

DATA SCIENCE

latihan dasar
by : Muhammad Lefy Hidayat

```
!pip install pandas
```

- '!' digunakan di Jupyter Notebook untuk menjalankan perintah shell.
- pip install pandas meminta pip untuk mengunduh dan menginstal pustaka pandas, yang berguna untuk analisis dan manipulasi data dalam bentuk tabel (DataFrame).

```
import pandas as pd
```

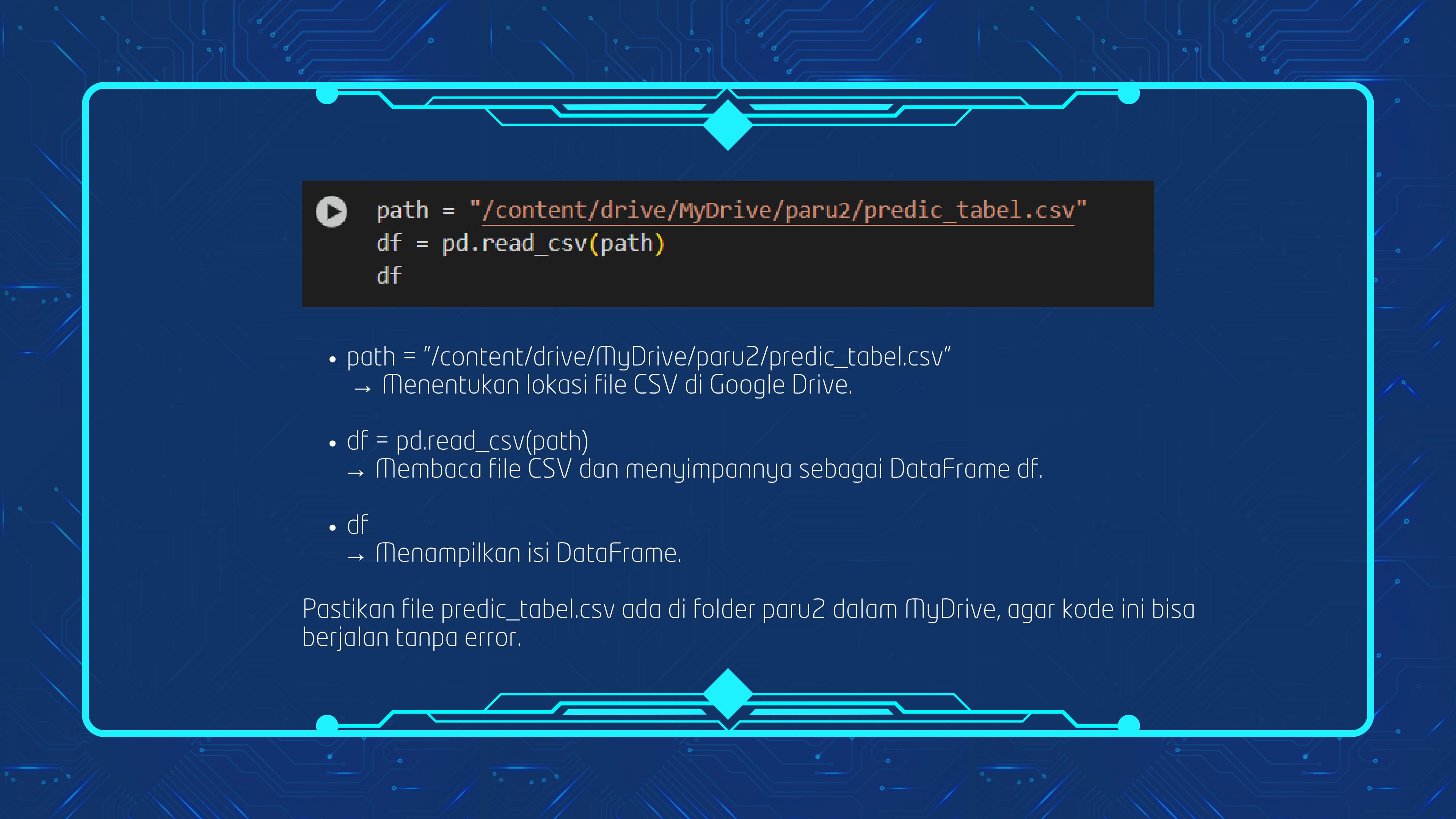
digunakan untuk mengimpor pustaka pandas dengan alias pd, yang berguna untuk mengolah dan menganalisis data dalam bentuk tabel (DataFrame) atau array (Series).



```
from google.colab import drive  
drive.mount('/content/drive')
```

- from google.colab import drive : Mengimpor modul untuk mengakses Drive di Google Colab.
- drive.mount('/content/drive') : Memasang (mount) Google Drive ke direktori /content/drive, sehingga file di Drive bisa digunakan dalam Colab.

Setelah menjalankan kode ini, Colab akan meminta izin akses ke akun Google.



```
 path = "/content/drive/MyDrive/paru2/predic_tabel.csv"  
df = pd.read_csv(path)  
df
```

- path = "/content/drive/MyDrive/paru2/predic_tabel.csv"
→ Menentukan lokasi file CSV di Google Drive.
- df = pd.read_csv(path)
→ Membaca file CSV dan menyimpannya sebagai DataFrame df.
- df
→ Menampilkan isi DataFrame.

Pastikan file `predic_tabel.csv` ada di folder `paru2` dalam `MyDrive`, agar kode ini bisa berjalan tanpa error.

No	Usia	Jenis_Kelamin	Merokok	Bekerja	Rumah_Tangga	Aktivitas_Begadang	Aktivitas_Olahraga	Asuransi	Penyakit_Bawaan	Hasil
0	1	Tua	Pria	Pasif	Tidak	Ya	Ya	Sering	Ada	Tidak Ya
1	2	Tua	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Ada Tidak
2	3	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak Tidak
3	4	Tua	Pria	Aktif	Ya	Tidak	Tidak	Jarang	Ada	Ada Tidak
4	5	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada Ya
...
29995	29996	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak Tidak
29996	29997	Tua	Wanita	Aktif	Ya	Tidak	Ya	Jarang	Ada	Ada Tidak
29997	29998	Tua	Wanita	Aktif	Ya	Tidak	Ya	Jarang	Ada	Ada Tidak
29998	29999	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada Tidak
29999	30000	Tua	Wanita	Pasif	Tidak	Ya	Tidak	Sering	Tidak	Tidak Ya

30000 rows × 11 columns

Jika sudah di run pada kode slide sebelumnya maka hasil tampilan akan seperti ini.

d df.head()

	No	Usia	Jenis_Kelamin	Merokok	Bekerja	Rumah_Tangga	Aktivitas_Begadang	Aktivitas_Olahraga	Asuransi	Penyakit_Bawaan	Hasil
0	1	Tua	Pria	Pasif	Tidak	Ya	Ya	Sering	Ada	Tidak	Ya
1	2	Tua	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Ada	Tidak
2	3	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
3	4	Tua	Pria	Aktif	Ya	Tidak	Tidak	Jarang	Ada	Ada	Tidak
4	5	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Ya

df.head() : Di gunakan untuk 5 baris pertama pada data

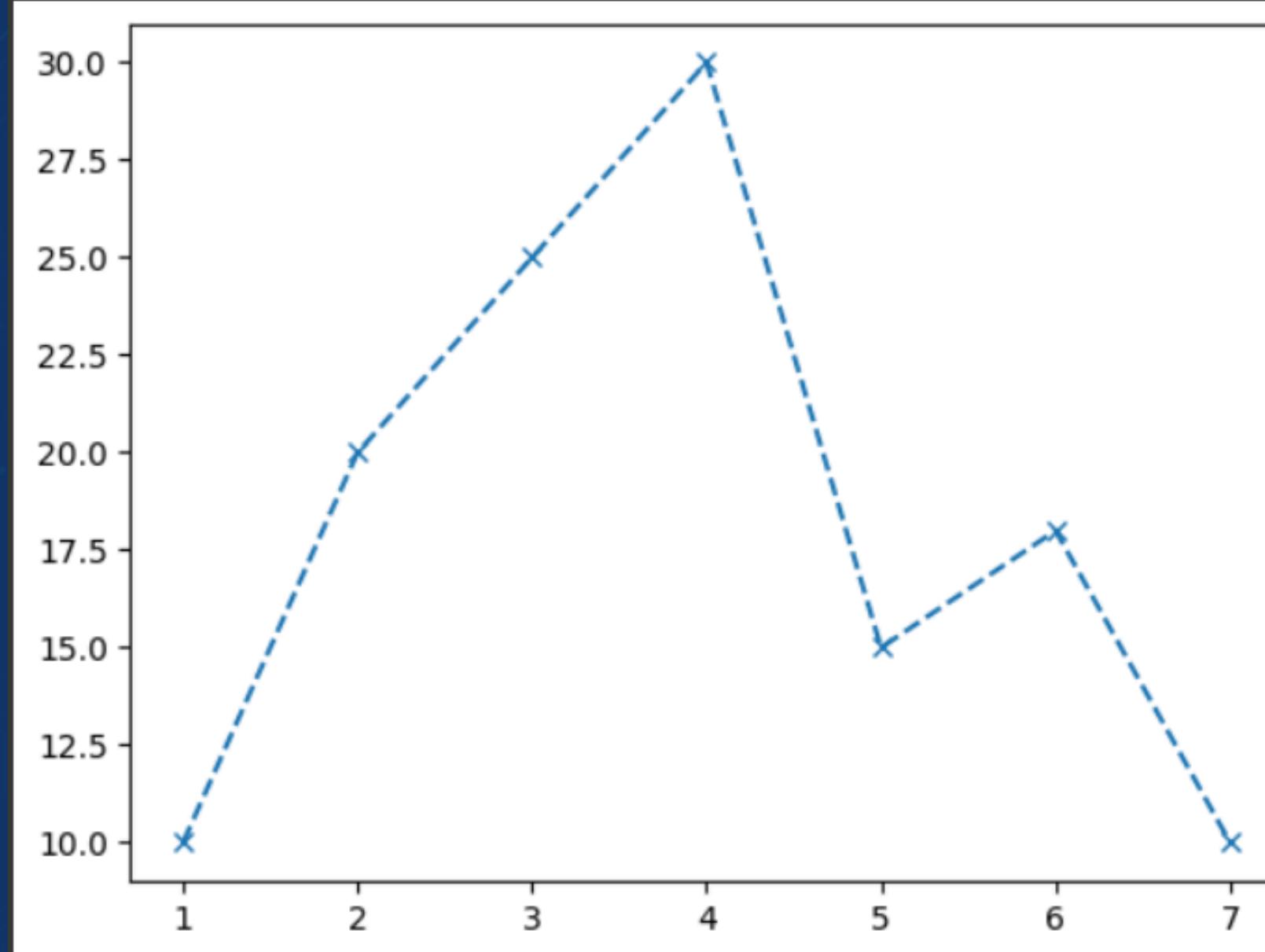
```
import matplotlib.pyplot as plt

# membuat component figure dan axis
fig, ax = plt.subplots()

data_x = [1, 2, 3, 4, 5, 6, 7]
data_y = [10, 20, 25, 30, 15, 18, 10]

# memberikan data kedalam axis
ax.plot(data_x, data_y, marker='x', linestyle='dashed')
plt.show()
```

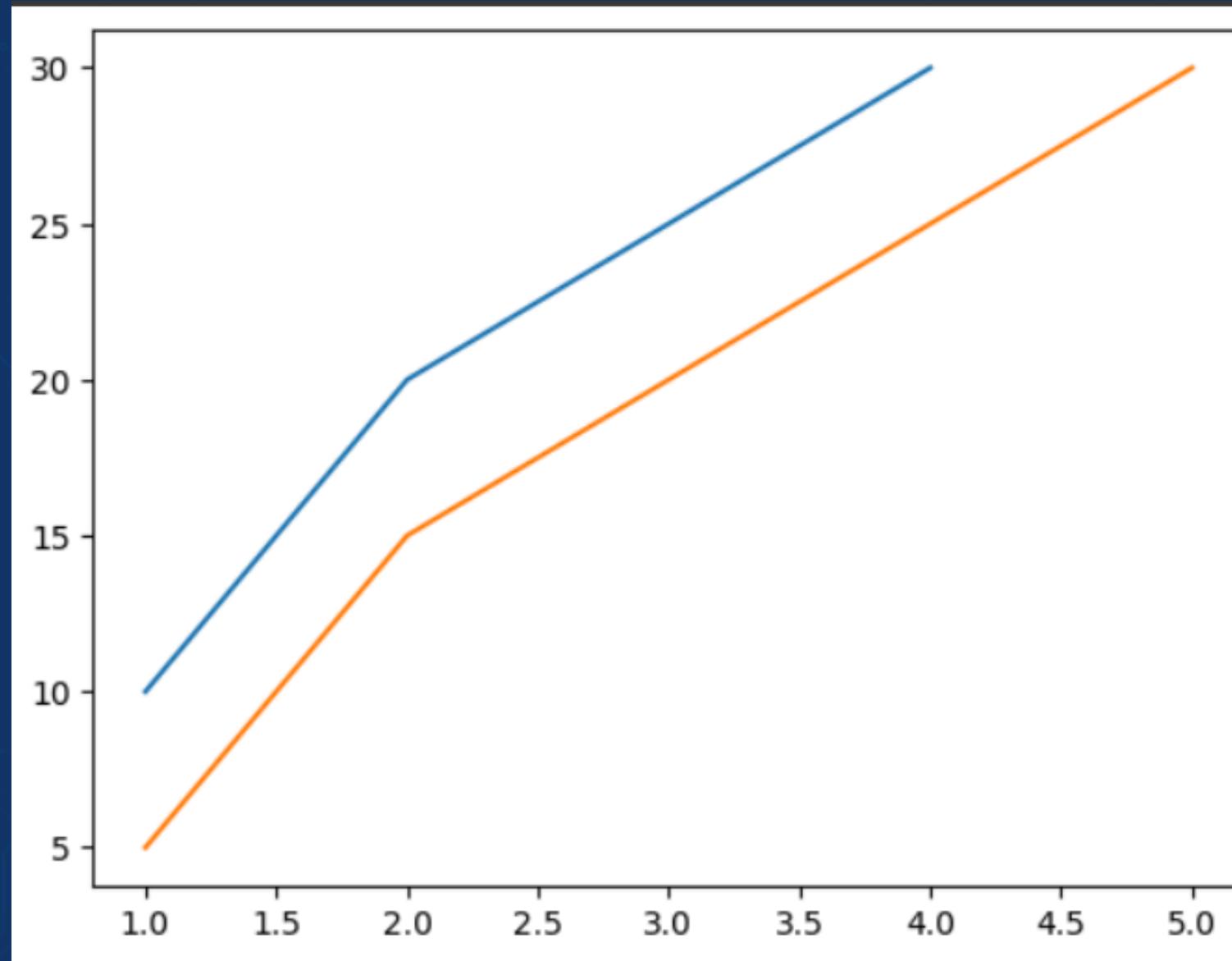
- `import matplotlib.pyplot as plt` → Mengimpor pustaka Matplotlib untuk visualisasi data.
- `fig, ax = plt.subplots()` → Membuat figure (area gambar) dan axis (sumbu grafik).
- Mendefinisikan data:
`data_x = [1, 2, 3, 4, 5, 6, 7]`
`data_y = [10, 20, 25, 30, 15, 18, 10]`
- Plot data dengan gaya tertentu:
`ax.plot(data_x, data_y, marker='x', linestyle='dashed')`
- `marker='x'` → Menandai titik data dengan simbol "x".
- `linestyle='dashed'` → Menggunakan garis putus-putus.
- `plt.show()` → Menampilkan plot.



Jika sudah di run pada kode slide sebelumnya maka hasil tampilan akan seperti ini.

```
fig, ax = plt.subplots()  
data_x1 = [1, 2, 3, 4]  
data_y1 = [10, 20, 25, 30]  
data_x2 = [1, 2, 3, 4, 5]  
data_y2 = [5, 15, 20, 25, 30]  
ax.plot(data_x1, data_y1)  
ax.plot(data_x2, data_y2)  
plt.show()
```

- `fig, ax = plt.subplots()` → Membuat figure (area gambar) dan axis (sumbu grafik).
- Mendefinisikan dua set data:
`data_x1 = [1, 2, 3, 4]`
`data_y1 = [10, 20, 25, 30]`
`data_x2 = [1, 2, 3, 4, 5]`
`data_y2 = [5, 15, 20, 25, 30]`
- Menambahkan dua garis ke dalam plot:
`ax.plot(data_x1, data_y1)`
`ax.plot(data_x2, data_y2)`
- Garis pertama menggunakan `data_x1` dan `data_y1`.
- Garis kedua menggunakan `data_x2` dan `data_y2`.
- `plt.show()` → Menampilkan plot dengan dua garis.



Jika sudah di run pada kode slide sebelumnya maka hasil tampilan akan seperti ini.

**SEKIAN
TERIMAKASIH**