

Import packages

```
In [1]: 1 import pandas as pd
```

Step:1 Import dataset

```
In [2]: 1 df= pd.read_csv("diabetes.csv")
```

In [3]: 1 df

Out[3]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	
0	6	148	72	35	0	33.6		0.6
1	1	85	66	29	0	26.6		0.3
2	8	183	64	0	0	23.3		0.6
3	1	89	66	23	94	28.1		0.1
4	0	137	40	35	168	43.1		2.2
5	5	116	74	0	0	25.6		0.2
6	3	78	50	32	88	31.0		0.2
7	10	115	0	0	0	35.3		0.1
8	2	197	70	45	543	30.5		0.1
9	8	125	96	0	0	0.0		0.2
10	4	110	92	0	0	37.6		0.1
11	10	168	74	0	0	38.0		0.5
12	10	139	80	0	0	27.1		1.4
13	1	189	60	23	846	30.1		0.3
14	5	166	72	19	175	25.8		0.5
15	7	100	0	0	0	30.0		0.4
16	0	118	84	47	230	45.8		0.5
17	7	107	74	0	0	29.6		0.2
18	1	103	30	38	83	43.3		0.1
19	1	115	70	30	96	34.6		0.5
20	3	126	88	41	235	39.3		0.7
21	8	99	84	0	0	35.4		0.3
22	7	196	90	0	0	39.8		0.4
23	9	119	80	35	0	29.0		0.2
24	11	143	94	33	146	36.6		0.2
25	10	125	70	26	115	31.1		0.2
26	7	147	76	0	0	39.4		0.2
27	1	97	66	15	140	23.2		0.4
28	13	145	82	19	110	22.2		0.2
29	5	117	92	0	0	34.1		0.3
...
738	2	99	60	17	160	36.6		0.4
739	1	102	74	0	0	39.5		0.2
740	11	120	80	37	150	42.3		0.7

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	
741	3	102	44	20	94	30.8	0.4	
742	1	109	58	18	116	28.5	0.2	
743	9	140	94	0	0	32.7	0.7	
744	13	153	88	37	140	40.6	1.1	
745	12	100	84	33	105	30.0	0.4	
746	1	147	94	41	0	49.3	0.3	
747	1	81	74	41	57	46.3	1.0	
748	3	187	70	22	200	36.4	0.4	
749	6	162	62	0	0	24.3	0.1	
750	4	136	70	0	0	31.2	1.1	
751	1	121	78	39	74	39.0	0.2	
752	3	108	62	24	0	26.0	0.2	
753	0	181	88	44	510	43.3	0.2	
754	8	154	78	32	0	32.4	0.4	
755	1	128	88	39	110	36.5	1.0	
756	7	137	90	41	0	32.0	0.3	
757	0	123	72	0	0	36.3	0.2	
758	1	106	76	0	0	37.5	0.1	
759	6	190	92	0	0	35.5	0.2	
760	2	88	58	26	16	28.4	0.7	
761	9	170	74	31	0	44.0	0.4	
762	9	89	62	0	0	22.5	0.1	
763	10	101	76	48	180	32.9	0.1	
764	2	122	70	27	0	36.8	0.3	
765	5	121	72	23	112	26.2	0.2	
766	1	126	60	0	0	30.1	0.3	
767	1	93	70	31	0	30.4	0.3	

768 rows × 9 columns

In [4]: 1 df.head()

Out[4]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
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 [5]: 1 df.tail()

Out[5]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
763	10	101	76	48	180	32.9		0.171
764	2	122	70	27	0	36.8		0.340
765	5	121	72	23	112	26.2		0.245
766	1	126	60	0	0	30.1		0.349
767	1	93	70	31	0	30.4		0.315



In [6]: 1 df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
Pregnancies          768 non-null int64
Glucose              768 non-null int64
BloodPressure        768 non-null int64
SkinThickness        768 non-null int64
Insulin              768 non-null int64
BMI                  768 non-null float64
DiabetesPedigreeFunction 768 non-null float64
Age                  768 non-null int64
Outcome              768 non-null int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

In [7]: 1 df.columns

Out[7]: Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'],
dtype='object')

```
In [8]: 1 type(df)
```

```
Out[8]: pandas.core.frame.DataFrame
```

```
In [9]: 1 df.shape
```

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

```
In [10]: 1 df.info
```

15		7	100		0	0	0	30.0
16		0	118	84	47	230	45.8	
17		7	107	74	0	0	29.6	
18		1	103	30	38	83	43.3	
19		1	115	70	30	96	34.6	
20		3	126	88	41	235	39.3	
21		8	99	84	0	0	35.4	
22		7	196	90	0	0	39.8	
23		9	119	80	35	0	29.0	
24		11	143	94	33	146	36.6	
25		10	125	70	26	115	31.1	
26		7	147	76	0	0	39.4	
27		1	97	66	15	140	23.2	
28		13	145	82	19	110	22.2	
29		5	117	92	0	0	34.1	
..
738		2	99	60	17	160	36.6	
739		1	102	74	0	0	39.5	
740		11	120	80	37	150	42.3	
741		3	102	44	20	94	30.8	

In [11]: 1 df.isnull()

Out[11]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False	False
10	False	False	False	False	False	False	False	False
11	False	False	False	False	False	False	False	False
12	False	False	False	False	False	False	False	False
13	False	False	False	False	False	False	False	False
14	False	False	False	False	False	False	False	False
15	False	False	False	False	False	False	False	False
16	False	False	False	False	False	False	False	False
17	False	False	False	False	False	False	False	False
18	False	False	False	False	False	False	False	False
19	False	False	False	False	False	False	False	False
20	False	False	False	False	False	False	False	False
21	False	False	False	False	False	False	False	False
22	False	False	False	False	False	False	False	False
23	False	False	False	False	False	False	False	False
24	False	False	False	False	False	False	False	False
25	False	False	False	False	False	False	False	False
26	False	False	False	False	False	False	False	False
27	False	False	False	False	False	False	False	False
28	False	False	False	False	False	False	False	False
29	False	False	False	False	False	False	False	False
...
738	False	False	False	False	False	False	False	False
739	False	False	False	False	False	False	False	False
740	False	False	False	False	False	False	False	False

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	Diabetes	PedigreeFunction
741	False	False	False	False	False	False	False	False
742	False	False	False	False	False	False	False	False
743	False	False	False	False	False	False	False	False
744	False	False	False	False	False	False	False	False
745	False	False	False	False	False	False	False	False
746	False	False	False	False	False	False	False	False
747	False	False	False	False	False	False	False	False
748	False	False	False	False	False	False	False	False
749	False	False	False	False	False	False	False	False
750	False	False	False	False	False	False	False	False
751	False	False	False	False	False	False	False	False
752	False	False	False	False	False	False	False	False
753	False	False	False	False	False	False	False	False
754	False	False	False	False	False	False	False	False
755	False	False	False	False	False	False	False	False
756	False	False	False	False	False	False	False	False
757	False	False	False	False	False	False	False	False
758	False	False	False	False	False	False	False	False
759	False	False	False	False	False	False	False	False
760	False	False	False	False	False	False	False	False
761	False	False	False	False	False	False	False	False
762	False	False	False	False	False	False	False	False
763	False	False	False	False	False	False	False	False
764	False	False	False	False	False	False	False	False
765	False	False	False	False	False	False	False	False
766	False	False	False	False	False	False	False	False
767	False	False	False	False	False	False	False	False

768 rows × 9 columns

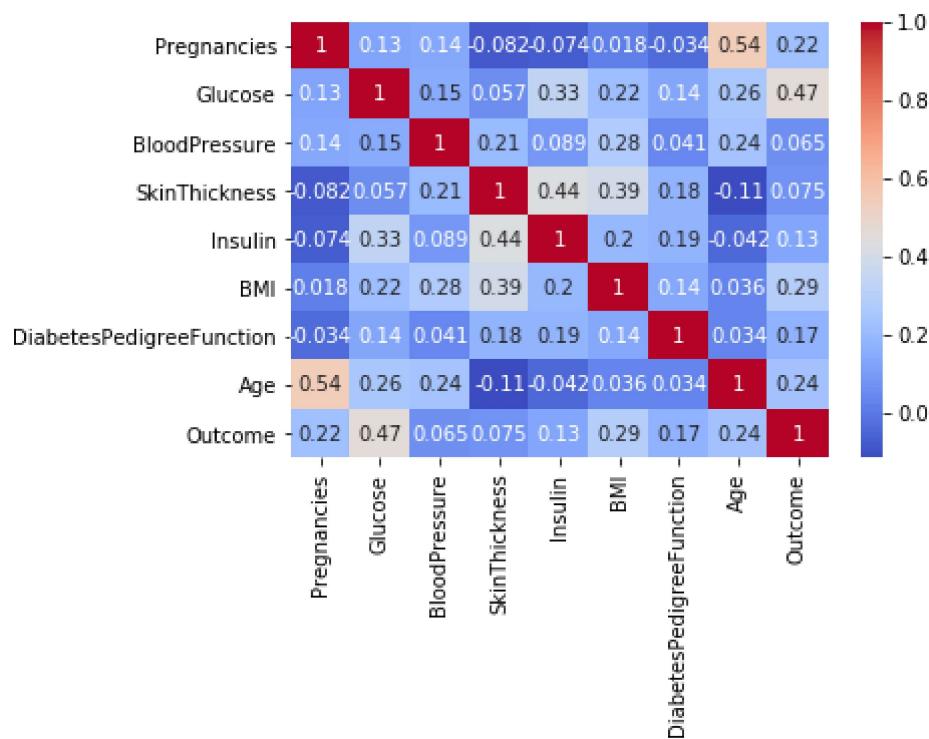
Step:2 Identity relationship between feature

In [17]:

```

1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4
5 # Load the diabetes dataset
6 diabetes = pd.read_csv('diabetes.csv')
7
8 # Create a correlation matrix
9 corr = diabetes.corr()
10
11 # Create a heatmap using seaborn
12 sns.heatmap(corr, annot=True, cmap='coolwarm')
13
14 # Show the plot
15 plt.show()

```



Step 3: Prediction using one feature

In [13]:

```

1 from sklearn.model_selection import train_test_split
2 from sklearn.linear_model import LogisticRegression

```

In [14]:

```

1 X=df[['Age']]
2 y=df[['Outcome']]

```

```
In [15]: 1 X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=.25,random_state=42)
          2 LOR=LogisticRegression()
          3 LOR.fit(X_train,y_train)
```

```
C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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)
```

```
In [16]: 1 LOR.predict(X_test)
```

```
In [18]: 1 print("coef_ : ",LOR.coef_)  
2 print("intercept_ : ",LOR.intercept_)
```

```
coef_ : [[0.04741434]]  
intercept : [-2.213557]
```

```
In [21]: 1 LOR.predict([[60]])
```

Out[21]: array([1], dtype=int64)

```
In [24]: 1 lrf=LOR.coef_*60+LOR.intercept_
          2 from scipy.special import expit
          3 a=expit(lrf)
          4 a
```

Out[24]: array([[0.65278491]])

```
In [27]: 1 if a>.5:  
2     print('Yes he will become diabetic')  
3 else:  
4     print('No, he will not be diabetic')
```

Yes he will become diabetic

Step-4. Prediction using many features

```
In [28]: 1 X1=df[['Glucose','BMI','Age']]
2 y1=df[['Outcome']]
3 X1_train,X1_test,y1_train,y1_test=train_test_split(X1,y1,test_size=.25,random_
4 LOR1=LogisticRegression()
5 LOR1.fit(X1_train,y1_train)
6 LOR1.predict(X1_test)
```

C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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[28]: array([0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0,
0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1,
0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0,
0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
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, 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, 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], dtype=int64)
```

```
In [30]: 1 print("coef_ : ",LOR1.coef_)
2 print("intercept_ : ",LOR1.intercept_)
```

coef_ : [[0.0244408 0.04999901 0.02717807]]
intercept_ : [-6.15377769]

```
In [31]: 1 lrf=LOR1.coef_*150*30*40+LOR1.intercept_
2 from scipy.special import expit
3 expit(lrf)
```

Out[31]: array([[1., 1., 1.]])

```
In [32]: 1 LOR1.predict([[150,30,40]])
```

Out[32]: array([1], dtype=int64)

```
In [33]: 1 LOR1.predict_proba([[150,30,40]])
```

Out[33]: array([[0.47516153, 0.52483847]])

Step5. Build LOR model with all features

```
In [34]: 1 X2=df.drop(['Outcome'],axis=1)
2 y2=df[['Outcome']]
3 X2_train,X2_test,y2_train,y2_test=train_test_split(X2,y2,test_size=.25,random
4 LOR2=LogisticRegression()
5 LOR2.fit(X2_train,y2_train)
6 LOR2.predict(X2_test)
```

C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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[34]: array([0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,
1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0,
0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1,
0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0,
0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1,
0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1,
0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1,
0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0,
0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1,
```

```
In [37]: 1 from sklearn.metrics import roc_auc_score
2 lor_aucy3=roc_auc_score(y2_test,y2_pred)
3 lor_aucy3
```

```
Out[37]: 0.7180982679392012
```

Step6. Forward selection procedure

```
In [40]: 1 def get_auc(var,tar,df):
2     X4=df[var]
3     y4=df[tar]
4     LOR4=LogisticRegression()
5     LOR4.fit(X4,y4)
6     pred=LOR4.predict_proba(X4)[:,1]
7     auc_val=roc_auc_score(y2,pred)
8     return auc_val
9 get_auc(['Glucose','BMI'],['Outcome'],df)
```

C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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[40]: 0.8102835820895523
```

In [41]: 1 get_auc(['Pregnancies','BloodPressure','SkinThickness'],['Outcome'],df)

```
C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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[41]: 0.6420932835820895

In [44]: 1 def best_next(current,cand,tar,df):

```
2     best_auc=-1
3     best_var=None
4     for i in cand:
5         auc_v=get_auc(current+[i],tar,df)
6         if auc_v>=best_auc:
7             best_auc=auc_v
8             best_var=i
9     return best_var
```

In [45]: 1 tar=['Outcome']

```
2 current=['Insulin','BMI','DiabetesPedigreeFunction','Age']
3 cand=['Pregnancies','Glucose','BloodPressure','SkinThickness']
4 next_var=best_next(current,cand,tar,df)
5 print(next_var)
```

Pregnancies

```
C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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)
```

In [48]:

```

1 tar = ["Outcome"]
2 current = []
3 cand =['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness',
4 'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age']
5 max_num = 7
6 num_it = min(max_num, len(cand))
7 for i in range(0, num_it):
8     next_var = best_next(current, cand, tar, df)
9     current += [next_var]
10    cand.remove(next_var)
11    print("Variable added in Step "+str(i+1) +' is ' + next_var +".")

```

Variable added in Step 1 is Pregnancies.
 Variable added in Step 2 is Glucose.
 Variable added in Step 3 is BloodPressure.
 Variable added in Step 4 is SkinThickness.
 Variable added in Step 5 is Insulin.
 Variable added in Step 6 is BMI.
 Variable added in Step 7 is DiabetesPedigreeFunction.

C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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)
```

```
In [47]: 1 print(current)
```

```
['Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age']
```

Step7. PLOT LINE GRAPH OF AUC VALUES AND SELCET CUT-OFF

```
In [50]: 1 X2_train,X2_test,y2_train,y2_test=train_test_split(X2,y2,stratify=y2,test_size=0.2)
```

```
In [51]: 1 prediction1 = LOR2.predict_proba(X2_test)
```

In [52]:

```

1 train = pd.concat([X2_train, y2_train], axis =1)
2 test = pd.concat([X2_test, y2_test], axis =1)
3 def auc_train_test (variables, target, train, test):
4     X_train = train [variables]
5     X_test = test[variables]
6     Y_train = train[target]
7     Y_test = test[target]
8     LR3 = LogisticRegression()
9     LR3.fit(X_train, Y_train)
10    predictions_train = LR3.predict_proba(X_train)[:,1]
11    predictions_test = LR3.predict_proba(X_test)[:,1]
12    auc_train = roc_auc_score(Y_train, predictions_train)
13    auc_test = roc_auc_score(Y_train, predictions_test)
14    return (auc_train, auc_test)
15 auc_values_train =[]
16 auc_values_test =[]
17 variables_evaluate=[]
18 for v in X2.columns:
19     variables_evaluate.append(v)
20     auc_train, auc_test = auc_train_test(variables_evaluate, ['Outcome'], train)
21     auc_values_train.append(auc_train)
22     auc_values_test.append(auc_test)

```

C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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().

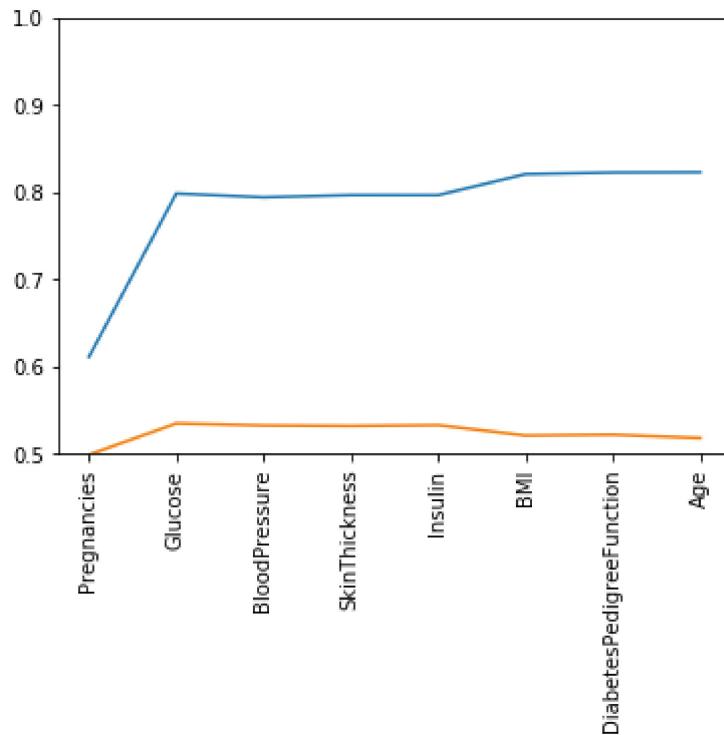
y = column_or_1d(y, warn=True)

C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\sklearn\utils\validation.py:578: 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)
```

In [53]:

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 x = np.array(range(0,len(auc_values_train)))
4 my_train = np.array(auc_values_train)
5 my_test = np.array(auc_values_test)
6 plt.xticks(x, X2.columns, rotation =90)
7
8 plt.plot(x, my_train)
9 plt.plot(x, my_test)
10 plt.ylim(0.5, 1)
11 plt.show()
```



Step8. DRAW CUMULATIVE GAIN CHART AND LIFT CHART

In [55]: 1 !pip install scikit-plot
2

```
Defaulting to user installation because normal site-packages is not writeable
Collecting scikit-plot
  Downloading scikit_plot-0.3.7-py3-none-any.whl (33 kB)
Collecting joblib>=0.10
  Downloading joblib-1.1.1-py2.py3-none-any.whl (309 kB)
Requirement already satisfied: matplotlib>=1.4.0 in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from scikit-plot) (2.2.2)
Requirement already satisfied: scipy>=0.9 in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from scikit-plot) (1.1.0)
Requirement already satisfied: scikit-learn>=0.18 in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from scikit-plot) (0.19.1)
Requirement already satisfied: numpy>=1.7.1 in c:\users\1mscdsa51\appdata\roaming\python\python36\site-packages (from matplotlib>=1.4.0->scikit-plot) (1.19.5)
Requirement already satisfied: cycler>=0.10 in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from matplotlib>=1.4.0->scikit-plot) (0.10.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from matplotlib>=1.4.0->scikit-plot) (2.2.0)
Requirement already satisfied: python-dateutil>=2.1 in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from matplotlib>=1.4.0->scikit-plot) (2.7.3)
Requirement already satisfied: pytz in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from matplotlib>=1.4.0->scikit-plot) (2018.4)
Requirement already satisfied: six>=1.10 in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from matplotlib>=1.4.0->scikit-plot) (1.11.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from matplotlib>=1.4.0->scikit-plot) (1.0.1)
Requirement already satisfied: setuptools in c:\program files (x86)\microsoft visual studio\shared\anaconda3_64\lib\site-packages (from kiwisolver>=1.0.1->matplotlib>=1.4.0->scikit-plot) (39.1.0)
Installing collected packages: joblib, scikit-plot
Successfully installed joblib-1.1.1 scikit-plot-0.3.7
```

WARNING: pip is being invoked by an old script wrapper. This will fail in a future version of pip.

Please see <https://github.com/pypa/pip/issues/5599> (<https://github.com/pypa/pip/issues/5599>) for advice on fixing the underlying issue.

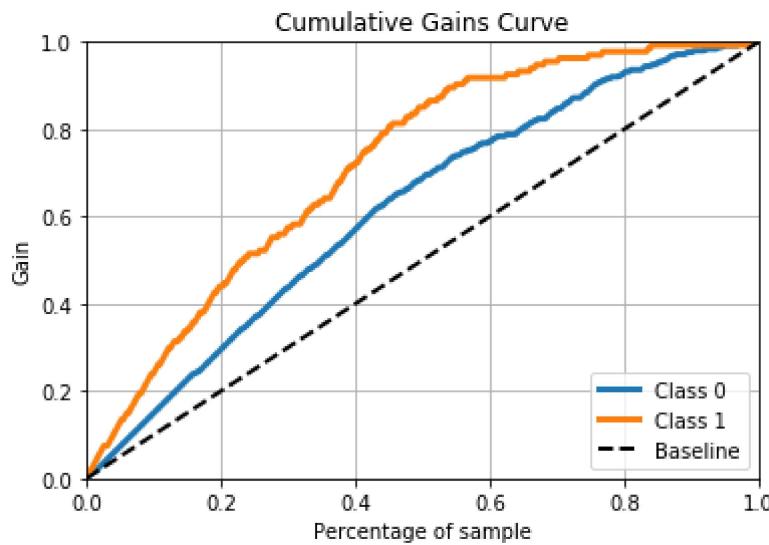
To avoid this problem you can invoke Python with '-m pip' instead of running pip directly.

In []: 1

In [58]:

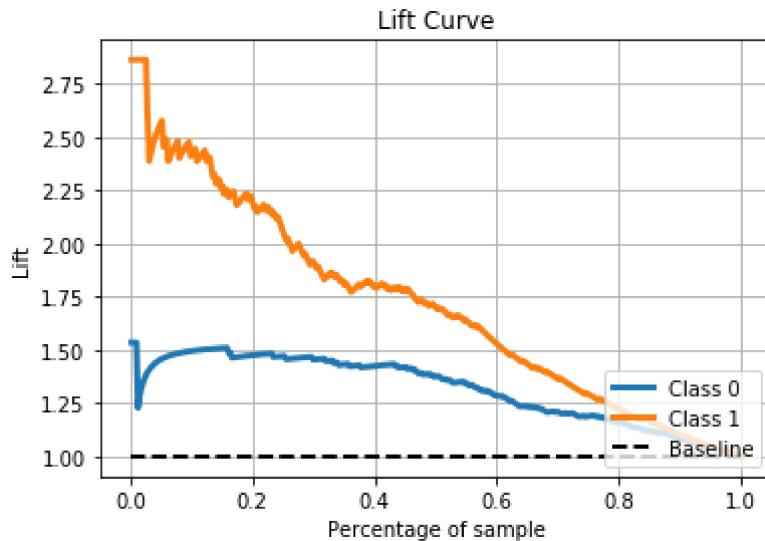
```
1 import scikitplot as skplt
2 skplt.metrics.plot_cumulative_gain(y2_test, prediction1)
3 plt.show()
4 plt.figure(figsize=(7,7))
5 skplt.metrics.plot_lift_curve(y2_test, prediction1)
6 plt.show()
```

C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\matplotlib\cbook\deprecation.py:107: MatplotlibDeprecationWarning: Passing one of 'on', 'true', 'off', 'false' as a boolean is deprecated; use an actual boolean (True/False) instead.
warnings.warn(message, mplDeprecation, stacklevel=1)



C:\Program Files (x86)\Microsoft Visual Studio\Shared\Anaconda3_64\lib\site-packages\matplotlib\cbook\deprecation.py:107: MatplotlibDeprecationWarning: Passing one of 'on', 'true', 'off', 'false' as a boolean is deprecated; use an actual boolean (True/False) instead.
warnings.warn(message, mplDeprecation, stacklevel=1)

<Figure size 504x504 with 0 Axes>



In []:

1