1. Lakukan clasifikasi dengan menggunakan Algoritma Naïve Bayes dalam soal No. 1 dan carilah knowledge sebanyak-banyak yang dapat Anda ketahui dari data tersebut.
   Jawaban :

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:

```
dataset = pd.read_csv('D:/UAS Data Mining/dataset_soal No.1.csv', delimiter = ';')
```

In [3]:

```
dataset.head()
```

Out[3]:

	Age	Income	Student	Credit_rating	Class (buy_computer)
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

In [4]:

```
dataset.shape
```

Out[4]:

```
(51, 5)
```

In [5]:

```
dataset['Class (buy_computer)'].value_counts()
```

Out[5]:

```
Yes    29
No     22
Name: Class (buy_computer), dtype: int64
```

In [6]:

```
PYes = 29/51
PNo = 22/51
```

In [7]:

```
print(PYes)
```

```
0.5686274509803921
```

In [8]:

```
print(PNo)
```

```
0.43137254901960786
```

In [9]:

```
pd.crosstab(dataset['Age'],dataset['Class (buy_computer)'])
```

Out[9]:

Class (buy_computer)	No	Yes
Age		
31..40	4	10
<= 30	7	8
<=30	3	2
> 40	8	9

In [10]:

```
pd.crosstab(dataset['Income'], dataset['Class (buy_computer)'])
```

Out[10]:

Class (buy_computer)	No	Yes
Income		
High	6	5
Low	11	10
Medium	5	14

In [17]:

```
PHighNo = 6/22
PLowNo = 11/22
PMediumNo = 5/22

PHighYes = 5/29
PLowYes = 10/29
PMediumYes = 14/29

PHigh = 11/51
PLow = 21/51
PMedium = 19/51

print(PMediumYes)
```

```
0.4827586206896552
```

In [18]:

```
PNoMedium = (PMediumNo*PNo)/PMedium
print(PNoMedium)
```

```
0.2631578947368421
```

In [19]:

```
PYesMedium = (PMediumYes*PYes)/PMedium
print(PYesMedium)
```

```
0.736842105263158
```

In [20]:

```
print(PNoMedium)
print(PYesMedium)
```

```
0.2631578947368421
0.736842105263158
```

In [13]:

```
pd.crosstab(dataset['Student'], dataset['Class (buy_computer)'])
```

Out[13]:

Class (buy_computer)	No	Yes
Student		
No	10	14
Yes	12	15

In [14]:

```
PNoStudentNo = 10/22
PYesStudentNo = 12/22

PNoStudentYes = 14/29
PYesStudentYes = 15/29

PNoStudent = 24/51
PYesStudent = 27/51
```

In [15]:

```
pd.crosstab(dataset['Credit_rating'], dataset['Class (buy_computer)'])
```

Out[15]:

Class (buy_computer)	No	Yes
Credit_rating		
Excellent	8	12
Fair	14	17

In [16]:

```
PExcellentNo = 8/22
PFairNo = 14/22

PExcellentYes = 12/29
PFairYes = 17/29

PExcellent = 20/51
PFair = 31/51
```

In [21]:

```
PNoClass = ((PMediumNo*PYesStudentNo*PFairNo)*PNo)/(PMedium*PYesStudent*PFair)
print(PNoClass)
```

```
0.2838541301267031
```

In [22]:

```
PClass = ((PMediumYes*PYesStudentYes*PFairYes)*PYes)/(PMedium*PYesStudent*PFair)
print(PClass)
```

```
0.694278175589332
```

In [23]:

```
PJum = PNoClass+PClass
print(PJum)
```

```
0.9781323057160352
```

In [24]:

```
PClass1 = (PClass/PJum) *100
print(PClass1)
```

```
70.9799861973775
```

In [25]:

```
PNoClass1 = (PNoClass*PJum) *100
print(PNoClass1)
```

```
27.764689478785154
```