# ASSIGNMENT 2

Name : Abhishek Badgujar

Class : BE - A

Roll no : 49

Problem Statement : Download Pima Indians Diabetes dataset. Use Naive Bayes" Algorithm for classification

In [13]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib as plt
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
```

In [2]:

```
df_read = pd.read_csv("diabetes.csv")
```

In [4]:

```
df_read.head()
```

Out[4]:

| | Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction |
|---|---|---|---|---|---|---|---|
| **0** | 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 |
| **1** | 1 | 85 | 66 | 29 | 0 | 26.6 | 0.35 |
| **2** | 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 |
| **3** | 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 |
| **4** | 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 |

In [6]:

```
print(df_read.dtypes)
```

```
Pregnancies int64
Glucose int64
BloodPressure int64
SkinThickness int64
Insulin int64
BMI float64
DiabetesPedigreeFunction float64
Age int64 Outcome
int64 dtype:
object In
[8]:
```

```
print(df_read.isnull().sum())
```

```
Pregnancies 0
Glucose 0
BloodPressure 0
SkinThickness 0
Insulin 0
BMI 0
DiabetesPedigreeFunction 0
Age 0 Outcome
0 dtype: int64
```

# Load the data from CSV file and split it into training and test datasets.

In [12]:

```
X = df_read.drop("Outcome",axis = 1) Y
= df_read['Outcome']
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size = 0.2,random_state = 0)
```

# Summarize the properties in the training dataset so that

```
array([0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0,
 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0,   1,
 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0,   0, 0,
 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1,   1, 1, 0,
 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0,   0, 1, 0, 0,
 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1,  0, 1, 0, 0, 0,
 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0],
 dtype=int64)
```

# Classify samples from a test dataset and a summarized training dataset.

In [18]:

[0]

In [ ]: