

**TUGAS KONSEP APLIKASI DATA MAINING ALGORITMA KNN  
MENGGUNAKAN JUPYTER NOTEBOOK**



**Nama : Himatus Yulvi A.S**

**Jurusan :Sistem Infomasi**

**Nim :17.51.0005**

**KEMENTRIAN RISET DAN TEKNOLOGI PENDIDIKAN TINGGI  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA KOMPUTER  
PRADNYA PARAMITA  
MALANG  
2020**

1. Yang pertama adalah persiapan instalasi (mengimplementasikan KKN dengan menggunakan jupyter notebook) antara lain pandas nanti di sebut sebagai pd, numpy disebut sebagai np, matplotlib.pyplot nanti di sebut sebagai plt, Kemudian %matplotlib inline.

```
In [30]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
%matplotlib inline
```

2. Pertama-tama adalah cek versi pandas. Dimana nama yang di gunakan adalah pd.\_\_version\_\_ script yang sedang kita jalankan, jika sudah berhasil maka akan muncul seperti data di bawah ini.

```
In [31]: pd.__version__
```

Output:

```
Out[31]: '1.0.1'
```

3. Selanjutnya ingin memasukan insert data array.

```
data=np.array ([[7.,5, 'BAD'], [7.6,4, 'BAD'], [7.5,2, 'BAD'], [3.5,5, 'GOOD'], [1.6,8, 'GOOD'], [3.,6, 'GOOD'], [3.,5, 'NN']])
query= [3.0,5, 'NN']
```

```
df=pd.DataFrame(data)
df.columns = ['daya', 'kekuatan', 'kat']
df
```

Output:

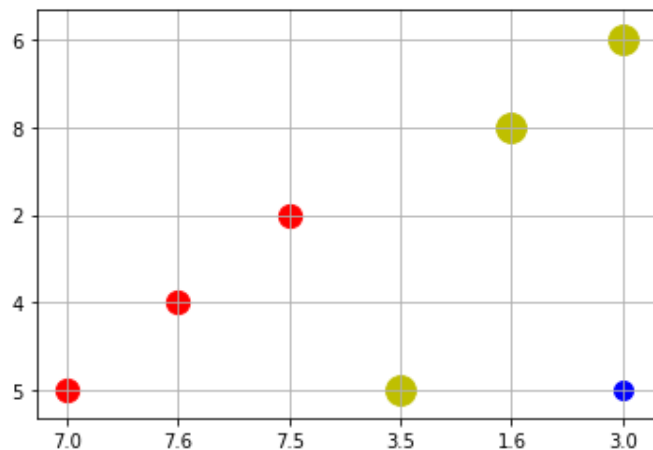
	daya	kekuatan	kat
0	7.0	5	BAD
1	7.6	4	BAD
2	7.5	2	BAD
3	3.5	5	GOOD
4	1.6	8	GOOD
5	3.0	6	GOOD
6	3.0	5	NN

4. lalu jika kita ingin setting / membuat graph untuk menampilkan grafiknya Jalankan skrip berikut:

```
for i in range (7):
    if(df.iloc[i]['kat'] == 'GOOD'):
        plt.scatter(df.loc[i] ['daya'],df.iloc[i] ['kekuatan'], s=250, c='y')
    elif(df.iloc[i]['kat'] == 'BAD'):
        plt.scatter (df.iloc[i] ['daya'],df.iloc[i] ['kekuatan'], s=150, c='r')
    else:
        plt.scatter (df.iloc[i] ['daya'],df.iloc[i] ['kekuatan'], s=100, c='b')

plt.grid()
plt.show()
```

Output:



5. Berikut ini penghitungan distance jalankan skrip berikut .

```
import math
dis= []
for i in range (7):
    dis.append(math.sqrt ((float (df.iloc[i]['daya']) - query [0]) **2 + (float(df.iloc[i]['kekuatan']) - query [1]) **2))
```

```
df['dis'] = dis
df
```

Output:

	daya	kekuatan	kat	dis
0	7.0	5	BAD	4.000000
1	7.6	4	BAD	4.707441
2	7.5	2	BAD	5.408327
3	3.5	5	GOOD	0.500000
4	1.6	8	GOOD	3.310589
5	3.0	6	GOOD	1.000000
6	3.0	5	NN	0.000000

6. Selajutnya kita akan menyortir terlebih dahulu jarak terdekatnya.

```
df.sort_values('dis')
```

Output:

	daya	kekuatan	kat	dis
6	3.0	5	NN	0.000000
3	3.5	5	GOOD	0.500000
5	3.0	6	GOOD	1.000000
4	1.6	8	GOOD	3.310589
0	7.0	5	BAD	4.000000
1	7.6	4	BAD	4.707441
2	7.5	2	BAD	5.408327

7. .Habis ada hasil terakhir meneksport ke bentuk excel dengan menggunakan script berikut ini :

```
df.to_excel ('C:/Users/yulvavi/outputs_KKN.xls')
```

## Kesimpulan

**Algoritma kNN** (*k-Nearest Neighbor*) ini adalah algoritma klasifikasi berdasarkan tetangga terdekat. Dari table di atas akan di ambil 3 tetangga terdekat antara lain 0.500000, 1.000000, dan 3.310589 , jadi di dalam 3 tetangga terdekat adalah semua GOOD maka kategori misterius tersebut adalah GOOD.