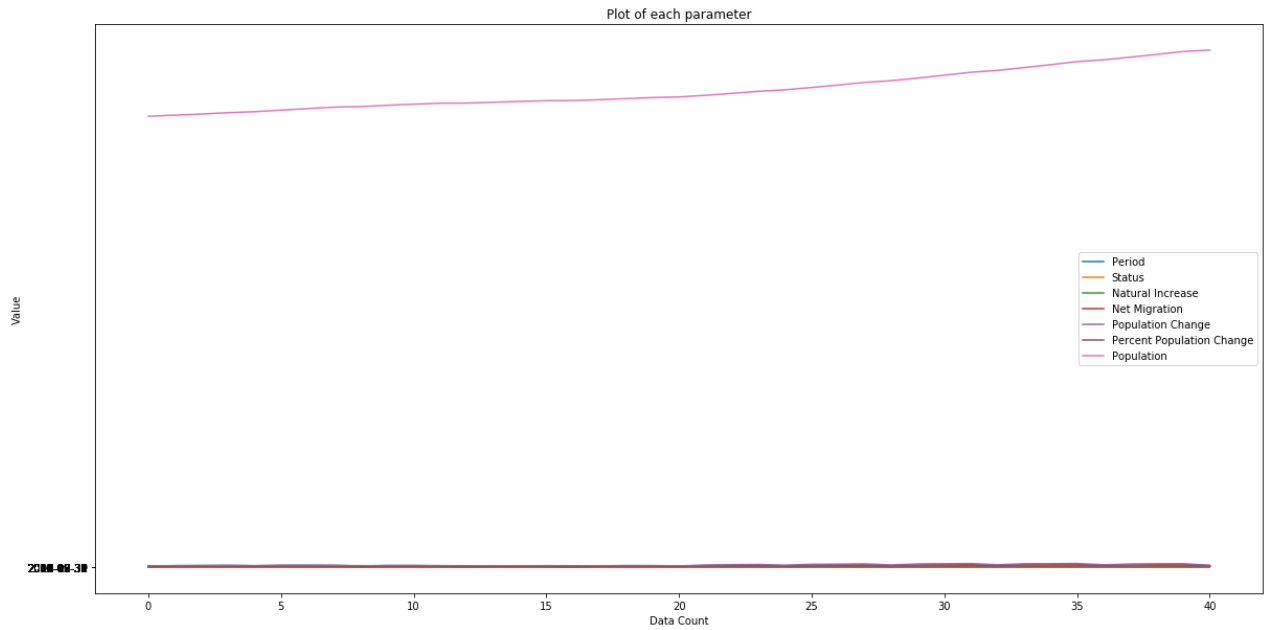
Gambar 2.3 Script untuk Grafik

```
In [5]: import matplotlib.pyplot as plt
```

```
In [7]: plt.figure(figsize=(20,10))
plt.plot(population['period'], label = 'Period')
plt.plot(population['status'], label = 'Status')
plt.plot(population['natural_increase'], label = 'Natural Increase')
plt.plot(population['net_migration'], label = 'Net Migration')
plt.plot(population['population_change'], label = 'Population Change')
plt.plot(population['percent_population_change'], label = 'Percent Population Change')
plt.plot(population['population'], label = 'Population')
plt.xlabel('Data Count')
plt.ylabel('Value')
plt.title('Plot of each parameter')
plt.legend();
plt.show()
```

Gambar 2.4 Script untuk Menampilkan Grafik

THOHARINNISA PAMBAYUN
03411940000001
DEPARTEMEN TEKNIK GEOFISIKA – ITS



Grafik 2.1 Hasil Plot yang diinginkan