

Laporan Praktikum Pertemuan 5
Data Science Lanjut
Exploratory Data Analitics With Python For Beginner

Dibuat Oleh

Nama : Muhamad Faisal Halim
NIM : 19.240.0163
Kelas : -
Mata Kuliah : Data Science Lanjut

Mahasiswa Pertukaran Mahasiswa.
Universitas Muhammadiyah Kalimantan Timur
~ STMIK Widya Pratama Pekalongan

Note

Data pada praktikum ini disamakan dengan data yang ada pada contoh yang diberikan di Openlearning UMKT.

Materi dan Praktikum

TITLE : Exploratory Data Analitics With Python For Beginner

Library adalah kumpulan kode pra-gabungan yang dapat digunakan secara berulang untuk mengurangi waktu yang diperlukan untuk membuat kode. Mereka sangat berguna untuk mengakses kode pra-tulis yang sering digunakan, daripada menuliskannya dari awal setiap saat. Mirip dengan perpustakaan fisik, ini adalah kumpulan sumber daya yang dapat digunakan kembali, yang berarti setiap perpustakaan memiliki sumber root.

NumPy

Adalah singkatan dari numerical python, dimana library ini berfokus pada pemrosesan komputansi numeric dan array multidimensional.

Pandas

Library ini memudahkan kita dalam melakukan manipulasi, cleansing, maupun analisis struktur data. Pandas menggunakan konsep array numpy namun memberikan index pada array tersebut sehingga bisa disebut series atau dataframe.

Pandas Series

Adalah array berlabel satu dimensi yang mampu menampung data jenis apa pun (integer, string, float, objek python, dll.). Label sumbu secara kolektif disebut indeks. Seri Panda tidak lain adalah kolom dalam lembar excel.

<pre>import pandas as pd x = pd.Series([6,3,4,6]) print(x) print(type(x))</pre> <p>✓ 0.7s</p> <table border="1"><tr><td>0</td><td>6</td></tr><tr><td>1</td><td>3</td></tr><tr><td>2</td><td>4</td></tr><tr><td>3</td><td>6</td></tr></table> <p>dtype: int64 <class 'pandas.core.series.Series'></p>	0	6	1	3	2	4	3	6	<pre>import pandas as pd import numpy as np data = np.array(['a','b','c','d']) karakter = pd.Series(data,index = ['satu','dua','tiga','empat']) print(karakter) print(type(karakter))</pre> <p>✓ 0.1s</p> <table border="1"><tr><td>satu</td><td>a</td></tr><tr><td>dua</td><td>b</td></tr><tr><td>tiga</td><td>c</td></tr><tr><td>empat</td><td>d</td></tr></table> <p>dtype: object <class 'pandas.core.series.Series'></p>	satu	a	dua	b	tiga	c	empat	d
0	6																
1	3																
2	4																
3	6																
satu	a																
dua	b																
tiga	c																
empat	d																

Pandas DataFrame

Adalah sebuah koleksi berurutan dengan nama dan jenis, atau bisa dikatakan dataframe pada pandas adalah koleksi data tabular yang memiliki sumbu berlabel (baris dan kolom).

Dengan adanya dataframe kita dapat dimudahkan dalam melakukan pembacaan data yang akan diubahnya menjadi tabel. Kita juga dapat menggunakan beberapa operasi table pada umumnya, seperti join, distinct, group by, agregasi dan teknik lain yang terdapat pada SQL.

```
import pandas as pd

df =pd.DataFrame({'tipe_int':[50,21],'tipe_string':['a','b']})

print(df)
print(type(df))
```

✓ 0.5s

	tipe_int	tipe_string
0	50	a
1	21	b

<class 'pandas.core.frame.DataFrame'>

SciPy

Dibuat untuk bekerja dengan array NumPy, dan menyediakan komputansi numeric yang raman bagi pengguna dan efisien, seperti rutinitas integrasi, differensiasi dan optimasi numeric

Matplotlib

Sebuah library yang khusus digunakan dalam pembuatan visualisasi data, library ini memiliki kapabilitas untuk memvisualisasikan data dalam 2 dimensi, contoh visualisasi yang dapat dibuat oleh library ini.

1. Line Chart
2. Pie Chart
3. Bar Chart
4. Box Plot Chart
5. Scatter Chart
6. Vionlin Chart
7. Error Bar Chart

Membaca file menggunakan pandas

Dalam melakukannya kita bisa membaca file secara offline seperti pada kode dibawah ini atau secara online. Agar bisa membaca file secara online, pada bagian `.read_csv(...)` isikan dengan url file kita.

```
import pandas as pd
order_df = pd.read_csv("./Dataset/usia_karyawan.csv")
print(order_df.head())
```

✓ 0.7s

	Kelompok Usia	Laki-laki	Perempuan
0	21-25 tahun	45	53
1	26-30 tahun	124	156
2	31-35 tahun	198	208
3	36-40 tahun	353	332
4	41-45 tahun	516	420

Kode dibawah ini menampilkan output jumlah baris dan kolom pada file csv kita

```
import pandas as pd
order_df = pd.read_csv("./Dataset/usia_karyawan.csv")
print(order_df.shape)
```

✓ 0.1s

(7, 3)

Sebenarnya pada laporan ini masih banyak yang perlu di tuliskan dan dipraktikan, namun karena sudah pernah di praktikan pada module-module sebelumnya jadi sekian dari laporan kali ini.

Mohon maaf atas kekurangannya,

