



Minggu ke-3

Praktikum Data Preprocessing (Data Normalization)

Ali Ridho Barakbah, Entin Martiana

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



Pengisian Nilai Missing Values

```
import pandas as pd

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

print('Dataset\n', dataset)

dataset = dataset.fillna(dataset.groupby('CLASS').transform('mean'))

print("\n\nDataset setelah pengisian missing value\n", dataset)
```

Menggunakan nilai rata-rata pada class

Dataset				
	#	X	Y	CLASS
0	1	4.0	53.0	1
1	2	5.0	63.0	1
2	3	10.0	59.0	1
3	4	9.0	NaN	1
4	5	13.0	49.0	1
5	6	13.0	69.0	1
6	7	12.0	88.0	1
7	8	15.0	75.0	1
8	9	18.0	61.0	1
9	10	19.0	65.0	1
10	11	22.0	74.0	1
11	12	NaN	72.0	1
12	13	28.0	76.0	1
13	14	24.0	58.0	1
14	15	27.0	NaN	1
15	16	28.0	60.0	1
16	17	30.0	52.0	1
17	18	NaN	60.0	1
18	19	32.0	61.0	1
19	20	36.0	72.0	1
20	44	86.0	132.0	2
21	45	85.0	115.0	2
22	46	85.0	96.0	2
23	47	78.0	NaN	2



Dataset setelah pengisian missing value				
	#	X	Y	CLASS
0	1	4.000000	53.000000	1
1	2	5.000000	63.000000	1
2	3	10.000000	59.000000	1
3	4	9.000000	64.833333	1
4	5	13.000000	49.000000	1
5	6	13.000000	69.000000	1
6	7	12.000000	88.000000	1
7	8	15.000000	75.000000	1
8	9	18.000000	61.000000	1
9	10	19.000000	65.000000	1
10	11	22.000000	74.000000	1
11	12	19.166667	72.000000	1
12	13	28.000000	76.000000	1
13	14	24.000000	58.000000	1
14	15	27.000000	64.833333	1
15	16	28.000000	60.000000	1
16	17	30.000000	52.000000	1
17	18	19.166667	60.000000	1
18	19	32.000000	61.000000	1
19	20	36.000000	72.000000	1
20	44	86.000000	132.000000	2
21	45	85.000000	115.000000	2
22	46	85.000000	96.000000	2
23	47	78.000000	116.200000	2



Mencari Nilai Statistik pada Atribut

```
import pandas as pd

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

dataX = dataset['X']

minval = dataX.min()
maxval = dataX.max()
avgval = dataX.mean()
stdval = dataX.std()

print('Nilai minimum = ', minval)
print('Nilai maksimum = ', maxval)
print('Nilai rata-rata = ', avgval)
print('Nilai deviasi standar = ', stdval)
```

```
Nilai minimum = 4
Nilai maksimum = 117
Nilai rata-rata = 54.88
Nilai deviasi standar = 30.502529802039774
```



Latihan Pertemuan 3

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 (Age, Fare), dan tampilkan
4. class \leftarrow ambil dataset kolom kelas (Survived)
5. Lakukan pengisian missing value pada fitur Age dengan nilai mean dari masing-masing class

Lakukan normalisasi pada data dengan algoritma berikut, dan tampilkan:

6. Min-Max (0-1)
7. Z-Score
8. Sigmoidal



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_M3_NRP_namadepan.pdf
- Upload file tersebut ke ETHOL
- Deadline upload: Kamis, 5 September 2024 pk 18.00

