U18ISI6204 – Machine Learning Techniques

LAB EXPERIMENT- 7

NAME: Aaron Mathew

ROLL_NO: 20BIS001

Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.

```
import pandas as pd
df=pd.read_csv("tennis.csv")
df
```

```
In [1]: import pandas as pd

In [2]: df=pd.read_csv("tennis.csv")
        df

Out[2]:
```

	outlook	temp	humidity	windy	play
0	sunny	hot	high	False	no
1	sunny	hot	high	True	no
2	overcast	hot	high	False	yes
3	rainy	mild	high	False	yes
4	rainy	cool	normal	False	yes
5	rainy	cool	normal	True	no
6	overcast	cool	normal	True	yes
7	sunny	mild	high	False	no
8	sunny	cool	normal	False	yes
9	rainy	mild	normal	False	yes
10	sunny	mild	normal	True	yes
11	overcast	mild	high	True	yes
12	overcast	hot	normal	False	yes
13	rainy	mild	high	True	no

```
X_train = pd.get_dummies(df[['outlook', 'temp', 'humidity', 'windy']])
y_train = pd.DataFrame(df['play'])
print(X_train.info())
print(X_train.head())
```

```
In [3]: X_train = pd.get_dummies(df[['outlook', 'temp', 'humidity', 'windy']])
        y_train = pd.DataFrame(df['play'])

In [4]: print(X_train.info())
        print(X_train.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 9 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   windy             14 non-null     bool
 1   outlook_overcast  14 non-null     uint8
 2   outlook_rainy     14 non-null     uint8
 3   outlook_sunny     14 non-null     uint8
 4   temp_cool         14 non-null     uint8
 5   temp_hot          14 non-null     uint8
 6   temp_mild         14 non-null     uint8
 7   humidity_high     14 non-null     uint8
 8   humidity_normal   14 non-null     uint8
dtypes: bool(1), uint8(8)
memory usage: 254.0 bytes
None
   windy  outlook_overcast  outlook_rainy  outlook_sunny  temp_cool  temp_hot  \
0  False                 0              0              1          0         1
1   True                 0              0              1          0         1
2  False                 1              0              0          0         1
3  False                 0              1              0          0         0
4  False                 0              1              0          1         0

   temp_mild  humidity_high  humidity_normal
```

print(y_train.info())

print(y_train)

```
In [5]: print(y_train.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   play    14 non-null     object
dtypes: object(1)
memory usage: 240.0+ bytes
None
```

```
In [6]: print(y_train)
```

```
   play
0    no
1    no
2   yes
3   yes
4   yes
5    no
6   yes
7    no
8   yes
9   yes
10  yes
11  yes
12  yes
13   no
```

```
from sklearn.naive_bayes import GaussianNB

classifier=GaussianNB()

classifier.fit(X_train,y_train)

classifier.score(X_train,y_train)

X_train.head()

classifier.predict([[True,0,0,1,0,1,0,1,0]])
```

In [7]:
```
from sklearn.naive_bayes import GaussianNB
classifier=GaussianNB()
classifier.fit(X_train,y_train)
classifier.score(X_train,y_train)
X_train.head()
```

```
C:\Users\dhars\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed
when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Out[7]:

	windy	outlook_overcast	outlook_rainy	outlook_sunny	temp_cool	temp_hot	temp_mild	humidity_high	humidity_normal
0	False	0	0	1	0	1	0	1	0
1	True	0	0	1	0	1	0	1	0
2	False	1	0	0	0	1	0	1	0
3	False	0	1	0	0	0	1	1	0
4	False	0	1	0	1	0	0	0	1

In [8]:
```
classifier.predict([[True,0,0,1,0,1,0,1,0]])
```

```
C:\Users\dhars\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but GaussianN
B was fitted with feature names
  warnings.warn(
```

Out[8]:
```
array(['no'], dtype='<U3')
```

y_train.head()

a=classifier.predict([[True,0,0,1,0,1,1,1,1]])

if(a[0]=="yes"):

    print("yOU CAN PLAY!!!!")

else:

    print("You cant play!!!")

In [9]:
```
y_train.head()
```

Out[9]:

	play
0	no
1	no
2	yes
3	yes
4	yes

In [10]:
```
a=classifier.predict([[True,0,0,1,0,1,1,1,1]])
```

```
C:\Users\dhars\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but GaussianN
B was fitted with feature names
  warnings.warn(
```

In [11]:
```
if(a[0]=="yes"):
    print("yOU CAN PLAY!!!!")
else:
    print("You cant play!!!")
```

```
You cant play!!!
```