

MODUL PRAKTIKUM KECERDASAN BUATAN

PERTEMUAN 1 – *REPRESENTASI PENGETAHUAN*

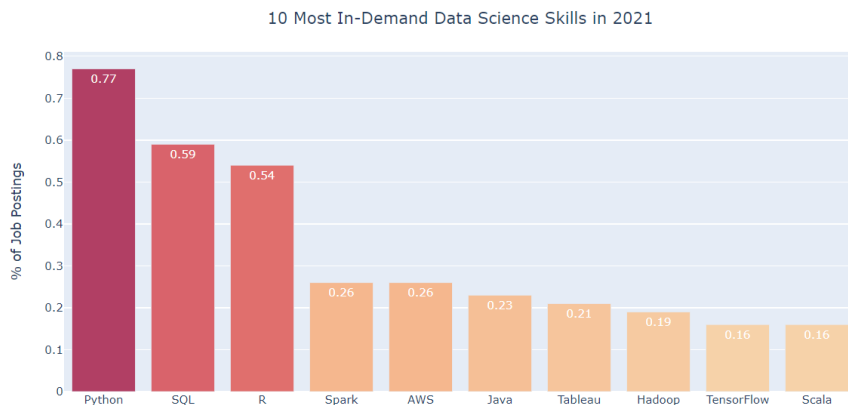
Tools : Jupyter notebook

Bahasa Pemrograman : Python

1. Pengenalan Python dan Google Colab

Python1 didesain oleh Guido van Rossum2 pada 1991 sebagai general-purpose language dan bersifat open-source. Python merupakan bahasa pemrograman populer tetapi juga memiliki properti yang sangat baik untuk programmer pemula, sehingga ideal untuk orang yang belum pernah memprogram sebelumnya. Python juga bersifat cross-platform, artinya bisa digunakan pada sistem operasi Windows, Mac OS, dan Linux.

Beberapa kelebihan dari properti tersebut adalah kode yang mudah dibaca, pengetikan dan penggunaan memori yang dinamis. Python adalah bahasa interpreter, sehingga kode dieksekusi segera di konsol Python tanpa memerlukan langkah kompilasi ke bahasa mesin. Selain konsol Python (yang disertakan dengan instalasi Python apa pun), Anda dapat menemukan konsol interaktif lainnya, seperti IPython (Interactive Python)³ dan Google Colab⁴ yang memberi Anda lingkungan yang lebih interaktif untuk mengeksekusi kode Python Anda.



Gambar 1 10 Most In-Demand Data Science Skills in 2021 (<https://towardsdatascience.com/>)

2. Library Dasar Python

Komunitas Python adalah salah satu komunitas pemrograman paling aktif dengan banyaknya pengembang library. Library Python yang paling populer untuk proyek data science adalah NumPy, SciPy, Pandas, Scikit-Learn, dan Matplotlib.

a. NumPy and SciPy: Numeric and Scientific Computation

- Library dasar untuk perhitungan saintifik (scientific computing) dengan Python
<https://numpy.org/>
- Alternatif untuk Python List: Numpy Array untuk n dimensi
- Mudah digunakan dan bersifat open source
- Jika library belum terpasang, tuliskan perintah instalasi :
`pip install numpy`
- Kemudian impor
`import numpy as np`

b. SciPy

- SciPy (dibaca “Sigh Pie ””) merupakan library yang bersifat open source dan tersedia di <https://www.scipy.org/>
- SciPy dibangun untuk untuk bekerja dengan NumPy array dan menyediakan kumpulan algoritma
- numerik , termasuk pemrosesan sinyal , optimasi , statistika , dan library Matplotlib untuk visualisasi data.
- Jika library belum terpasang , tuliskan perintah instalasi :
`pip install scipy`

c. Pandas

- Pandas (Panel Data) merupakan libraryvpopular di Python yang digunakan untuk data structure dan data analysis
- Bersifat open source dan tersedia di <https://pandas.pydata.org/>
- Pandas sangat berkaitan dengan NumPy
- Jika library belum terpasang, tuliskan perintah instalasi :
`pip install pandas`
- Kemudian impor
`import pandas as pd`
- Pandas dapat mengimpor data dari berbagai format: comma separated value (CSV), file tek, Microsoft Excel, database SQL, dan format HDF5

d. Matplotlib

- Matplotlib adalah library Python untuk visualisasi data dengan dua dimensi
- Bersifat open source dan tersedia di <https://matplotlib.org/>
- Matplotlib berkaitan dengan NumPy dan Pandas
- Jika library belum terpasang, tuliskan perintah instalasi :
`pip install matplotlib`
- Kemudian impor
`import matplotlib.pyplot as plt`

e. Seaborn

- Seaborn adalah library visualisasi data Python serupa dengan Matplotlib) yang menyediakan high level interface untuk menggambar grafik statistika yang menarik dan informatif
- Library ini bersifat open source dan tersedia di <https://seaborn.pydata.org/>
- Jika library belum terpasang, tuliskan perintah instalasi :
`pip install seaborn`
- Kemudian impor
`import seaborn as sns`

f. Scikit-learn

- Scikit-learn adalah library untuk mempraktikkan machine learning dan membuat model. Bersifat open source dan tersedia di <https://scikit-learn.org/>
- Scikit-learn diawali dari project SciPy (Scientific Python) yang berisi fungsi-fungsi matematis
- Jika library belum terpasang, tuliskan perintah instalasi:
`pip install sklearn`
- Kemudian impor:
`import sklearn`

3. Jupyter Notebook

Dengan munculnya aplikasi web, generasi baru IDE untuk bahasa interaktif seperti Python telah dikembangkan. Dimulai di komunitas akademisi dan e-learning, IDE berbasis web dikembangkan dengan mempertimbangkan agar kode dan seluruh environment Anda dapat disimpan di server. Salah satu aplikasi pertama dari jenis WIDE ini dikembangkan oleh William Stein pada awal 2005 menggunakan Python 2.3 sebagai bagian dari perangkat lunak matematika SageMath. Di SageMath, server dapat diatur di pusat, seperti universitas atau sekolah, dan kemudian siswa dapat mengerjakan pekerjaan rumah mereka baik di kelas maupun di rumah. Selain itu, siswa dapat menjalankan semua langkah sebelumnya berulang kali, dan kemudian mengubah beberapa sel kode tertentu (segmen dokumen yang mungkin berisi kode sumber yang dapat dieksekusi) dan menjalankan operasi lagi. Instruktur juga dapat memiliki akses ke sesi siswa dan meninjau kemajuan atau hasil mereka.

Saat ini, sesi semacam itu disebut notebook dan tidak hanya digunakan di ruang kelas tetapi juga digunakan untuk menunjukkan hasil dalam presentasi atau di dasbor bisnis. Penggunaan notebook adalah bagian dari pengembangan IPython. Sejak Desember 2011, IPython telah dirilis sebagai versi browser dari konsol interaktifnya, yang disebut notebook IPython, yang menunjukkan hasil eksekusi Python dengan sangat jelas dan ringkas melalui sel baris. Sel dapat berisi konten selain kode. Misalnya, sel markdown yang dapat ditambahkan pada sel untuk menjelaskan algoritma yang dibuat. Dimungkinkan juga untuk menyisipkan gambar plot Matplotlib untuk mengilustrasikan contoh atau bahkan halaman web. Belakangan ini, beberapa jurnal ilmiah mulai menerima notebook untuk menunjukkan hasil eksperimen, dilengkapi dengan kode dan sumber datanya. Dengan cara ini, eksperimen menjadi lengkap dan dapat direplikasi.

Sejak proyek data science berkembang pesat, notebook IPython telah dipisahkan dari perangkat lunak IPython dan sekarang telah menjadi bagian dari proyek yang lebih besar: Jupyter¹⁶. Jupyter (untuk Julia, Python dan R) bertujuan untuk menggunakan kembali WIDE yang sama untuk semua bahasa interpreter ini dan tidak hanya Python. Semua notebook IPython lama diimpor secara otomatis ke versi baru saat dibuka dengan platform Jupyter; tetapi setelah dikonversi ke versi baru, notebook tidak dapat digunakan lagi di versi notebook IPython lama.

4. Representasi Pengetahuan

a. Data Frame

Data frame merupakan tabel/data tabular dengan array dua dimensi yaitu baris dan kolom. Struktur data ini merupakan cara paling standar untuk menyimpan data. Setiap kolom pada data frame merupakan objek dari Series, dan baris terdiri dari elemen yang ada pada Series.

Series merupakan struktur data dasar dalam Pandas. Series diibaratkan sebagai array satu dimensi sama halnya dengan numpy array, hanya bedanya mempunyai index dan index tersebut dapat kita kontrol dari setiap elemen tersebut.

REPRESENTASI DATA FRAME

a. Membuat Data Frame

```
In [35]: import numpy as np
import pandas as pd

In [36]: my_data = np.array([[5.1,3.5,1.4,0.2],[4.9,3.0,1.4,0.2],[4.7,3.2,1.3,0.2],[4.6,3.1,1.5,0.2],[5.0,3.6,1.4,0.2]
])

In [37]: column_names = ['sepal.length', 'sepal.width', 'petal.length', 'petal.width']

In [38]: data_frame = pd.DataFrame(data=my_data, columns=column_names)
print(data_frame)
```

	sepal.length	sepal.width	petal.length	petal.width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

b. Menambahkan Kolom Baru

```
In [39]: data_frame ["species"] = 'Iris-setosa'

In [40]: print(data_frame)
```

	sepal.length	sepal.width	petal.length	petal.width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

c. Menentukan Subset

```
In [44]: print("rows 0, 1, and 2 : ")
print(data_frame.head(3), '\n')

print("row 3 : ")
print(data_frame.iloc[[3]], '\n')

print("row 1, 2, and 3 : ")
print(data_frame[1:4], '\n')

print("Column 'species' : ")
print(data_frame['species'])
```

rows 0, 1, and 2 :

	sepal.length	sepal.width	petal.length	petal.width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa

row 3 :

	sepal.length	sepal.width	petal.length	petal.width	species
3	4.6	3.1	1.5	0.2	Iris-setosa

row 1, 2, and 3 :

	sepal.length	sepal.width	petal.length	petal.width	species
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa

Column 'species' :

	species
0	Iris-setosa
1	Iris-setosa
2	Iris-setosa
3	Iris-setosa
4	Iris-setosa

Name: species, dtype: object

Latihan

1. Buat DataFrame panda 3x4 (3 baris x 4 kolom) yang kolomnya diberi nama **Zahra**, **Lily**, **Dito**, dan **Naufal**. Isi masing-masing dari 12 sel di DataFrame dengan bilangan bulat acak antara 0 dan 100.
2. Tampilkan output sebagai berikut:
 - Seluruh DataFrame
 - Nilai dalam sel baris #1 dari kolom Lily (Untuk kelas A), kolom Naufal (Untuk kelas B), dan kolom Zahra (Untuk kelas C).
3. Buat kolom kelima bernama Janet, yang diisi dengan jumlah baris dari kolom Zahra dan Dito.