Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To obtain the appropriate class for the given features using Naïve Bayes algorithm**	
Experiment No: 06	**Date:**	**Enrollment No:92000133018**

Aim: **To obtain the appropriate class for the given features using Naïve Bayes algorithm**

IDE: Google Colab

Theory:

Machine learning is a method of data analysis that automates analytical model building of data set. Using the implemented algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look. Naive bayes algorithm is one of the most popular machines learning technique.

Conditional Probability is just what is the probability that something will happen, given that something else has already happened. Let say we have a collection of people. Some of them are singers. They are either male or female. If we select a random sample, what is the probability that this person is a male? what is the probability that this person is a male and singer? Conditional Probability is the best option here.

We can calculate probability like,
P(Singer & Male) = P(Male) x P(Singer / Male)

What is Bayes rule ?
We can simply define Bayes rule like this. Let A1, A2, …, An be a set of mutually exclusive events that together form the sample space S. Let B be any event from the same sample space, such that P(B) > 0. Then, P(Ak | B) = P(Ak ∩ B) / P(A1 ∩ B) + P(A2 ∩ B) + . . . + P(An ∩ B)

What is Bayes classifier?
Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features in machine learning. Basically, we can use above theories and equations for classification problem.

Methodology:

1. Load the basic libraries and packages
2. Load the dataset
3. Analyse the dataset
4. Normalize the data
5. Pre-process the data
6. Visualize the Data
7. Separate the training and testing data
8. Apply the Bernoulli Naïve Bayes algorithm
9. Predict the testing dataset
10. Obtain the confusion matrix
11. Obtain the accuracy score

Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To obtain the appropriate class for the given features using Naïve Bayes algorithm**	
Experiment No: 06	**Date:**	**Enrollment No:92000133018**

12. Visualize the classified dataset
13. Apply the Gaussian Naïve Bayes algorithm
14. Predict the testing dataset
15. Obtain the confusion matrix
16. Obtain the accuracy score
17. Visualize the classified dataset

Program (Code):

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

dataset = pd.read_csv('/content/naive.csv')

dataset.head()

dataset.describe()

dataset.shape

X = dataset.iloc[:, [0,1]].values
y = dataset.iloc[:, 2].values

X

y

#Training and Testing Data (divide the data into two part)
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test =train_test_split(X,y,test_size=0.25, random_state=0)

from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.fit_transform(X_test)

from sklearn.naive_bayes import BernoulliNB
classifer=BernoulliNB()
```

```
classifer.fit(X_train,y_train)

y_pred = classifer.predict(X_test)

from sklearn.metrics import confusion_matrix
cm =confusion_matrix(y_test, y_pred)
print(cm)

from sklearn.metrics import accuracy_score
accuracy_score(y_test, y_pred)

# Visualising the Training set results
from matplotlib.colors import ListedColormap
X_set, y_set = X_train, y_train
X1, X2 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                     np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X2, classifer.predict(np.array([X1.ravel(), X2.ravel()]).T).reshape(X1.shape),
             alpha = 0.75, cmap = ListedColormap(('red', 'green')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
for i, j in enumerate(np.unique(y_set)):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(('blue', 'yellow'))(i), label = j)
plt.title('Classifier NBC (Training set)')
plt.xlabel('Age')
plt.ylabel('Estimated Salary')
plt.legend()
plt.show()

# Visualising the Test set results
from matplotlib.colors import ListedColormap
X_set, y_set = X_test, y_test
X1, X2 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                     np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X2, classifer.predict(np.array([X1.ravel(), X2.ravel()]).T).reshape(X1.shape),
             alpha = 0.75, cmap = ListedColormap(('red', 'green')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
for i, j in enumerate(np.unique(y_set)):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(('blue', 'yellow'))(i), label = j)
plt.title('Classifier NBC(Test set)')
plt.xlabel('Age')
```

```
plt.ylabel('Estimated Salary')
plt.legend()
plt.show()

from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(X_train, y_train)

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

accuracy_score(y_test, y_pred)
```

Results:

To be attached with

a. Classified dataset using Bernoulli Naïve Bayes

```
from sklearn.naive_bayes import BernoulliNB
classifer=BernoulliNB()
classifer.fit(X_train,y_train)
```

```
BernoulliNB()
```

```
[12] y_pred = classifer.predict(X_test)
```

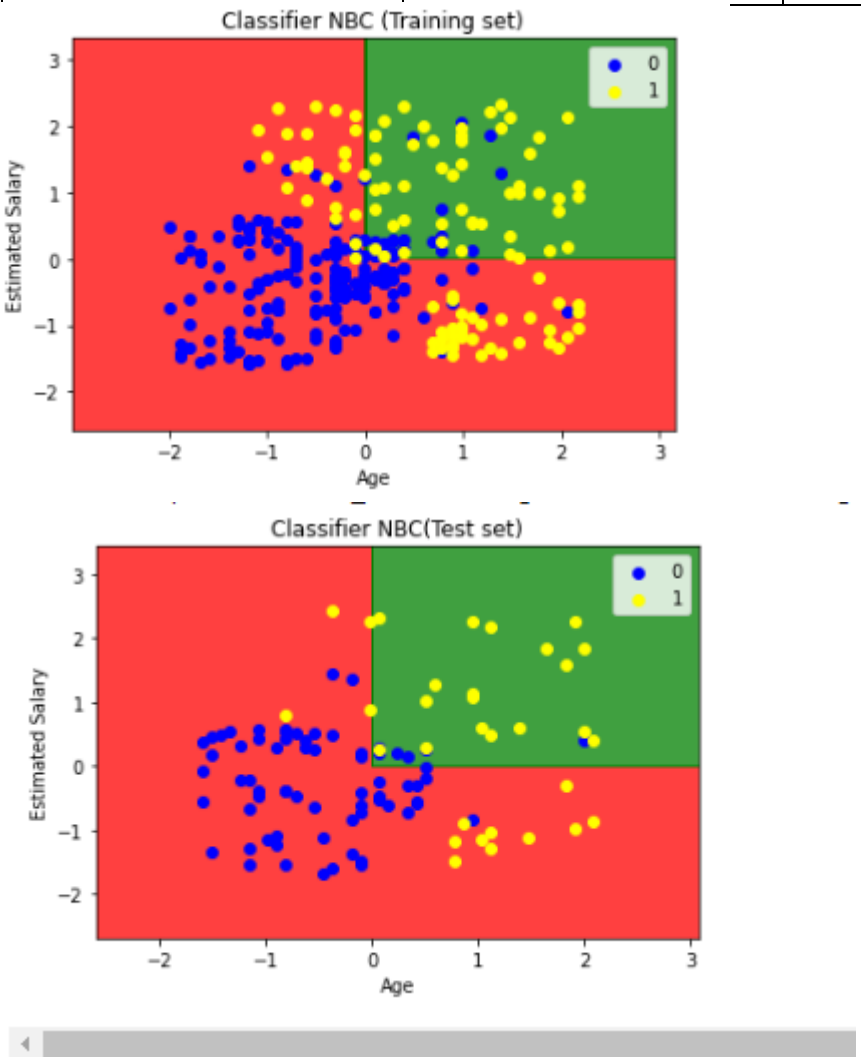
```
[13] from sklearn.metrics import confusion_matrix
     cm =confusion_matrix(y_test, y_pred)
     print(cm)

     [[62  6]
      [14 18]]
```

```
[14] from sklearn.metrics import accuracy_score
     accuracy_score(y_test, y_pred)

     0.8
```

Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Machine Learning (01CT0519)	**Aim: To obtain the appropriate class for the given features using Naïve Bayes algorithm**	
Experiment No: 06	**Date:**	**Enrollment No:92000133018**



b. Classified dataset using Gaussian Naïve Bayes

Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To obtain the appropriate class for the given features using Naïve Bayes algorithm**	
Experiment No: 06	**Date:**	**Enrollment No:92000133018**

```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(X_train, y_train)
```

```
GaussianNB()
```

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

```
[[62  6]
 [14 18]]
```

```
accuracy_score(y_test, y_pred)
```

```
0.8
```

Observation and Result Analysis:

a. Nature of the dataset

b. During Training Process

c. After the training Process

Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To obtain the appropriate class for the given features using Naïve Bayes algorithm**	
Experiment No: 06	**Date:**	**Enrollment No:92000133018**

d. Observation over the classification using both the approaches

Post Lab Exercise:

a. Which are the types of Naïve Bayes algorithms?

b. Which are the limitations of the Naïve Bayes algorithm?

c. Give three different applications of Naïve Bayes algorithm.

d. Explain Bernoulli Naïve Bayes

e. Explain Gaussian Naïve Bayes

Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To obtain the appropriate class for the given features using Naïve Bayes algorithm**	
Experiment No: 06	**Date:**	**Enrollment No:92000133018**

f. Explain Multinomial Naïve Bayes

Post Lab Activity:

Consider any dataset from **https://archive.ics.uci.edu/ml/datasets.php** and perform the classification of various classes using Naïve Bayes algorithm. Make sure that the dataset is not matching with your classmates. You can also select the dataset from other ML repositories with prior permission from your concerned subject faculty.