

Library Pandas

Ledy Elsera Astrianty



Pengenalan

- library Python untuk pengolahan dan analisis data.
- Fleksibel
- Banyak fitur untuk pengolahan data

Struktur Data Pandas

- **DataFrames**, terdiri dari baris dan kolom
- **Series**, terdiri dari nilai-nilai yang berurutan
- **Index**, label yang digunakan untuk mengakses data

```
In [2]: ▶ # Membuat DataFrame dengan data
data = {'Nama': ['Andi', 'Budi', 'Cici'],
        'Umur': [20, 25, 30]}
df = pd.DataFrame(data)
```

```
# Menampilkan DataFrame
print(df)
```

	Nama	Umur
0	Andi	20
1	Budi	25
2	Cici	30

Tipe data Pandas

- **Int64**, bilangan bulat 64-bit
- **Float64**, bilangan pecahan 64-bit
- **Object**: tipe data string
- **Datetime64**: tipe data tanggal dan waktu

```
In [3]: ▶ # Membuat DataFrame dengan data
data = {'Nama': ['Andi', 'Budi', 'Cici'],
        'Umur': [20, 25, 30]}
df = pd.DataFrame(data)

# Menampilkan tipe data kolom 'Umur'
print(df['Umur'].dtype)

int64
```

Membaca Data dengan Pandas

- **Fungsi**, read_csv(), read_excel(), read_json()
- **Parameter**: file_path, delimiter, header

```
In [4]: ▶ # Membaca data dari file CSV
df = pd.read_csv('Heart_Disease_Prediction.csv')

# Menampilkan DataFrame
print(df)
```

	Age	Sex	Chest pain type	BP	Cholesterol	FBS over 120	EKG results
0	70	1	4	130	322	0	2
1	67	0	3	115	564	0	2
2	57	1	2	124	261	0	0
3	64	1	4	128	263	0	0
4	74	0	2	120	269	0	2

Mengolah Data dengan Pandas

- **Fungsi**, drop(), rename(), groupby()
- **Parameter**, columns, index, groupby

```
In [6]: ▶ # Membuat DataFrame dengan data
data = {'Nama': ['Andi', 'Budi', 'Cici'],
        'Umur': [20, 25, 30]}
df = pd.DataFrame(data)

# Menghapus kolom 'Umur'
df = df.drop('Umur', axis=1)

# Menampilkan DataFrame
print(df)
```

```
      Nama
0    Andi
1    Budi
2    Cici
```

```
In [12]: df.sort_values(by=['EKG results']).head(3)
```

Out[12]:

	Age	Sex	Chest pain type	BP	Cholesterol	FBS over 120	EKG results	Max HR	Exercise angina	ST depression	Slope of ST	Number of vessels fluro	Thallium	Heart Disease
178	67	1	4	125	254	1	0	163	0	0.2	2	2	7	Presence
100	44	0	3	108	141	0	0	175	0	0.6	2	0	3	Absence
99	50	0	2	120	244	0	0	162	0	1.1	1	0	3	Absence

```
In [13]: df[['Age', 'EKG results']].groupby(df['Sex']).max()
```

Out[13]:

	Age	EKG results
Sex		
0	76	2
1	77	2

Menganalisis Data dengan Pandas

- **Fungsi**, mean(), median(), std()
- **Parameter**, axis, skipna

```
In [7]: ▶ # Membuat DataFrame dengan data
data = {'Nama': ['Andi', 'Budi', 'Cici'],
        'Umur': [20, 25, 30]}
df = pd.DataFrame(data)

# Menghitung rata-rata umur
rata_rata_umur = df['Umur'].mean()

# Menampilkan rata-rata umur
print(rata_rata_umur)
```

25.0

Menggabungkan Data dengan Pandas

- **Fungsi:** concat(), merge(), join()
- **Parameter:** dfs, axis, how

```
In [8]: data1 = {'Nama': ['Andi', 'Budi'],  
                'Umur': [20, 25]}  
df1 = pd.DataFrame(data1)  
  
data2 = {'Nama': ['Cici', 'Dedi'],  
        'Umur': [30, 35]}  
df2 = pd.DataFrame(data2)  
  
# Menggabungkan DataFrame  
df = pd.concat([df1, df2])  
  
# Menampilkan DataFrame  
print(df)
```

	Nama	Umur
0	Andi	20
1	Budi	25
0	Cici	30
1	Dedi	35

Mengubah Data dengan Pandas

- Fungsi: rename(), replace()
- Parameter: columns, index, values

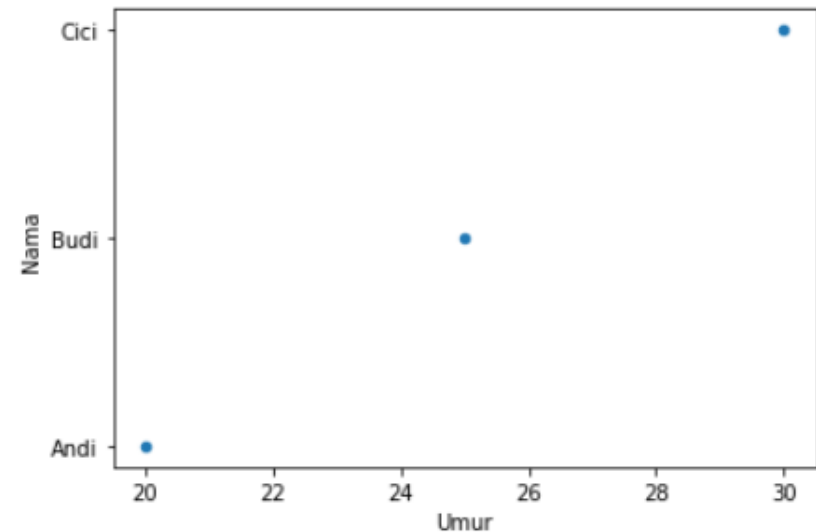
```
In [10]: ▶ data = {'Nama': ['Andi', 'Budi', 'Cici'],  
                  'Umur': [20, 25, 30]}  
df = pd.DataFrame(data)  
df['Umur'] = df['Umur'] + 5  
print(df)
```

	Nama	Umur
0	Andi	25
1	Budi	30
2	Cici	35

Membuat Grafik dengan Pandas

- Fungsi: plot()
- Parameter: kind, x, y

```
In [19]: ► data = {'Nama': ['Andi', 'Budi', 'Cici'],  
                  'Umur': [20, 25, 30]}  
df = pd.DataFrame(data)  
df.plot(  
    x='Umur',  
    y='Nama',  
    kind='scatter'  
);
```



Thank You



Library Pandas