

```
In [4]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import csv

In [5]: data =pd.read_csv('E:/fahmi/uas_mining/noi/dataset_soal No.1.csv',delimiter=',')

In [6]: data.head()

Out[6]:
   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 [7]: data.tail(10)

Out[7]:
   Age  Income  Student  Credit_rating  Class (buy_computer)
41  > 40     Low      Yes           Fair                No
42  31..40    Low      Yes           Fair                Yes
43  31..40    Low      Yes      Excellent                No
44  <= 30    High      No      Excellent                No
45  <= 30    Medium  Yes      Excellent                Yes
46  > 40     Low      Yes           Fair                Yes
47  <= 30     Low      Yes           Fair                Yes
48  31..40    Medium  No           Fair                No
49  31..40    High      Yes      Excellent                Yes
50  > 40    Medium      No      Excellent                No

In [9]: data['Age'].value_counts()
Out[9]:
> 40      17
<= 30     15
31..40     14
<=30       5
Name: Age, dtype: int64

In [11]: data['Income'].value_counts()
Out[11]:
Low      21
Medium   19
High     11
Name: Income, dtype: int64

In [12]: data['Student'].value_counts()
Out[12]:
Yes      27
No       24
Name: Student, dtype: int64

In [13]: data['Credit_rating'].value_counts()
Out[13]:
Fair      31
Excellent  28
Name: Credit_rating, dtype: int64

In [14]: data['Class (buy_computer)'].value_counts()
Out[14]:
Yes      29
No       22
Name: Class (buy_computer), dtype: int64

In [15]: data.shape
Out[15]: (51, 5)

In [17]: PYes = 27/51
PNo = 24/51

In [18]: pd.crosstab(data['Age'],data['Income'])
Out[18]:
Income  High  Low  Medium
Age
31..40     5     5     4
<= 30      1     6     8
<=30       5     0     0
> 40       0    10     7

In [19]: pd.crosstab(data['Age'],data['Student'])
Out[19]:
Student  No  Yes
Age
31..40     7     7
<= 30      6     9
<=30       5     0
> 40       6    11

In [20]: pd.crosstab(data['Age'],data['Credit_rating'])
Out[20]:
Credit_rating  Excellent  Fair
Age
31..40             7     7
<= 30             4    11
<=30              3     2
> 40             6    11

In [21]: pd.crosstab(data['Age'],data['Class (buy_computer)'])
Out[21]:
Class (buy_computer)  No  Yes
Age
31..40             4    10
<= 30             7     8
<=30              3     2
> 40             8     9

In [22]: pd.crosstab(data['Age'],data['Income'])
Out[22]:
Income  High  Low  Medium
Age
31..40     5     5     4
<= 30      1     6     8
<=30       5     0     0
> 40       0    10     7

In [23]: pd.crosstab(data['Age'],data['Income'])
Out[23]:
Income  High  Low  Medium
Age
31..40     5     5     4
<= 30      1     6     8
<=30       5     0     0
> 40       0    10     7

In [24]: pd.crosstab(data['Age'],data['Income'])
Out[24]:
Income  High  Low  Medium
Age
31..40     5     5     4
<= 30      1     6     8
<=30       5     0     0
> 40       0    10     7

In [25]: pd.crosstab(data['Age'],data['Class (buy_computer)'])
Out[25]:
Class (buy_computer)  No  Yes
Age
31..40             4    10
<= 30             7     8
<=30              3     2
> 40             8     9

In [26]: pd.crosstab(data['Income'],data['Class (buy_computer)'])
Out[26]:
Class (buy_computer)  No  Yes
Income
High              6     5
Low             11    10
Medium           5    14

In [27]: pd.crosstab(data['Income'],data['Credit_rating'])
Out[27]:
Credit_rating  Excellent  Fair
Income
High              5     6
Low              8    13
Medium           7    12

In [28]: pd.crosstab(data['Income'],data['Age'])
Out[28]:
Age  31..40  <= 30  <=30  > 40
Income
High      5      1      5      0
Low       5      6      0    10
Medium    4      8      0     7

In [32]: PHighNo = 6/22
PLOWNo = 11/22
PMediumNo = 5/22

PHighYes = 5/29
PLOWYes = 10/29
PMediumYes = 5/29

PHigh = 11/52
PLOW = 21/51
PMedium = 19/51

print (PHighNo)
0.2727272727272727

In [33]: print (PHighYes)
0.1724137931034483

In [34]: print (PHigh)
0.21153846153846154

In [36]: print (PLOWNo)
0.5

In [38]: print (PLOWYes)
0.3448275862068966

In [39]: print (PLOW)
0.4117647058823529

In [40]: print (PMediumYes)
0.1724137931034483

In [41]: print (PMediumNo)
0.22727272727272727

In [42]: print (PMedium)
0.37254901960784315

In [43]: data.describe()
Out[43]:
   Age  Income  Student  Credit_rating  Class (buy_computer)
count  51      51      51           51                51
unique   4       3       2             2                 2
top    > 40     Low      Yes           Fair                Yes
freq     17      21      27             31                29

In [48]: pd.crosstab(data['Income'], data['Student'])
Out[48]:
Student  No  Yes
Income
High      9     2
Low       1    20
Medium    14     5

In [49]: PHighNo = 9/24
PLOWNo = 1/24
PMediumNo= 14/24

PHighYes = 2/27
PLOWYes = 20/27
PMediumYes = 5/27

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

print(PHighNo)
0.375

In [51]: print(PLOWNo)
0.041666666666666664

In [52]: print(PMediumNo)
0.5833333333333334

In [53]: print(PHighYes)
0.07407407407407407

In [54]: print(PLOWYes)
0.7407407407407407

In [55]: print(PMediumYes)
0.18518518518518517

In [56]: print(PHigh)
0.21568627450980393

In [57]: print(PLOW)
0.4117647058823529

In [58]: print(PMedium)
0.37254901960784315

In [59]: pd.crosstab(data['Credit_rating'], data['Student'])
Out[59]:
Student  No  Yes
Credit_rating
Excellent  8    12
Fair      16    15

In [60]: PExcellentNo = 8/24
PFairNo = 16/24

PExcellentYes = 12/27
PFairYes = 15/27

PExcellent = 20/51
PFair = 31/51

print(PExcellentNo)
0.3333333333333333

In [61]: print(PFairNo)
0.6666666666666666

In [62]: print(PExcellentYes)
0.4444444444444444

In [63]: print(PFairYes)
0.5555555555555556

In [64]: print(PExcellent)
0.39215686274509803

In [65]: print(PFair)
0.6078431372549019

In [66]: pd.crosstab(data['Income'], data['Class (buy_computer)'])
Out[66]:
Class (buy_computer)  No  Yes
Income
High              6     5
Low             11    10
Medium           5    14

In [67]: PHighNo = 6/22
PLOWNo = 11/22
PMediumNo= 5/22

PHighYes = 5/29
PLOWYes = 10/29
PMediumYes = 24/29

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

print(PHighNo)
0.2727272727272727

In [68]: print(PLOWNo)
0.5

In [69]: print(PMediumNo)
0.22727272727272727

In [70]: print(PHighYes)
0.1724137931034483

In [71]: print(PLOWYes)
0.3448275862068966

In [72]: print(PMediumYes)
0.8275862068965517

In [73]: pd.crosstab(data['Credit_rating'], data['Class (buy_computer)'])
Out[73]:
Class (buy_computer)  No  Yes
Credit_rating
Excellent      8    12
Fair          14    17

In [74]: PExcellentNo = 8/22
PFairNo = 14/22

PExcellentYes = 12/29
PFairYes = 17/29

PExcellent = 20/51
PFair = 31/51

print(PExcellentNo)
0.36363636363636365

In [75]: print(PFairNo)
0.6363636363636364

In [76]: print(PExcellentYes)
0.41379310344827586

In [77]: print(PFairYes)
0.5862068965517241

In [ ] :
```