

**TUGAS AKHIR SEMESTER**  
**KONSEP DAN APLIKASI DATA MINNING**  
**KNN**  
**OLEH :**  
**SANTOSO [17.52.0006]**



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

- A. Berdasarkan data tersebut bagaimana perlakuan dengan kondisi Ibu hamil dengan Usia 30 Tahun, yang merupakan Kelahiran ke -1, dengan Waktu kelahiran sesuai dengan HPL, Memiliki tekanan darah Normal? Carilah KNN dengan menggunakan Key = 5

1. Melakukan import library yang dibutuhkan untuk mengerjakan soal.

```
import numpy as np
import pandas as pd
```

2. Selanjutnya membaca file yang akan digunakan untuk data set. Kebetulan file telah saya berada di direktori D di dalam folder dataset “D:\dataset\dataset2.xlsx”

```
df = pd.read_csv('D:\dataset\dataset4.txt',
                 delimiter=',')
```

3. Kemudian kita lakukan pengecekan pada ini dataset yang kita panggil

df

	Usia	Kelahiran_ke-	Waktu_Kelahiran	Tekanan_darah	Kelainan_jantung	Caesarian
0	22	1	0	2	0	0
1	26	2	0	1	0	1
2	26	2	1	1	0	0
3	28	1	0	2	0	0
4	22	2	0	1	0	1
...	...	...	...	...	...	...
75	27	2	1	1	0	0
76	33	4	0	1	0	1
77	29	2	1	2	0	1
78	25	1	2	0	0	1
79	24	2	2	1	0	0

80 rows × 6 columns

4. Selanjutnya lakukan perhitungan Algoritma KNN dengan kondisi pada Nomer 4A yaitu

Usia 30 Tahun, yang merupakan Kelahiran ke -1, dengan Waktu kelahiran sesuai dengan HPL, Memiliki tekanan darah Normal? Carilah KNN dengan menggunakan Key = 5, dengan syntax berikut ini :

```
import math
dis = []
for i in range(80):
    dis.append(math.sqrt((float(df.iloc[i]['Usia'])-30)**2+
                        (float(df.iloc[i]['Kelahiran_ke-'])- 1)**2+
                        (float(df.iloc[i]['Waktu_Kelahiran'])-0)**2+
                        (float(df.iloc[i]['Tekanan_darah'])-1)**2))
```

5. Lalu menampung hasilnya di variable dis

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

	Usia	Kelahiran_ke-	Waktu_Kelahiran	Tekanan_darah	Kelainan_jantung	Caesarian	dis
0	22	1	0	2	0	0	8.062258
1	26	2	0	1	0	1	4.123106
2	26	2	1	1	0	0	4.242641
3	28	1	0	2	0	0	2.236068
4	22	2	0	1	0	1	8.062258
...	...	...	...	...	...	...	...
75	27	2	1	1	0	0	3.316625
76	33	4	0	1	0	1	4.242641
77	29	2	1	2	0	1	2.000000
78	25	1	2	0	0	1	5.477226
79	24	2	2	1	0	0	6.403124

80 rows × 7 columns

6. Lalu kita lakukan penyortiran.

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

	Usia	Kelahiran_ke-	Waktu_Kelahiran	Tekanan_darah	Kelainan_jantung	Caesarian	dis
27	30	1	0	1	0	0	0.000000
38	31	1	0	1	0	0	1.000000
67	29	2	0	1	1	0	1.414214
54	29	2	0	1	1	1	1.414214
59	30	2	1	2	1	1	1.732051
...	...	...	...	...	...	...	...
41	19	1	0	1	0	1	11.000000
61	19	1	0	1	0	1	11.000000
25	18	1	0	1	0	0	12.000000
26	18	1	1	2	1	1	12.083046
70	17	1	0	0	0	1	13.038405

80 rows x 7 columns

7. Berikutnya kita kan mengambil 5 teratas dari hasil tersebut dan kita tam[ung di variable y

```
y = df.sort_values('dis').head(5)
y
```

	Usia	Kelahiran_ke-	Waktu_Kelahiran	Tekanan_darah	Kelainan_jantung	Caesarian	dis
27	30	1	0	1	0	0	0.000000
38	31	1	0	1	0	0	1.000000
67	29	2	0	1	1	0	1.414214
54	29	2	0	1	1	1	1.414214
59	30	2	1	2	1	1	1.732051

8. Lalu kita akan mengambil data rerata dari Qty untuk menentukan banyak roti yang harus diproduksi.

```
In [23]: z = y["Caesarian"]
z
Out[23]: 27    0
        38    0
        67    0
        54    1
        59    1
        Name: Caesarian, dtype: int64
```

---

```
In [24]: np.mean(z)
Out[24]: 0.4
```

9. Algoritma KNN untuk soal dengan kondisi Ibu hamil dengan Usia 30 Tahun, yang merupakan Kelahiran ke -1, dengan Waktu kelahiran sesuai dengan HPL, Memiliki tekanan darah Normal maka ditemukan bahwa ibu hamil tersebut akan dilakukan 0,4 atau mendekati 0 maka akan bersalin secara normal.

- A. Bagaimana Apabila Ibu hamil dengan Usia 29 Tahun, yang merupakan Kelahiran ke -2, dengan Waktu kelahiran sesuai dengan HPL, Memiliki tekanan darah Tinggi? Carilah KNN dengan menggunakan Key = 5

1. Melakukan import library yang dibutuhkan untuk mengerjakan soal.

```
import numpy as np
import pandas as pd
```

2. Selanjutnya membaca file yang akan digunakan untuk data set. Kebetulan file telah saya berada di direktori D di dalam folder dataset “D:\dataset\dataset2.xlsx”

```
df = pd.read_csv('D:\dataset\dataset4.txt',
                 delimiter=',')
```

3. Kemudian kita lakukan pengecekan pada ini dataset yang kita panggil

df

	Usia	Kelahiran_ke-	Waktu_Kelahiran	Tekanan_darah	Kelainan_jantung	Caesarian
0	22	1	0	2	0	0
1	26	2	0	1	0	1
2	26	2	1	1	0	0
3	28	1	0	2	0	0
4	22	2	0	1	0	1
...	...	...	...	...	...	...
75	27	2	1	1	0	0
76	33	4	0	1	0	1
77	29	2	1	2	0	1
78	25	1	2	0	0	1
79	24	2	2	1	0	0

80 rows × 6 columns

4. Selanjutnya lakukan perhitungan Algoritma KNN dengan kondisi pada Nomer 4A yaitu

Usia 30 Tahun, yang merupakan Kelahiran ke -1, dengan Waktu kelahiran sesuai dengan HPL, Memiliki tekanan darah Normal? Carilah KNN dengan menggunakan Key = 5, dengan syntax berikut ini :

```
import math
dis = []
for i in range(80):
    dis.append(math.sqrt((float(df.iloc[i]['Usia'])-29)**2
                        +(float(df.iloc[i]['Kelahiran_ke-'])- 2)**2
                        +(float(df.iloc[i]['Waktu_Kelahiran'])-0)**2
                        +(float(df.iloc[i]['Tekanan_darah'])-2)**2))
```

5. Lalu menampung hasilnya di variable dis

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

	Usia	Kelahiran_ke-	Waktu_Kelahiran	Tekanan_darah	Kelainan_jantung	Caesarian	dis
0	22	1	0	2	0	0	7.071068
1	26	2	0	1	0	1	3.162278
2	26	2	1	1	0	0	3.316625
3	28	1	0	2	0	0	1.414214
4	22	2	0	1	0	1	7.071068
...	...	...	...	...	...	...	...
75	27	2	1	1	0	0	2.449490
76	33	4	0	1	0	1	4.582576
77	29	2	1	2	0	1	1.000000
78	25	1	2	0	0	1	5.000000
79	24	2	2	1	0	0	5.477226

6. Lalu kita lakukan penyortiran.

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

	Usia	Kelahiran_ke-	Waktu_Kelahiran	Tekanan_darah	Kelainan_jantung	Caesarian	dis
54	29	2	0	1	1	1	1.000000
77	29	2	1	2	0	1	1.000000
67	29	2	0	1	1	0	1.000000
59	30	2	1	2	1	1	1.414214
3	28	1	0	2	0	0	1.414214
...	...	...	...	...	...	...	...
61	19	1	0	1	0	1	10.099505
26	18	1	1	2	1	1	11.090537
31	40	1	0	1	1	1	11.090537
25	18	1	0	1	0	0	11.090537
70	17	1	0	0	0	1	12.206556

80 rows × 7 columns

7. Berikutnya kita kan mengambil 5 teratas dari hasil terseebut dan kita tam[ung di variable y

```
y = df.sort_values('dis').head(5)
y
```

	Usia	Kelahiran_ke-	Waktu_Kelahiran	Tekanan_darah	Kelainan_jantung	Caesarian	dis
54	29	2	0	1	1	1	1.000000
77	29	2	1	2	0	1	1.000000
67	29	2	0	1	1	0	1.000000
59	30	2	1	2	1	1	1.414214
3	28	1	0	2	0	0	1.414214



8. Lalu kita akan mengambil data rerata dari Qty untuk menentukan banyak roti yang harus diproduksi

```
z = y["Caesarian"]  
z
```

```
54    1  
77    1  
67    0  
59    1  
3     0  
Name: Caesarian, dtype: int64
```

```
np.mean(z)
```

```
0.6
```

9. Algoritma KNN untuk soal dengan Ibu hamil dengan Usia 29 Tahun, yang merupakan Kelahiran ke -2, dengan Waktu kelahiran sesuai dengan HPL, Memiliki tekanan darah Tinggi maka di temukan bahwa ibu hamil tersebut akan di berlakukan 0,6 atau mendekati 1 maka akan bersalin secara Caesarian