Assignment 6

Name: Khalate Shubham

Class: AIDS A

Rollno: 23107062

Batch: C

```
import pandas as pd

df=pd.read_csv("iris.csv")

df
```

```
     Unnamed: 0  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
\
0             1           5.1          3.5           1.4          0.2

1             2           4.9          3.0           1.4          0.2

2             3           4.7          3.2           1.3          0.2

3             4           4.6          3.1           1.5          0.2

4             5           5.0          3.6           1.4          0.2

..          ...           ...          ...           ...          ...

145         146           6.7          3.0           5.2          2.3

146         147           6.3          2.5           5.0          1.9

147         148           6.5          3.0           5.2          2.0

148         149           6.2          3.4           5.4          2.3

149         150           5.9          3.0           5.1          1.8


       Species
0       setosa
1       setosa
2       setosa
3       setosa
4       setosa
..         ...
145  virginica
146  virginica
147  virginica
```

```
148  virginica
149  virginica

[150 rows x 6 columns]
```

```
df.count()
```

```
Unnamed: 0      150
Sepal.Length    150
Sepal.Width     150
Petal.Length    150
Petal.Width     150
Species         150
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Unnamed: 0    150 non-null    int64
 1   Sepal.Length  150 non-null    float64
 2   Sepal.Width   150 non-null    float64
 3   Petal.Length  150 non-null    float64
 4   Petal.Width   150 non-null    float64
 5   Species       150 non-null    object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

```
df.describe()
```

```
       Unnamed: 0  Sepal.Length  Sepal.Width  Petal.Length
Petal.Width
count  150.000000    150.000000   150.000000    150.000000
150.000000
mean    75.500000      5.843333     3.057333      3.758000
1.199333
std     43.445368      0.828066     0.435866      1.765298
0.762238
min      1.000000      4.300000     2.000000      1.000000
0.100000
25%     38.250000      5.100000     2.800000      1.600000
0.300000
50%     75.500000      5.800000     3.000000      4.350000
1.300000
75%    112.750000      6.400000     3.300000      5.100000
1.800000
max    150.000000      7.900000     4.400000      6.900000
2.500000
```

```
df.notnull()
```

```
     Unnamed: 0  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Species
0          True          True         True          True         True
True
1          True          True         True          True         True
True
2          True          True         True          True         True
True
3          True          True         True          True         True
True
4          True          True         True          True         True
True
..          ...           ...          ...           ...          ...
...
145        True          True         True          True         True
True
146        True          True         True          True         True
True
147        True          True         True          True         True
True
148        True          True         True          True         True
True
149        True          True         True          True         True
True

[150 rows x 6 columns]
```

```
df.isnull()
```

```
     Unnamed: 0  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Species
0         False         False        False         False        False
False
1         False         False        False         False        False
False
2         False         False        False         False        False
False
3         False         False        False         False        False
False
4         False         False        False         False        False
False
..          ...           ...          ...           ...          ...
...
145       False         False        False         False        False
False
146       False         False        False         False        False
False
147       False         False        False         False        False
```

```
False
148       False          False         False          False         False
False
149       False          False         False          False         False
False

[150 rows x 6 columns]
```

```
df.dtypes
```

```
Unnamed: 0        int64
Sepal.Length    float64
Sepal.Width     float64
Petal.Length    float64
Petal.Width     float64
Species          object
dtype: object
```

```
df.shape
```

```
(150, 6)
```

```
df.ndim
```

```
2
```

```
df.value_counts()
```

```
Unnamed: 0  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Species
1           5.1           3.5          1.4           0.2
setosa        1
95          5.6           2.7          4.2           1.3
versicolor    1
97          5.7           2.9          4.2           1.3
versicolor    1
98          6.2           2.9          4.3           1.3
versicolor    1
99          5.1           2.5          3.0           1.1
versicolor    1

..
51          7.0           3.2          4.7           1.4
versicolor    1
52          6.4           3.2          4.5           1.5
versicolor    1
53          6.9           3.1          4.9           1.5
versicolor    1
54          5.5           2.3          4.0           1.3
versicolor    1
150         5.9           3.0          5.1           1.8
```

```
virginica     1
Name: count, Length: 150, dtype: int64
```

```
x=df.drop(['Species'],axis=1)
```

```
x
```

```
     Unnamed: 0  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
0             1           5.1          3.5           1.4          0.2
1             2           4.9          3.0           1.4          0.2
2             3           4.7          3.2           1.3          0.2
3             4           4.6          3.1           1.5          0.2
4             5           5.0          3.6           1.4          0.2
..          ...           ...          ...           ...          ...
145         146           6.7          3.0           5.2          2.3
146         147           6.3          2.5           5.0          1.9
147         148           6.5          3.0           5.2          2.0
148         149           6.2          3.4           5.4          2.3
149         150           5.9          3.0           5.1          1.8

[150 rows x 5 columns]
```

```
y=df["Species"]
```

```
y
```

```
0         setosa
1         setosa
2         setosa
3         setosa
4         setosa
         ...
145    virginica
146    virginica
147    virginica
148    virginica
149    virginica
Name: Species, Length: 150, dtype: object
```

```
from sklearn import linear_model
from sklearn.model_selection import train_test_split

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

```
x_train
```

```
     Unnamed: 0  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
90           91           5.5          2.6           4.4          1.2
101         102           5.8          2.7           5.1          1.9
14           15           5.8          4.0           1.2          0.2
51           52           6.4          3.2           4.5          1.5
71           72           6.1          2.8           4.0          1.3
```

```
..          ...           ...           ...           ...           ...
130         131           7.4           2.8           6.1           1.9
27           28           5.2           3.5           1.5           0.2
19           20           5.1           3.8           1.5           0.3
145         146           6.7           3.0           5.2           2.3
0             1           5.1           3.5           1.4           0.2

[120 rows x 5 columns]
```

x_test

```
     Unnamed: 0  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
43           44           5.0           3.5           1.6           0.6
82           83           5.8           2.7           3.9           1.2
84           85           5.4           3.0           4.5           1.5
98           99           5.1           2.5           3.0           1.1
80           81           5.5           2.4           3.8           1.1
18           19           5.7           3.8           1.7           0.3
94           95           5.6           2.7           4.2           1.3
125         126           7.2           3.2           6.0           1.8
110         111           6.5           3.2           5.1           2.0
55           56           5.7           2.8           4.5           1.3
100         101           6.3           3.3           6.0           2.5
143         144           6.8           3.2           5.9           2.3
119         120           6.0           2.2           5.0           1.5
37           38           4.9           3.6           1.4           0.1
32           33           5.2           4.1           1.5           0.1
91           92           6.1           3.0           4.6           1.4
58           59           6.6           2.9           4.6           1.3
78           79           6.0           2.9           4.5           1.5
87           88           6.3           2.3           4.4           1.3
23           24           5.1           3.3           1.7           0.5
29           30           4.7           3.2           1.6           0.2
111         112           6.4           2.7           5.3           1.9
11           12           4.8           3.4           1.6           0.2
86           87           6.7           3.1           4.7           1.5
142         143           5.8           2.7           5.1           1.9
138         139           6.0           3.0           4.8           1.8
54           55           6.5           2.8           4.6           1.5
139         140           6.9           3.1           5.4           2.1
93           94           5.0           2.3           3.3           1.0
40           41           5.0           3.5           1.3           0.3
```

y_train

```
90     versicolor
101     virginica
14         setosa
51     versicolor
71     versicolor
```

```
          ...
130     virginica
27         setosa
19         setosa
145     virginica
0          setosa
Name: Species, Length: 120, dtype: object
```

y_test

```
43         setosa
82     versicolor
84     versicolor
98     versicolor
80     versicolor
18         setosa
94     versicolor
125     virginica
110     virginica
55     versicolor
100     virginica
143     virginica
119     virginica
37         setosa
32         setosa
91     versicolor
58     versicolor
78     versicolor
87     versicolor
23         setosa
29         setosa
111     virginica
11         setosa
86     versicolor
142     virginica
138     virginica
54     versicolor
139     virginica
93     versicolor
40         setosa
Name: Species, dtype: object
```

```
len(x_train)
```

```
120
```

```
len(x_test)
```

```
30
```

```
len(y_train)
```

```
120
```

```
len(y_test)
```

```
30
```

```
from sklearn.naive_bayes import GaussianNB,MultinomialNB,BernoulliNB
```

```
GNB=GaussianNB()
```

```
GNB.fit(x_train,y_train)
```

```
GaussianNB()
```

```
y_pred=GNB.predict(x_test)
```

```
y_pred
```

```
array(['setosa', 'versicolor', 'versicolor', 'versicolor',
'versicolor',
       'setosa', 'versicolor', 'virginica', 'virginica', 'versicolor',
       'virginica', 'virginica', 'virginica', 'setosa', 'setosa',
       'versicolor', 'versicolor', 'versicolor', 'versicolor',
'setosa',
       'setosa', 'virginica', 'setosa', 'versicolor', 'virginica',
       'virginica', 'versicolor', 'virginica', 'versicolor',
'setosa'],
      dtype='<U10')
```

```
from sklearn.metrics import confusion_matrix,accuracy_score
```

```
cm=confusion_matrix(y_test,y_pred)
```

```
cm
```

```
array([[ 8,  0,  0],
       [ 0, 13,  0],
       [ 0,  0,  9]])
```

```
GNB.score(x_train,y_train)
```

```
0.9916666666666667
```

```
from sklearn.metrics import accuracy_score,classification_report
```

```
from sklearn import metrics
```

```
accuracy_g=accuracy_score(y_test,y_pred)
```

```
accuracy_g
```

```
1.0
```

```
error_rate=1-accuracy_g
```

```
error_rate
```

```
0.0
```

```
cr_g=classification_report(y_pred,y_test)
```

```
print(cr_g)
```

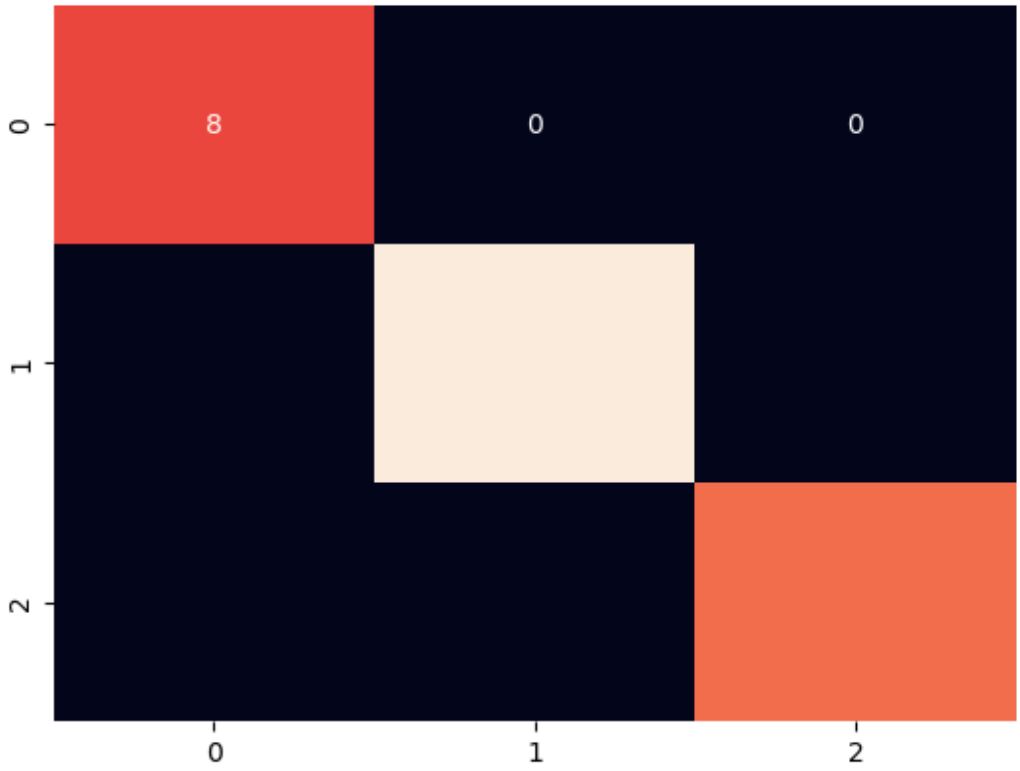
```
              precision    recall  f1-score   support

      setosa       1.00      1.00      1.00         8
  versicolor       1.00      1.00      1.00        13
   virginica       1.00      1.00      1.00         9

    accuracy                           1.00        30
   macro avg       1.00      1.00      1.00        30
weighted avg       1.00      1.00      1.00        30
```

```
import seaborn as sns
import matplotlib.pyplot as plt

sns.heatmap(cm,annot=True,cbar=False)
plt.xlabel = ("actual value")
plt.ylabel = ("predicted value")
plt.show()
```



```
M=BernoulliNB()

M.fit(x_train,y_train)

BernoulliNB()

M.fit

<bound method _BaseDiscreteNB.fit of BernoulliNB()>

y_pred1=M.predict(x_test)

cm1=confusion_matrix(y_test,y_pred1)

cm1

array([[ 8,  0,  0],
       [13,  0,  0],
       [ 9,  0,  0]])

v=MultinomialNB()

v

MultinomialNB()

v.fit(x_train,y_train)

MultinomialNB()

y_pred2=v.predict(x_test)

cm2=confusion_matrix(y_test,y_pred2)

cm2

array([[7, 1, 0],
       [0, 6, 7],
       [0, 1, 8]])
```