

```
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
```

Assignment 3

All Variables

```
In [29]: def getCoeffLogit( MODEL, TRAIN_DATA ) :
    varNames = list( TRAIN_DATA.columns.values )
    coef_dict = {}
    coef_dict["INTERCEPT"] = MODEL.intercept_[0]
    for coef, feat in zip(MODEL.coef_[0],varNames):
        coef_dict[feat] = coef
    print("\nFLAG")
    print("-----")
    print("Total Variables: ", len( coef_dict ) )
    for i in coef_dict :
        print( i, " = ", coef_dict[i] )

WHO = "REG_ALL"

CLM = LogisticRegression( solver='newton-cg', max_iter=1000 )
CLM = CLM.fit( X_train, Y_train[ FLAG ] )

TRAIN_CLM = getProbAccuracyScores( WHO + "_Train", CLM, X_train, Y_train[ FLAG ] )
TEST_CLM = getProbAccuracyScores( WHO, CLM, X_test, Y_test[ FLAG ] )

print_Accuracy( WHO + " RMSE ACCURACY", [ TRAIN_CLM, TEST_CLM ] )

print_ROC_Curve( WHO, [ TRAIN_CLM, TEST_CLM ] )
print_Accuracy( WHO + " CLASSIFICATION ACCURACY", [ TRAIN_CLM, TEST_CLM ] )

#LOSS Regression

def getCoeffLinear( MODEL, TRAIN_DATA ) :
    varNames = list( TRAIN_DATA.columns.values )
    coef_dict = {}
    coef_dict["INTERCEPT"] = MODEL.intercept_
    for coef, feat in zip(MODEL.coef_,varNames):
        coef_dict[feat] = coef
    print("\nLOSS")
    print("-----")
    print("Total Variables: ", len( coef_dict ) )
    for i in coef_dict :
        print( i, " = ", coef_dict[i] )

AMT = LinearRegression()
AMT = AMT.fit( W_train, Z_train[LOSS] )

TRAIN_AMT = getAmtAccuracyScores( WHO + "_Train", AMT, W_train, Z_train[LOSS] )
TEST_AMT = getAmtAccuracyScores( WHO, AMT, W_test, Z_test[LOSS] )
print_Accuracy( WHO + " RMSE ACCURACY", [ TRAIN_AMT, TEST_AMT ] )

varNames = list( X_train.columns.values )

REG_ALL_CLM_COEF = getCoeffLogit( CLM, X_train )
REG_ALL_AMT_COEF = getCoeffLinear( AMT, X_train )

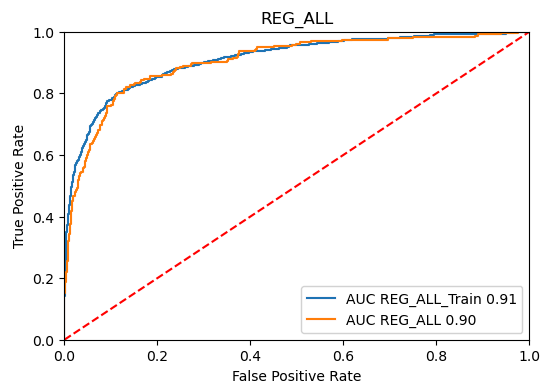
REG_ALL_CLM = TEST_CLM.copy()
REG_ALL_AMT = TEST_AMT.copy()
```

```

C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:466: LineSearchWarning: The line search algorithm did not converge
    warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:314: LineSearchWarning: The line search algorithm did not converge
    warn('The line search algorithm did not converge', LineSearchWarning)
REG_ALL RMSE ACCURACY
=====
REG_ALL_Train = 0.8930369127516778
REG_ALL      = 0.8808724832214765
-----

```

[illegible]



```
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
```

```

REG_ALL CLASSIFICATION ACCURACY
=====
REG_ALL_Train = 0.8930369127516778
REG_ALL = 0.8808724832214765
-----

```

```

REG_ALL RMSE ACCURACY
=====
REG_ALL_Train = 3673.227061891806
REG_ALL = 3754.7356551934763
-----

```

```

FLAG
-----
Total Variables: 29
INTERCEPT = -5.435082439095795
Unnamed: 0 = -6.934396183211991e-05
LOAN = -1.9839825794254507e-06
OHE_REASON_HOMEIMP = 0.33292937091822955
OHE_REASON_DEBTCON = 0.14893139057546884
OHE_REASON_MISSING = -0.5043834492795197
OHE_JOB_OFFICE = -0.4403793898454696
OHE_JOB_OTHER = 0.2762334630907767
OHE_JOB_MGR = 0.21155628136113833
OHE_JOB_PROFEXEC = 0.0
OHE_JOB_SALES = 1.344023654051649
M_MORTDUE = 0.5024512734920682
IMP_MORTDUE = -3.5012147941512535e-06
M_VALUE = 3.522712756317599
IMP_VALUE = 4.317989578325467e-06
M_YOJ = -0.6265276813729354
IMP_YOJ = -0.01747638464527435
M_DEROG = -1.8996042047864978
IMP_DEROG = 0.5007936717589831
M_DELIQ = -0.4685279468411719
IMP_DELIQ = 0.8045444006431522
M_CLAGE = 1.2623238375095556
IMP_CLAGE = -0.005588743043030741
M_NINQ = 0.15837056498304586
IMP_NINQ = 0.15533913876650177
M_CLNO = 1.668387938668998
IMP_CLNO = -0.009406071096347472
M_DEBTINC = 2.624415110141342
IMP_DEBTINC = 0.0970698587362559

```

```

LOSS
-----
Total Variables: 29
INTERCEPT = -1209.768528457298
Unnamed: 0 = 0.6171536565904032
LOAN = -0.0009594887314227451
OHE_REASON_HOMEIMP = -214.01204012113615
OHE_REASON_DEBTCON = 186.48144035296838
OHE_REASON_MISSING = 27.530599832632085
OHE_JOB_OFFICE = -153.8318058506309
OHE_JOB_OTHER = 45.34030341378558
OHE_JOB_MGR = -108.57772971283828
OHE_JOB_PROFEXEC = -2.6147972675971687e-12
OHE_JOB_SALES = 2717.1693534070973
M_MORTDUE = 826.3961867277318
IMP_MORTDUE = -0.007713324099253575
M_VALUE = 4290.232575456508
IMP_VALUE = 0.008599099402534158
M_YOJ = -690.4502832153239
IMP_YOJ = -22.09012083528172
M_DEROG = -1381.5771658825563
IMP_DEROG = 812.9588059202737
M_DELIQ = 317.1857290953153
IMP_DELIQ = 1164.4367148771412
M_CLAGE = -52.300400760190136
IMP_CLAGE = -6.877845376036764
M_NINQ = 120.76829651830703
IMP_NINQ = 191.96664438677308
M_CLNO = 2037.8450168796653
IMP_CLNO = 45.30765005782466
M_DEBTINC = 4226.338509859905
IMP_DEBTINC = 74.00335697515517

```

```

C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):

```

Decision Tree

```

In [30]: WHO = "REG_TREE"
CLM = LogisticRegression( solver='newton-cg', max_iter=1000 )
CLM = CLM.fit( X_train[vars_tree_flag], Y_train[ FLAG ] )

TRAIN_CLM = getProbAccuracyScores( WHO + "_Train", CLM, X_train[vars_tree_flag], Y_train[ FLAG ] )
TEST_CLM = getProbAccuracyScores( WHO, CLM, X_test[vars_tree_flag], Y_test[ FLAG ] )

```

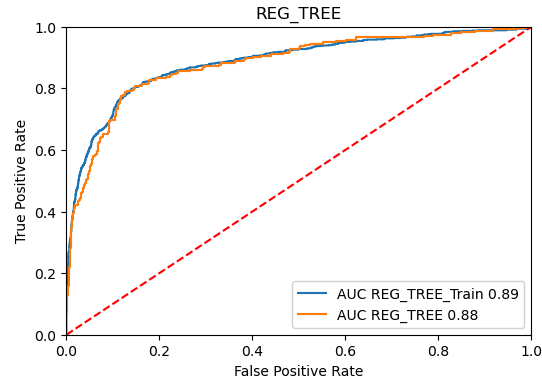
```
print_ROC_Curve( WHO, [ TRAIN_CLM, TEST_CLM ] )  
print_Accuracy( WHO + " CLASSIFICATION ACCURACY", [ TRAIN_CLM, TEST_CLM ] )
```

rs/jamia/Downloads/Assignment 2 (2).html

```

if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):

```



```

REG_TREE CLASSIFICATION ACCURACY
=====
REG_TREE_Train = 0.8810822147651006
REG_TREE = 0.8666107382550335
-----

```

```

In [31]: def getCoeffLinear( MODEL, TRAIN_DATA ) :
varNames = list( TRAIN_DATA.columns.values )
coeff_dict = {}
coeff_dict["INTERCEPT"] = MODEL.intercept_
for coef, feat in zip(MODEL.coef_, varNames):
coeff_dict[feat] = coef
print("\nLOSS")
print("-----")
print("Total Variables: ", len( coeff_dict ) )
for i in coeff_dict :
print( i, " = ", coeff_dict[i] )

AMT = LinearRegression()
AMT = AMT.fit( W_train[vars_tree_amt], Z_train[LOSS] )

TRAIN_AMT = getAmtAccuracyScores( WHO + "_Train", AMT, W_train[vars_tree_amt], Z_train[LOSS] )
TEST_AMT = getAmtAccuracyScores( WHO, AMT, W_test[vars_tree_amt], Z_test[LOSS] )
print_Accuracy( WHO + " RMSE ACCURACY", [ TRAIN_AMT, TEST_AMT ] )

varNames = list( X_train.columns.values )

REG_TREE_CLM_COEF = getCoeffLogit( CLM, X_train[vars_tree_flag] )
REG_TREE_AMT_COEF = getCoeffLinear( AMT, X_train[vars_tree_amt] )

REG_TREE_CLM = TEST_CLM.copy()
REG_TREE_AMT = TEST_AMT.copy()

```

```

REG_TREE RMSE ACCURACY
=====
REG_TREE_Train = 3786.240357467525
REG_TREE = 3850.540793930161
-----

```

```

FLAG
-----
Total Variables: 8
INTERCEPT = -4.889917121878428
IMP_VALUE = 9.94440257953066e-07
M_DEROG = -0.70941571448698
IMP_DEROG = 0.5401551121471723
IMP_DELIQ = 0.6867383723426624
IMP_CLAGE = -0.006498991436989625
M_DEBTINC = 2.779816830618247
IMP_DEBTINC = 0.08835364489046318

LOSS
-----
Total Variables: 10
INTERCEPT = -1047.2688586231825
Unnamed: 0 = 0.651552609634051
OHE_REASON_DEBTCON = 258.41093862951607
IMP_MORTDUE = 0.0017472255112593819
IMP_DEROG = 968.0601775079234
IMP_DELIQ = 1135.4770251933098
IMP_CLAGE = -7.397511908425877
IMP_CLNO = 43.37993381691741
M_DEBTINC = 4677.094409277137
IMP_DEBTINC = 71.38674889287049

```



```
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
```

```
In [32]: #variables tree liked
def getTreeVars( TREE, varNames ) :
    tree_ = TREE.tree_
    varName = [ varNames[i] if i != _tree.TREE_UNDEFINED else "undefined!" for i in tree_.feature ]
    nameSet = set()
    for i in tree_.feature :
        if i != _tree.TREE_UNDEFINED :
            nameSet.add( i )
    nameList = list( nameSet )
    parameter_list = list()
    for i in nameList :
        parameter_list.append( varNames[i] )
    return parameter_list

theTree = tree.DecisionTreeRegressor( max_depth=4 )
theTree = theTree.fit( X, Y )
treeList = getTreeVars( theTree, varNames )
print( treeList )
```

```
['Unnamed: 0', 'LOAN', 'IMP_YOJ', 'IMP_DELIQ', 'IMP_CLNO', 'M_DEBTINC', 'IMP_DEBTINC']
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
```

Random Forest

```
In [33]: WHO = "REG_RF"

print("\n\n")
RF_flag = []
for i in vars_RF_flag :
    print(i)
    theVar = i[0]
    RF_flag.append( theVar )

print("\n\n")
RF_amt = []
for i in vars_RF_amt :
    print(i)
    theVar = i[0]
    RF_amt.append( theVar )

CLM = LogisticRegression( solver='newton-cg', max_iter=1000 )
CLM = CLM.fit( X_train[RF_flag], Y_train[ FLAG ] )

TRAIN_CLM = getProbAccuracyScores( WHO + "_Train", CLM, X_train[RF_flag], Y_train[ FLAG ] )
TEST_CLM = getProbAccuracyScores( WHO, CLM, X_test[RF_flag], Y_test[ FLAG ] )

print_ROC_Curve( WHO, [ TRAIN_CLM, TEST_CLM ] )
```

```

print_Accuracy( WHO + " CLASSIFICATION ACCURACY", [ TRAIN_CLM, TEST_CLM ] )

#Regression RF LOSS

AMT = LinearRegression()
AMT = AMT.fit( W_train[RF_amt], Z_train[LOSS] )

TRAIN_AMT = getAmtAccuracyScores( WHO + "_Train", AMT, W_train[RF_amt], Z_train[LOSS] )
TEST_AMT = getAmtAccuracyScores( WHO, AMT, W_test[RF_amt], Z_test[LOSS] )
print_Accuracy( WHO + " RMSE ACCURACY", [ TRAIN_AMT, TEST_AMT ] )

REG_RF_CLM_COEF = getCoefLogit( CLM, X_train[RF_flag] )
REG_RF_AMT_COEF = getCoefLinear( AMT, X_train[RF_amt] )

REG_RF_CLM = TEST_CLM.copy()
REG_RF_AMT = TEST_AMT.copy()

('M_DEBTINC', 100)
('IMP_DEBTINC', 59)
('IMP_CLAGE', 37)
('IMP_DELIQ', 33)
('Unnamed: 0', 32)
('IMP_VALUE', 30)
('IMP_CLNO', 27)
('IMP_MORTDUE', 27)
('LOAN', 25)
('IMP_Y0J', 23)

('Unnamed: 0', 100)
('M_DEBTINC', 90)
('IMP_DEBTINC', 39)
('IMP_DELIQ', 36)
('IMP_CLAGE', 31)
('IMP_CLNO', 23)
('LOAN', 22)
('IMP_VALUE', 15)

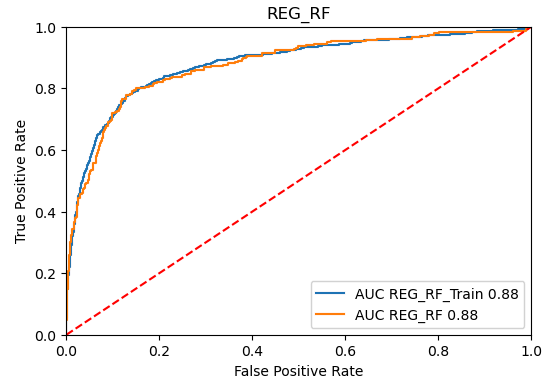
```

file:///C:/Users/jamia/Downloads/Assignment 2 (2).html

```

if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):

```



```

C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
REG_RF CLASSIFICATION ACCURACY
=====
REG_RF_Train = 0.8745805369127517
REG_RF = 0.8657718120805369
-----

```

```

REG_RF RMSE ACCURACY
=====
REG_RF_Train = 3859.486871231465
REG_RF = 3845.8514817467126
-----

```

```

FLAG
-----
Total Variables: 11
INTERCEPT = -4.424805758618589
M_DEBTINC = 2.767422452997574
IMP_DEBTINC = 0.08929834073884058
IMP_CLAGE = -0.006063300572911903
IMP_DELIQ = 0.7319345819448747
Unnamed: 0 = -9.276597190858906e-05
IMP_VALUE = 3.1671116342963647e-06
IMP_CLNO = -0.007430708110113477
IMP_MORTDUE = -3.169478983925164e-06
LOAN = 4.864688405827578e-06
IMP_YOJ = -0.02002740757298695

```

```

LOSS
-----
Total Variables: 9
INTERCEPT = -828.9433818804428
Unnamed: 0 = 0.6609921676553199
M_DEBTINC = 4945.216367851677
IMP_DEBTINC = 71.01238638641999
IMP_DELIQ = 1238.026691409141
IMP_CLAGE = -8.467223272849122
IMP_CLNO = 47.372886643776404
LOAN = 0.00062665488180444
IMP_VALUE = 0.0027837209280382686

```

```
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
```

Gradient Boosting

```
In [34]: WHO = "REG_GB"

print("\n\n")
GB_flag = []
for i in vars_GB_flag :
    print(i)
    theVar = i[0]
    GB_flag.append( theVar )

print("\n\n")
GB_amt = []
for i in vars_GB_amt :
    print(i)
    theVar = i[0]
    GB_amt.append( theVar )

CLM = LogisticRegression( solver='newton-cg', max_iter=1000 )
CLM = CLM.fit( X_train[GB_flag], Y_train[ FLAG ] )

TRAIN_CLM = getProbAccuracyScores( WHO + "_Train", CLM, X_train[GB_flag], Y_train[ FLAG ] )
TEST_CLM = getProbAccuracyScores( WHO, CLM, X_test[GB_flag], Y_test[ FLAG ] )

print_ROC_Curve( WHO, [ TRAIN_CLM, TEST_CLM ] )
print_Accuracy( WHO + " CLASSIFICATION ACCURACY", [ TRAIN_CLM, TEST_CLM ] )

#Regression or GB LOSS

AMT = LinearRegression()
AMT = AMT.fit( W_train[GB_amt], Z_train[LOSS] )

TRAIN_AMT = getAmtAccuracyScores( WHO + "_Train", AMT, W_train[GB_amt], Z_train[LOSS] )
TEST_AMT = getAmtAccuracyScores( WHO, AMT, W_test[GB_amt], Z_test[LOSS] )
print_Accuracy( WHO + " RMSE ACCURACY", [ TRAIN_AMT, TEST_AMT ] )

REG_GB_CLM_COEF = getCofLogit( CLM, X_train[GB_flag] )
REG_GB_AMT_COEF = getCofLinear( AMT, X_train[GB_amt] )

REG_GB_CLM = TEST_CLM.copy()
REG_GB_AMT = TEST_AMT.copy()

('M_DEBTINC', 100)
('IMP_DEBTINC', 29)
('IMP_DELIQ', 19)
('IMP_CLAGE', 16)
('IMP_DEROG', 7)

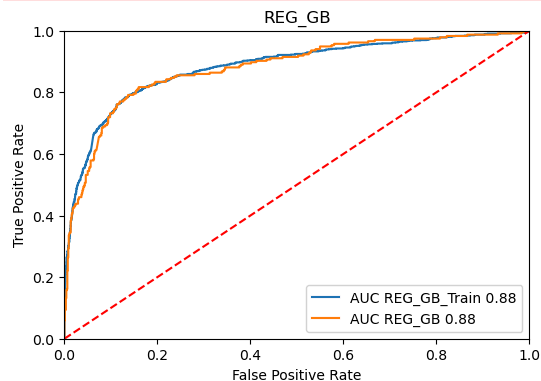
('M_DEBTINC', 100)
('Unnamed: 0', 92)
('IMP_DELIQ', 44)
('IMP_DEBTINC', 36)
('IMP_CLAGE', 16)
('IMP_CLNO', 14)
('IMP_DEROG', 12)
```

file:///C:/Users/jamia/Downloads/Assignment 2 (2).html

```

instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):

```



```

REG_GB CLASSIFICATION ACCURACY
=====
REG_GB_Train = 0.8766778523489933
REG_GB      = 0.8708053691275168
-----

```

```

REG_GB RMSE ACCURACY
=====
REG_GB_Train = 3788.5348779902947
REG_GB      = 3850.885647456625
-----

```

```

FLAG
-----
Total Variables: 6
INTERCEPT = -4.904836811894744
M_DEBTINC   = 2.760147831230137
IMP_DEBTINC = 0.08948343127742956
IMP_DELTINC = 0.66320967870209
IMP_CLAGE   = -0.006427666524814138
IMP_DEROG   = 0.5706988321413794

```

```

LOSS
-----
Total Variables: 8
INTERCEPT = -918.3303690087341
M_DEBTINC   = 4673.87885657157
Unnamed: 0  = 0.6788285148069945
IMP_DELTINC = 1132.1771879487108
IMP_DEBTINC = 72.7061873797417
IMP_CLAGE   = -7.536097089895807
IMP_CLNO    = 46.846302214793866
IMP_DEROG   = 959.3183391334528

```

```

C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):

```

Stepwise Selection

```

In [35]: U_train = X_train[ vars_tree_flag ]
stepVarNames = list( U_train.columns.values )
maxCols = U_train.shape[1]

sfs = SFS( LogisticRegression( solver='newton-cg', max_iter=1000 ),
           k_features=( 1, maxCols ),
           forward=True,
           floating=False,
           cv=3
           )
sfs.fit(U_train.values, Y_train[ FLAG ].values)

theFigure = plot_sfs(sfs.get_metric_dict(), kind=None )
plt.title('Default PROBABILITY Sequential Forward Selection (w. StdErr)')
plt.grid()
plt.show()

#get metrics, what variables it liked and average error
dfm = pd.DataFrame.from_dict( sfs.get_metric_dict()).T
dfm_names = dfm.columns.values
print( dfm_names )
dfm = dfm[ ['feature_names', 'avg_score'] ]

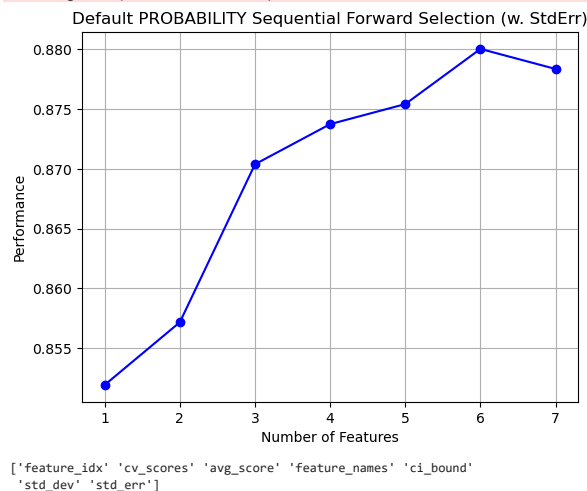
```


[illegible]

```

C:\Anaconda\Lib\site-packages\sklearn\utils\optimize.py:203: UserWarning: Line Search failed
warnings.warn("Line Search failed")
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:466: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:314: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:425: LineSearchWarning: Rounding errors prevent the line search from converging
warn(msg, LineSearchWarning)
C:\Anaconda\Lib\site-packages\sklearn\utils\optimize.py:203: UserWarning: Line Search failed
warnings.warn("Line Search failed")
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:466: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:314: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:425: LineSearchWarning: Rounding errors prevent the line search from converging
warn(msg, LineSearchWarning)
C:\Anaconda\Lib\site-packages\sklearn\utils\optimize.py:203: UserWarning: Line Search failed
warnings.warn("Line Search failed")
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:466: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:314: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:425: LineSearchWarning: Rounding errors prevent the line search from converging
warn(msg, LineSearchWarning)
C:\Anaconda\Lib\site-packages\sklearn\utils\optimize.py:203: UserWarning: Line Search failed
warnings.warn("Line Search failed")
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:466: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:314: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:425: LineSearchWarning: Rounding errors prevent the line search from converging
warn(msg, LineSearchWarning)
C:\Anaconda\Lib\site-packages\sklearn\utils\optimize.py:203: UserWarning: Line Search failed
warnings.warn("Line Search failed")
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:466: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:314: LineSearchWarning: The line search algorithm did not converge
warn('The line search algorithm did not converge', LineSearchWarning)
C:\Anaconda\Lib\site-packages\scipy\optimize\_linesearch.py:425: LineSearchWarning: Rounding errors prevent the line search from converging
warn(msg, LineSearchWarning)
C:\Anaconda\Lib\site-packages\sklearn\utils\optimize.py:203: UserWarning: Line Search failed
warnings.warn("Line Search failed")

```



In [36]:

dfm

Out[36]:

	feature_names	avg_score
1	(5,)	0.851927
2	(4, 5)	0.857171
3	(3, 4, 5)	0.870384
4	(3, 4, 5, 6)	0.87374
5	(2, 3, 4, 5, 6)	0.875417
6	(1, 2, 3, 4, 5, 6)	0.88003
7	(0, 1, 2, 3, 4, 5, 6)	0.878352

In [37]:

```

print(dfm.dtypes)
dfm.avg_score = dfm.avg_score.astype(float)
print(dfm.dtypes)
print(dfm)

```

```

feature_names      object
avg_score           object
dtype: object
feature_names      object
avg_score           float64
dtype: object
      feature_names  avg_score
1              (5,)    0.851927
2              (4, 5)    0.857171
3            (3, 4, 5)    0.870384
4          (3, 4, 5, 6)    0.873740
5        (2, 3, 4, 5, 6)    0.875417
6      (1, 2, 3, 4, 5, 6)    0.880030
7    (0, 1, 2, 3, 4, 5, 6)    0.878352

```

```

In [38]: def getVariables( DFM, INDEX, NAMES ) :
          theVars = DFM.iloc[ INDEX, ]
          theVars = theVars.feature_names
          print( INDEX,"=",theVars )
          theVarNames = []
          for i in theVars :
              index = int(i)
              try :
                  theName = NAMES[ index ]
                  theVarNames.append( theName )
              except :
                  pass
          return theVarNames

          print(" ..... ")

          maxIndex = dfm.avg_score.argmax()
          print( maxIndex )
          regList = getVariables( dfm, maxIndex, varNames )
          print( regList )

          .....
          5 = ('1', '2', '3', '4', '5', '6')
          ['LOAN', 'OHE_REASON_HOMEIMP', 'OHE_REASON_DEBTCON', 'OHE_REASON_MISSING', 'OHE_JOB_OFFICE', 'OHE_JOB_OTHER']

```

```

In [41]: stepVars = dfm.iloc[ maxIndex, ]
          stepVars = stepVars.feature_names
          print( stepVars )

          finalStepVars = []
          for i in stepVars :
              index = int(i)
              try :
                  theName = stepVarNames[ index ]
                  finalStepVars.append( theName )
              except :
                  pass

          for i in finalStepVars :
              print(i)

          U_train = X_train[ finalStepVars ]
          U_test = X_test[ finalStepVars ]

          ('1', '2', '3', '4', '5', '6')
          M_DEROG
          IMP_DEROG
          IMP_DELIQ
          IMP_CLAGE
          M_DEBTINC
          IMP_DEBTINC

```

```

In [44]: V_train = W_train[ GB_amt ]
          stepVarNames = list( V_train.columns.values )
          maxCols = V_train.shape[1]

          sfs = SFS( LinearRegression(),
                      k_features=( 1, maxCols ),
                      forward=True,
                      floating=False,
                      scoring = 'r2',
                      cv=5
                    )
          sfs.fit(V_train.values, Z_train[ LOSS ].values)

          theFigure = plot_sfs(sfs.get_metric_dict(), kind=None )
          plt.title('LOSS Forward Selection')
          plt.grid()
          plt.show()

          dfm = pd.DataFrame.from_dict( sfs.get_metric_dict()).T
          dfm = dfm[ ['feature_names', 'avg_score'] ]
          dfm.avg_score = dfm.avg_score.astype(float)

          print(" ..... ")
          maxIndex = dfm.avg_score.argmax()
          print("argmax")
          print( dfm.iloc[ maxIndex, ] )
          print(" ..... ")

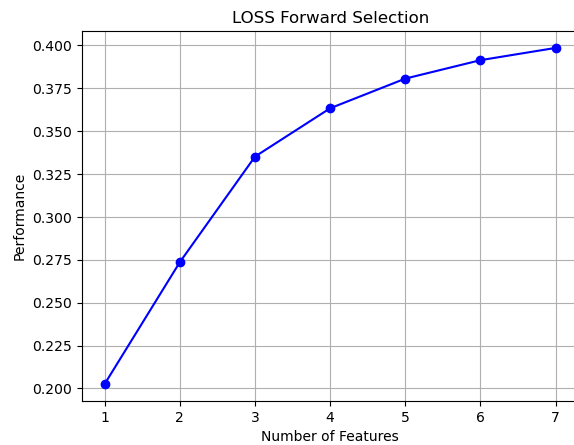
          stepVars = dfm.iloc[ maxIndex, ]
          stepVars = stepVars.feature_names
          print( stepVars )

          finalStepVars = []
          for i in stepVars :
              index = int(i)
              try :
                  theName = stepVarNames[ index ]
                  finalStepVars.append( theName )
              except :
                  pass

          for i in finalStepVars :
              print(i)

```

```
V_train = W_train[ finalStepVars ]
V_test = W_test[ finalStepVars ]
```



```
.....
argmax
feature_names      (0, 1, 2, 3, 4, 5, 6)
avg_score          0.398375
Name: 7, dtype: object
.....
('0', '1', '2', '3', '4', '5', '6')
M_DEBTINC
Unnamed: 0
IMP_DELTINC
IMP_DEBTINC
IMP_CLAGE
IMP_CLNO
IMP_DEROG
```

```
In [46]: WHO = "REG_STEPWISE"

CLM = LogisticRegression( solver='newton-cg', max_iter=1000 )
CLM = CLM.fit( U_train, Y_train[ FLAG ] )

TRAIN_CLM = getProbAccuracyScores( WHO + "_Train", CLM, U_train, Y_train[ FLAG ] )
TEST_CLM = getProbAccuracyScores( WHO, CLM, U_test, Y_test[ FLAG ] )

print_ROC_Curve( WHO, [ TRAIN_CLM, TEST_CLM ] )
print_Accuracy( WHO + " CLASSIFICATION ACCURACY", [ TRAIN_CLM, TEST_CLM ] )

# DAMAGES

AMT = LinearRegression()
AMT = AMT.fit( V_train, Z_train[FLAG] )

TRAIN_AMT = getAmtAccuracyScores( WHO + "_Train", AMT, V_train, Z_train[LOSS] )
TEST_AMT = getAmtAccuracyScores( WHO, AMT, V_test, Z_test[FLAG] )
print_Accuracy( WHO + " RMSE ACCURACY", [ TRAIN_AMT, TEST_AMT ] )

REG_STEP_CLM_COEF = getCofLogit( CLM, U_train )
REG_STEP_AMT_COEF = getCofLinear( AMT, V_train )

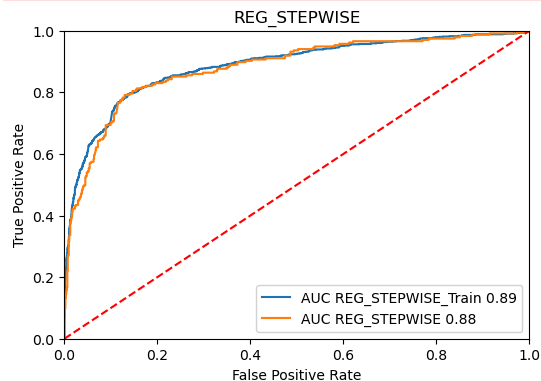
REG_STEP_CLM = TEST_CLM.copy()
REG_STEP_AMT = TEST_AMT.copy()
```

file:///C:/Users/jamia/Downloads/Assignment 2 (2).html

```

instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
  if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)`
instead.
  if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):

```



REG_STEPWISE CLASSIFICATION ACCURACY

```

=====
REG_STEPWISE_Train = 0.8802432885906041
REG_STEPWISE      = 0.8666107382550335
-----

```

REG_STEPWISE RMSE ACCURACY

```

=====
REG_STEPWISE_Train = 6926.149285430292
REG_STEPWISE      = 0.31089901163192085
-----

```

FLAG

```

-----
Total Variables: 7
INTERCEPT = -4.855706069405142
M_DEROG    = -0.7136124293178427
IMP_DEROG  = 0.5359661653690089
IMP_DELIHQ = 0.6841654129272896
IMP_CLAGE  = -0.006326639479117751
M_DEBTINC  = 2.773981174414131
IMP_DEBTINC = 0.08953090425878735

```

LOSS

```

-----
Total Variables: 8
INTERCEPT = -0.023145071413099322
M_DEBTINC   = 0.4409632154020521
Unnamed: 0  = -8.000769509228112e-06
IMP_DELIHQ  = 0.08121939721255265
IMP_DEBTINC = 0.0065440419659819815
IMP_CLAGE   = -0.00046950884511877205
IMP_CLNO    = -0.0015854317038726576
IMP_DEROG   = 0.0746411235248708

```

```

C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
C:\Anaconda\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):

```

```

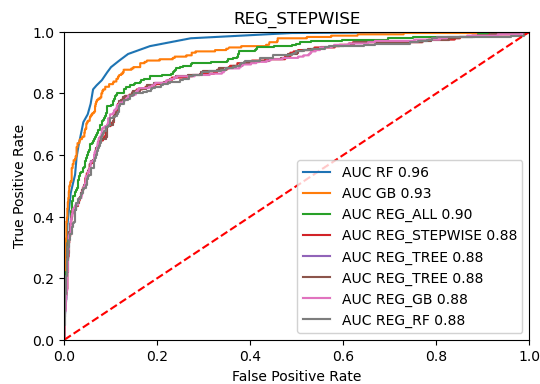
In [48]: ALL_CLM = [ REG_TREE_CLM, RF_CLM, GB_CLM, REG_ALL_CLM, REG_TREE_CLM, REG_RF_CLM, REG_GB_CLM, REG_STEP_CLM ]

ALL_CLM = sorted( ALL_CLM, key = lambda x: x[4], reverse=True )
print_ROC_Curve( WHO, ALL_CLM )

ALL_CLM = sorted( ALL_CLM, key = lambda x: x[1], reverse=True )
print_Accuracy( "ALL FLAG ACCURACY", ALL_CLM )

ALL_AMT = [ TREE_AMT, RF_AMT, GB_AMT, REG_ALL_AMT, REG_TREE_AMT, REG_RF_AMT, REG_GB_AMT, REG_STEP_AMT ]
ALL_AMT = sorted( ALL_AMT, key = lambda x: x[1] )
print_Accuracy( "ALL LOSS ACCURACY", ALL_AMT )

```



```

ALL FLAG ACCURACY
=====
RF      = 0.9093959731543624
GB      = 0.8993288590604027
REG_ALL = 0.8808724832214765
REG_GB  = 0.8708053691275168
REG_STEPWISE = 0.8666107382550335
REG_TREE = 0.8666107382550335
REG_TREE = 0.8666107382550335
REG_RF  = 0.8657718120805369
-----

ALL LOSS ACCURACY
=====
REG_STEPWISE = 0.31089901163192085
RF      = 2730.1689152417407
GB      = 2767.579215901141
TREE    = 3301.3240853093553
REG_ALL  = 3754.7356551934763
REG_RF   = 3845.8514817467126
REG_TREE = 3850.540793930161
REG_GB   = 3850.885647456625
-----

```