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

```
%matplotlib inline
%matplotlib inline
from sklearn.model_selection import train_test_split,cross_val_score
from sklearn.model_selection import RandomizedSearchCV,GridSearchCV
from sklearn.metrics import confusion_matrix,classification_report
from sklearn.metrics import precision_score,recall_score,f1_score
from sklearn.metrics import roc_curve,auc
```

```
dataset.head_tail("diabetes.csv")
```

```
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  DiabetesPedigreeFunction  Age  Outcome
0            6      148             72             35        0 33.6                     0.627   50        1
1            1       85             66             29        0 26.6                     0.351   31        0
2            8      183             64              0        0 23.3                     0.672   32        1
3            1       89             66             23       94 28.1                     0.167   21        0
4            0      137             40             35      168 43.1                     2.288   33        1
...        ...      ...            ...            ...      ...  ...                       ...  ...      ...
762          9      171             78             40      540 34.5                     0.573   51        0
763          7      100             76              0        0 30.8                     0.270   26        0
764          1      121             78             39        0 39.0                     0.261   28        0
765          1      108             60             46      180 35.5                     0.415   24        0
767          1       93             70             31        0 30.4                     0.315   23        0

700 rows * 9 columns.
```

```
data.info, data.describe for grade
data.grade
```

```
<bound method pandas.core.frame.DataFrame.describe of       Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  DiabetesPedigreeFunction  Age  Outcome
0              6      148             72             35        0 33.6                     0.627   50        1
1              1       85             66             29        0 26.6                     0.351   31        0
2              8      183             64              0        0 23.3                     0.672   32        1
3              1       89             66             23       94 28.1                     0.167   21        0
4              0      137             40             35      168 43.1                     2.288   33        1
...          ...      ...            ...            ...      ...  ...                       ...  ...      ...
762            9      171             78             40      540 34.5                     0.573   51        0
763            7      100             76              0        0 30.8                     0.270   26        0
764            1      121             78             39        0 39.0                     0.261   28        0
765            1      108             60             46      180 35.5                     0.415   24        0
767            1       93             70             31        0 30.4                     0.315   23        0
```

```
<information about the dataset
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 700 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               700 non-null    int64
 1   Glucose                   700 non-null    int64
 2   BloodPressure             700 non-null    int64
 3   SkinThickness             700 non-null    int64
 4   Insulin                   700 non-null    int64
 5   BMI                       700 non-null    float64
 6   DiabetesPedigreeFunction  700 non-null    float64
 7   Age                       700 non-null    int64
 8   Outcome                   700 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 54.7 KB
```

```
<shape of the project in rows and columns
data.shape
```

```
(700, 9)
```

```
<first five rows data in the data set
data.head()
```

```
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  DiabetesPedigreeFunction  Age  Outcome
0            6      148             72             35        0 33.6                     0.627   50        1
1            1       85             66             29        0 26.6                     0.351   31        0
2            8      183             64              0        0 23.3                     0.672   32        1
3            1       89             66             23       94 28.1                     0.167   21        0
4            0      137             40             35      168 43.1                     2.288   33        1
```

```
<last five rows in the data set
data.tail()
```

```
     Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  DiabetesPedigreeFunction  Age  Outcome
762            9      171             78             40      540 34.5                     0.573   51        0
763            7      100             76              0        0 30.8                     0.270   26        0
764            1      121             78             39        0 39.0                     0.261   28        0
765            1      108             60             46      180 35.5                     0.415   24        0
767            1       93             70             31        0 30.4                     0.315   23        0
```

```
<unique values
data["Outcome"].value_counts()
```

```
Outcome
0    500
1    200
Name: count, dtype: int64
```

```
<of all unique values in the data set
data.nunique()
```

```
Pregnancies                  17
Glucose                     135
BloodPressure               ...
SkinThickness               ...
Insulin                     ...
BMI                         ...
DiabetesPedigreeFunction    ...
Age                         ...
Outcome                       2
dtype: int64
```

```
<of null values
data.isnull()
```

```
     Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  DiabetesPedigreeFunction  Age  Outcome
0          False    False          False          False    False  ...                     ...  ...      ...
1          False    False          False          False    False  ...                     ...  ...      ...
2          False    False          False          False    False  ...                     ...  ...      ...
3          False    False          False          False    False  ...                     ...  ...      ...
4          False    False          False          False    False  ...                     ...  ...      ...
...          ...      ...            ...            ...      ...  ...                     ...  ...      ...
762        False    False          False          False    False  ...                     ...  ...      ...
763        False    False          False          False    False  ...                     ...  ...      ...
764        False    False          False          False    False  ...                     ...  ...      ...
765        False    False          False          False    False  ...                     ...  ...      ...
767        False    False          False          False    False  ...                     ...  ...      ...
```

```
<count of null values
data.isnull().sum()
```

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

```
<count of the project in rows and columns
data.shape
```

```
(700, 9)
```

```
<first five rows in the data set
data.head()
```

```
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  DiabetesPedigreeFunction  Age  Outcome
0            6      148             72             35        0 33.6                     0.627   50        1
1            1       85             66             29        0 26.6                     0.351   31        0
2            8      183             64              0        0 23.3                     0.672   32        1
3            1       89             66             23       94 28.1                     0.167   21        0
4            0      137             40             35      168 43.1                     2.288   33        1
```

```
<last five rows in the data set
data.tail()
```

```
     Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  DiabetesPedigreeFunction  Age  Outcome
762            9      171             78             40      540 34.5                     0.573   51        0
763            7      100             76              0        0 30.8                     0.270   26        0
764            1      121             78             39        0 39.0                     0.261   28        0
765            1      108             60             46      180 35.5                     0.415   24        0
767            1       93             70             31        0 30.4                     0.315   23        0
```

```
<unique values
data["Outcome"].value_counts()
```

```
Outcome
0    500
1    200
Name: count, dtype: int64
```

```
<of all unique values in the data set
data.nunique()
```

```
Pregnancies                  17
Glucose                     135
BloodPressure               ...
SkinThickness               ...
Insulin                     ...
BMI                         ...
DiabetesPedigreeFunction    ...
Age                         ...
Outcome                       2
dtype: int64
```

```
<of null values
data.isnull()
```

```
     Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  DiabetesPedigreeFunction  Age  Outcome
0          False    False          False          False    False  ...                     ...  ...      ...
1          False    False          False          False    False  ...                     ...  ...      ...
2          False    False          False          False    False  ...                     ...  ...      ...
3          False    False          False          False    False  ...                     ...  ...      ...
4          False    False          False          False    False  ...                     ...  ...      ...
...          ...      ...            ...            ...      ...  ...                     ...  ...      ...
762        False    False          False          False    False  ...                     ...  ...      ...
763        False    False          False          False    False  ...                     ...  ...      ...
764        False    False          False          False    False  ...                     ...  ...      ...
765        False    False          False          False    False  ...                     ...  ...      ...
767        False    False          False          False    False  ...                     ...  ...      ...
```

```
<count of null values
data.isnull().sum()
```

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

```
<count of the project in rows and columns
data.shape
```

```
(700, 9)
```

```
<information about the dataset
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 700 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               700 non-null    int64
 1   Glucose                   700 non-null    int64
 2   BloodPressure             700 non-null    int64
 3   SkinThickness             700 non-null    int64
 4   Insulin                   700 non-null    int64
 5   BMI                       700 non-null    float64
 6   DiabetesPedigreeFunction  700 non-null    float64
 7   Age                       700 non-null    int64
 8   Outcome                   700 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 54.7 KB
```

```
<unique values
data.nunique()
```

```
Pregnancies                  17
Glucose                     135
BloodPressure               ...
SkinThickness               ...
Insulin                     ...
BMI                         ...
DiabetesPedigreeFunction    ...
Age                         ...
Outcome                       2
dtype: int64
```

```
<count of null values
data.isnull().sum()
```

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

```
<count of the project in rows and columns
data.shape
```

```
(700, 9)
```

```
<first five rows data
```