

**Ismet Okan Celik CWID: 10472265**

*I pledge my honor that I have abided by the Stevens Honor System.*

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import metrics
from sklearn.linear_model import LogisticRegression
from sklearn.feature_selection import RFE
from sklearn.model_selection import train_test_split
```

```
In [2]: data=pd.read_csv('dataSet_2.csv',header=None)
data
```

```
Out[2]:
```

	0	1	2	3	4	5	6	7	8	9	...	914	915	916	917
0	1.85050	0.44413	-0.26072	-0.92156	-0.52009	2.73230	-1.5232	-1.2479	1.4870	-0.52449	...	-0.403160	-0.555010	-0.590640	-0.63190
1	2.13930	0.67815	-0.34487	-0.93527	-0.52009	3.08680	-1.5549	-1.2048	1.4753	-0.41132	...	-1.182500	1.527000	0.267590	2.79870
2	2.19670	0.63249	-0.28908	-0.89806	-0.52009	3.08970	-1.5624	-1.1764	1.5184	-0.37138	...	0.258950	-0.537660	1.899500	0.56850
3	2.21590	0.68735	-0.31650	-0.90001	-0.52009	3.15650	-1.5627	-1.0686	1.5353	-0.34309	...	1.780200	2.820300	0.268560	-0.15080
4	2.26450	0.69541	-0.31650	-0.93331	-0.52009	3.09840	-1.5769	-1.1232	1.4557	-0.32478	...	-0.991740	-0.520880	-0.426940	0.50630
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5917	-0.49460	0.25691	1.60670	1.64210	1.92240	0.24546	1.9545	1.6000	1.6307	2.14820	...	-0.132260	-0.004755	-0.801210	0.04200
5918	-0.48429	0.29911	1.63510	1.67340	1.92240	0.26870	1.9817	1.6059	1.6307	2.15980	...	1.373600	-0.004755	0.982660	0.09490
5919	-0.43715	0.28530	1.66250	1.68120	1.92240	0.29775	2.0059	1.6073	1.6424	2.20810	...	-0.898300	-0.004755	-0.693790	1.14710
5920	-0.47545	0.31983	1.60670	1.71850	1.92240	0.26289	2.0425	1.6059	1.6816	2.20140	...	-0.971120	-0.004755	0.982660	-0.98700
5921	-0.46514	0.28339	1.60670	1.73220	1.92240	0.26579	2.0516	1.6146	1.6660	2.23140	...	0.003648	-0.004755	0.000621	-0.00110

5922 rows × 924 columns

```
In [3]: predict=data.iloc[:,0:923]
response=data[923]
```

```
In [4]: x_train, x_test, y_train, y_test=train_test_split(predict, response, test_size=0.24,random_state=0,shuffle=False)
print('X_Train--->(Number of Row, Number of Column)=' ,np.shape(x_train))
print('Y_Train--->(Number of Row, Number of Column)=' ,np.shape(x_test))
print('X_Test--->(Number of Row, Number of Column)=' ,np.shape(y_train))
print('Y_Test--->(Number of Row, Number of Column)=' ,np.shape(y_test))
```

```
X_Train--->(Number of Row, Number of Column)= (4500, 923)
Y_Train--->(Number of Row, Number of Column)= (1422, 923)
X_Test--->(Number of Row, Number of Column)= (4500,)
Y_Test--->(Number of Row, Number of Column)= (1422,)
```

```
In [5]: #Training with Original Matrix
```

```
model=LogisticRegression(solver='liblinear',C=1.0,random_state=0)
model_fit=model.fit(x_train,y_train)

prob_pred_train=model_fit.predict_proba(x_train)[: ,1]

FP_Rate_Train,TP_Rate_Train,_=metrics.roc_curve(y_train,prob_pred_train)
AUC_Train=metrics.roc_auc_score(y_train,prob_pred_train)
```

```
In [6]: #Testing with Original Matrix
```

```
prob_pred_test=model_fit.predict_proba(x_test)[: ,1]

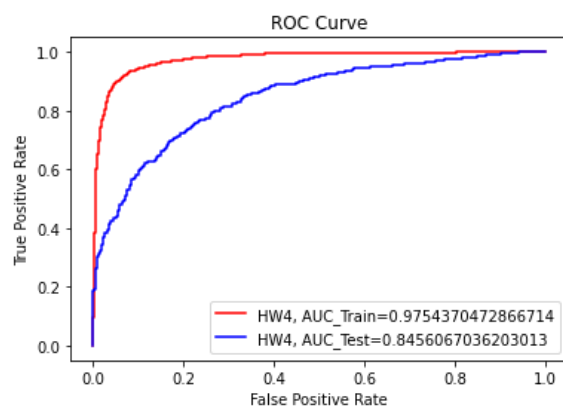
FP_Rate_Test,TP_Rate_Test,_=metrics.roc_curve(y_test,prob_pred_test)
AUC_Test=metrics.roc_auc_score(y_test,prob_pred_test)
```

In [7]: *# Plotting ROC for Training and Testing with Original Matrix*

```
difference1=AUC_Train-AUC_Test
print('Difference Between AUC_Train and AUC_Test=',difference1)
plt.plot(FP_Rate_Train,TP_Rate_Train,label='HW4, AUC_Train='+str(AUC_Train),color='r')
plt.plot(FP_Rate_Test,TP_Rate_Test,label='HW4, AUC_Test='+str(AUC_Test),color='b')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve')
plt.legend(loc=4)
plt.show
```

Difference Between AUC\_Train and AUC\_Test= 0.12983034366637003

Out[7]: <function matplotlib.pyplot.show(close=None, block=None)>



In [97]: *# Feature Selection From Original Data by Using Recursive Feature Elimination (RFE)*

```
rfe_selector=RFE(estimator=LogisticRegression(solver='liblinear',C=1.0,random_state=0),n_features_to_select=10,step=1)
rfe_selector=rfe_selector.fit(x_train,y_train);
rfe_support=rfe_selector.get_support();
rfe_feature=x_train.loc[:,rfe_support].columns.tolist();
```

Fitting estimator with 923 features.  
Fitting estimator with 913 features.  
Fitting estimator with 903 features.  
Fitting estimator with 893 features.  
Fitting estimator with 883 features.  
Fitting estimator with 873 features.  
Fitting estimator with 863 features.  
Fitting estimator with 853 features.  
Fitting estimator with 843 features.  
Fitting estimator with 833 features.  
Fitting estimator with 823 features.  
Fitting estimator with 813 features.  
Fitting estimator with 803 features.  
Fitting estimator with 793 features.  
Fitting estimator with 783 features.  
Fitting estimator with 773 features.  
Fitting estimator with 763 features.  
Fitting estimator with 753 features.  
Fitting estimator with 743 features.  
Fitting estimator with 733 features.  
Fitting estimator with 723 features.  
Fitting estimator with 713 features.  
Fitting estimator with 703 features.  
Fitting estimator with 693 features.  
Fitting estimator with 683 features.  
Fitting estimator with 673 features.  
Fitting estimator with 663 features.  
Fitting estimator with 653 features.  
Fitting estimator with 643 features.  
Fitting estimator with 633 features.  
Fitting estimator with 623 features.  
Fitting estimator with 613 features.  
Fitting estimator with 603 features.  
Fitting estimator with 593 features.  
Fitting estimator with 583 features.  
Fitting estimator with 573 features.  
Fitting estimator with 563 features.  
Fitting estimator with 553 features.  
Fitting estimator with 543 features.  
Fitting estimator with 533 features.  
Fitting estimator with 523 features.  
Fitting estimator with 513 features.  
Fitting estimator with 503 features.  
Fitting estimator with 493 features.  
Fitting estimator with 483 features.  
Fitting estimator with 473 features.  
Fitting estimator with 463 features.  
Fitting estimator with 453 features.  
Fitting estimator with 443 features.  
Fitting estimator with 433 features.  
Fitting estimator with 423 features.  
Fitting estimator with 413 features.  
Fitting estimator with 403 features.  
Fitting estimator with 393 features.  
Fitting estimator with 383 features.  
Fitting estimator with 373 features.  
Fitting estimator with 363 features.  
Fitting estimator with 353 features.  
Fitting estimator with 343 features.  
Fitting estimator with 333 features.  
Fitting estimator with 323 features.  
Fitting estimator with 313 features.  
Fitting estimator with 303 features.  
Fitting estimator with 293 features.  
Fitting estimator with 283 features.  
Fitting estimator with 273 features.  
Fitting estimator with 263 features.  
Fitting estimator with 253 features.  
Fitting estimator with 243 features.

```

Fitting estimator with 233 features.
Fitting estimator with 223 features.
Fitting estimator with 213 features.
Fitting estimator with 203 features.
Fitting estimator with 193 features.
Fitting estimator with 183 features.
Fitting estimator with 173 features.
Fitting estimator with 163 features.
Fitting estimator with 153 features.
Fitting estimator with 143 features.
Fitting estimator with 133 features.
Fitting estimator with 123 features.
Fitting estimator with 113 features.
Fitting estimator with 103 features.
Fitting estimator with 93 features.
Fitting estimator with 83 features.
Fitting estimator with 73 features.
Fitting estimator with 63 features.
Fitting estimator with 53 features.
Fitting estimator with 43 features.
Fitting estimator with 33 features.
Fitting estimator with 23 features.
Fitting estimator with 13 features.

```

```

In [98]: #Selected Features From Original Matrix
selected_original_train=x_train[rfe_feature]
selected_original_test=x_test[rfe_feature]
selected_original_train

```

```

Out[98]:

```

	9	10	14	31	67	334	348	369	370	824
0	-0.52449	0.092423	0.128900	0.33626	0.80399	0.718540	-1.22810	3.130600	-0.262270	0.19500
1	-0.41132	0.097491	0.126020	0.69238	0.86677	0.939780	-0.79481	3.442700	-0.172230	0.05466
2	-0.37138	0.050184	0.082772	0.68583	0.86258	0.004308	-0.42913	0.161190	0.007712	-0.87080
3	-0.34309	0.070458	0.077006	0.70112	0.87514	0.102430	-0.23053	0.250020	0.614220	-0.09877
4	-0.32478	0.018083	0.071239	0.70549	0.83747	-0.020251	-0.19732	0.316050	0.500950	-0.10414
...	...	...	...	...	...	...	...	...	...	...
4495	0.86844	1.197400	1.397400	0.77322	1.13880	-1.048900	-1.25250	-0.759570	-0.390140	-0.46121
4496	0.84847	1.167000	1.354200	0.74263	1.09700	-0.779230	-0.64141	-0.059978	-0.120240	-1.47020
4497	0.83182	1.151800	1.354200	0.72078	1.07390	-0.656960	-0.18379	-0.533220	-0.348900	0.04782
4498	0.80520	1.129800	1.345500	0.67709	1.03840	-0.534780	0.09119	-0.842820	-0.744600	-0.39424
4499	0.80353	1.094300	1.316700	0.62028	1.00280	0.175530	0.15233	-0.595740	-0.328510	-1.24280

4500 rows × 10 columns

```
In [99]: ''' Generating new predictor matrix by using [X(t)- X(t-1)]/X(t-1) and adding this matrix
to the 10 selected features from original data'''

new_predictor_train=np.subtract(x_train[2:4500],x_train[1:4499])
new_predictor_train=np.divide(new_predictor_train,x_train[1:4499])
all_predictor_train=pd.concat([selected_original_train[2:4500],new_predictor_train],axis=1,ignore_index=True)

new_response_train=y_train[2:4500]

new_predictor_test=np.subtract(x_test[2:1422],x_test[1:1421])
new_predictor_test=np.divide(new_predictor_test,x_test[1:1421])
all_predictor_test=pd.concat([selected_original_test[2:1422],new_predictor_test],axis=1,ignore_index=True)

new_response_test=y_test[2:4500]

print('Shape of all Predictor Train Matrix')
print('(Number of Row, Number of Column)=' ,np.shape(all_predictor_train),'\n')

print('Shape of all Predictor Test Matrix')
print('(Number of Row, Number of Column)=' ,np.shape(all_predictor_test))
```

Shape of all Predictor Train Matrix  
(Number of Row, Number of Column)= (4498, 933)

Shape of all Predictor Test Matrix  
(Number of Row, Number of Column)= (1420, 933)

C:\Users\okanc\AppData\Local\Temp\ipykernel\_13196\380629413.py:4: FutureWarning: Calling a ufunc on non-aligned DataFrames (or DataFrame/Series combination). Currently, the indices are ignored and the result takes the index/columns of the first DataFrame. In the future , the DataFrames/Series will be aligned before applying the ufunc. Convert one of the arguments to a NumPy array (eg 'ufunc(df1, np.asarray(df2))' to keep the current behaviour, or align manually (eg 'df1, df2 = df1.align(df2)') before passing to the ufunc to obtain the future behaviour and silence this warning.

```
new_predictor_train=np.subtract(x_train[2:4500],x_train[1:4499])
```

C:\Users\okanc\AppData\Local\Temp\ipykernel\_13196\380629413.py:5: FutureWarning: Calling a ufunc on non-aligned DataFrames (or DataFrame/Series combination). Currently, the indices are ignored and the result takes the index/columns of the first DataFrame. In the future , the DataFrames/Series will be aligned before applying the ufunc. Convert one of the arguments to a NumPy array (eg 'ufunc(df1, np.asarray(df2))' to keep the current behaviour, or align manually (eg 'df1, df2 = df1.align(df2)') before passing to the ufunc to obtain the future behaviour and silence this warning.

```
new_predictor_train=np.divide(new_predictor_train,x_train[1:4499])
```

C:\Users\okanc\AppData\Local\Temp\ipykernel\_13196\380629413.py:10: FutureWarning: Calling a ufunc on non-aligned DataFrames (or DataFrame/Series combination). Currently, the indices are ignored and the result takes the index/columns of the first DataFrame. In the future , the DataFrames/Series will be aligned before applying the ufunc. Convert one of the arguments to a NumPy array (eg 'ufunc(df1, np.asarray(df2))' to keep the current behaviour, or align manually (eg 'df1, df2 = df1.align(df2)') before passing to the ufunc to obtain the future behaviour and silence this warning.

```
new_predictor_test=np.subtract(x_test[2:1422],x_test[1:1421])
```

C:\Users\okanc\AppData\Local\Temp\ipykernel\_13196\380629413.py:11: FutureWarning: Calling a ufunc on non-aligned DataFrames (or DataFrame/Series combination). Currently, the indices are ignored and the result takes the index/columns of the first DataFrame. In the future , the DataFrames/Series will be aligned before applying the ufunc. Convert one of the arguments to a NumPy array (eg 'ufunc(df1, np.asarray(df2))' to keep the current behaviour, or align manually (eg 'df1, df2 = df1.align(df2)') before passing to the ufunc to obtain the future behaviour and silence this warning.

```
new_predictor_test=np.divide(new_predictor_test,x_test[1:1421])
```

```
In [100]: model_new=LogisticRegression(solver='liblinear',C=1.0,random_state=0)
model_fit_new=model.fit(all_predictor_train,new_response_train)

prob_pred_train_new=model_fit_new.predict_proba(all_predictor_train)[: ,1]
```

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

```
In [101]: # Testing with New Matrix(Selected 10 Fetures From Original Matrix + New Generated Features)

prob_pred_test_new=model_fit_new.predict_proba(all_predictor_test)[: ,1]

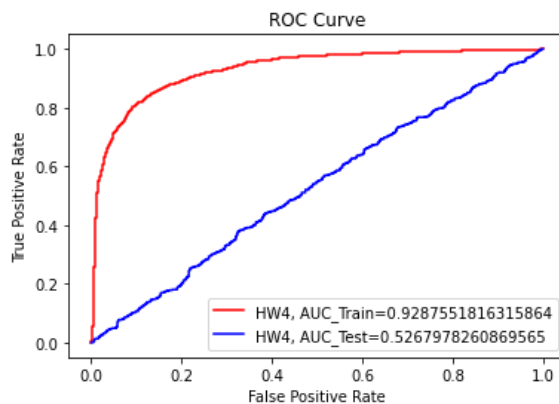
FP_Rate_Test_new,TP_Rate_Test_new,_=metrics.roc_curve(new_response_test,prob_pred_test_new)
AUC_Test_new=metrics.roc_auc_score(new_response_test,prob_pred_test_new)
```

In [102]: # Plotting ROC - Training and Testing for New Matrix

```
difference_new=AUC_Train_new-AUC_Test_new
print('Difference Between New AUC_Train and AUC_Test=',difference_new)
plt.plot(FP_Rate_Train_new,TP_Rate_Train_new,label='HW4, AUC_Train='+str(AUC_Train_new),color='r')
plt.plot(FP_Rate_Test_new,TP_Rate_Test_new,label='HW4, AUC_Test='+str(AUC_Test_new),color='b')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve')
plt.legend(loc=4)
plt.show
```

Difference Between New AUC\_Train and AUC\_Test= 0.40195735554462997

Out[102]: <function matplotlib.pyplot.show(close=None, block=None)>



We can see how the new generated matrix affected the AUC\_Train and AUC\_Test. The difference between them got bigger and the AUC\_Test value decreased drastically. Model performance with testing data is bad.

```
In [103]: ''' Feature Selection From New Matrix (Selected 10 Fetures From Original Matrix + New Generated Features)
by Using Recursive Feature Elimination (RFE).This time 20 Features are Selected'''

rfe_selector=RFE(estimator=LogisticRegression(solver='liblinear',C=1.0,random_state=0),n_features_to_select=20,step=1)
rfe_selector=rfe_selector.fit(all_predictor_train,new_response_train);
rfe_support=rfe_selector.get_support();
rfe_feature=all_predictor_train.loc[:,rfe_support].columns.tolist();
```

Fitting estimator with 933 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 923 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 913 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 903 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 893 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 883 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 873 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 863 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 853 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 843 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 833 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 823 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.  
warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 813 features.

```
C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 803 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 793 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 783 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 773 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 763 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 753 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 743 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 733 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 723 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 713 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 703 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 693 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 683 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 673 features.
```



```
C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 663 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 653 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 643 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 633 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 623 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 613 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 603 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 593 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 583 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 573 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 563 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 553 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 543 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 533 features.
```

```
C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 523 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 513 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 503 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 493 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 483 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 473 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 463 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 453 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 443 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 433 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 423 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 413 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 403 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase the number of iterations.")

Fitting estimator with 393 features.
```

```
C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 383 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 373 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 363 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 353 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 343 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 333 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 323 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 313 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 303 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 293 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 283 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 273 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 263 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 253 features.
```

```
C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 243 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 233 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 223 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 213 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 203 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 193 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 183 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 173 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 163 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 153 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 143 features.
Fitting estimator with 133 features.
Fitting estimator with 123 features.

C:\Users\okanc\anaconda3\lib\site-packages\sklearn\svm\_base.py:985: ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "

Fitting estimator with 113 features.
Fitting estimator with 103 features.
Fitting estimator with 93 features.
Fitting estimator with 83 features.
Fitting estimator with 73 features.
Fitting estimator with 63 features.
Fitting estimator with 53 features.
Fitting estimator with 43 features.
Fitting estimator with 33 features.
Fitting estimator with 23 features.
```

```
In [104]: # New Selected Features From New Matrix
all_predictor_train=all_predictor_train[rfe_feature]
all_predictor_test=all_predictor_test[rfe_feature]
all_predictor_train[rfe_feature]
```

```
Out[104]:
```

	0	5	6	7	8	9	10	22	26	39	43	44	46
2	-0.37138	0.004308	-0.42913	0.161190	0.007712	-0.87080	0.026831	-0.000000	-0.005378	-0.000000	0.554651	0.018707	0.018282
3	-0.34309	0.102430	-0.23053	0.250020	0.614220	-0.09877	0.008740	-0.065934	-0.007179	-0.000000	0.050059	-0.039789	-0.060792
4	-0.32478	-0.020251	-0.19732	0.316050	0.500950	-0.10414	0.021932	-0.028249	-0.001784	-0.000000	-0.017882	0.082887	0.054425
5	-0.33477	0.371970	0.69930	0.915790	0.905200	-0.16757	-0.021462	-0.184111	0.010818	-0.000000	-0.218493	0.002945	0.066980
6	-0.37305	1.256600	0.33464	1.339800	1.378800	-1.24620	0.004648	-0.166274	-0.007166	-0.000000	-0.124249	0.058698	0.033995
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4495	0.86844	-1.048900	-1.25250	-0.759570	-0.390140	-0.46121	-0.120108	-0.014022	0.000000	0.000000	0.010017	-0.019001	-0.013323
4496	0.84847	-0.779230	-0.64141	-0.059978	-0.120240	-1.47020	-0.409495	-0.008888	0.000000	0.000000	-0.024786	-0.006485	-0.072600
4497	0.83182	-0.656960	-0.18379	-0.533220	-0.348900	0.04782	-0.693489	-0.012555	0.000000	-1.082239	-0.050816	-0.017262	-0.027323
4498	0.80520	-0.534780	0.09119	-0.842820	-0.744600	-0.39424	-1.914422	-0.019967	0.000000	-0.000000	-0.053554	-0.013196	0.037441
4499	0.80353	0.175530	0.15233	-0.595740	-0.328510	-1.24280	-0.000000	0.011120	0.052320	-0.000000	0.016972	-0.053578	-0.115466

4498 rows × 20 columns

```
In [105]: # Training with New Selected Features

model_new=LogisticRegression(solver='liblinear',C=1.0,random_state=0)
model_fit_new=model.fit(all_predictor_train,new_response_train)

prob_pred_train_new=model_fit_new.predict_proba(all_predictor_train)[:,1]

FP_Rate_Train_new,TP_Rate_Train_new,_,_=metrics.roc_curve(new_response_train,prob_pred_train_new)
AUC_Train_new=metrics.roc_auc_score(new_response_train,prob_pred_train_new)
```

```
In [106]: # Testing with New Selected Features

prob_pred_test_new=model_fit_new.predict_proba(all_predictor_test)[:,1]

FP_Rate_Test_new,TP_Rate_Test_new,_,_=metrics.roc_curve(new_response_test,prob_pred_test_new)
AUC_Test_new=metrics.roc_auc_score(new_response_test,prob_pred_test_new)
```

In [107]: *# Plotting ROC for New Selected Features*

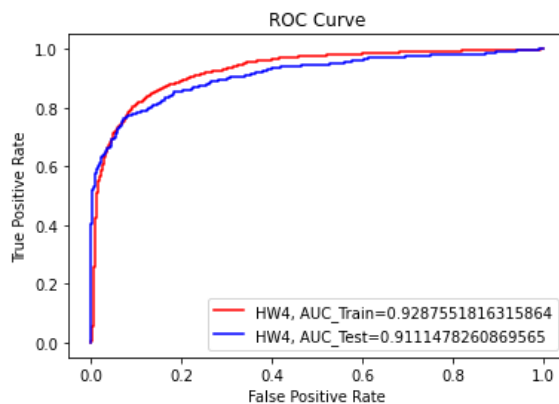
```

difference_new=AUC_Train_new-AUC_Test_new
print('Difference Between New AUC_Train and AUC_Test=',difference_new)
plt.plot(FP_Rate_Train_new,TP_Rate_Train_new,label='HW4, AUC_Train='+str(AUC_Train_new),color='r')
plt.plot(FP_Rate_Test_new,TP_Rate_Test_new,label='HW4, AUC_Test='+str(AUC_Test_new),color='b')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve')
plt.legend(loc=4)
plt.show

```

Difference Between New AUC\_Train and AUC\_Test= 0.017607355544629888

Out[107]: <function matplotlib.pyplot.show(close=None, block=None)>



As we can see on the graph, the difference between AUC\_Train and AUC\_Test is 0.017607355544629888; it is really small after the new feature selection.

The new predictor matrix has affected the feature selection; we can see that the last feature selection first 6 columns are from original data, and the rest of the 14 columns are from the newly generated matrix. When we compare the three graphs, we can see that the last performance is way better than the other two. Because in the last part we eliminated most of the irrelevant columns and because of that our model performed better.

#### **Training and Testing:**

1-Original Data ----> AUC\_Train - AUC\_Test=0.12983034366637003

2-New Matrix (Original Data + Generated Matrix) ----> AUC\_Train - AUC\_Test= 0.40244435932705347

3-Selected Features From New Matrix---->AUC\_Train - AUC\_Test= 0.017607355544629888

We reached the high accuracy by the help of feature engineering and feature selection.