

Configuration

In [1]:

```
# Parameters
ENABLE_COLAB = False

PROJECT_NAME = 'ML1010-Group-Project'
EXPERIMENT_NAME = 'ReviewText_Lemma_Bert2 (Logistic Regression)'
FILE_NAME = '01_ML1010_GP_LR_Bert2'
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 16:13
 Hello sir. Reminder, no more coffee.

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
```

```
2022-01-25 16:13:43.347915: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0'; dlerror: libcudart.so.11.0: cannot open shared object file: No such file or directory
2022-01-25 16:13:43.347954: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlerror if you do not have a GPU set up on your machine.
```

```
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/utility_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 = LogisticRegression(max_iter=200, verbose=0)
ANALYSIS_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_WORK + "/01_NL_ReviewText_All(new

    testDfBert = cwutils.getBertEncodeFrame(df=testDf,
                                              bertColumn=ANALYSIS_COL,
                                              uniqueColumn=UNIQUE_COL,
                                              otherColumns=[TARGET_COL]
                                              )

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

```

DataExperiment summary:

```

---> projectName: ML1010-Group-Project
---> experimentName: ReviewText_Lemma_Bert2 (Logistic Regression)
---> isDataPackageLoaded: True
---> isBaseModelLoaded: False
---> isBaseModelPredicted: False
---> isBaseModelLearningCurveCreated: False
---> isFinalModelLoaded: False
---> isFinalModelPredicted: False
---> isFinalModelLearningCurveCreated: False
---> isClassifierLoaded: True
LogisticRegression(max_iter=200)

```

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: 13440



	overall_posneg	ttlCol
0	0	13440
1	1	13440

Completed train/test split (train_size = 0.8):

---> Original data size: 26880

---> Training data size: 21504

---> Testing data size: 5376

---> Stratified on column: overall_posneg

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

Base Model Stats:

Accuracy: 0.8

Precision: 0.8

Recall: 0.8

F1 Score: 0.8

Cohen kappa: 0.61

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/linear_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):

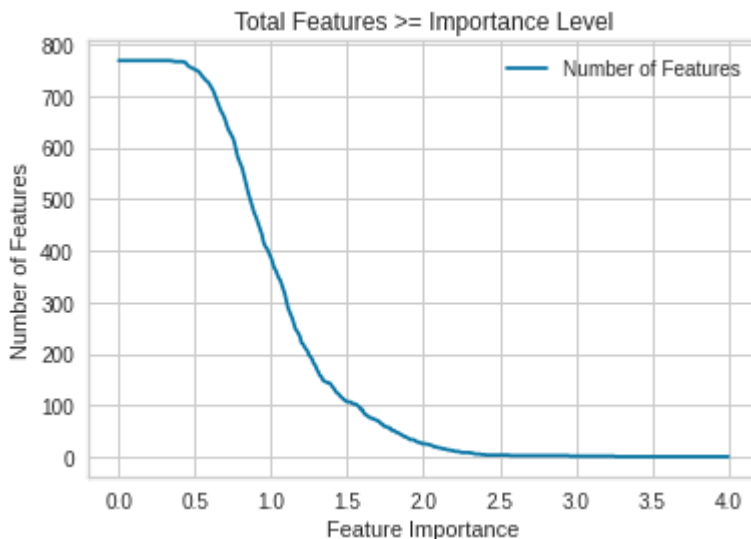
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

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

```
In [10]: _ = myExp.analyzeBaseModelFeatureImportance(startValue=0,
                                                increment=0.01,
                                                upperValue=4,
                                                returnAbove=1.25,
                                                showSummary=True)

#myExp.showBaseLimeGlobalImportance()
```

```
0%|          | 0/402 [00:00<?, ?it/s]
Feature Importance Summary:
--> Original feature count: 768
--> Returned feature count: 197
--> Removed feature count: 571
--> Return items above (including): 1.25
```



```
In [11]: %%time
myExp.createFinalModel(featureImportanceThreshold=1.25)
```

```
0%|          | 0/101 [00:00<?, ?it/s]
0%|          | 0/101 [00:00<?, ?it/s]
```

Final Model Stats:

Accuracy: 0.8

Precision: 0.8

Recall: 0.8

F1 Score: 0.8

Cohen kappa: 0.6

CPU times: user 4.47 s, sys: 5 s, total: 9.48 s

Wall time: 683 ms

/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

In [12]:

```
%time
myExp.createBaseModelLearningCurve()
```

[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.

/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,

[Parallel(n_jobs=1)]: Done 1 out of 1 | elapsed: 0.1s remaining: 0.0s

[learning_curve] Training set sizes: [1720 3440 8601 17203]

[CV] END, score=(train=0.913, test=0.784) total time=0.1s

/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,

[Parallel(n_jobs=1)]: Done 2 out of 2 | elapsed: 0.3s remaining: 0.0s

[CV] END, score=(train=0.876, test=0.789) total time=0.2s

/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,

[Parallel(n_jobs=1)]: Done 3 out of 3 | elapsed: 1.1s remaining: 0.0s

[CV] END, score=(train=0.849, test=0.802) total time=0.7s

```
/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,
/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,
[CV] END, score=(train=0.835, test=0.807) total time=1.8s
[CV] END, score=(train=0.905, test=0.780) total time=0.1s
/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,
[CV] END, score=(train=0.876, test=0.794) total time=0.2s
/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,
[CV] END, score=(train=0.845, test=0.809) total time=0.8s
/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,
 /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
 [CV] END, score=(train=0.833, test=0.819) total time=1.8s
 [CV] END, score=(train=0.908, test=0.792) total time=0.1s
 /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,
 [CV] END, score=(train=0.878, test=0.800) total time=0.3s
 /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,
 [CV] END, score=(train=0.846, test=0.812) total time=0.7s
 /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,
 /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.

```

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=0.834, test=0.817) total time=
1.9s
[CV] END ....., score=(train=0.915, test=0.781) total time=
0.1s
/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=0.877, test=0.796) total time=
0.2s
/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=0.847, test=0.811) total time=
0.8s
/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,
/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-regression
 extra warning msg= LOGISTIC SOLVER CONVERGENCE MSG,
 [CV] END, score=(train=0.833, test=0.816) total time=1.8s
 [CV] END, score=(train=0.910, test=0.774) total time=0.1s
 /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,
 [CV] END, score=(train=0.869, test=0.794) total time=0.2s
 /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,
 [CV] END, score=(train=0.848, test=0.800) total time=0.8s
 [CV] END, score=(train=0.833, test=0.809) total time=1.9s
 CPU times: user 2min 43s, sys: 1min 9s, total: 3min 52s
 Wall time: 14.9 s
 /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,
 [Parallel(n_jobs=1)]: Done 20 out of 20 | elapsed: 14.8s finished

In [13]:

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

[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
 /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   1 out of   1 | elapsed:   0.1s 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,
[Parallel(n_jobs=1)]: Done   2 out of   2 | elapsed:   0.1s remaining:   0.
0s
[learning_curve] Training set sizes: [ 1720  3440  8601 17203]
[CV] END ....., score=(train=0.856, test=0.793) total time=
0.1s
[CV] END ....., score=(train=0.831, test=0.799) total time=
0.1s
/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,
[Parallel(n_jobs=1)]: Done   3 out of   3 | elapsed:   0.4s remaining:   0.
0s
[CV] END ....., score=(train=0.828, test=0.808) total time=
0.3s
/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,
/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-regression
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/linear_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-regression
[CV] END, score=(train=0.820, test=0.806) total time=
0.4s
[CV] END, score=(train=0.851, test=0.798) total time=
0.0s
[CV] END, score=(train=0.828, test=0.810) total time=
0.1s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/linear_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-regression
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[CV] END, score=(train=0.822, test=0.814) total time=
0.2s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/linear_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-regression
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/linear_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-regression

```

sion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
[CV] END ....., score=(train=0.816, test=0.814) total time=
0.4s
[CV] END ....., score=(train=0.842, test=0.798) total time=
0.0s
[CV] END ....., score=(train=0.823, test=0.810) total time=
0.1s
/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=0.818, test=0.818) total time=
0.2s
/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,
/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,
/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=0.816, test=0.822) total time=
0.4s
[CV] END ....., score=(train=0.858, test=0.791) total time=
0.0s

```

```
[CV] END ..... , score=(train=0.833, test=0.799) 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-regression
 extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

```
[CV] END ..... , score=(train=0.822, test=0.807) total time=
0.2s
```

```
/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-regression
 extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

```
/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-regression
 extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

```
/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-regression
 extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

```
[CV] END ..... , score=(train=0.818, test=0.811) total time=
0.5s
[CV] END ..... , score=(train=0.845, test=0.785) total time=
0.0s
[CV] END ..... , score=(train=0.829, test=0.795) total time=
0.1s
```

```
/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=0.821, test=0.806) total time=
0.2s
[CV] END ....., score=(train=0.819, test=0.811) total time=
0.5s
CPU times: user 24 s, sys: 35.9 s, total: 59.9 s
Wall time: 3.89 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,
[Parallel(n_jobs=1)]: Done 20 out of 20 | elapsed: 3.9s finished

```

In [14]:

```

myExp.showBaseModelReport(axisLabels=axis_labels,
                           startValue=0,
                           increment=0.01,
                           upperValue=4)

```

Base Model Stats:

Accuracy: 0.8

Precision: 0.8

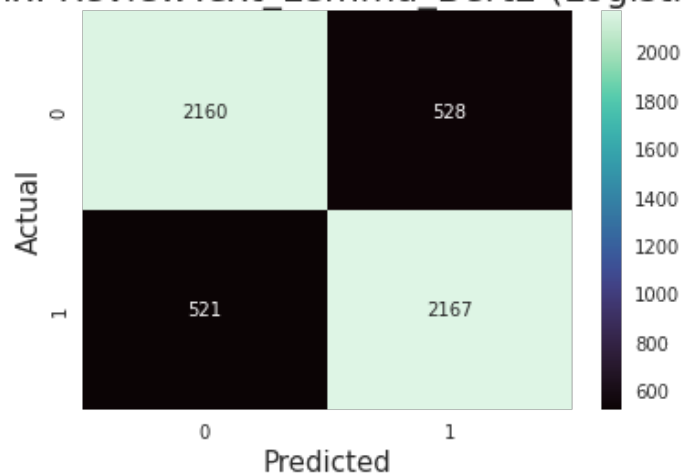
Recall: 0.8

F1 Score: 0.8

Cohen kappa: 0.61

	precision	recall	f1-score	support
0	0.81	0.80	0.80	2688
1	0.80	0.81	0.81	2688
accuracy			0.80	5376
macro avg	0.80	0.80	0.80	5376
weighted avg	0.80	0.80	0.80	5376

Confusion Matrix: ReviewText_Lemma_Bert2 (Logistic Regression)

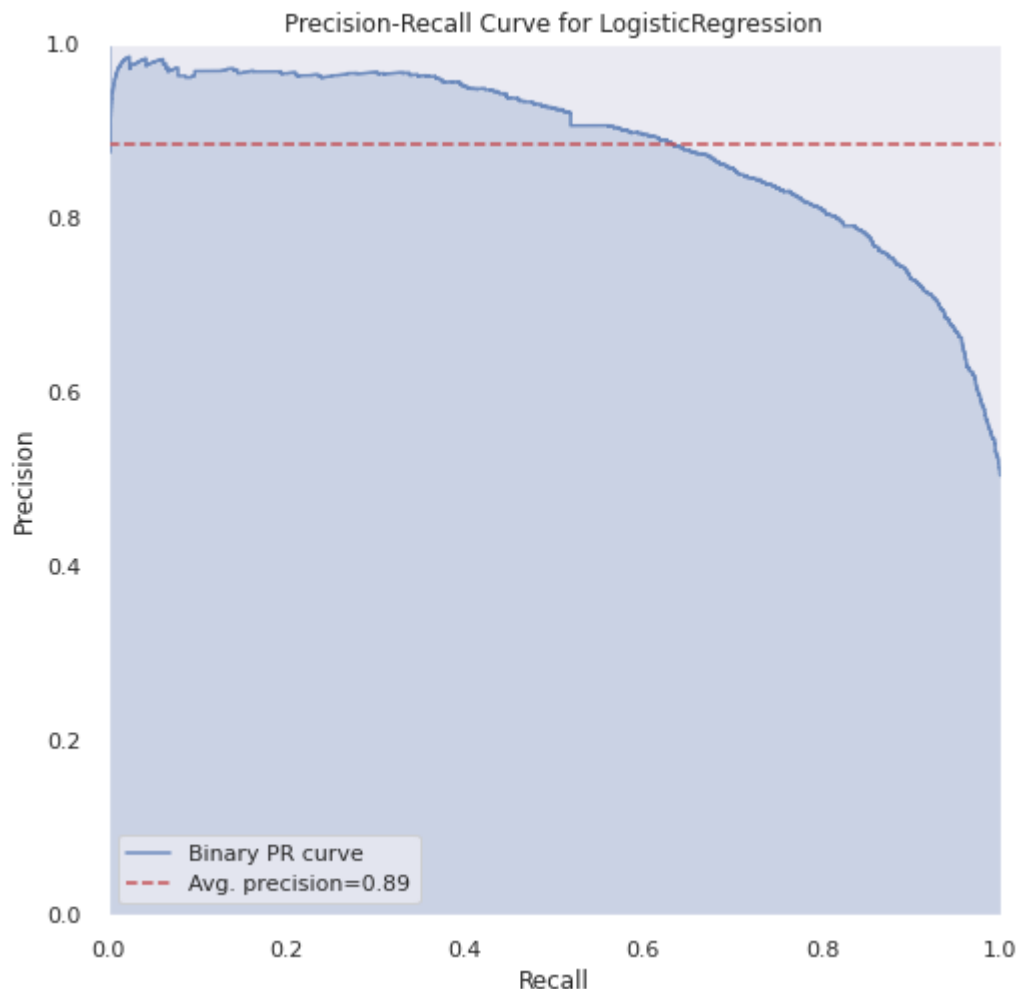


```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/base.py:444: UserWarning: X has feature names, but LogisticRegression was fitted without feature names
```

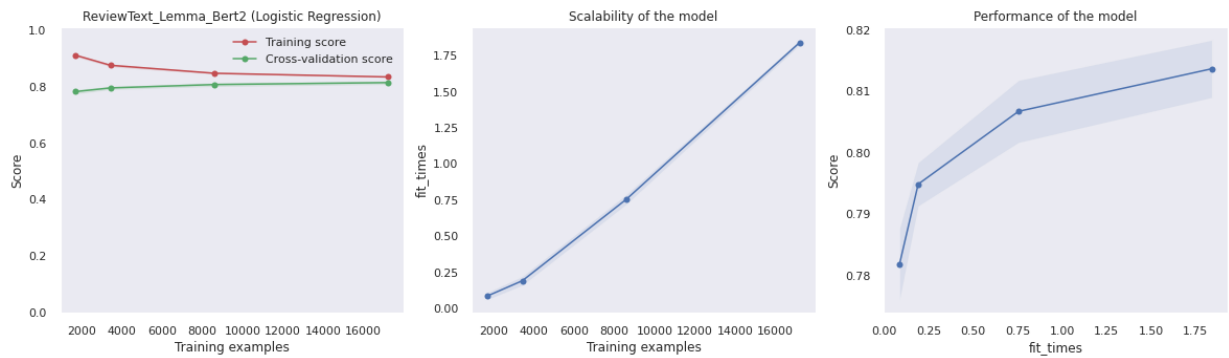
```
f"X has feature names, but {self.__class__.__name__} was fitted without"
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/base.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.
```

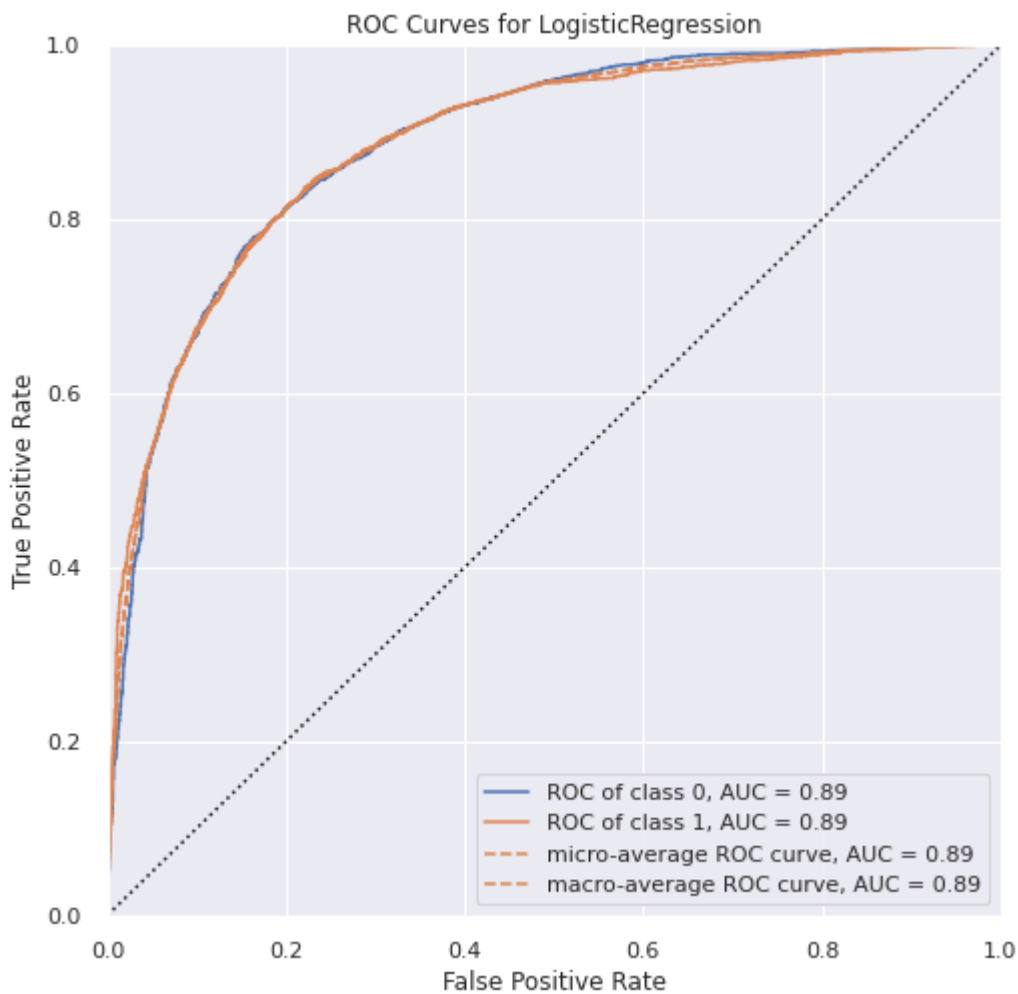
Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

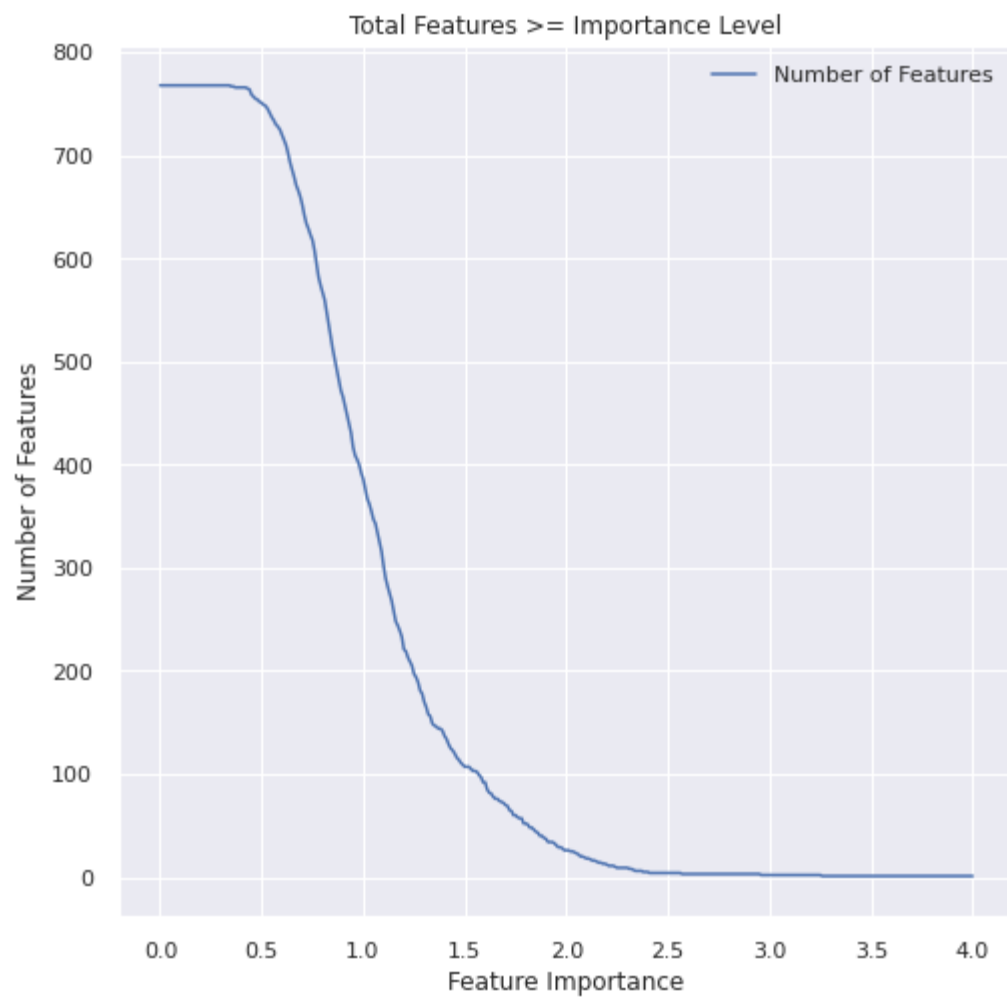
Please also refer to the documentation for alternative solver options:

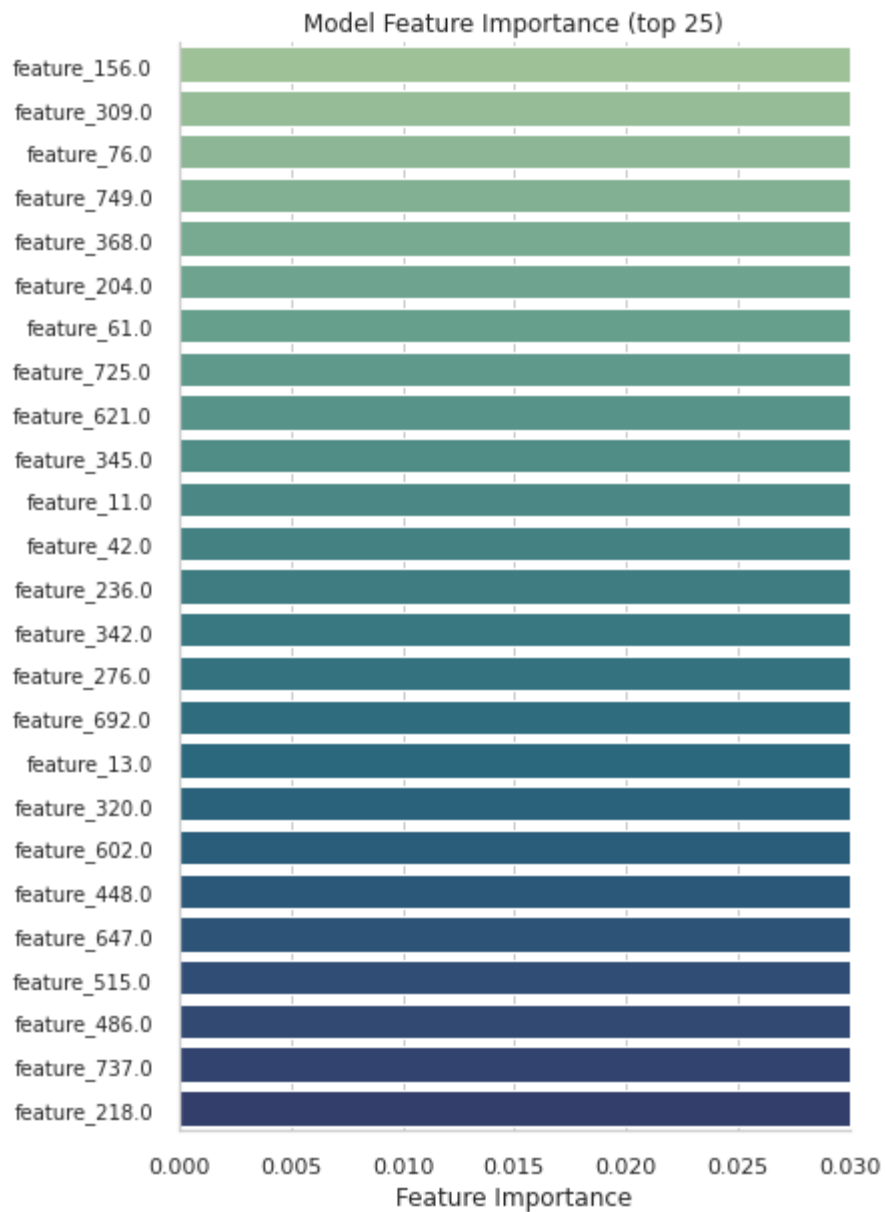
https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

```
extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
```

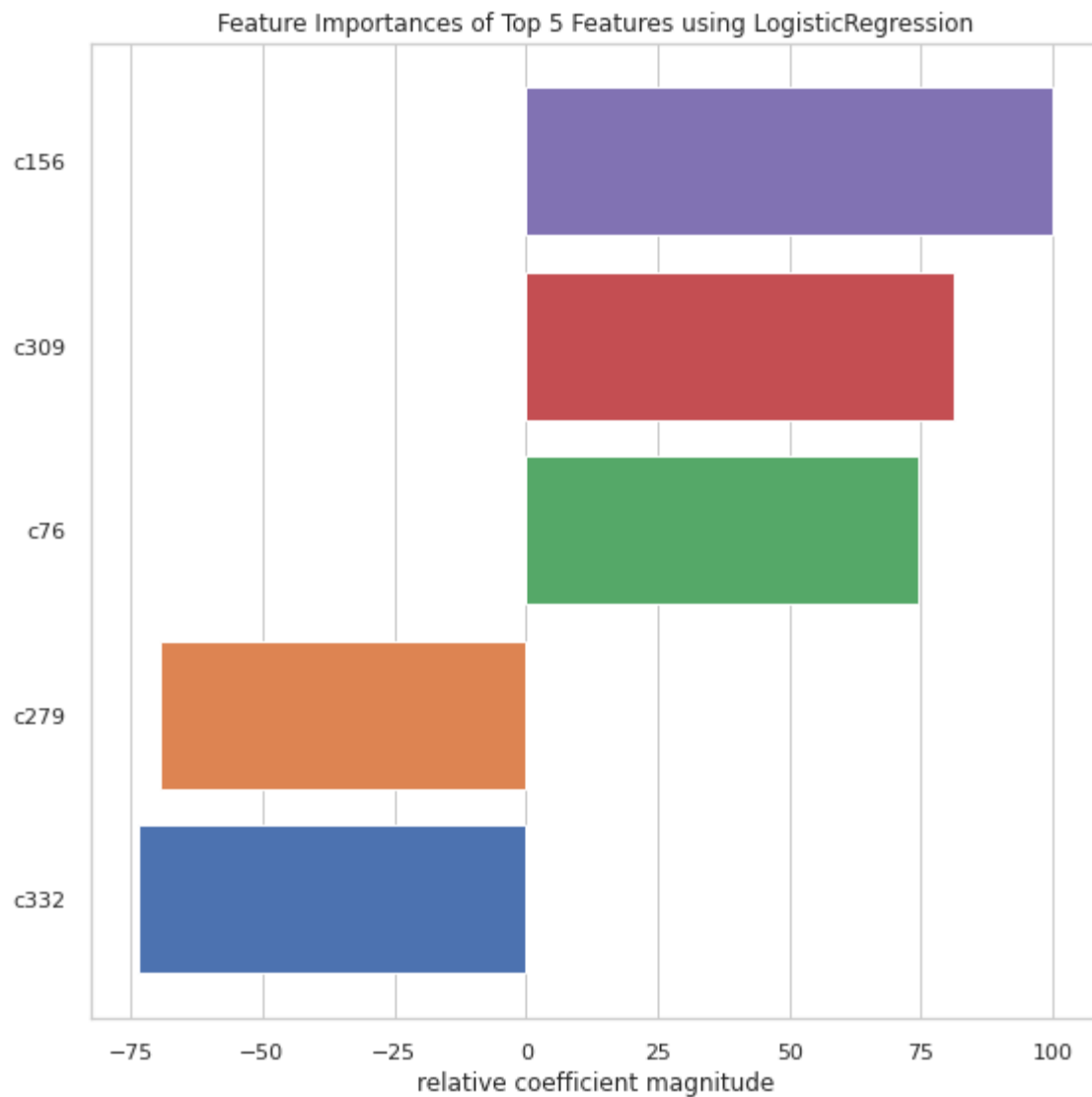


```
0%|          | 0/402 [00:00<?, ?it/s]
```





```
/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/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,
```

```
<Figure size 576x576 with 0 Axes>
```



In [15]:

```
myExp.showFinalModelReport(axisLabels=axis_labels,
                             startValue=0,
                             increment=0.01,
                             upperValue=4)
```

Final Model Stats:

Accuracy: 0.8

Precision: 0.8

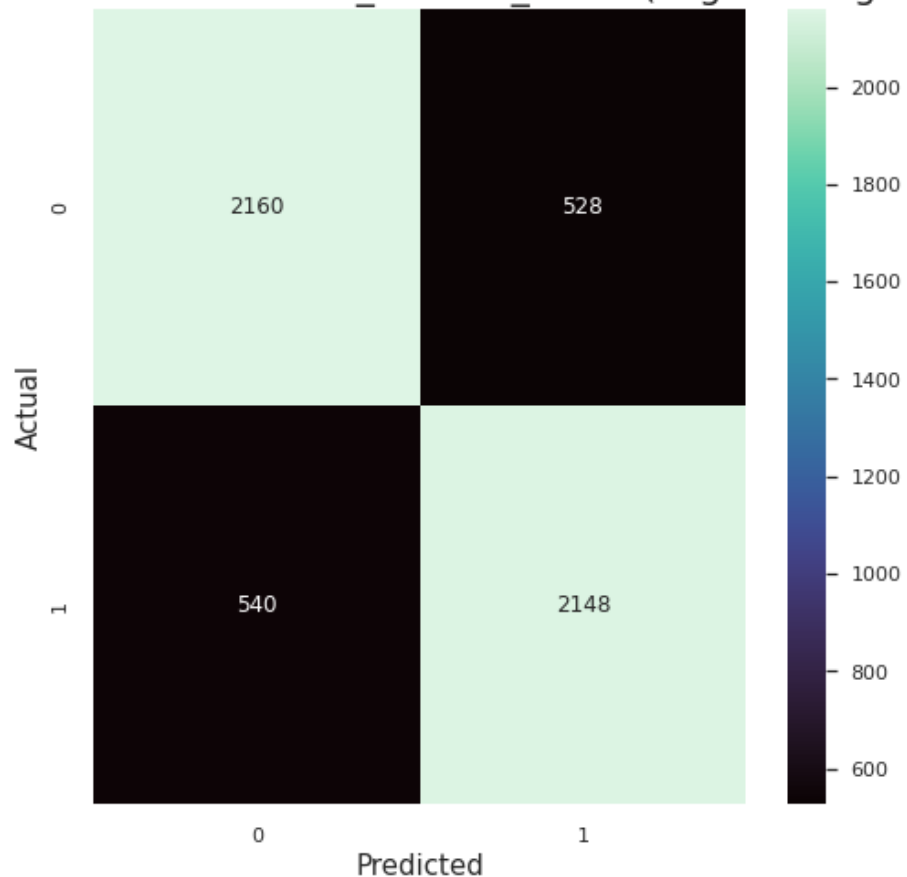
Recall: 0.8

F1 Score: 0.8

Cohen kappa: 0.6

	precision	recall	f1-score	support
0	0.80	0.80	0.80	2688
1	0.80	0.80	0.80	2688
accuracy			0.80	5376
macro avg	0.80	0.80	0.80	5376
weighted avg	0.80	0.80	0.80	5376

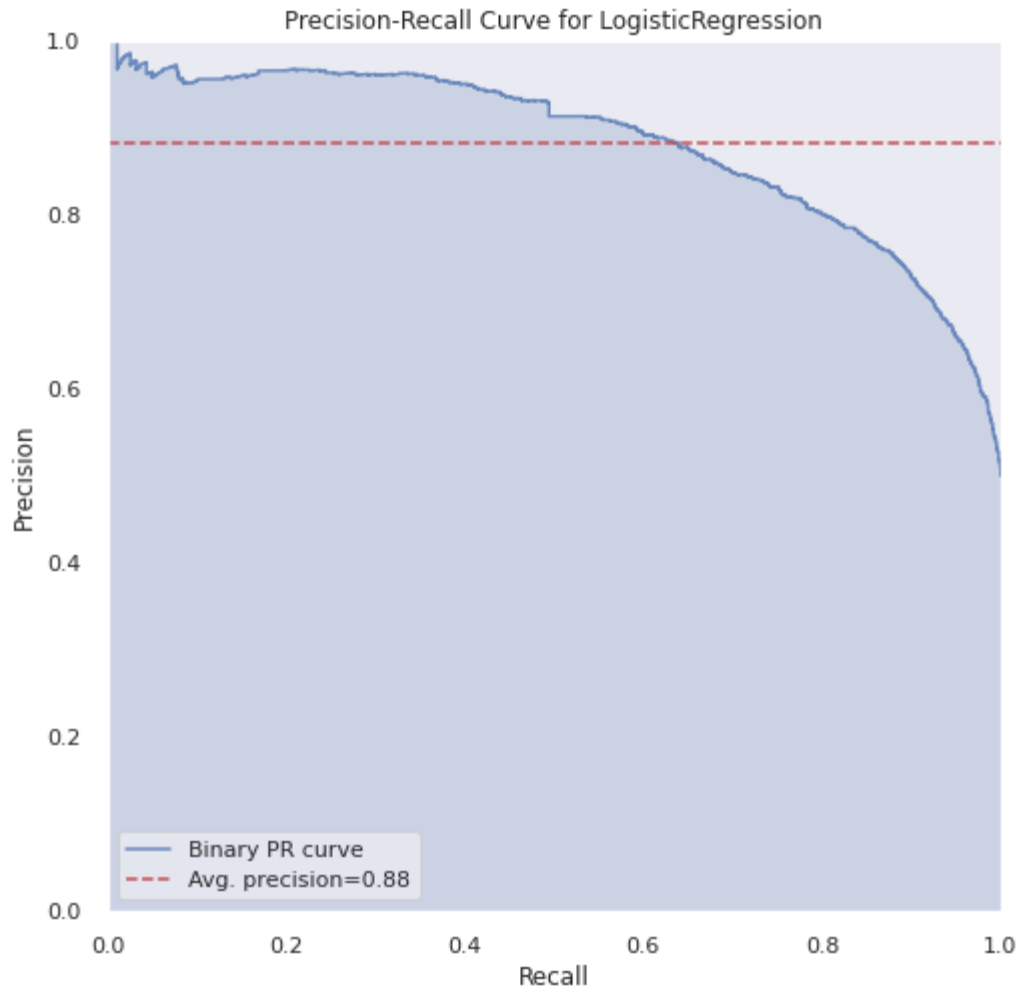
Confusion Matrix: ReviewText_Lemma_Bert2 (Logistic Regression)



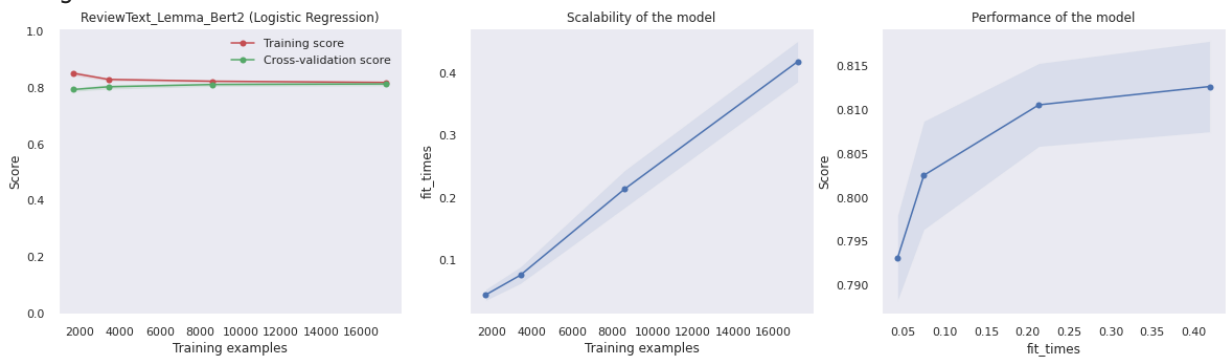
```

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/base.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/base.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>



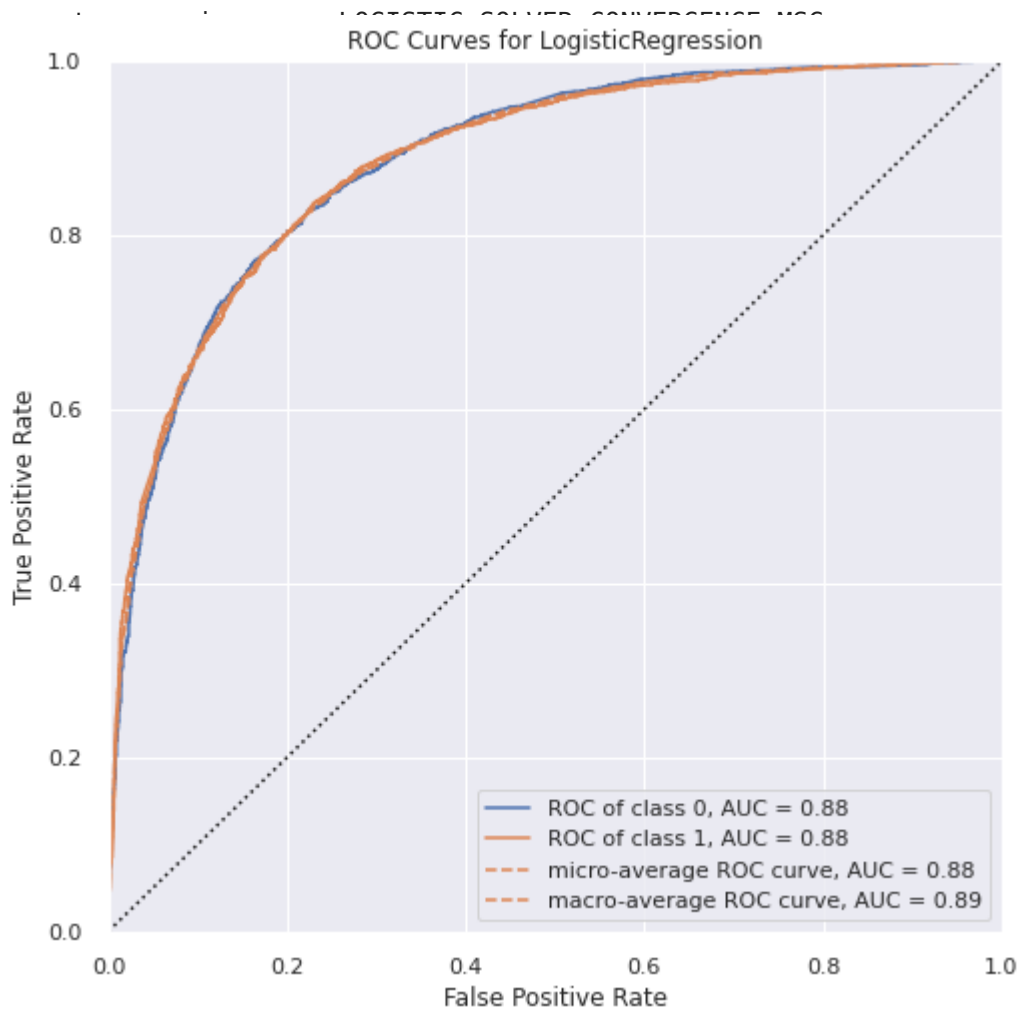
Final model ROCAUC not calculated. Starting now

```

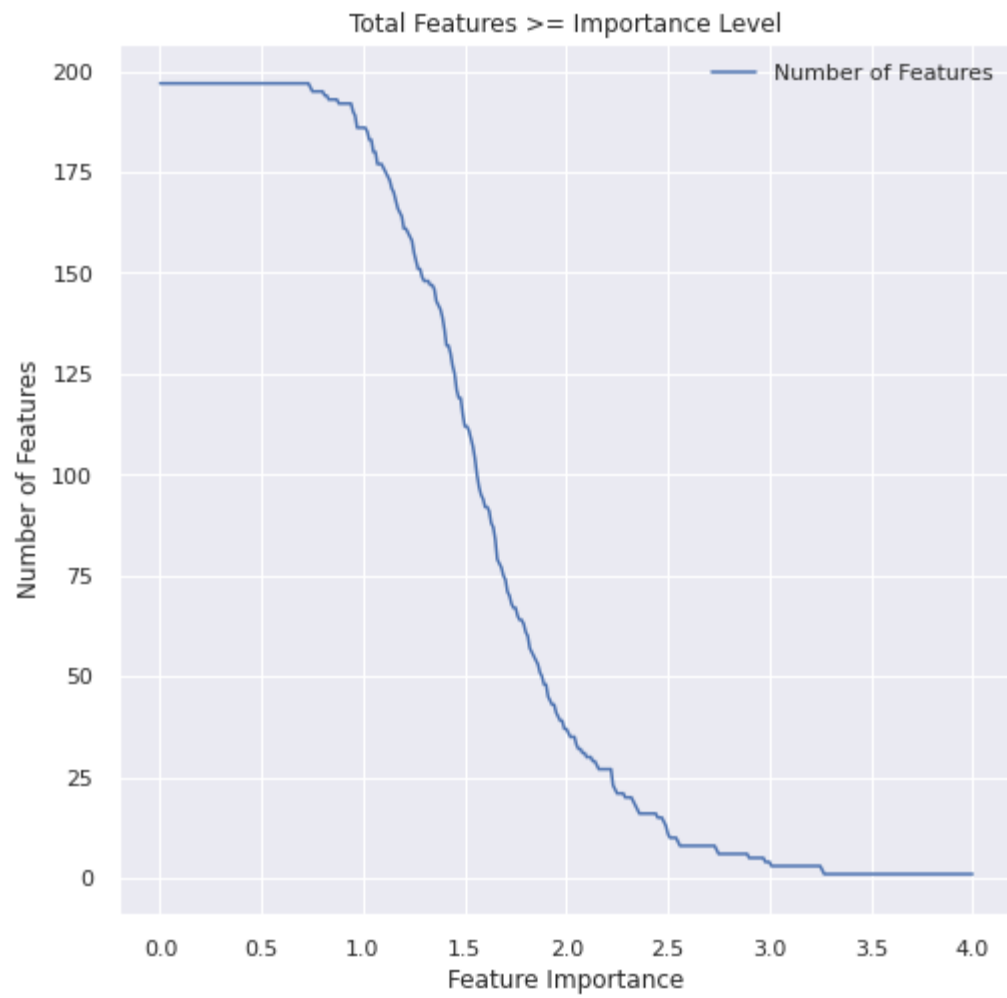
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/sklearn/linear_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

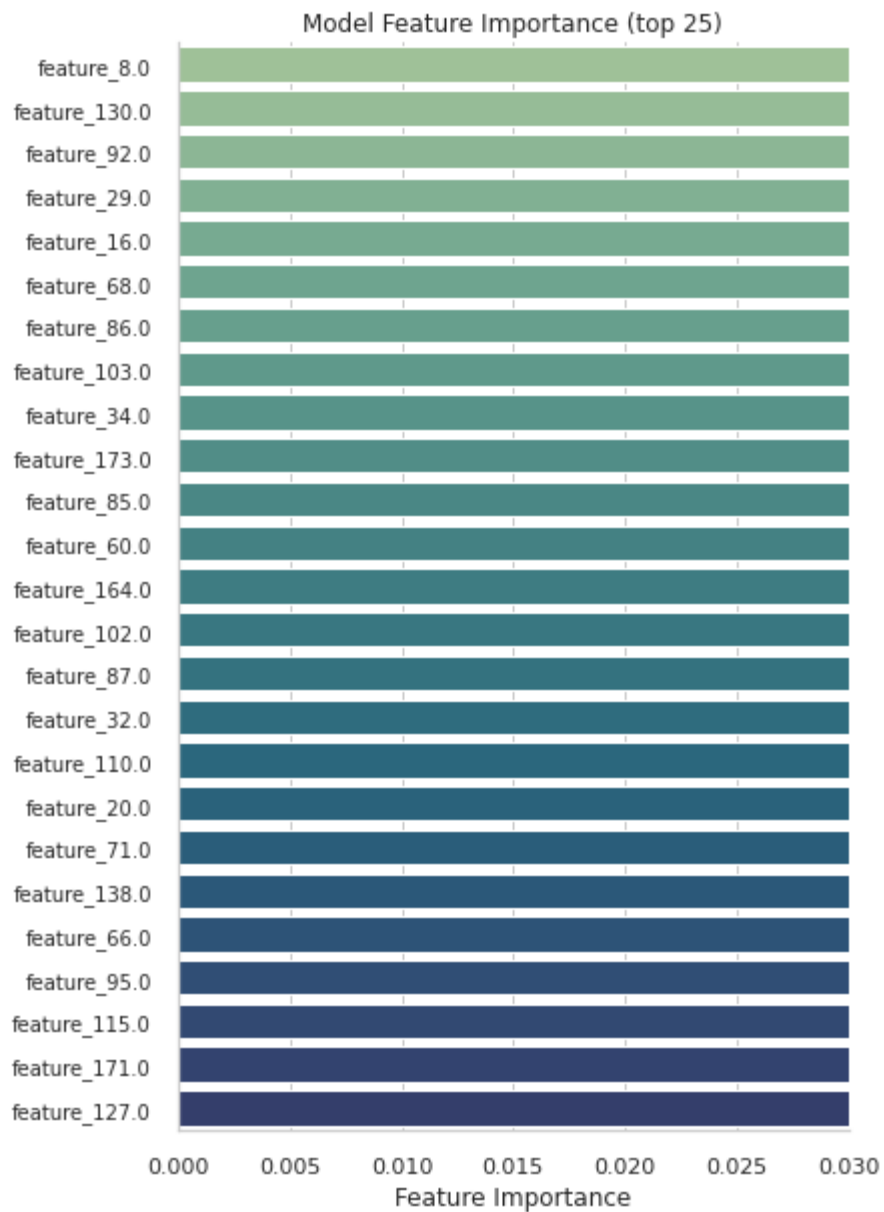
```

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

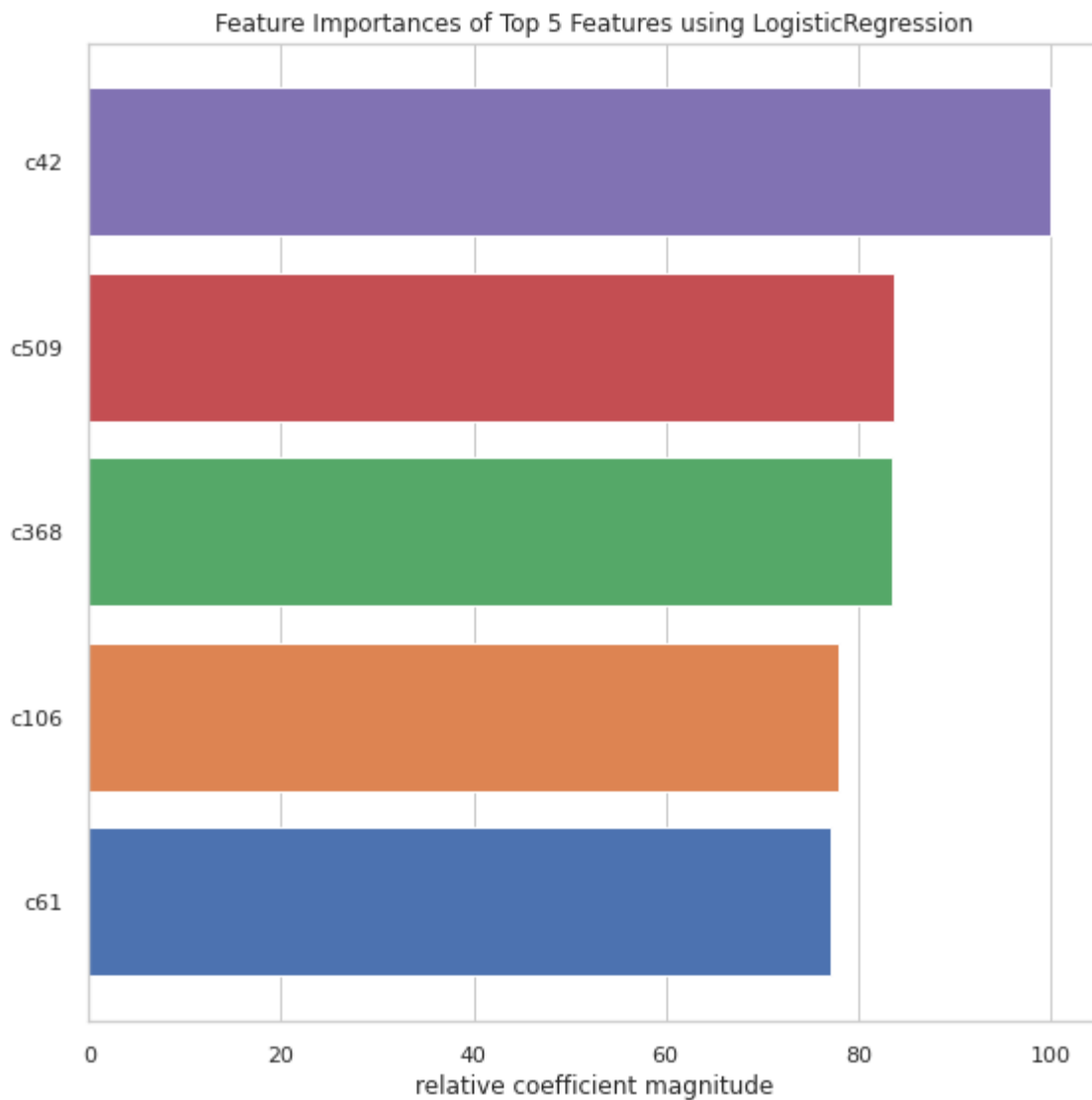


0%| | 0/402 [00:00<?, ?it/s]





```
/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/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,
```

```
<Figure size 576x576 with 0 Axes>
```

Global Importance: Weights



In [16]:

```
myExp.display()
```

DataExperiment summary:

```

---> projectName: ML1010-Group-Project
---> experimentName: ReviewText_Lemma_Bert2 (Logistic Regression)
---> isDataPackageLoaded: True
---> isBaseModelLoaded: True
---> isBaseModelPredicted: True
---> isBaseModelLearningCurveCreated: True
---> isFinalModelLoaded: True
---> isFinalModelPredicted: True
---> isFinalModelLearningCurveCreated: True
---> isClassifierLoaded: True
LogisticRegression(max_iter=200)

```

DataPackage summary:

Attributes:

```

---> uniqueColumn: uuid
---> targetColumn: overall_posneg

```

Process:

```

---> isBalanced: True
---> isTrainTestSplit: True

```

Data:

```

---> isOrigDataLoaded: False
---> isTrainDataLoaded: True
---> isTestDataLoaded: True

```

Save Experiment

In [17]:

```
jarvis.saveExperiment(myExp, FILE_NAME)
```

Scratchpad

In []:

In []: