



Minggu ke-2

## Praktikum Data Preprocessing (Data Manipulation & Visualisation)

Ali Ridho Barakbah, Entin Martiana

Knowledge Engineering Research Group  
Department of Information and Computer Engineering  
Politeknik Elektronika Negeri Surabaya



ruspini.csv

## Memanggil & Menampilkan Dataset

ruspini - Notepad

File Edit Format View Help

#,X,Y,CLASS

1,4,53,1

2,5,63,1

3,10,59,1

4,9,77,1

5,13,49,1

6,13,69,1

7,12,88,1

8,15,75,1

9,18,61,1

10,19,65,1

11,22,74,1

12,27,72,1

13,28,76,1

14,24,58,1

```
import pandas as pd

dataset = pd.read_csv('ruspini.csv')

print(dataset)
```

	#	X	Y	CLASS
0	1	4	53	1
1	2	5	63	1
2	3	10	59	1
3	4	9	77	1
4	5	13	49	1
5	6	13	69	1
6	7	12	88	1
7	8	15	75	1
8	9	18	61	1
9	10	19	65	1
10	11	22	74	1
11	12	27	72	1
12	13	28	76	1
13	14	24	58	1
14	15	27	55	1
15	16	28	60	1
16	17	30	52	1
17	18	31	60	1
18	19	32	61	1
19	20	36	72	1
20	44	86	132	2
21	45	85	115	2

# Mengetahui Jumlah baris & Kolom

---

```
import pandas as pd

dataset = pd.read_csv('ruspini.csv')
rows, cols= dataset.shape

print('Jumlah baris', rows)
print('Jumlah kolom', cols)
```

```
Jumlah baris 75
Jumlah kolom 4
```

# Mengambil Data pada Kolom tertentu

```
import pandas as pd

dataset = pd.read_csv('ruspini.csv', )
data = pd.DataFrame(dataset, columns=['X', 'Y'])

print(data)
```

```
import pandas as pd

dataset = pd.read_csv('ruspini.csv', )
data = dataset.loc[:, ['X', 'Y']]

print(data)
```

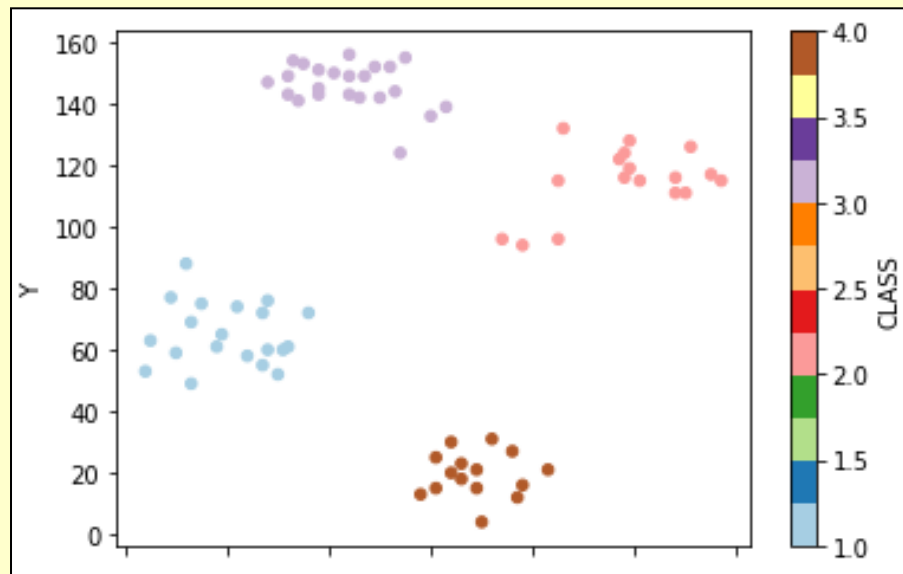
	X	Y
0	4	53
1	5	63
2	10	59
3	9	77
4	13	49
5	13	69
6	12	88
7	15	75
8	18	61
9	19	65
10	22	74
11	27	72
12	28	76
13	24	58
14	27	55
15	28	60
16	30	52
17	31	60
18	32	61

# Visualisasi data

```
import pandas as pd

dataset = pd.read_csv('ruspini.csv', )

dataset.plot(x='X', y='Y', kind='scatter', c='CLASS', colormap='Paired')
```



# Latihan Pertemuan 2

## Titanic Dataset

---

- Disediakan oleh Kaggle
- Memberikan informasi dari penumpang Titanic yang selamat dan tidak.
- Terdiri dari :
  - PassengerId : urutan nomer data dari penumpang
  - Survived : status selamat (0:meninggal, 1:selamat)
  - Pclass : kelas kamar dari penumpang (1: highclass, 2:midclass, 3:lowclass)
  - Name : nama penumpang
  - Sex : jenis kelamin penumpang (male, female)
  - Age : umur penumpang
  - SibSp : jumlah saudara kandung dan pasangan dari penumpang yang ada di kapal
  - Parch : jumlah orangtua dan anak dari penumpang
  - Ticket : kode tiket penumpang
  - Fare : ongkos tiket yang dibeli penumpang
  - Cabin : Kode kabin
  - Embarked : Kota keberangkatan penumpang (C:Cherbourg, Q:Queenstown, S:Southampton)



# Assignment #

---

1. dataset  $\leftarrow$  titanic.csv, dan tampilkan
2. rows, cols  $\leftarrow$  jumlah baris dan kolom pada dataset, dan tampilkan
3. data  $\leftarrow$  ambil dataset kolom fitur (Name, Sex, Age, Pclass, Fare), dan tampilkan
4. class  $\leftarrow$  ambil dataset kolom kelas (Survived), dan tampilkan
5. data  $\leftarrow$  data + fitur Relatives (jumlah akumulasi dari kolom fitur SibSp+Parch dari dataset), dan tampilkan
6. Hitunglah berapa penumpang pada masing-masing nilai Pclass
7. Hitunglah berapa penumpang pada masing-masing nilai Sex
8. Hitunglah berapa penumpang yang tidak selamat (Survived=0) dan selamat (Survived=1) untuk masing-masing Pclass
9. Visualisasi dengan warna yang berbeda untuk class dimana sumbu x=urutan data dan sumbu y=Sex
10. Visualisasi dengan warna yang berbeda untuk class dimana sumbu x=urutan data dan sumbu y=Age (abaikan data yang terdapat missing value)

# Pengumpulan Tugas

---

- Buatlah coding dengan Bahasa pemrograman/tools apapun untuk semua assignment
- Buatlah laporan dalam slide ppt. Laporan terdiri dari screenshot coding dan hasil running untuk setiap assignment.
- Simpan laporan dalam file pdf dengan format penamaan:  
DM\_M2\_NRP\_namadepan.pdf
- Upload file tersebut ke ETHOL
- Deadline upload: Selasa, 31 Agustus 2024

