Configuration

In [1]:
```
# Parameters
ENABLE_COLAB = False

PROJECT_NAME = 'ML1010-Group-Project'
EXPERIMENT_NAME = 'ReviewText_Lemma_TFIDF2_FullData (LR)'
FILE_NAME = '01_ML1010_GP_LR_TFIDF2_FullData'
LOAD_FROM_EXP = False

#Root Machine Learning Directory. Projects appear underneath
GOOGLE_DRIVE_MOUNT = '/content/gdrive'
COLAB_ROOT_DIR = GOOGLE_DRIVE_MOUNT + '/MyDrive/Colab Notebooks'
COLAB_INIT_DIR = COLAB_ROOT_DIR + '/utility_files'

LOCAL_ROOT_DIR = '/home/magni//ML_Root/project_root'
LOCAL_INIT_DIR = LOCAL_ROOT_DIR + '/utility_files'
```

Bootstrap Environment

In [2]:
```
#add in support for utility file directory and importing
import sys
import os

if ENABLE_COLAB:
  #Need access to drive
  from google.colab import drive
  drive.mount(GOOGLE_DRIVE_MOUNT, force_remount=True)

  #add in utility directory to syspath to import
  INIT_DIR = COLAB_INIT_DIR
  sys.path.append(os.path.abspath(INIT_DIR))

  #Config environment variables
  ROOT_DIR = COLAB_ROOT_DIR

else:
  #add in utility directory to syspath to import
  INIT_DIR = LOCAL_INIT_DIR
  sys.path.append(os.path.abspath(INIT_DIR))

  #Config environment variables
  ROOT_DIR = LOCAL_ROOT_DIR

#Import Utility Support
from jarvis import Jarvis
jarvis = Jarvis(ROOT_DIR, PROJECT_NAME)

import mv_python_utils as mvutils
```

```
Wha...where am I?
I am awake now.
```

```
I have set your current working directory to /home/magni/ML_Root/project_root
/ML1010-Group-Project
The current time is 21:13
Hello sir. I see you are having a productive evening.
```

Setup Runtime Environment

In [3]:

```
if ENABLE_COLAB:
  #!pip install scipy -q
  #!pip install scikit-learn -q
  #!pip install pycaret -q
  #!pip install matplotlib -q
  #!pip install joblib -q
  #!pip install pandasql -q
  !pip install umap_learn -q
  !pip install sentence_transformers -q
  !pip install spacytextblob -q
  !pip install flair -q
  display('Google Colab enabled')
else:
  display('Google Colab not enabled')

#Common imports
import json
import pandas as pd
import numpy as np
import matplotlib
import re
import nltk
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn import metrics
from sklearn.datasets import load_digits
from sklearn.model_selection import train_test_split as tts
#from yellowbrick.classifier import ConfusionMatrix
#from sklearn.linear_model import LogisticRegression
from yellowbrick.target import ClassBalance
from xgboost import XGBClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier

nltk.download('stopwords')
%matplotlib inline
```

```
'Google Colab not enabled'
```

```
[nltk_data] Downloading package stopwords to /home/magni/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

In [4]:
```
import importlib
import cw_df_metric_utils as cwutils
import DataPackage as dp
import DataPackageSupport as dps
import DataExperiment
import DataExperimentSupport
```

```
2022-01-26 21:13:25.501006: W tensorflow/stream_executor/platform/default/dso
_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0'; dlerror: l
ibcudart.so.11.0: cannot open shared object file: No such file or directory
2022-01-26 21:13:25.501037: I tensorflow/stream_executor/cuda/cudart_stub.cc:
29] Ignore above cudart dlerror if you do not have a GPU set up on your machi
ne.
```

In [5]:
```
importlib.reload(dp)
importlib.reload(dps)
importlib.reload(DataExperiment)
importlib.reload(DataExperimentSupport)
```

Out[5]:
```
<module 'DataExperimentSupport' from '/home/magni/ML_Root/project_root/utilit
y_files/DataExperimentSupport.py'>
```

Load Data

In [6]:
```
from sklearn.linear_model import LogisticRegression
#axis_labels=[1,2,3,4,5]
axis_labels=[0,1]
#classifier = XGBClassifier(eval_metric='mlogloss')
classifier = LogisticRegression(verbose=0)
#ANALSYSIS_COL = 'reviewText_lemma_bert'
UNIQUE_COL = 'uuid'
TARGET_COL = 'overall_posneg'
```

In [7]:
```
if LOAD_FROM_EXP:
    #start from saved state
    myExp = jarvis.loadExperiment(FILE_NAME)
    myExp.display()

else:
    #start from source file and regenerate
    testDf = pd.read_pickle(jarvis.DATA_DIR_EXP + '/reviewText_TF-IDF_Full.pk

    myExp = DataExperiment.DataExperiment(projectName=PROJECT_NAME,
                                          experimentName=EXPERIMENT_NAME,
                                          origData=testDf,
                                          uniqueColumn=UNIQUE_COL,
                                          targetColumn=TARGET_COL,
                                          classifier=classifier)
```

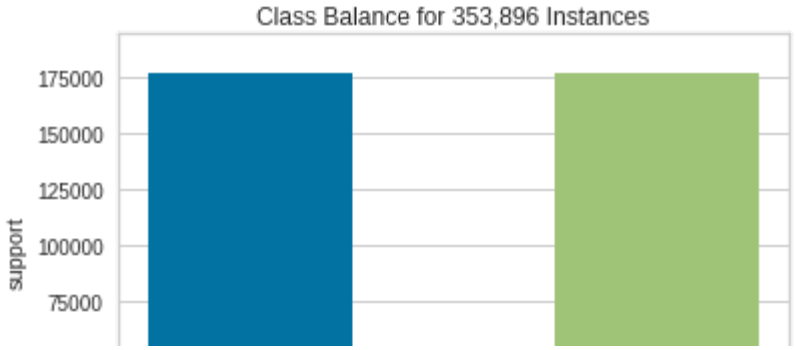
```
DataExperiment summary:
---> projectName: ML1010-Group-Project
---> experimentName: ReviewText_Lemma_TFIDF2_FullData (LR)
---> isDataPackageLoaded: True
```

```
---> isBaseModelLoaded: False
---> isBaseModelPredicted: False
---> isBaseModelLearningCurveCreated: False
---> isFinalModelLoaded: False
---> isFinalModelPredicted: False
---> isFinalModelLearningCurveCreated: False
---> isClassifierLoaded: True
LogisticRegression()

    DataPackage summary:
    Attributes:
    ---> uniqueColumn: uuid
    ---> targetColumn: overall_posneg
    Process:
    ---> isBalanced: False
    ---> isTrainTestSplit: False
    Data:
    ---> isOrigDataLoaded: True
    ---> isTrainDataLoaded: False
    ---> isTestDataLoaded: False
```

In [8]:

```
#myExp.processDataPackage()
myExp.dataPackage.classBalanceUndersample()
myExp.dataPackage.splitTrainTest()
```


```
Undersampling data to match min class: 0 of size: 176948
```

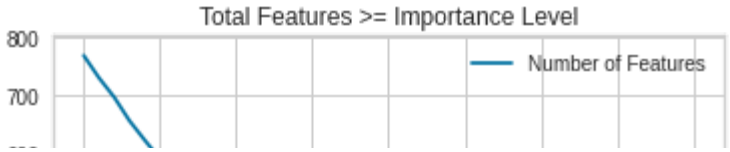


In [9]:
```
%%time
myExp.createBaseModel()
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
Base Model Stats:
Accuracy: 0.85
Precision: 0.85
Recalll: 0.85
F1 Score: 0.85
Cohen kappa:: 0.7
CPU times: user 2min 23s, sys: 25.1 s, total: 2min 48s
Wall time: 18.9 s
```

In [10]:
```
impFeatures = myExp.analyzeBaseModelFeatureImportance(returnAbove=1,
                                                      increment=0.1,
                                                     upperValue=4)
```

```
  0%|          | 0/42 [00:00<?, ?it/s]
Feature Importance Summary:
---> Original feature count: 768
---> Returned feature count: 403
---> Removed feature count: 365
---> Return items above (including): 1
```



In [11]:
```
%%time
myExp.createFinalModel(featureImportanceThreshold=0.5)
```

```
  0%|          | 0/101 [00:00<?, ?it/s]
  0%|          | 0/101 [00:00<?, ?it/s]
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
Final Model Stats:
Accuracy: 0.83
Precision: 0.83
Recalll: 0.83
F1 Score: 0.83
Cohen kappa:: 0.65
CPU times: user 2min 2s, sys: 20.1 s, total: 2min 23s
Wall time: 16.1 s
```

In [12]:
```
%%time
myExp.createBaseModelLearningCurve()
```

```
[learning_curve] Training set sizes: [ 22649  45298 113246 226492]
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent worke
rs.
[CV] END ....................., score=(train=0.858, test=0.845) total time=
1.0s
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    1.0s remaining:    0.
0s
[CV] END ....................., score=(train=0.856, test=0.848) total time=
2.1s
[Parallel(n_jobs=1)]: Done   2 out of   2 | elapsed:    3.1s remaining:    0.
0s
[CV] END ....................., score=(train=0.854, test=0.851) total time=
5.2s
[Parallel(n_jobs=1)]: Done   3 out of   3 | elapsed:    8.5s remaining:    0.
0s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
```

```
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[CV] END ....................., score=(train=0.853, test=0.852) total time=  1
4.9s
[CV] END ....................., score=(train=0.858, test=0.843) total time=
1.0s
[CV] END ....................., score=(train=0.854, test=0.846) total time=
2.8s
[CV] END ....................., score=(train=0.853, test=0.848) total time=
4.8s
[CV] END ....................., score=(train=0.852, test=0.848) total time=  1
2.4s
[CV] END ....................., score=(train=0.855, test=0.848) total time=
1.2s
[CV] END ....................., score=(train=0.853, test=0.850) total time=
2.0s
[CV] END ....................., score=(train=0.853, test=0.851) total time=
6.0s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[CV] END ....................., score=(train=0.851, test=0.852) total time=  1
4.4s
[CV] END ....................., score=(train=0.855, test=0.843) total time=
1.3s
[CV] END ....................., score=(train=0.853, test=0.847) total time=
2.1s
[CV] END ....................., score=(train=0.853, test=0.849) total time=
6.1s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[CV] END ....................., score=(train=0.852, test=0.850) total time=  1
4.1s
[CV] END ....................., score=(train=0.858, test=0.844) total time=
0.9s
[CV] END ....................., score=(train=0.854, test=0.848) total time=
2.2s
[CV] END ....................., score=(train=0.853, test=0.850) total time=
6.0s
[CV] END ....................., score=(train=0.852, test=0.850) total time=  1
```

```
5.0s
CPU times: user 15min 23s, sys: 2min 21s, total: 17min 45s
Wall time: 1min 57s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[Parallel(n_jobs=1)]: Done  20 out of  20 | elapsed:  1.9min finished
```

In [13]:

```
%%time
myExp.createFinalModelLearningCurve()
```

```
[learning_curve] Training set sizes: [ 22649  45298 113246 226492]
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent worke
rs.
[CV] END ....................., score=(train=0.829, test=0.817) total time=
0.9s
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    0.9s remaining:    0.
0s
[CV] END ....................., score=(train=0.825, test=0.820) total time=
1.7s
[Parallel(n_jobs=1)]: Done   2 out of   2 | elapsed:    2.6s remaining:    0.
0s
[CV] END ....................., score=(train=0.823, test=0.822) total time=
4.4s
[Parallel(n_jobs=1)]: Done   3 out of   3 | elapsed:    7.1s remaining:    0.
0s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[CV] END ....................., score=(train=0.824, test=0.822) total time=  1
2.4s
[CV] END ....................., score=(train=0.829, test=0.816) total time=
0.8s
[CV] END ....................., score=(train=0.826, test=0.820) total time=
2.2s
[CV] END ....................., score=(train=0.825, test=0.821) total time=
5.5s
[CV] END ....................., score=(train=0.824, test=0.822) total time=  1
2.0s
[CV] END ....................., score=(train=0.829, test=0.821) total time=
```

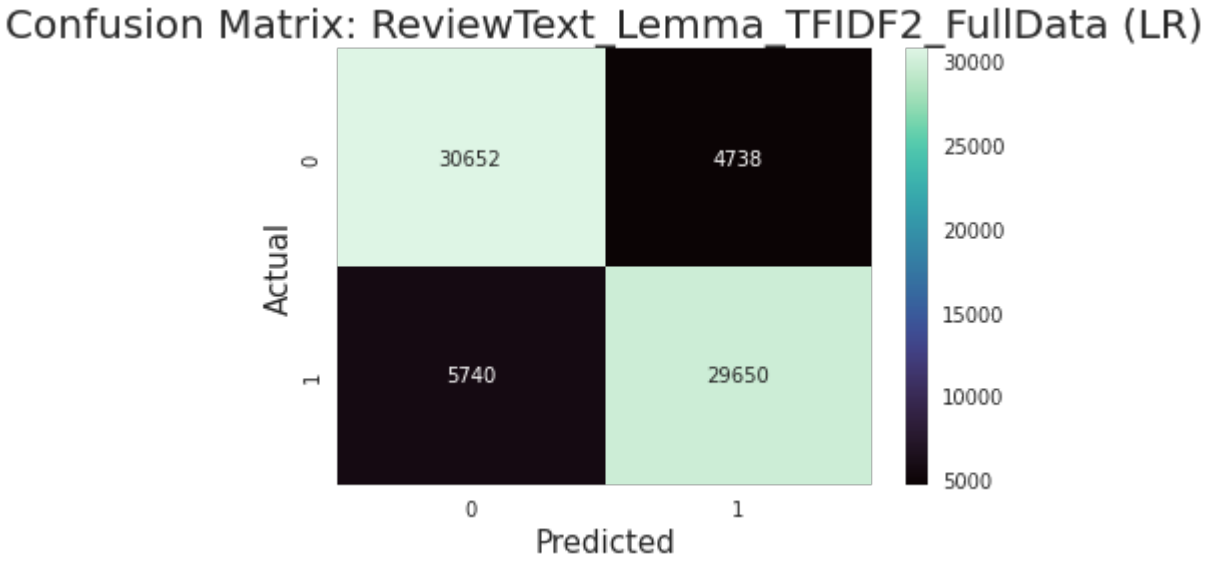
```
0.8s
[CV] END ....................., score=(train=0.826, test=0.823) total time=
1.9s
[CV] END ....................., score=(train=0.824, test=0.824) total time=
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[CV] END ....................., score=(train=0.823, test=0.825) total time=  1
2.5s
[CV] END ....................., score=(train=0.825, test=0.820) total time=
1.0s
[CV] END ....................., score=(train=0.825, test=0.822) total time=
2.3s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[CV] END ....................., score=(train=0.825, test=0.823) total time=
6.1s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[CV] END ....................., score=(train=0.824, test=0.824) total time=  1
2.1s
[CV] END ....................., score=(train=0.828, test=0.817) total time=
0.7s
[CV] END ....................., score=(train=0.826, test=0.818) total time=
1.9s
[CV] END ....................., score=(train=0.825, test=0.820) total time=
5.6s
[CV] END ....................., score=(train=0.825, test=0.820) total time=  1
2.7s
CPU times: user 13min 27s, sys: 2min 24s, total: 15min 52s
Wall time: 1min 43s
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra warning msg= LOGISTIC SOLVER CONVERGENCE MSG,
```

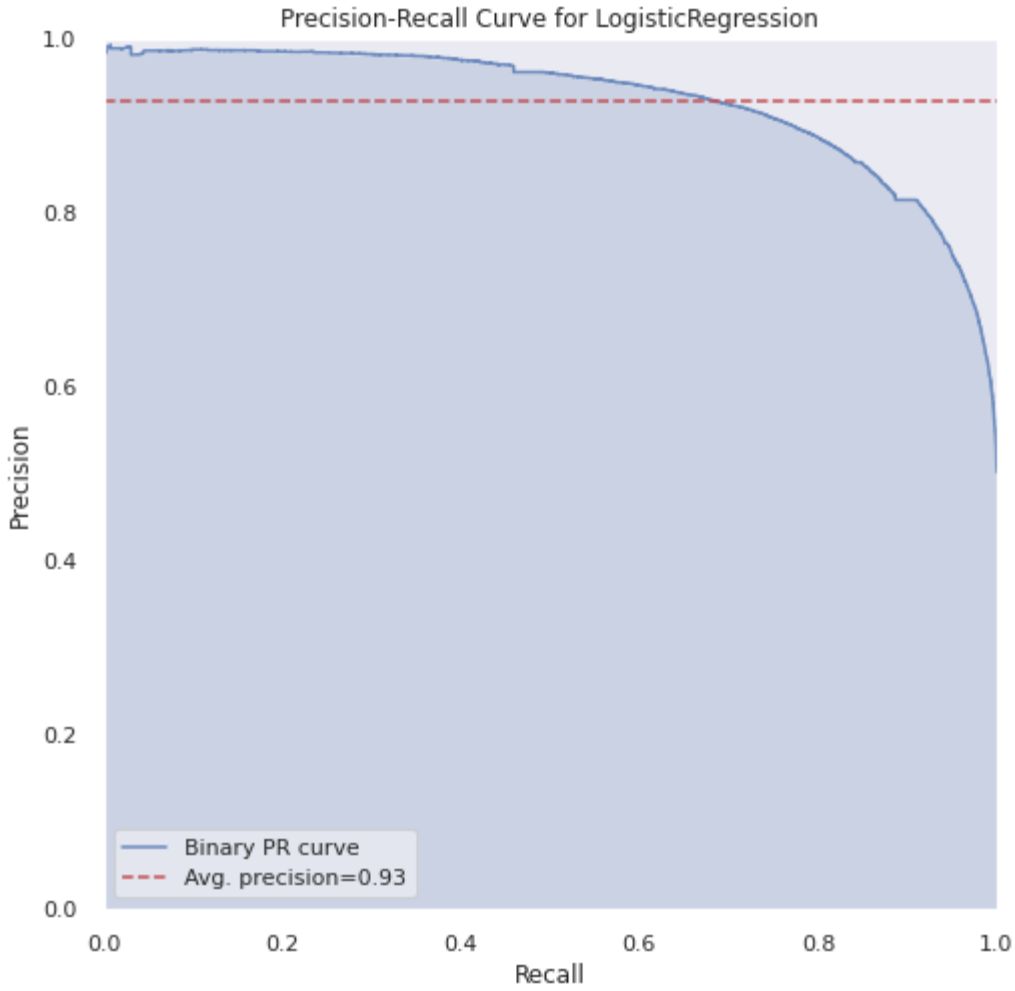
In [14]:

```
myExp.showBaseModelReport(axis_labels,
                          upperValue=0.025,
                          topn=10)
```

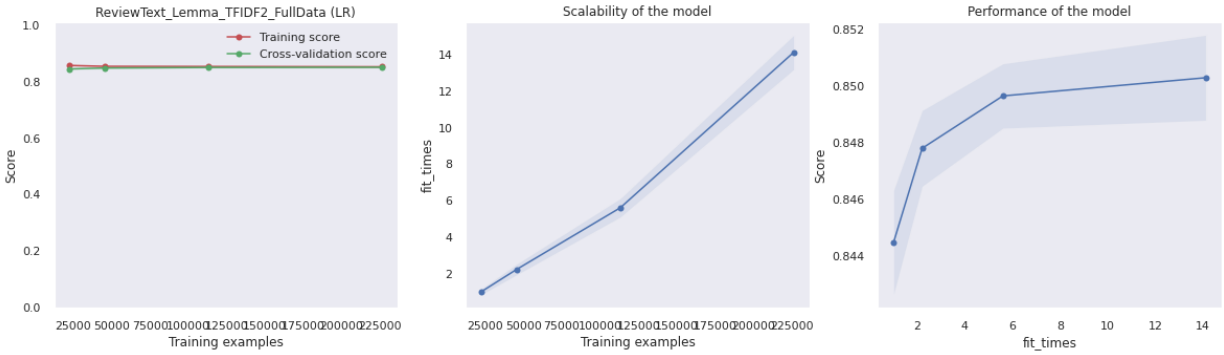
```
Base Model Stats:
Accuracy: 0.85
Precision: 0.85
Recalll: 0.85
F1 Score: 0.85
Cohen kappa:: 0.7
              precision    recall  f1-score   support

           0       0.84      0.87      0.85     35390
           1       0.86      0.84      0.85     35390

    accuracy                           0.85     70780
   macro avg       0.85      0.85      0.85     70780
weighted avg       0.85      0.85      0.85     70780
```




```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/bas
e.py:444: UserWarning: X has feature names, but LogisticRegression was fitted
without feature names
  f"X has feature names, but {self.__class__.__name__} was fitted without"
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/bas
e.py:444: UserWarning: X has feature names, but LogisticRegression was fitted
without feature names
  f"X has feature names, but {self.__class__.__name__} was fitted without"
```



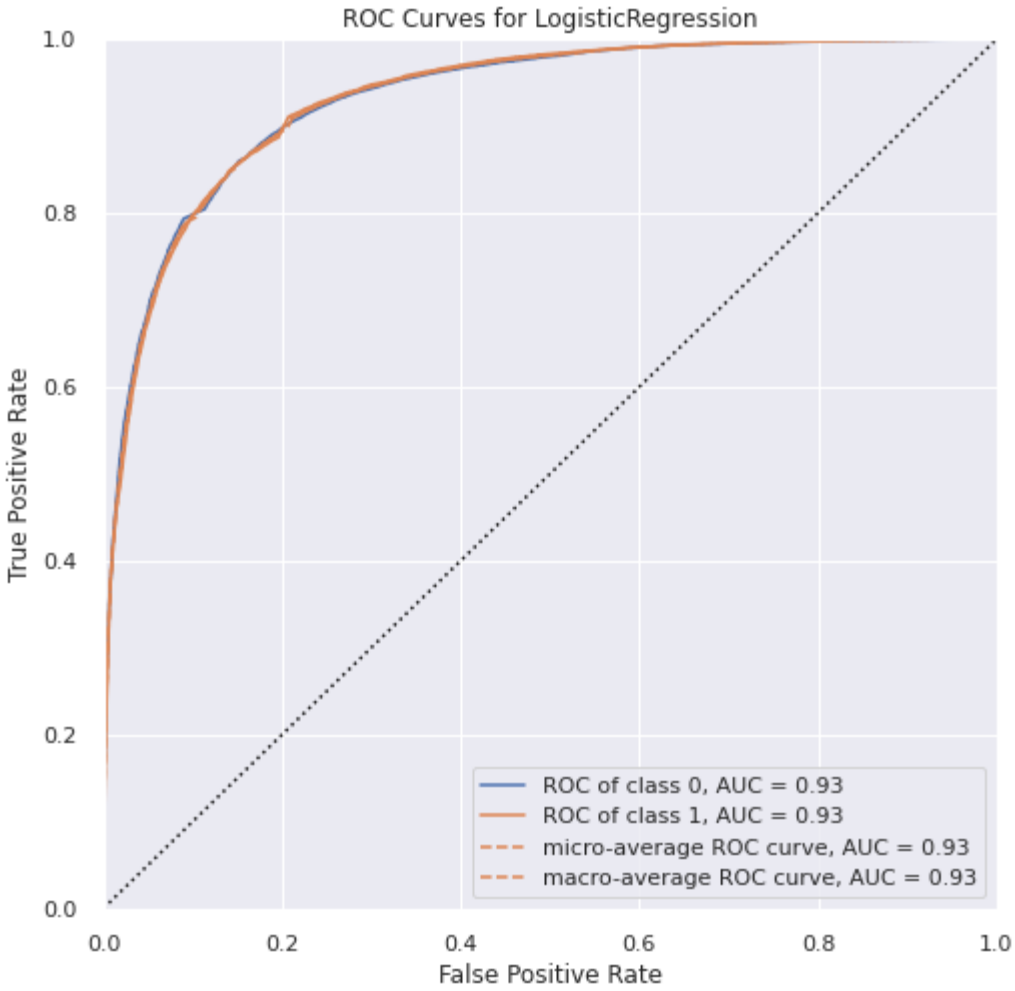
<Figure size 576x576 with 0 Axes>



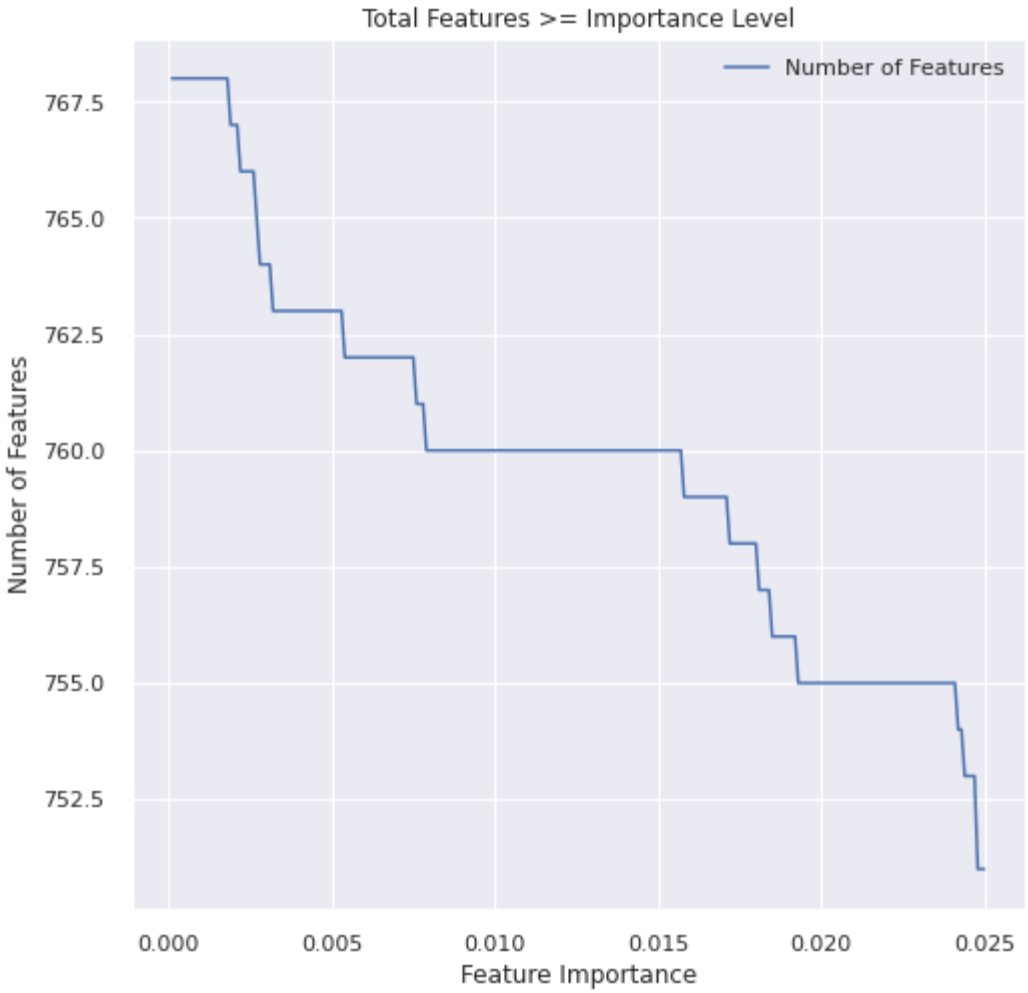
Base model ROCAUC not calculated. Starting now

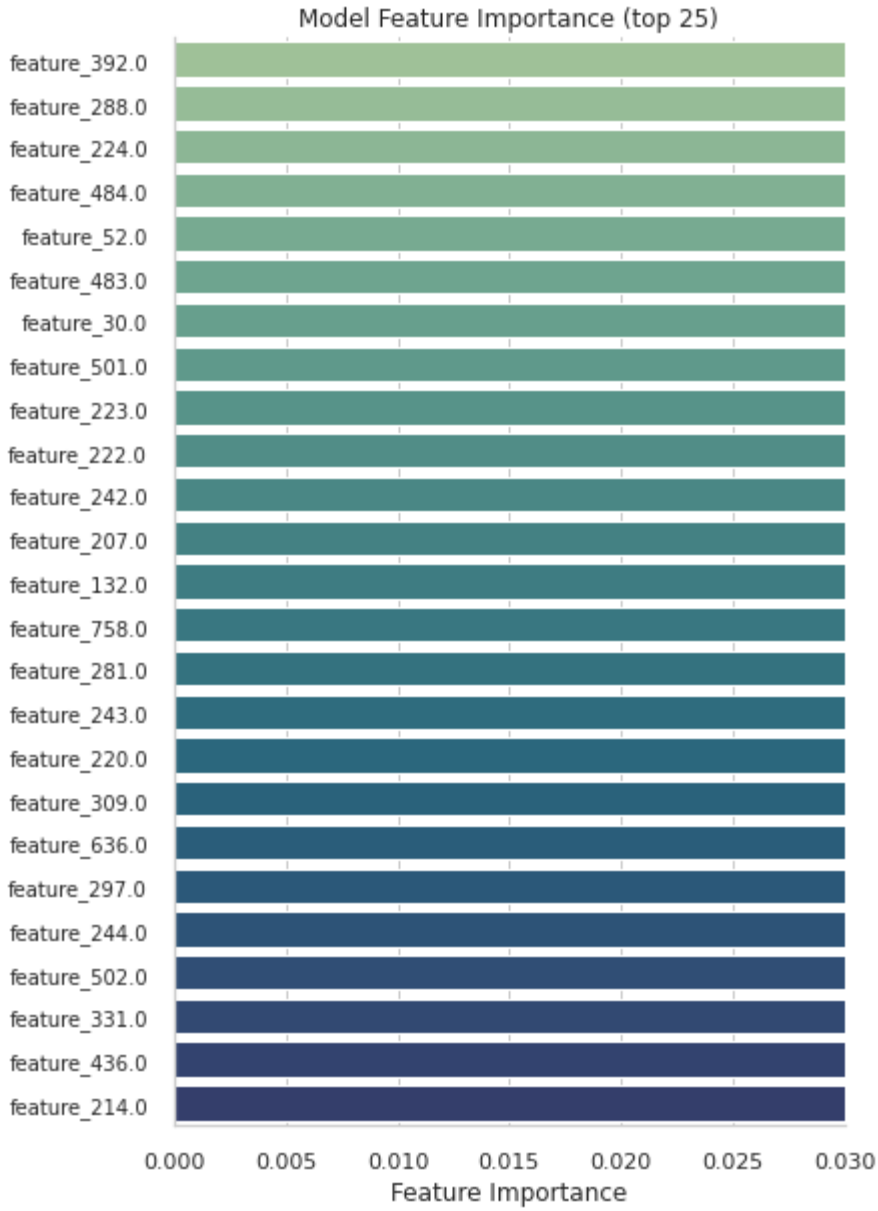
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/linear_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

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

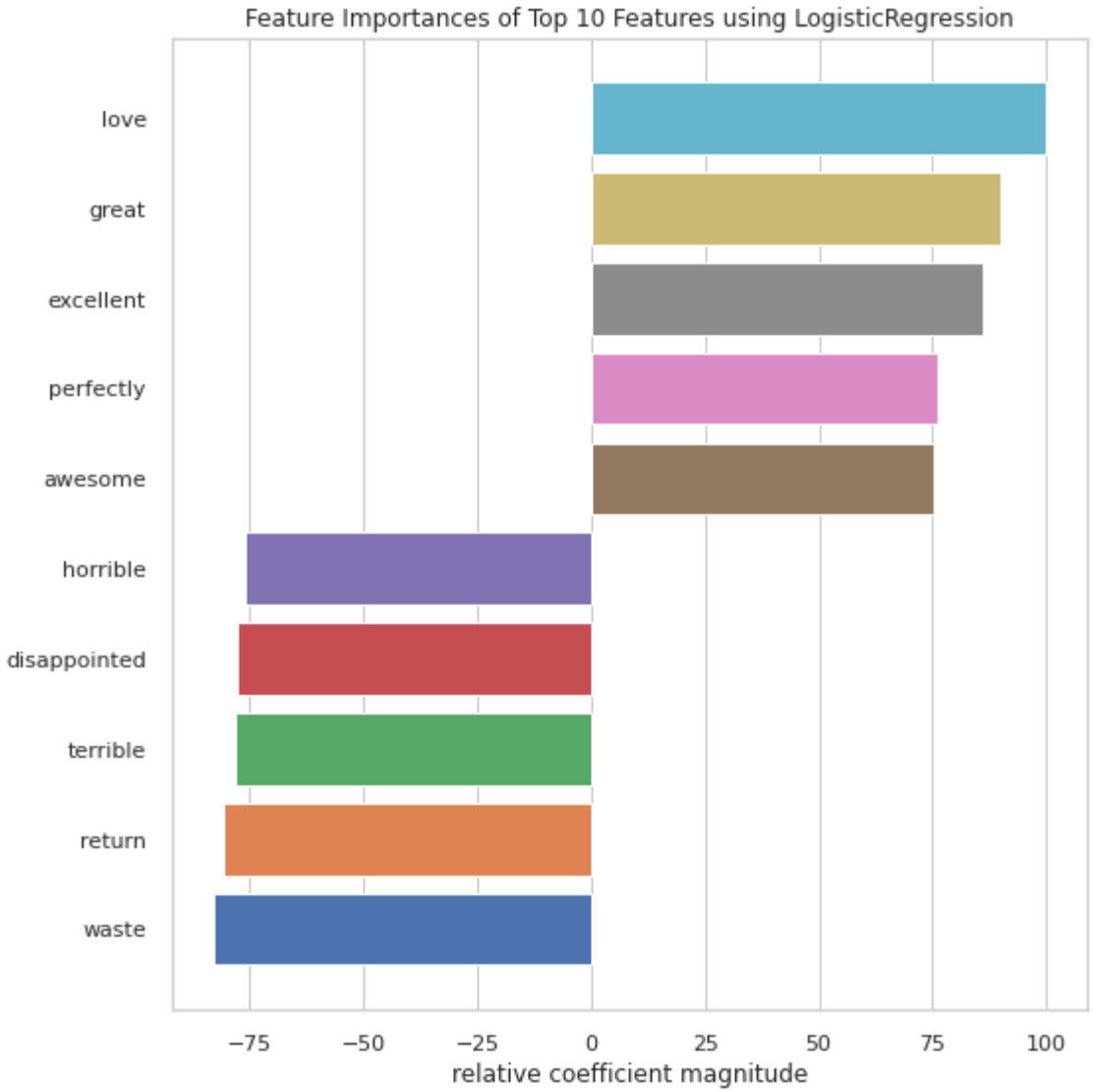



```
0%|          | 0/251 [00:00<?, ?it/s]
```


Total Features >= Importance Level
Number of Features
767.5
765.0
762.5
760.0
757.5
755.0
752.5
Number of Features
0.000
0.005
0.010
0.015
0.020
0.025
Feature Importance




```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/yellowbrick/
model_selection/importances.py:199: YellowbrickWarning: detected multi-dimens
ional feature importances but stack=False, using mean to aggregate them.
  YellowbrickWarning,
```

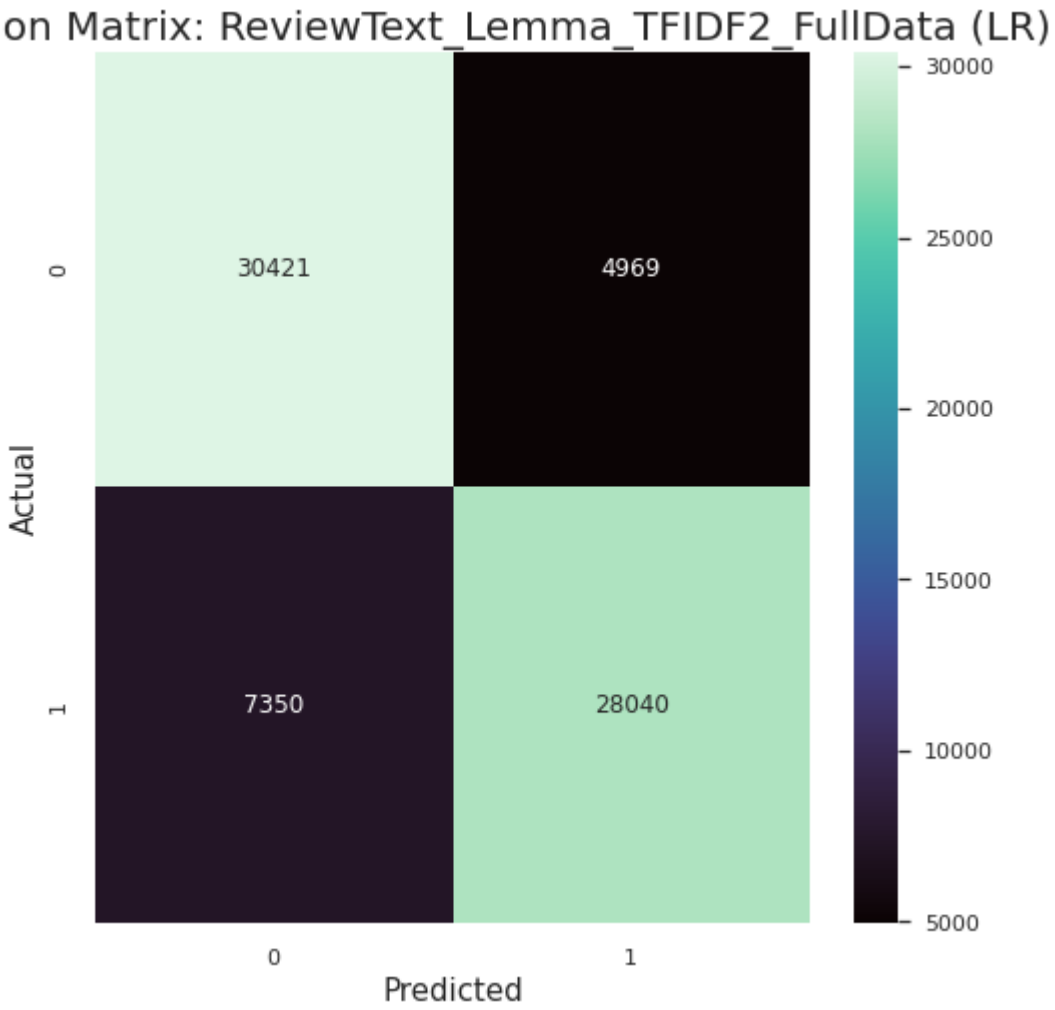



```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
<Figure size 576x576 with 0 Axes>
```

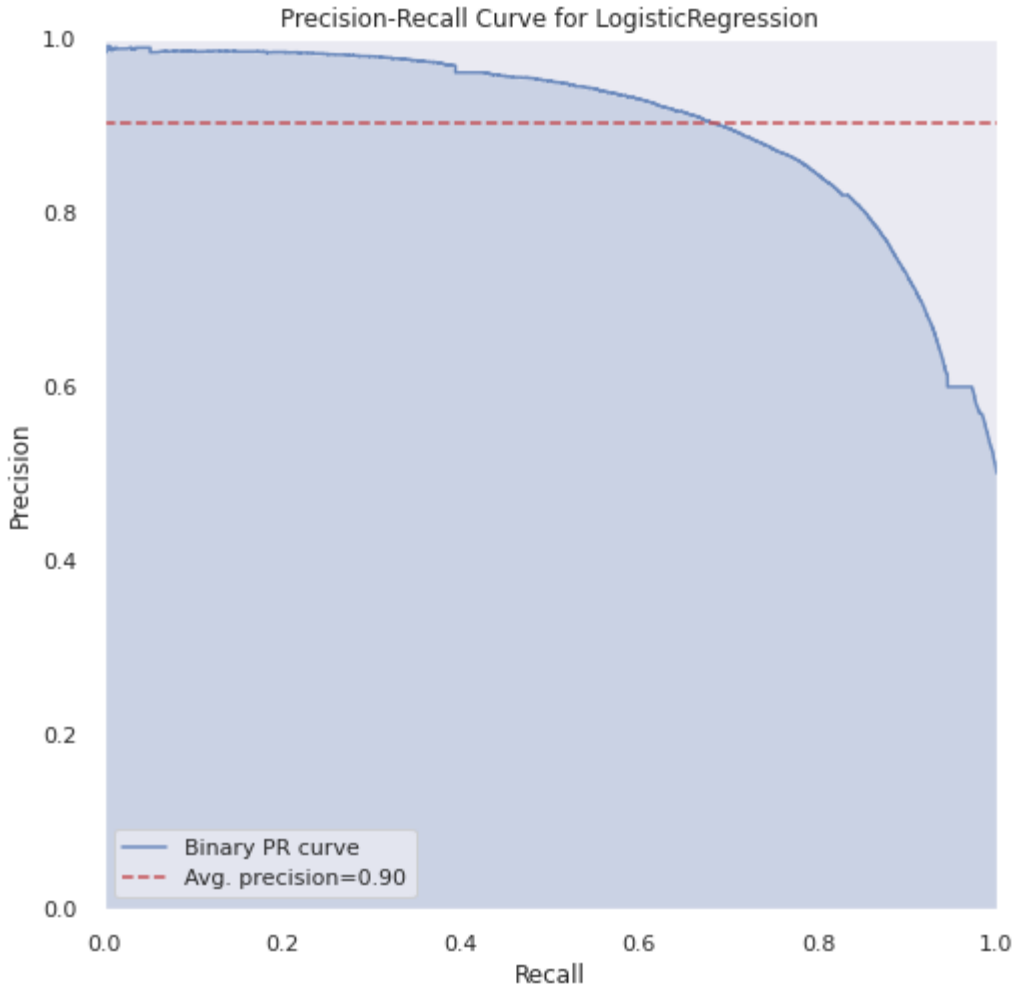
In [15]:

```
myExp.showFinalModelReport(axis_labels,
                           startValue=0.01,
                           increment=0.001,
                           upperValue=0.03,
                           topn=10)
```

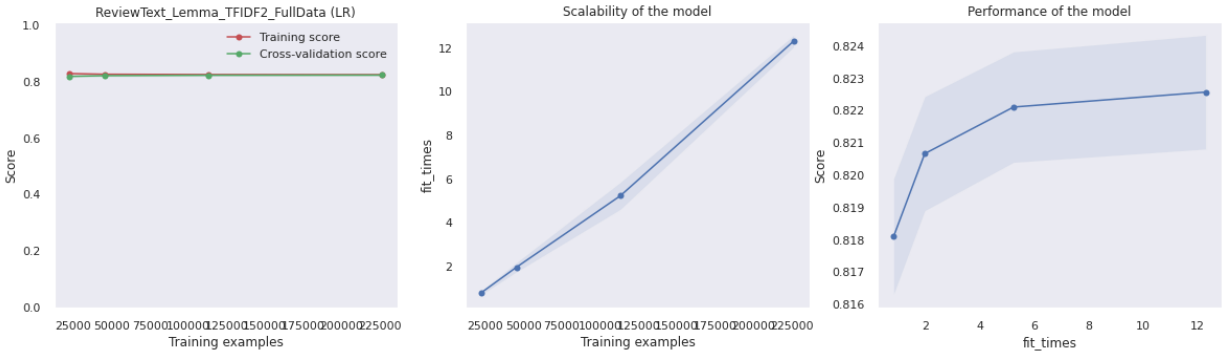
```
Final Model Stats:
Accuracy: 0.83
Precision: 0.83
Recalll: 0.83
F1 Score: 0.83
Cohen kappa:: 0.65
              precision    recall  f1-score   support

           0       0.81      0.86      0.83     35390
           1       0.85      0.79      0.82     35390

    accuracy                           0.83     70780
   macro avg       0.83      0.83      0.83     70780
weighted avg       0.83      0.83      0.83     70780
```



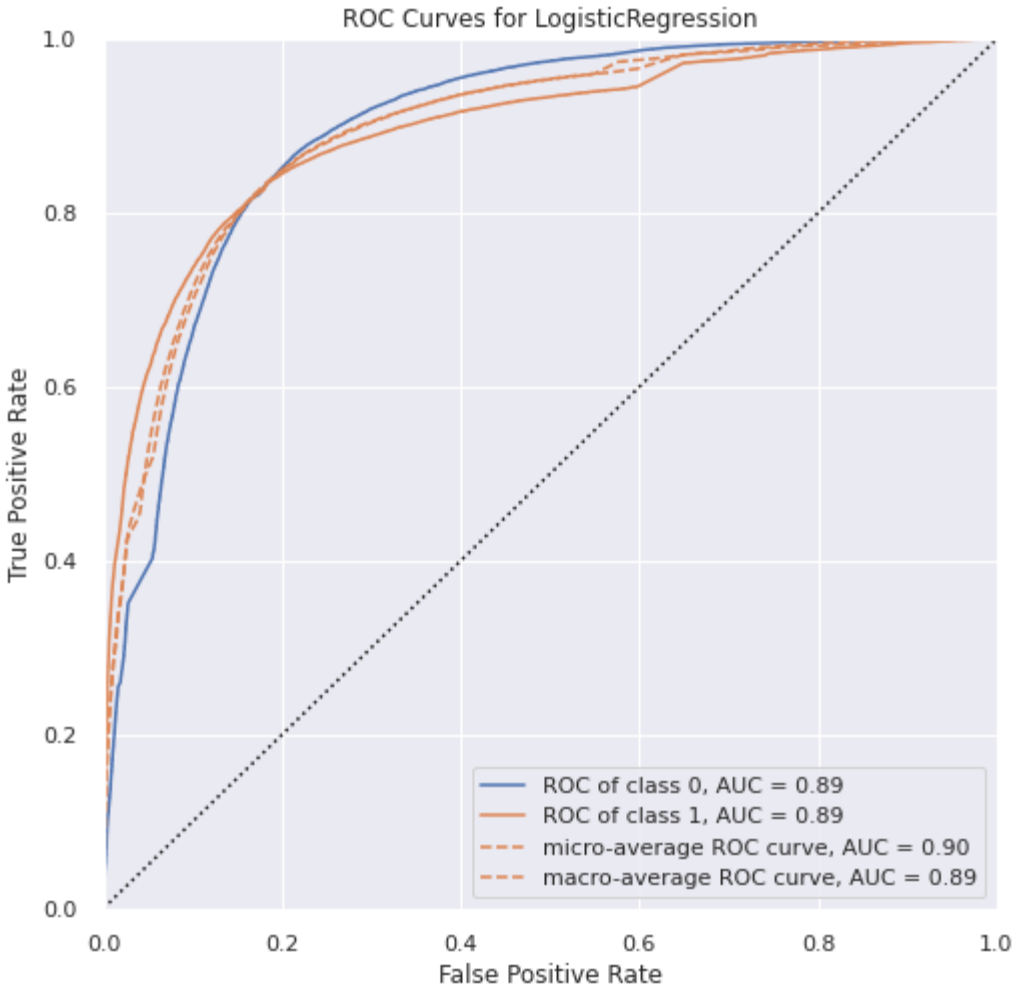

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/bas
e.py:444: UserWarning: X has feature names, but LogisticRegression was fitted
without feature names
  f"X has feature names, but {self.__class__.__name__} was fitted without"
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/bas
e.py:444: UserWarning: X has feature names, but LogisticRegression was fitted
without feature names
  f"X has feature names, but {self.__class__.__name__} was fitted without"
```



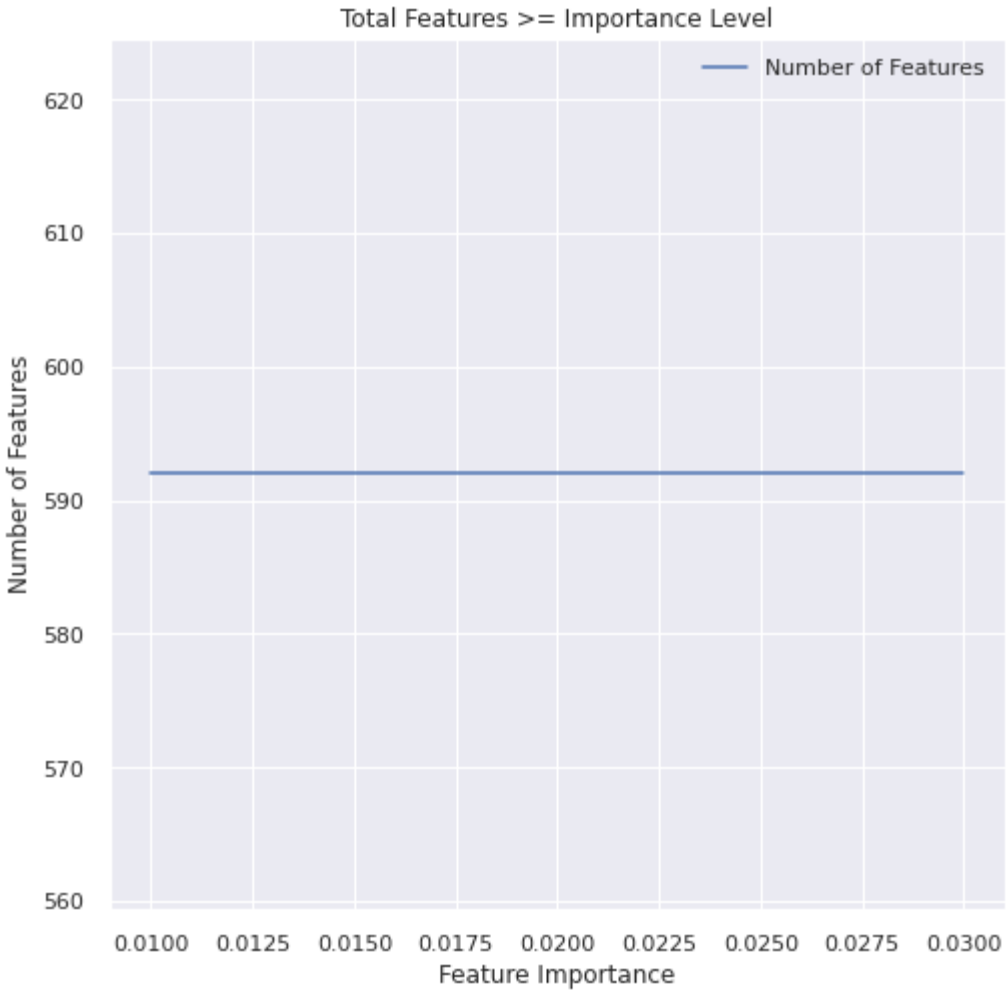
<Figure size 576x576 with 0 Axes>

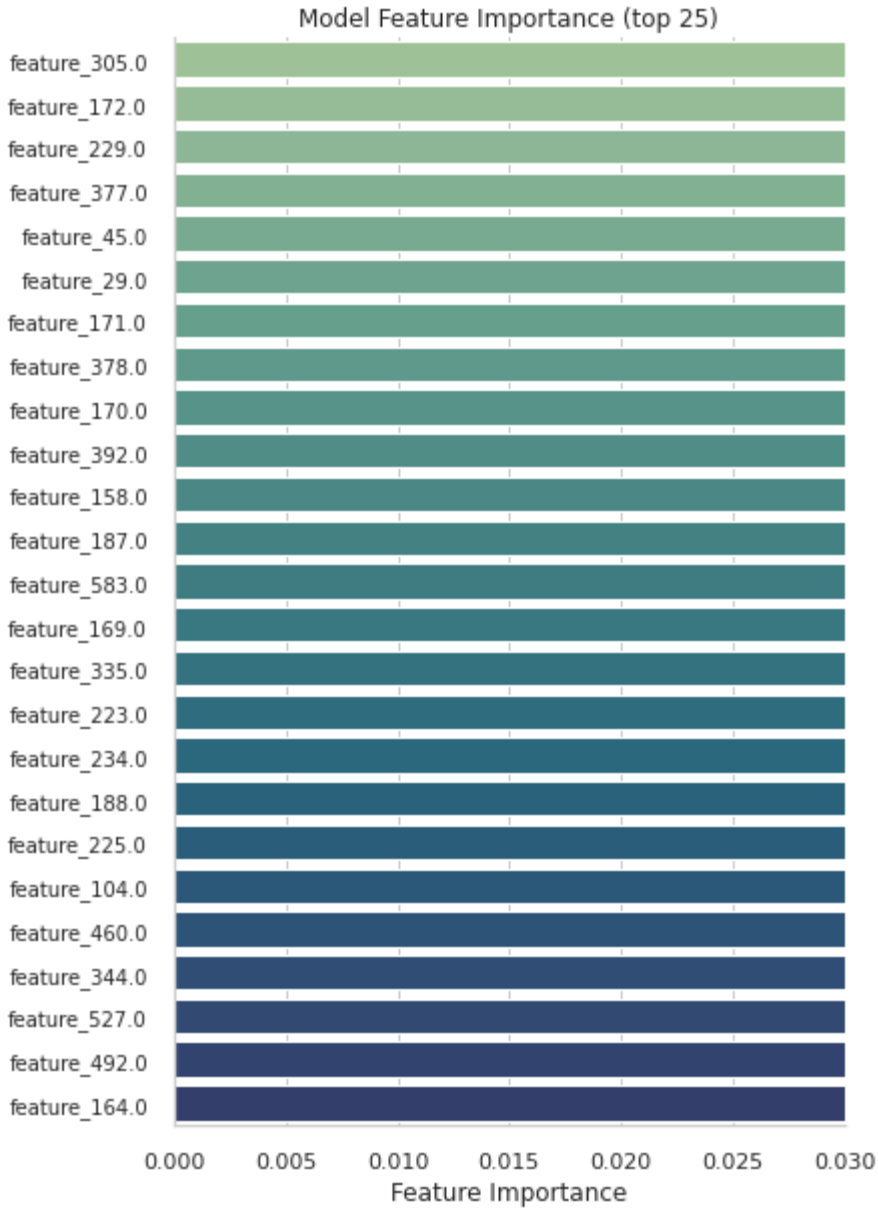



```
Final model ROCAUC not calculated. Starting now
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

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

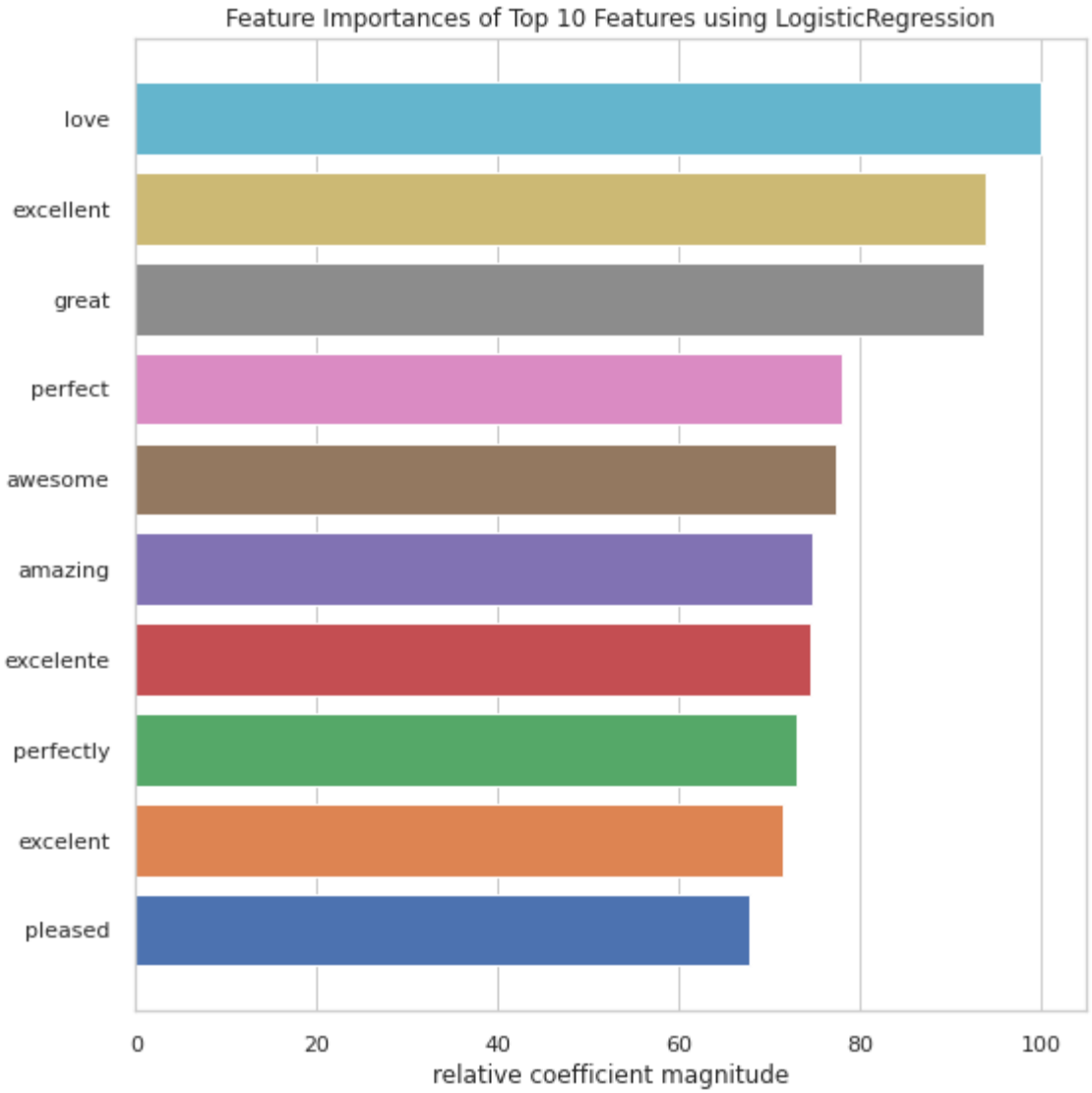



```
0%|          | 0/22 [00:00<?, ?it/s]
```


Total Features >= Importance Level
Number of Features
620
610
600
590
580
570
560
Number of Features
0.0100
0.0125
0.0150
0.0175
0.0200
0.0225
0.0250
0.0275
0.0300
Feature Importance




```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/yellowbrick/
model_selection/importances.py:199: YellowbrickWarning: detected multi-dimens
ional feature importances but stack=False, using mean to aggregate them.
  YellowbrickWarning,
```




```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/line
ar_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (stat
us=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regres
sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
<Figure size 576x576 with 0 Axes>
```

In [16]:
```
myExp.display()
```

```
DataExperiment summary:
---> projectName: ML1010-Group-Project
---> experimentName: ReviewText_Lemma_TFIDF2_FullData (LR)
---> isDataPackageLoaded: True
---> isBaseModelLoaded: True
---> isBaseModelPredicted: True
---> isBaseModelLearningCurveCreated: True
```

```
---> isFinalModelLoaded: True
---> isFinalModelPredicted: True
---> isFinalModelLearningCurveCreated: True
---> isClassifierLoaded: True
LogisticRegression()

    DataPackage summary:
    Attributes:
    ---> uniqueColumn: uuid
    ---> targetColumn: overall_posneg
    Process:
    ---> isBalanced: True
    ---> isTrainTestSplit: True
    Data:
    ---> isOrigDataLoaded: False
    ---> isTrainDataLoaded: True
```

Save Experiment

In [17]:
```
jarvis.saveExperiment(myExp, FILE_NAME)
```

In [ ]:
```
```

Scratchpad

In [ ]:
```
```

In [ ]:
```
```

In [ ]:
```
```

In [18]:
```
importlib.reload(dp)
importlib.reload(dps)
importlib.reload(DataExperiment)
importlib.reload(DataExperimentSupport)
```

Out[18]:
```
<module 'DataExperimentSupport' from '/home/magni/ML_Root/project_root/utilit
y_files/DataExperimentSupport.py'>
```