Pima Indians diabetes XGB

Pima Indians are a native american tribe in Arizona, United states who are susceptible to diabetes. We are going to use Ensemble learning technique XGB or Extreme gradient Boosting to predict which test subjects are more susceptible to diabetes.

In [1]:
```
#importing packages
import pandas as pd
import numpy as np


from sklearn.model_selection import train_test_split
from sklearn import model_selection
from sklearn.ensemble import AdaBoostClassifier
from sklearn import svm

from xgboost import XGBClassifier

from sklearn.metrics import accuracy_score, auc, balanced_accuracy_score, confusion_matrix, f1_score, precision_
```

```
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\compat.py:36: FutureWarning: pandas.Int64Index is depre
cated and will be removed from pandas in a future version. Use pandas.Index with the appropriate dtype instead.
  from pandas import MultiIndex, Int64Index
```

In [2]:
```
conda install -c anaconda py-xgboost
```

```
Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... done

# All requested packages already installed.


Note: you may need to restart the kernel to use updated packages.
```

In [3]:
```
#loading the dataset
pimadf=pd.read_csv("C:\\Users\\sujoydutta\\Desktop\\Data analysis\\Python\\Datasets_Py\\Ensemble\\pima-indians-
pimadf.head()
```

Out[3]:

	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

In [9]:
```
#understanding the dataset
pimadf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Pregnancies               768 non-null    int64  
 1   Glucose                   768 non-null    int64  
 2   BloodPressure             768 non-null    int64  
 3   SkinThickness             768 non-null    int64  
 4   Insulin                   768 non-null    int64  
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64  
 8   Outcome                   768 non-null    int64  
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

In [5]:
```
#extracting the values
array= pimadf.values
array
```

Out[5]:
```
array([[  6.   , 148.   ,  72.   , ...,   0.627,  50.   ,   1.   ],
       [  1.   ,  85.   ,  66.   , ...,   0.351,  31.   ,   0.   ],
       [  8.   , 183.   ,  64.   , ...,   0.672,  32.   ,   1.   ],
       ...,
       [  5.   , 121.   ,  72.   , ...,   0.245,  30.   ,   0.   ],
       [  1.   , 126.   ,  60.   , ...,   0.349,  47.   ,   1.   ],
       [  1.   ,  93.   ,  70.   , ...,   0.315,  23.   ,   0.   ]])
```

In [6]:
```
#Splitting the dataset X- independent variables y- dependent variable
X= array[:,0:8]
y= array[:,8]
```

In [7]:
```
#setting parameters
seed=None
num_trees=30
```

In [8]:
```
#trying Adaboost
kfold= model_selection.KFold(n_splits=10, random_state=seed)
model=AdaBoostClassifier(n_estimators=num_trees)
results= model_selection.cross_val_score(model,X,y, cv= kfold)
print(results.mean())
```

```
0.760457963089542
```

In [36]:
```
#trying Xgb
kfold= model_selection.KFold(n_splits=10, random_state=None)
model=XGBClassifier(n_estimators=num_trees)
results= model_selection.cross_val_score(model,X,y, cv= kfold)
print(results.mean())
```

```
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
```

```
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
[20:42:23] WARNING: ..\src\learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric used with
the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval_metric if you'd like
to restore the old behavior.
0.7499487354750513
```

```
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
C:\Users\sujoydutta\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder
in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the followin
g: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y)
as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: