**Experiment No.: 5**

**Name: JESNAMOL THOMAS**

**Roll No:6**

**Batch:MCA B**

**Date:15/10/22**

**Aim**

Gaussian Naive Bayes classifier

**Program and Output**

import pandas as pd

import io

from sklearn.model\_selection import train\_test\_split

df1 = pd.read\_csv('/content/golf-dataset.csv')

print(df1)

df1.head(10)

X = df1.iloc[:, [0,1,2, 3]].values

y = df1.iloc[:, -1].values

print(X)

print(y)

X\_train, X\_test, y\_train, y\_test = train\_test\_split( X, y, test\_size = 0.2, random\_state=42)

print(X\_test)

df1['Outlook'].replace(['Rainy'],[0], inplace=True)

df1['Outlook'].replace(['Overcast'],[1], inplace=True)

df1['Outlook'].replace(['Sunny'],[2], inplace=True)

print(df1)

**Output**

Outlook Temp Humidity Windy Play Golf

0 Rainy Hot High False No

1 Rainy Hot High True No

2 Overcast Hot High False Yes

3 Sunny Mild High False Yes

4 Sunny Cool Normal False Yes

5 Sunny Cool Normal True No

6 Overcast Cool Normal True Yes

7 Rainy Mild High False No

8 Rainy Cool Normal False Yes

9 Sunny Mild Normal False Yes

10 Rainy Mild Normal True Yes

11 Overcast Mild High True Yes

12 Overcast Hot Normal False Yes

13 Sunny Mild High True No

[['Rainy' 'Hot' 'High' False]

['Rainy' 'Hot' 'High' True]

['Overcast' 'Hot' 'High' False]

['Sunny' 'Mild' 'High' False]

['Sunny' 'Cool' 'Normal' False]

['Sunny' 'Cool' 'Normal' True]

['Overcast' 'Cool' 'Normal' True]

['Rainy' 'Mild' 'High' False]

['Rainy' 'Cool' 'Normal' False]

['Sunny' 'Mild' 'Normal' False]

['Rainy' 'Mild' 'Normal' True]

['Overcast' 'Mild' 'High' True]

['Overcast' 'Hot' 'Normal' False]

['Sunny' 'Mild' 'High' True]]

['No' 'No' 'Yes' 'Yes' 'Yes' 'No' 'Yes' 'No' 'Yes' 'Yes' 'Yes' 'Yes' 'Yes'

'No']

[['Sunny' 'Mild' 'Normal' False]

['Overcast' 'Mild' 'High' True]

['Rainy' 'Hot' 'High' False]]

Outlook Temp Humidity Windy Play Golf

0 0 Hot High False No

1 0 Hot High True No

2 1 Hot High False Yes

3 2 Mild High False Yes

4 2 Cool Normal False Yes

5 2 Cool Normal True No

6 1 Cool Normal True Yes

7 0 Mild High False No

8 0 Cool Normal False Yes

9 2 Mild Normal False Yes

10 0 Mild Normal True Yes

11 1 Mild High True Yes

12 1 Hot Normal False Yes

13 2 Mild High True No

**Another Method**

import pandas as pd

import numpy as np

from sklearn import preprocessing

data= pd.read\_csv('golf-dataset.csv')

data

label\_encoder = preprocessing.LabelEncoder()

data['Outlook']= label\_encoder.fit\_transform(data['Outlook'])

data['Temp']= label\_encoder.fit\_transform(data['Temp'])

data['Humidity']= label\_encoder.fit\_transform(data['Humidity'])

data['Windy']= label\_encoder.fit\_transform(data['Windy'])

data['Play Golf']= label\_encoder.fit\_transform(data['Play Golf'])

data

**Output**

|  | **Outlook** | **Temp** | **Humidity** | **Windy** | **Play Golf** |
| --- | --- | --- | --- | --- | --- |
| **0** | 1 | 1 | 0 | 0 | 0 |
| **1** | 1 | 1 | 0 | 1 | 0 |
| **2** | 0 | 1 | 0 | 0 | 1 |
| **3** | 2 | 2 | 0 | 0 | 1 |
| **4** | 2 | 0 | 1 | 0 | 1 |
| **5** | 2 | 0 | 1 | 1 | 0 |
| **6** | 0 | 0 | 1 | 1 | 1 |
| **7** | 1 | 2 | 0 | 0 | 0 |
| **8** | 1 | 0 | 1 | 0 | 1 |
| **9** | 2 | 2 | 1 | 0 | 1 |
| **10** | 1 | 2 | 1 | 1 | 1 |
| **11** | 0 | 2 | 0 | 1 | 1 |
| **12** | 0 | 1 | 1 | 0 | 1 |
| **13** | 2 | 2 | 0 | 1 | 0 |