Karanjot Singh

In [6]:

```
import pandas as pd

diabetes_data = pd.read_csv('diabetes_preprocess.csv')
df = pd.DataFrame(diabetes_data)
print("Replacing Missing Values By Mean\n\n",df.fillna(df.mean()))

print("Replacing missing Values By Median\n\n",df.fillna(df.median()))

```

```
Replacing Missing Values By Mean

      Pregnancies  Glucose  BloodPressure  SkinThickness     Insulin   BMI
\
0               6    148.0           72.0      35.000000  105.659898  33.6
1               1     85.0           66.0      29.000000  105.659898  26.6
2               8    183.0           64.0      25.876155  105.659898  23.3
3               1     89.0           66.0      23.000000   94.000000  28.1
4               0    137.0            4.0      35.000000  168.000000  43.1
..            ...      ...            ...            ...         ...   ...
763            10     11.0           76.0      48.000000   18.000000  32.9
764             2    122.0            7.0      27.000000  105.659898  36.8
765             5    121.0           72.0      23.000000  112.000000  26.2
766             1    126.0            6.0      25.876155  105.659898   3.1
767             1     93.0            7.0      31.000000  105.659898   3.4

     DiabetesPedigreeFunction  Age  Outcome
0                       0.627    5        1
1                       0.351   31        0
2                       0.672   32        1
3                       0.167   21        0
4                       2.288   33        1
..                        ...  ...      ...
763                     0.171   63        0
764                     0.340   27        0
765                     0.245    3        0
766                     0.349   47        1
767                     0.315   23        0

[768 rows x 9 columns]
Replacing missing Values By Median

      Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
0               6    148.0           72.0           35.0     71.0  33.6
1               1     85.0           66.0           29.0     71.0  26.6
2               8    183.0           64.0           27.0     71.0  23.3
3               1     89.0           66.0           23.0     94.0  28.1
4               0    137.0            4.0           35.0    168.0  43.1
..            ...      ...            ...            ...      ...   ...
763            10     11.0           76.0           48.0     18.0  32.9
764             2    122.0            7.0           27.0     71.0  36.8
765             5    121.0           72.0           23.0    112.0  26.2
766             1    126.0            6.0           27.0     71.0   3.1
767             1     93.0            7.0           31.0     71.0   3.4

     DiabetesPedigreeFunction  Age  Outcome
0                       0.627    5        1
1                       0.351   31        0
2                       0.672   32        1
3                       0.167   21        0
```

```
4                    2.288   33        1
..                     ...   ...     ...
763                  0.171   63        0
764                  0.340   27        0
765                  0.245    3        0
766                  0.349   47        1
767                  0.315   23        0

[768 rows x 9 columns]
```

In [71]:

```
import pandas as pd
import numpy as np
from sklearn import preprocessing
bank_data = pd.read_csv('bank.csv',delimiter =";")
df = pd.DataFrame(bank_data)
le = preprocessing.LabelEncoder()
df1 = df[['job','marital','education','default','housing','loan','contact','month','po
#print(df1)
ND = preprocessing.MinMaxScaler()
df2 = ND.fit_transform(df[['balance','pdays','duration']])
print(df2.round(2))
print("Saving to a File")
np.savetxt('preprocesses.csv',df2,delimiter =' ')

```

```
[[0.07 0.   0.02]
 [0.11 0.39 0.07]
 [0.06 0.38 0.06]
 ...
 [0.05 0.   0.05]
 [0.06 0.24 0.04]
 [0.06 0.29 0.11]]
Saving to a File
```