

TUGAS AKHIR SEMESTER
KONSEP DAN APLIKASI DATA MINNING
NAÏVE BAYES
OLEH :
SANTOSO [17.52.0006]



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

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\dataset1.xlsx”

```
In [34]: data = pd.read_excel('D:\dataset\dataset1.xlsx')
```

3. Kemudian kita lakukan pengecekan pada ini dataset yang kita panggil dengan “data.head()” atau “data.tail()”.

```
In [35]: data.head()
```

Out[35]:

	Age	Income	Student	Credit_rating	Class
0	<=30	High	No	Fair	No
1	<=30	High	No	Excellent	No
2	31..40	High	No	Fair	Yes
3	> 40	Medium	No	Fair	Yes
4	> 40	Low	Yes	Fair	Yes

4. Kemudian melihat ukuran data set yang dibaca

```
In [36]: data.shape
```

Out[36]: (51, 5)

5. Berikutnya kita akan mencari tahu jumlah pengunjung yang membeli dan tidak membeli di sini field yang di baca adalah field “Class”.

*)nama field di dataset soal adalah Class (buy_computer) di rasa cukup panjang jadi saya perpendek demi kemudahan mengerjakan

```
In [37]: data['Class'].value_counts()
```

```
Out[37]: Yes    29
         No     22
         Name: Class, dtype: int64
```

6. Langkah selanjutnya adalah mencari knowledge dari data set.
 - a. Pada bagian awal kita akan melakukan persiapan dataset dengan melakukan crosstab pada beberapa field dengan field lainnya.
Contohnya seperti ini

```
In [51]: pd.crosstab(data['Age'], data['Income'])
```

```
Out[51]:
```

	Income	High	Low	Medium
Age				
31..40		5	5	4
<=30		6	6	8
> 40		0	10	7

- b. Lalu kita akan memvariable kan data yang akan kita gunakan untuk mencari knowledge.

```
In [52]: P31_40High = 5/14  
P31_40Low = 5/14  
P31_40Medium = 4/14  
  
PKur30High = 6/20  
PKur30Low = 6/20  
PKur30Medium = 8/20  
  
PLEb40High = 0/17  
PLEb40Low = 10/17  
PLEb40Medium = 7/17  
  
P31_40 = 14/51  
PKur30 = 20/51  
PLEb40 = 17/51
```

- c. Kita lakukan pengecekan pada setiap variable yang kita definisikan di atas

```
In [57]: print("PNoStudentNo= ",PNoStudentNo)
print("PYesStudentNo= ",PYesStudentNo)

print("PNoStudentYes= ",PNoStudentYes)
print("PYesStudentYes= ",PYesStudentYes)

print("PNoStudent= ",PNoStudent)
print("PYesStudent= ",PYesStudent)

print("PExcellentNo= ",PExcellentNo)
print("PFairNo= ",PFairNo)

print("PExcellentYes= ",PExcellentYes)
print("PFairYes= ",PFairYes)

print("PExcellent= ",PExcellent)
print("PFair= ",PFair)

print("P31_40No= ",P31_40No)
print("PKur30No= ",PKur30No)
print("PLeb40No= ",PLeb40No)

print("P31_40Yes= ",P31_40Yes)
print("PKur30Yes= ",PKur30Yes)
print("PLeb40Yes= ",PLeb40Yes)

print("P31_40= ",P31_40)
print("PKur30= ",PKur30)
print("PLeb40= ",PLeb40)

print("PExcellentHigh= ",PExcellentHigh)
print("PExcellentLow= ",PExcellentLow)
print("PExcellentMedium= ",PExcellentMedium)

print("PFairHigh= ",PFairHigh)
print("PFairLow= ",PFairLow)
print("PFairMedium= ",PFairMedium)

print("PNoStudentExcellent= ",PNoStudentExcellent)
print("PNoStudentFair= ",PNoStudentFair)

print("PYesStudentExcellent= ",PYesStudentExcellent)
print("PYesStudentFair= ",PYesStudentFair)
```

```

PNoStudentNo= 0.45454545454545453
PYesStudentNo= 0.5454545454545454
PNoStudentYes= 0.4827586206896552
PYesStudentYes= 0.5172413793103449
PNoStudent= 0.47058823529411764
PYesStudent= 0.5294117647058824
PExcellentNo= 0.36363636363636365
PFairNo= 0.6363636363636364
PExcellentYes= 0.41379310344827586
PFairYes= 0.5862068965517241
PExcellent= 0.39215686274509803
PFair= 0.6078431372549019
P31_40No= 0.2857142857142857
PKur30No= 0.5
PLEb40No= 0.47058823529411764
P31_40Yes= 0.7142857142857143
PKur30Yes= 0.5
PLEb40Yes= 0.5294117647058824
P31_40= 0.27450980392156865
PKur30= 0.39215686274509803
PLEb40= 0.3333333333333333
PExcellentHigh= 0.25
PExcellentLow= 0.4
PExcellentMedium= 0.35
PFairHigh= 0.1935483870967742
PFairLow= 0.41935483870967744
PFairMedium= 0.3870967741935484
PNoStudentExcellent= 0.3333333333333333
PNoStudentFair= 0.6666666666666666
PYesStudentExcellent= 0.4444444444444444
PYesStudentFair= 0.5555555555555556

```

- d. Lalu melakukan perbandingan perhitungan salah satunya seperti ini

```

In [60]: #membuktika pembeli terbanyak bukan dari kalangan student
PYesNotStudent = (PNotStudentYes*PYes)/PNotStudent
PYesStudent = (PStudentYes*PYes)/PStudent

print(PYesNotStudent)
print(PStudent)

0.7048611111111111
0.5294117647058824

```

```

In [61]: #degan lebih besarnya nilai probabilitas dari not Student yang melaukakan
#pembelian disimpulkan bahwa mayorits pembli adalah notStudent

```

```
In [62]: data.describe()
```

```
Out[62]:
```

	Age	Income	Student	Credit_rating	Class
count	51	51	51	51	51
unique	3	3	2	2	2
top	<=30	Low	Yes	Fair	Yes
freq	20	21	27	31	29