

Nama : Sukmawati Dwi Lestari

NIM : 1301184345

Kelas : IF-42-02

LAPORAN OBSERVASI TUGAS PEMROGRAMAN 3

K-Nearest Neighbor(KNN)

1. Perhitungan Ukuran Jarak

Pada perhitungan jarak saya kNN saya menggunakan rumus Euclidean Distance, berikut adalah rumus dari Euclidean Distance :

$$= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$
$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

2. Prapemrosesan Data

Pertama kita membaca data sesuai dengan file csv yaitu "Diabetes.csv" untuk kebutuhan prapemrosesan data dengan cara mengimport file tersebut

```
[255] import pandas as pd
import numpy as np
from math import sqrt

[256] datadiabet = pd.read_csv('Diabetes.csv')
print(datadiabet)
```

	Pregnancies	Glucose	...	Age	Outcome
0	6	148	...	50	1
1	1	85	...	31	0
2	8	183	...	32	1
3	1	89	...	21	0
4	0	137	...	33	1
...
763	10	101	...	63	0
764	2	122	...	27	0
765	5	121	...	30	0
766	1	126	...	47	1
767	1	93	...	23	0

[768 rows x 9 columns]

Kemudian pada prapemrosesan data ini saya memilih suatu metode yang digunakan untuk normalisasi data. Metode yang saya pilih adalah Z-score atau disebut juga standard score, cara ini akan menghasilkan suatu nilai baru yang berkisar di angka 0 dengan rentang antara -3 dan 3 tetapi bisa juga lebih tinggi atau lebih rendah. Berikut adalah rumus dari Z-Score

$$x_{new} = \frac{x_{old} - \mu}{\sigma}$$

3. Klasifikasi kNN

Sebelumnya sudah menghitung jarak dengan menggunakan rumus Euclidean Distance, pada tahap tersebut terjadi perhitungan dan menentukan berapa jarak *neighbour* yang dimiliki. Selain itu saya membandingkan *neighbour* tersebut yang terdapat pada fungsi "klasifikasi" dengan membandingkan data *outcome* dengan keluaran(k-Fold kedua) untuk mendapatkan nilai kebenarannya.

Nama : Sukmawati Dwi Lestari

NIM : 1301184345

Kelas : IF-42-02

4. Pemilihan k terbaik

```
▶ k = 97
x = klasifikasi(trainset,testing,k)
print('Nilai k terbaik : ',k)
print('Hasil setiap K-Fold : ')
for i in range(len(x)):
    s = i+1
    print(s,x[i])
print('Rata-rata Persentase : ',sum(x)/len(x))
```

5. Perhitungan rata-rata akurasi kNN menggunakan 5-fold validation

Pada tahap ini saya melakukan perhitungan dengan k = 97 dan mendapatkan nilai

k-Fold 1	64.28571428571429
k-Fold 2	75.32467532467533
k-Fold 3	65.80645161290323
k-Fold 4	59.74025974025974
k-Fold 5	64.70588235294117

Dengan memperoleh rata-rata presentase : 65.97259666329875