Name:P.Asha Belcilda

Rollno:225229104

Predictive Analytics for Hospitals

Step 1.Import dataset

In [23]:
```
import pandas as pd
```

In [24]:
```
data=pd.read_csv("diabetes.csv")
print(data)
```

```
     Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
0              6      148             72             35        0  33.6   
1              1       85             66             29        0  26.6   
2              8      183             64              0        0  23.3   
3              1       89             66             23       94  28.1   
4              0      137             40             35      168  43.1   
..           ...      ...            ...            ...      ...   ...   
763           10      101             76             48      180  32.9   
764            2      122             70             27        0  36.8   
765            5      121             72             23      112  26.2   
766            1      126             60              0        0  30.1   
767            1       93             70             31        0  30.4   

     DiabetesPedigreeFunction  Age  Outcome  
0                       0.627   50        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   30        0  
766                     0.349   47        1  
767                     0.315   23        0  

[768 rows x 9 columns]
```

In [25]:
```
#head
data.head()
```

Out[25]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	A
0	6	148	72	35	0	33.6	0.627	
1	1	85	66	29	0	26.6	0.351	
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 [26]:
```
#shape
data.shape
```

Out[26]:
```
(768, 9)
```

In [27]:
```
#columns
data.columns
```

Out[27]:
```
Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'],
      dtype='object')
```

In [28]:
```
#dtype
data.dtypes
```

Out[28]:
```
Pregnancies                 int64
Glucose                     int64
BloodPressure               int64
SkinThickness               int64
Insulin                     int64
BMI                       float64
DiabetesPedigreeFunction  float64
Age                         int64
Outcome                     int64
dtype: object
```

In [29]:
```
#info
data.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 [30]:
```
# value_counts
data.BloodPressure.value_counts()
```

Out[30]:
```
70     57
74     52
78     45
68     45
72     44
64     43
80     40
76     39
60     37
0      35
62     34
66     30
82     30
88     25
84     23
90     22
86     21
58     21
50     13
56     12
52     11
54     11
75      8
92      8
65      7
85      6
94      6
48      5
96      4
44      4
100     3
106     3
98      3
110     3
55      2
108     2
104     2
46      2
30      2
122     1
95      1
102     1
61      1
24      1
38      1
40      1
114     1
Name: BloodPressure, dtype: int64
```

Step 2.Identify relationship between feature

In [31]:
```
data.style.background_gradient(cmap ='Blues')
```

Out[31]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigree
0	6	148	72	35	0	33.600000	
1	1	85	66	29	0	26.600000	
2	8	183	64	0	0	23.300000	
3	1	89	66	23	94	28.100000	
4	0	137	40	35	168	43.100000	
5	5	116	74	0	0	25.600000	
6	3	78	50	32	88	31.000000	
7	10	115	0	0	0	35.300000	
8	2	197	70	45	543	30.500000	
9	8	125	96	0	0	0.000000	

Step 3.Prediction using one feature

In [32]:
```
X = data[['Age']]
Y = data[['Outcome']]
```

In [33]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
```

In [34]:
```
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=.25,random_state=4
LOR=LogisticRegression()
LOR.fit(X_train,Y_train)
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Out[34]:
```
LogisticRegression()
```

In [35]:
```
LOR.predict(X_test)
```

Out[35]:
```
array([0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0,
       0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0,
       0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0], dtype=int64)
```

In [36]:
```
print("coef_ : ",LOR.coef_)
print("intercept_: ",LOR.intercept_)
```

```
coef_ :  [[0.05221912]]
intercept_:  [-2.39506398]
```

In [37]:
```
LOR.predict([[60]])
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X
does not have valid feature names, but LogisticRegression was fitted with featu
re names
  warnings.warn(
```

Out[37]:
```
array([1], dtype=int64)
```

In [38]:
```
lrf=LOR.coef_ * 60 + LOR.intercept_
from scipy.special import expit
d = expit(lrf)
d
```

Out[38]:
```
array([[0.67657656]])
```

In [39]:
```
if d > 0.5:
    print('Yes, he will become diabetic ')
else:
    print('No, he will not be diabetic')
```

```
Yes, he will become diabetic
```

Step-4 Prediction using many features

In [41]:
```
X1=data[['Glucose','BMI','Age']]
X1_train,X1_test,y1_train,y1_test = train_test_split(X1,Y,random_state=42,test_si
LOR1 = LogisticRegression()
LOR1.fit(X1_train,y1_train)
LOR1.predict(X1_test)
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Out[41]:
```
array([0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,
       0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0,
       0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1,
       0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0,
       0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1,
       0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0,
       0, 0, 0, 0, 1, 1, 0, 1, 1], dtype=int64)
```

In [42]:
```
print("coef_ : ",LOR1.coef_)
print("intercept_: ",LOR1.intercept_)
```

```
coef_ :  [[0.03292234 0.09635698 0.04398021]]
intercept_:  [-9.39683405]
```

In [43]:
```
lrf1=LOR1.coef_ * 150 * 30 * 40+ LOR1.intercept_
from scipy.special import expit
expit(lrf1)
```

Out[43]:
```
array([[1., 1., 1.]])
```

In [44]:
```
LOR1.predict([[150,30,40]])
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X
does not have valid feature names, but LogisticRegression was fitted with featu
re names
  warnings.warn(
```

Out[44]:
```
array([1], dtype=int64)
```

In [45]:
```
LOR1.predict_proba([[150,30,40]])
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X
does not have valid feature names, but LogisticRegression was fitted with featu
re names
  warnings.warn(
```

Out[45]:
```
array([[0.45228691, 0.54771309]])
```

Step 5.Build LoR model with all features

In [47]:
```
X2=data.drop(['Outcome'],axis=1)
X2_train,X2_test,y2_train,y2_test = train_test_split(X2,Y,test_size=.23,random_st
LOR2=LogisticRegression()
LOR2.fit(X2_train,y2_train)
LOR2.predict(X2_test)
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:81
4: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-
learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regressi
on (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regressi
on)
  n_iter_i = _check_optimize_result(
```

Out[47]:
```
array([1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0,
       1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0,
       1], dtype=int64)
```

In [48]:
```
y2_pred = LOR2.predict(X2_test)
y2_pred
```

Out[48]:
```
array([1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0,
       1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0,
       1], dtype=int64)
```

In [49]:
```
from sklearn.metrics import roc_auc_score
lor_auc = roc_auc_score(y2_test,y2_pred)
print("Auc: ",lor_auc)
```

```
Auc:  0.6670329670329671
```

Step 6.Forward Selected Procedure

In [56]:
```
def get_auc(var,tar,data):
    fx=data[var]
    fy=data[tar]
    LOR4=LogisticRegression()
    LOR4.fit(fx,fy)
    pred=LOR4.predict_proba(fx)[:,1]
    auc_val = roc_auc_score(Y,pred)
    return auc_val
get_auc(['Glucose','BMI'],['Outcome'],data)
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Out[56]:
```
0.8109328358208956
```

In [57]:
```
get_auc(['Pregnancies','BloodPressure','SkinThickness'],['Outcome'],data)
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Out[57]:
```
0.6444962686567164
```

In [58]:
```
def best_next(current,cand,tar,data):
    best_auc=-1
    best_var=None
    for i in cand:
        auc_v = get_auc(current+[i],tar,data)
        if auc_v>=best_auc:
            best_auc=auc_v
            best_var=i
    return best_var
```

In [59]:
```
current=['Insulin','BMI','DiabetesPedigreeFunction','Age']
cand=['Pregnancies','Glucose','BloodPressure','SkinThickness']
tar=['Outcome']
next_var = best_next(current,cand,tar,data)
next_var
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Out[59]: 'Glucose'

In [60]:
```
tar =['Outcome']
current=[]
cand=['Pregnancies','Glucose','BloodPressure','SkinThickness','Insulin','BMI','Di
max_num = 7
num_it = min(max_num,len(cand))
for i in range(0,num_it):
    next_var = best_next(current,cand,tar,data)
    current += [next_var]
    cand.remove(next_var)
    print("variable add in step "+str(i+1)+' is '+ next_var +" .")
```

```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
```

In [61]:
```
print(current)
```

```
['Glucose', 'BMI', 'Pregnancies', 'DiabetesPedigreeFunction', 'BloodPressure',
'Age', 'SkinThickness']
```

Step-7 Plot line graph of AUC values and select cut-off

In [82]:
```
X2_train,X2_test,y2_train,y2_test = train_test_split(X2,Y,stratify=Y,test_size=.5
```

In [83]:
```
prediction=LOR2.predict_proba(X2_test)
```

In [84]:
```
train = pd.concat([X2_train,y2_train],axis =1)
test = pd.concat([X2_test,y2_test],axis =1)
def auc_train_test (variables,target, train, test):
    X_train = train[variables]
    X_test = test[variables]
    Y_train =train[target]
    Y_test = test[target]
    Lor=LogisticRegression()
    Lor.fit(X_train,Y_train)
    prediction_train = Lor.predict_proba(X_train)[:,1]
    prediction_test = Lor.predict_proba(X_test)[:,1]
    auc_train = roc_auc_score(Y_train, prediction_train)
    auc_test = roc_auc_score(Y_train,prediction_test)
    return (auc_train,auc_test)
auc_values_train=[]
auc_values_test=[]
variable_evaluate=[]
for v in X2.columns:
        variable_evaluate.append(v)
        auc_train,auc_test = auc_train_test(variable_evaluate,['Outcome'],train,t
        auc_values_train.append(auc_train)
        auc_values_test.append(auc_test)
```

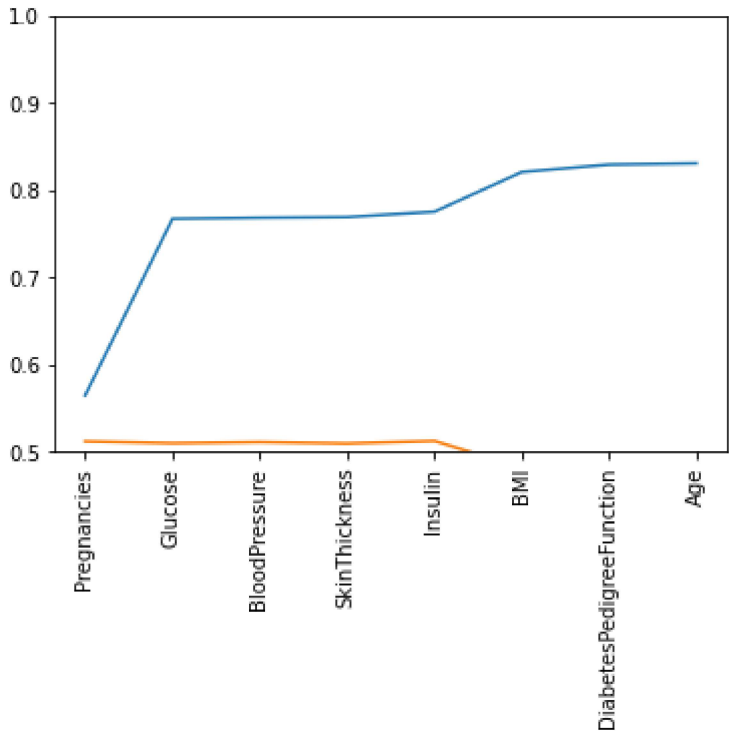
```
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:81
4: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
```

```
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-
learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regressi
on (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regressi
on)
  n_iter_i = _check_optimize_result(
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\utils\validation.py:993: Dat
aConversionWarning: A column-vector y was passed when a 1d array was expected.
 Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\ashac\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:81
4: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-
learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regressi
on (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regressi
on)
  n_iter_i = _check_optimize_result(
```

In [87]:
```
import matplotlib.pyplot as plt
import numpy as np
x =np.array(range(0,len(auc_values_train)))
my_train = np.array(auc_values_train)
my_test = np.array(auc_values_test)
plt.xticks(x,X2.columns,rotation=90)
plt.plot(x,my_train)
plt.plot(x,my_test)
plt.ylim(0.5,1)
plt.show()
```

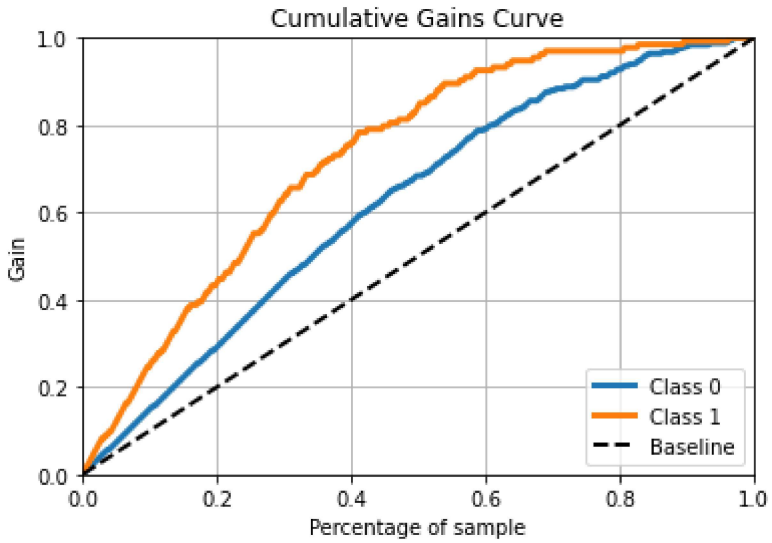


Step 8 Draw cumulative gain chart and lift chart

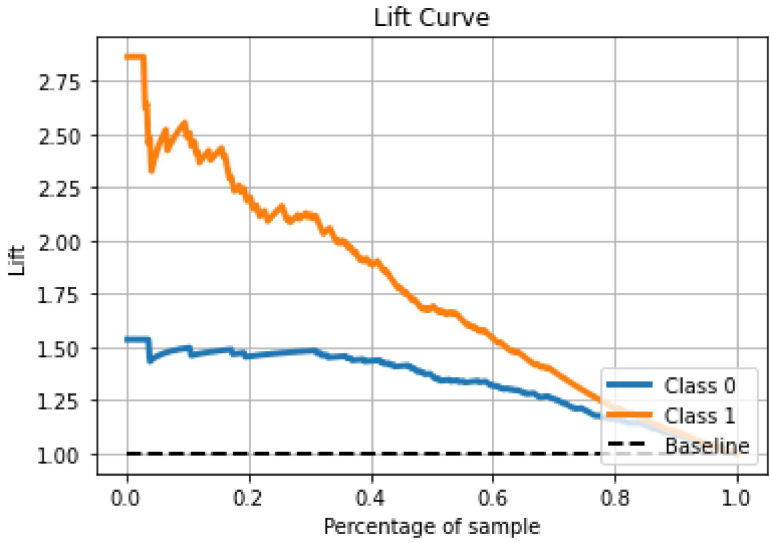
In [88]:
```
pip install scikit.plot
```

```
Requirement already satisfied: scikit.plot in c:\users\ashac\anaconda3\lib\site
-packages (0.3.7)
Requirement already satisfied: matplotlib>=1.4.0 in c:\users\ashac\anaconda3\li
b\site-packages (from scikit.plot) (3.5.1)
Requirement already satisfied: joblib>=0.10 in c:\users\ashac\anaconda3\lib\sit
e-packages (from scikit.plot) (1.1.0)
Requirement already satisfied: scikit-learn>=0.18 in c:\users\ashac\anaconda3\l
ib\site-packages (from scikit.plot) (1.0.2)
Requirement already satisfied: scipy>=0.9 in c:\users\ashac\anaconda3\lib\site-
packages (from scikit.plot) (1.7.3)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\ashac\anaconda3\li
b\site-packages (from matplotlib>=1.4.0->scikit.plot) (1.3.2)
Requirement already satisfied: numpy>=1.17 in c:\users\ashac\anaconda3\lib\site
-packages (from matplotlib>=1.4.0->scikit.plot) (1.21.5)
Requirement already satisfied: pillow>=6.2.0 in c:\users\ashac\anaconda3\lib\si
te-packages (from matplotlib>=1.4.0->scikit.plot) (9.0.1)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\ashac\anaconda3
\lib\site-packages (from matplotlib>=1.4.0->scikit.plot) (2.8.2)
Requirement already satisfied: packaging>=20.0 in c:\users\ashac\anaconda3\lib
\site-packages (from matplotlib>=1.4.0->scikit.plot) (21.3)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\ashac\anaconda3\lib
\site-packages (from matplotlib>=1.4.0->scikit.plot) (3.0.4)
Requirement already satisfied: cycler>=0.10 in c:\users\ashac\anaconda3\lib\sit
e-packages (from matplotlib>=1.4.0->scikit.plot) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\ashac\anaconda3\li
b\site-packages (from matplotlib>=1.4.0->scikit.plot) (4.25.0)
Requirement already satisfied: six>=1.5 in c:\users\ashac\anaconda3\lib\site-pa
ckages (from python-dateutil>=2.7->matplotlib>=1.4.0->scikit.plot) (1.16.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\ashac\anaconda3
\lib\site-packages (from scikit-learn>=0.18->scikit.plot) (2.2.0)
Note: you may need to restart the kernel to use updated packages.
```

In [89]:
```
import scikitplot as skplt
skplt.metrics.plot_cumulative_gain(y2_test,prediction)
plt.show()
plt.figure(figsize=(7,7))
skplt.metrics.plot_lift_curve(y2_test, prediction)
plt.show()
```




```
<Figure size 504x504 with 0 Axes>
```



In [ ]: