

UAS SISTEM APLIKASI DATA MAINING



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

Jawab:

A) Knowledge Menghasilkan 5 urutan awal dan 10 urutan akhir.

Hasil:

```
In [3]: df=pd.read_csv('C:/Users/Yulvavi/Documents/KULIAH/Himatus Yulvi/Sistem Data maining/dataset_soal_1.csv',delimiter=',')
```

```
In [4]: df.head()
```

```
Out[4]:
```

	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 [5]: df.tail(10)
```

```
Out[5]:
```

	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

B) Knowledge Menghasilkan Nilai dari Student , Age, Icome, Kredit_rating dan Class buy_computer.

Hasil:

```
In [6]: df.shape
```

```
Out[6]: (51, 5)
```

```
In [7]: df['Student'].value_counts()
```

```
Out[7]: Yes    27  
        No     24  
        Name: Student, dtype: int64
```

```
In [8]: df['Age'].value_counts()
```

```
Out[8]: > 40    17  
        < 30    15  
        31..40  14  
        <=30    5  
        Name: Age, dtype: int64
```

```
In [9]: df['Income'].value_counts()
```

```
Out[9]: Low     21  
        Medium  19  
        High    11  
        Name: Income, dtype: int64
```

```
In [10]: df['Credit_rating'].value_counts()
```

```
Out[10]: Fair     31  
         Excellent 20  
         Name: Credit_rating, dtype: int64
```

```
In [11]: df['Class (buy_computer)'].value_counts()
```

```
Out[11]: Yes     29  
        No      22
```

C)Knowledge Menghasilkan Nilai dari Pyes dan Pno di lihat dari data Student.
Hasil:

```
In [12]: PYES= 27/51  
        PNO=24/51  
  
In [13]: print (PYES)  
0.5294117647058824  
  
In [14]: print (PNO)  
0.47058823529411764
```

D)Knowledge Menghasilkan Nilai dari Plow,PMedium dan Phigh di lihat dari data Income
Hasil:

```
In [15]: PLOW=21/51  
        PMedium=19/51  
        PHigh=11/51  
  
In [16]: print (PLOW)  
0.4117647058823529  
  
In [17]: print (PMedium)  
0.37254901960784315  
  
In [18]: print (PHigh)  
0.21568627450980393
```

E)Knowledge Menghasilkan Nilai dari Pyes dan Pno di lihat dari data Class(buy_computer)
Hasil:

```
In [19]: PYES= 29/51  
        PNo=22/51  
  
In [20]: print (PYES)  
0.5686274509803921  
  
In [21]: print (PNo)  
0.43137254901960786
```

F)Nilai Knowledge ≤ 30 dan < 30 data Age yang masih campur .

Hasil:

```
In [22]: age_more_than_30 = df[ df["Age"] == "<=30" ]
```

```
In [23]: print(age_more_than_30)
```

	Age	Income	Student	Credit_rating	Class (buy_computer)
0	<=30	High	No	Fair	No
1	<=30	High	No	Excellent	No
25	<=30	High	No	Excellent	Yes
31	<=30	High	No	Fair	Yes
37	<=30	High	No	Excellent	No

```
In [24]: age_more_than_30 = df[ df["Age"] == "< 30" ]
```

```
In [25]: print(age_more_than_30)
```

	Age	Income	Student	Credit_rating	Class (buy_computer)
7	< 30	Medium	No	Fair	No
8	< 30	Low	Yes	Fair	No
10	< 30	Medium	Yes	Excellent	Yes
14	< 30	Medium	No	Fair	No
15	< 30	Low	Yes	Fair	No
16	< 30	Medium	No	Fair	Yes
17	< 30	Low	Yes	Fair	Yes
18	< 30	Medium	No	Fair	Yes
19	< 30	Low	Yes	Fair	No
20	< 30	Medium	No	Fair	Yes
21	< 30	Low	Yes	Fair	No
23	< 30	Medium	Yes	Excellent	Yes
44	< 30	High	No	Excellent	No
45	< 30	Medium	Yes	Excellent	Yes
47	< 30	Low	Yes	Fair	Yes

G)Knowledge Menghasilkan Nilai dari data Age 30..40 dan > 40 yang masih belum di kelompokkan.

Hasil:

```
In [26]: age_more_than_40 = df[ df["Age"] == "30..40" ]
```

```
In [27]: print(age_more_than_30)
```

	Age	Income	Student	Credit_rating	Class (buy_computer)
7	< 30	Medium	No	Fair	No
8	< 30	Low	Yes	Fair	No
10	< 30	Medium	Yes	Excellent	Yes
14	< 30	Medium	No	Fair	No
15	< 30	Low	Yes	Fair	No
16	< 30	Medium	No	Fair	Yes
17	< 30	Low	Yes	Fair	Yes
18	< 30	Medium	No	Fair	Yes
19	< 30	Low	Yes	Fair	No
20	< 30	Medium	No	Fair	Yes
21	< 30	Low	Yes	Fair	No
23	< 30	Medium	Yes	Excellent	Yes
44	< 30	High	No	Excellent	No
45	< 30	Medium	Yes	Excellent	Yes
47	< 30	Low	Yes	Fair	Yes

```
In [28]: age_more_than_40 = df[ df["Age"] == ">40" ]
```

```
In [29]: print(age_more_than_30)
```

	Age	Income	Student	Credit_rating	Class (buy_computer)
7	< 30	Medium	No	Fair	No
8	< 30	Low	Yes	Fair	No
10	< 30	Medium	Yes	Excellent	Yes
14	< 30	Medium	No	Fair	No
15	< 30	Low	Yes	Fair	No
16	< 30	Medium	No	Fair	Yes
17	< 30	Low	Yes	Fair	Yes
18	< 30	Medium	No	Fair	Yes
19	< 30	Low	Yes	Fair	No
20	< 30	Medium	No	Fair	Yes
21	< 30	Low	Yes	Fair	No
23	< 30	Medium	Yes	Excellent	Yes
44	< 30	High	No	Excellent	No
45	< 30	Medium	Yes	Excellent	Yes
47	< 30	Low	Yes	Fair	Yes

H)Knowledge Menghasilkan Nilai dari data Icome dan Student sudah tertara Nilai Yes dan No.

Hasil:

```
In [30]: pd.crosstab (df['Income'], df['Student'])
```

Out[30]:

	Student	No	Yes
Income			
High	9	2	
Low	1	20	
Medium	14	5	

Output Hasil dari pengabungan nilai Yes dan No. dan data diatas Menghasilkan Nilai:

```
In [31]: #student

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 [32]: print (PLowNo)

0.041666666666666664
```

```
In [33]: print (PMediumNo)

0.5833333333333334
```

I)Knowledge Menghasilkan Nilai dari data Icome dan Vredit_rating sudah tertara Nilai Excellent dan Fair
Hasil:

```
In [38]: pd.crosstab (df['Income'], df['Credit_rating'])
```

```
Out[38]:
```

	Credit_rating	
	Excellent	Fair
Income		
High	5	6
Low	8	13
Medium	7	12

Output Hasil dari pengabungan nilai Nilai Excellent dan Fair dari data diatas Menghasilkan Nilai:

```
In [54]: PHighExcellent=5/20
         PLOWExcellent=8/20
         PMediumExcellent=7/20

         PHighFair=6/31
         PLOWFair=13/31
         PMediumFair=12/31

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

         print (PHighExcellent)

0.25

In [56]: print (PHighFair)

0.1935483870967742

In [57]: print (PHigh)

0.21568627450980393

In [58]: print (PLOWExcellent)

0.4

In [59]: print (PLOWFair)

0.41935483870967744

In [60]: print (PLOW)

0.4117647058823529
```

J) Knowledge Menghasilkan Nilai dari data Icome dan Class (buy_Computer) sudah tertara Nilai Yes dan No.

Hasil:

```
In [39]: pd.crosstab (df['Income'], df['Class (buy_computer)'])

Out[39]:
```

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

Output Hasil dari pengabungan nilai Yes dan No. dari data diatas untuk Menghasilkan Nilai (high,low,dan Medium):

```
In [61]: PHighNo=6/22  
        PLowNo=11/22  
        PMediumNo=5/22  
  
        PHighYes=5/29  
        PLowYes=10/29  
        PMediumYes=5/29  
  
        PHigh =11/51  
        PLow=21/51  
        PMedium=19/51  
  
        print (PHighNo)
```

```
0.2727272727272727
```

```
In [62]: print (PLowNo)
```

```
0.5
```

```
In [64]: print (PMediumNo)
```

```
0.22727272727272727
```

```
In [66]: print (PHighYes)
```

```
0.1724137931034483
```

```
In [67]: print (PLowYes)
```

```
0.3448275862068966
```


K) Knowledge Menghasilkan Nilai dari data Age dan Student.

Hasil:

```
In [35]: pd.crosstab (df['Age'], df['Student'])
```

Out[35]:

	Student	
	No	Yes
Age		
31..40	7	7
< 30	6	9
<=30	5	0
> 40	6	11

```
In [ ]: P31.40No=7/24
P<30No =6/24
P<=30No=5/24
P>40No=6/24

P31.40Yes=7/27
P<30Yes=9/27
P<=30Yes=0/27
P>40Yes=11/27

P31.40=14/51
P<30=15/51
P<=30=5/51
P>40=17/51
```

L) Knowledge yang Menghasilkan Nilai dari data Age dan Kredit_rating beserta penjumlahannya

Hasil:

```
In [36]: pd.crosstab (df['Age'], df['Credit_rating'])
```

Out[36]:

	Credit_rating	
	Excellent	Fair
Age		
31..40	7	7
< 30	4	11
<=30	3	2
> 40	6	11

```
In [ ]: P31.40Excellent=7/20
P<30Excellent=4/20
P<=30Excellent=3/20
P>40Excellent=6/20

P31.40Fair=7/31
P<30Fair=11/31
P<=30Fair=2/31
P>40Fair=11/31

P31.40=14/51
P<30=15/51
P<=30=5/51
P>40=17/51
```

M) Knowledge yang Menghasilkan Nilai dari data Age dan Buy_computer berserta penjumlahannya
Hasil:

```
In [37]: pd.crosstab (df['Age'], df['Class (buy_computer)'])
```

```
Out[37]:
```

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

```
In [ ]: P31..40No=4/22
P<30No =7/22
P<=30No=3/22
P>40No=8/22

P31..40Yes=10/29
P<30Yes=8/29
P<=30Yes=2/29
P>40Yes=9/29

P31..40=14/51
P<30=15/51
P<=30=5/51
P>40=17/51
```

N) Knowledge yang Menghasilkan Nilai dari data Age dan Buy_computer berserta penjumlahannya.
Hasil:

```
In [65]: pd.crosstab (df['Income'], df['Age'])
```

```
Out[65]:
```

	Age			
	31..40	< 30	<=30	> 40
Income				
High	5	1	5	0
Low	5	6	0	10
Medium	4	8	0	7

```
In [ ]: PHigh31..40 =5/14
PLow31..40 =5/14
PMedium31..40 =4/14

PHigh<= 30 =1/15
PLow<= 30 =6/15
PMedium<= 30 =8/15

PHigh<=30 =5/5
PLow<=30 =0/5
PMedium<=30 =0/5

PHigh> 40 =0/17
PLow> 40 =10/17
PMedium> 40 =7/10

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

O) Knowledge Menghasilkan Nilai dari seluruh data.
Hasil:

In [72]: `df.describe()`

Out[72]:

	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