



Minggu ke-9

Praktikum Klasifikasi dengan Studi Kasus Water Potability

Entin Martiana, Ali Ridho Barakbah, Renovita

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')

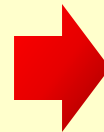
print('Dataset\n', dataset)

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



Assignment Minggu 9

Water Potability Dataset

- Disediakan oleh Kaggle
- Memberikan informasi kualitas air dengan target 1(bisa diminum), 0(tidak bisa diminum).
- Terdiri dari :
 1. ph: pH air (0 hingga 14).
 2. Hardness: Kapasitas air untuk mengendapkan sabun dalam mg/L.
 3. Solids: Total padatan terlarut dalam ppm.
 4. Kloramin: Jumlah Kloramin dalam ppm.
 5. Sulfat: Jumlah Sulfat yang dilarutkan dalam mg/L.
 6. Conductivity: Konduktivitas listrik air dalam S/cm.
 7. Organic_carbon: Jumlah karbon organik dalam ppm.
 8. Trihalomethanes: Jumlah Trihalomethanes dalam g/L.
 9. Turbidity: Ukuran sifat pemancar cahaya air di NTU.
 10. Potability: Menunjukkan apakah air aman untuk dikonsumsi manusia. Dapat diminum -1 dan Tidak dapat diminum -0



Assignment# - Klasifikasi Water Potability

1. dataset ← water_potability.csv
2. Buatlah skenario tertentu pada pemilihan fitur sehingga error ratio dari klasifikasi test_data dapat sekecil mungkin
3. Lakukan validation Model dengan metode Hold-out
4. Lakukan pengisian missing value jika atribut yang anda pilih ada yang nilainya kosong
5. Lakukan normalisasi terhadap train_data dengan metode min-max(0-1)
6. Lakukan normalisasi terhadap test_data dengan metode min-max(0-1) dengan nilai min & max dari train_data
7. Algoritma klasifikasi yang dipakai adalah k-NN, Naïve Bayes, Decision Tree, tunjukkan akurasi



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:
MLSDT_M7_NRP_namadepan.pdf
- Upload file tersebut ke ETHOL
- Deadline upload: Kamis, 19 Oktober 2023

