In [29]:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn import metrics
```

In [30]:

```
dfx = pd.read_csv('Diabetes_XTrain.csv')
dfy = pd.read_csv('Diabetes_YTrain.csv')
print(dfx.shape)
print(dfy.shape)
dfx.head(n=5)
```

```
(576, 8)
(576, 1)
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	A
0	7	168	88	42	321	38.2	0.787	4
1	8	110	76	0	0	27.8	0.237	5
2	7	147	76	0	0	39.4	0.257	4
3	2	100	66	20	90	32.9	0.867	2
4	4	129	86	20	270	35.1	0.231	2

In [31]:

```
dfy.head(n=5)
```

	Outcome
0	1
1	0
2	1
3	1
4	0

In [32]:

```
X_Train = dfx.values
Y_Train = dfy.values
Y_Train = Y_Train.reshape((-1))
print(X_Train.shape)
print(Y_Train.shape)
```

```
(576, 8)
(576,)
```

In [33]:

```
dfx = pd.read_csv('Diabetes_XTest.csv')
print(dfx.shape)
dfx.head(n=5)
```

```
(192, 8)
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	A
0	0	180	90	26	90	36.5	0.314	3
1	2	93	64	32	160	38.0	0.674	2
2	2	114	68	22	0	28.7	0.092	2
3	13	76	60	0	0	32.8	0.180	4
4	1	80	74	11	60	30.0	0.527	2

In [34]:

```
X_Test = dfx.values
print(X_Test.shape)
```

```
(192, 8)
```

In [35]:

```
def dist(x1,x2):
    return np.sqrt(sum((x1-x2)**2))

def KNN(X,Y,query_point,K=5):
    vals = []
    m = X.shape[0]

    for i in range(m):
        d = dist(query_point,X[i])
        vals.append((d,Y[i]))

    vals = sorted(vals)
    # Nearest/First K points
    vals = vals[:K]

    vals = np.array(vals)
    #print("Vals: ",vals)
    new_vals = np.unique(vals[:,1],return_counts=True)
    #print(new_vals)

    index = new_vals[1].argmax()
    pred = new_vals[0][index]

    return pred
```

In [36]:

```
pred = KNN(X_Train,Y_Train,X_Test[3])
print(int(pred))
```

```
0
```

In [37]:

```
# Accuracy of KNN
Y_Pred = []
for i in range(X_Test.shape[0]):
    p = int(KNN(X_Train,Y_Train,X_Test[i]))
    Y_Pred.append(p)

#print("Accuracy: ",np.mean(Y_Pred==Y_Tst))
```

In [38]:

```
#Saving File
df = pd.DataFrame(data = Y_Pred, columns=["Outcome"])
df.to_csv('Diabetes_Ytest.csv',index=False)
```

In [39]:

```
len(Y_Pred)
```

```
192
```

In [ ]: