Configuration

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

PROJECT_NAME = 'ML1010-Group-Project'
EXPERIMENT_NAME = 'ReviewText_Lemma_Bert2_Full (XGB)'
FILE_NAME = '01_ML1010_GP_XGB_Bert2_Full'
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]:
```python
#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 09:49
Hello sir. Extra caffeine may help.

Setup Runtime Environment

In [3]:
```python
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]:
```python
import importlib
import cw_df_metric_utils as cwutils
import DataPackage as dp
import DataPackageSupport as dps
import DataExperiment
import DataExperimentSupport
```

```
2022-01-25 09:49:09.776551: 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
2022-01-25 09:49:09.776579: 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]:
```python
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]:
```python
#axis_labels=[1,2,3,4,5]
axis_labels=[0,1]
#classifier = RandomForestClassifier()
classifier = XGBClassifier(eval_metric='mlogloss')
ANALSYSIS_COL = 'reviewText_lemma_bert'
UNIQUE_COL = 'uuid'
TARGET_COL = 'overall_posneg'
```

In [7]:

```python
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_Cellphone_full.pkl.gz

    testDfBert = cwutils.getBertEncodeFrame(df=testDf,
                                            bertColumn=ANALSYSIS_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_Full (XGB)
---> isDataPackageLoaded: True
---> isBaseModelLoaded: False
---> isBaseModelPredicted: False
---> isBaseModelLearningCurveCreated: False
---> isFinalModelLoaded: False
---> isFinalModelPredicted: False
---> isFinalModelLearningCurveCreated: False
---> isClassifierLoaded: True
XGBClassifier(base_score=None, booster=None, colsample_bylevel=None,
              colsample_bynode=None, colsample_bytree=None,
              enable_categorical=False, eval_metric='mlogloss', gamma=None,
              gpu_id=None, importance_type=None, interaction_constraints=Non
e,
              learning_rate=None, max_delta_step=None, max_depth=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, reg_alpha=None,
              reg_lambda=None, scale_pos_weight=None, subsample=None,
              tree_method=None, validate_parameters=None, verbosity=None)

    DataPackage summary:
    Attributes:
    ---> uniqueColumn: uuid
    ---> targetColumn: overall_posneg
    Process:
    ---> isBalanced: False
    ---> isTrainTestSplit: False
    Data:
    ---> isOrigDataLoaded: True
    ---> isTrainDataLoaded: False
```
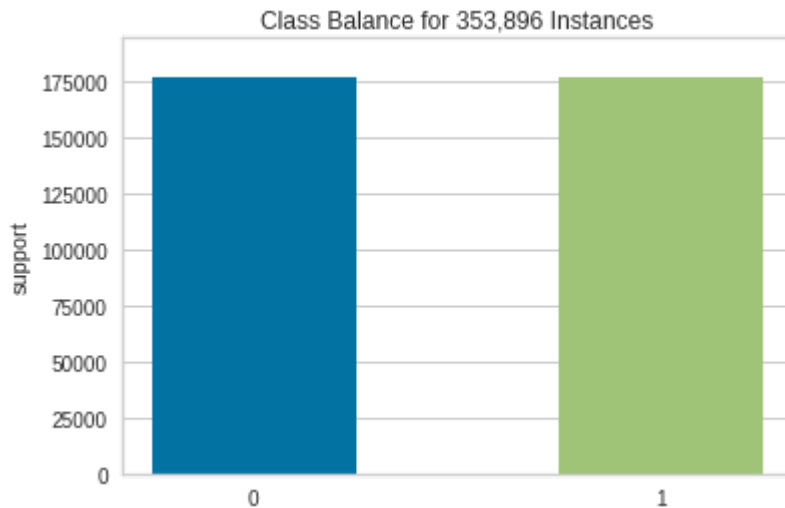
In [8]:
```
myExp.dataPackage.classBalanceUndersample()
myExp.dataPackage.splitTrainTest()
```

Class Balance for 506,438 Instances

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

Class Balance for 353,896 Instances

	overall_posneg	ttlCol
0	0	176948
1	1	176948

```
Completed train/test split (train_size = 0.8):
---> Original data size: 353896
---> Training data size: 283116
---> Testing data size: 70780
---> Stratified on column: overall_posneg
```

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

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi

fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
Base Model Stats:
Accuracy: 0.83
Precision: 0.83
Recalll: 0.83
F1 Score: 0.83
Cohen kappa:: 0.67
CPU times: user 48min 56s, sys: 1.77 s, total: 48min 58s
Wall time: 3min 8s

In [10]:
```python
impDf = myExp.analyzeBaseModelFeatureImportance(returnAbove=0.0010)
```

  0%| | 0/101 [00:00<?, ?it/s]
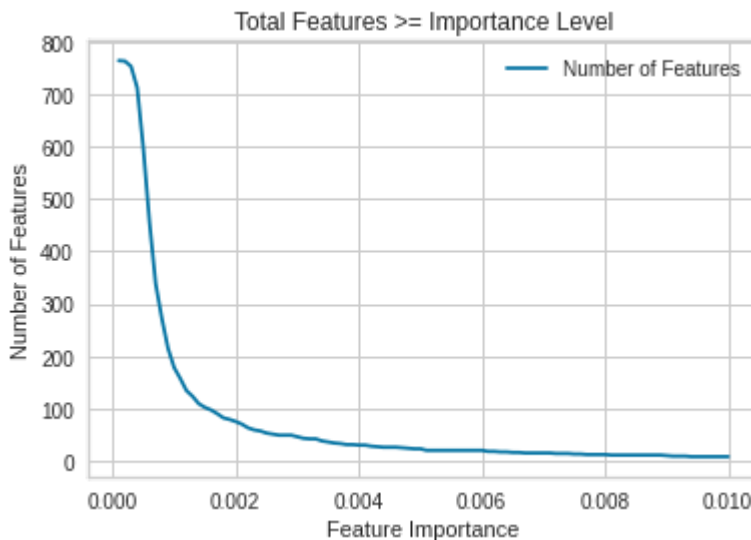Feature Importance Summary:
---> Original feature count: 768
---> Returned feature count: 178
---> Removed feature count: 590
---> Return items above (including): 0.001

In [11]:
```python
%%time
myExp.createFinalModel(featureImportanceThreshold=0.0010)
```

  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/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
Final Model Stats:
Accuracy: 0.83
Precision: 0.83
Recalll: 0.83
F1 Score: 0.83
Cohen kappa:: 0.65
CPU times: user 12min 2s, sys: 489 ms, total: 12min 2s

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

```
[learning_curve] Training set sizes: [ 22649  45298 113246 226492]
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent worke
rs.
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.989, test=0.812) total time=  1
4.3s
[Parallel(n_jobs=1)]: Done    1 out of    1 | elapsed:   14.3s remaining:    0.
0s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.966, test=0.819) total time=  2
7.6s
[Parallel(n_jobs=1)]: Done    2 out of    2 | elapsed:   42.0s remaining:    0.
0s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.920, test=0.826) total time= 1.2
min
[Parallel(n_jobs=1)]: Done    3 out of    3 | elapsed: 1.9min remaining:    0.
0s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.888, test=0.832) total time= 2.5
min
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
```

```
[CV] END ..................., score=(train=0.989, test=0.814) total time=  1
4.8s
```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
```
[CV] END ..................., score=(train=0.970, test=0.819) total time=  2
8.1s
```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
```
[CV] END ..................., score=(train=0.920, test=0.826) total time= 1.2
min
```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
```
[CV] END ..................., score=(train=0.888, test=0.832) total time= 2.5
min
```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
```
[CV] END ..................., score=(train=0.989, test=0.813) total time=  1
4.6s
```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
```
[CV] END ..................., score=(train=0.969, test=0.819) total time=  2
8.7s
```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
```
[CV] END ..................., score=(train=0.919, test=0.825) total time= 1.2
min
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.887, test=0.830) total time= 2.5
min
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.990, test=0.812) total time=  1
4.9s
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.969, test=0.816) total time=  2
7.8s
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.918, test=0.826) total time= 1.2
min
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.887, test=0.830) total time= 2.5
min
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.989, test=0.814) total time=  1
3.8s
```

```
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
```

```
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
[CV] END ...................., score=(train=0.968, test=0.822) total time=  2
6.7s

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.918, test=0.828) total time= 1.1
min

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.887, test=0.834) total time= 2.5
min
CPU times: user 5h 42min 8s, sys: 14.5 s, total: 5h 42min 23s
```

In [13]:
```python
%%time
myExp.createFinalModelLearningCurve()
```

```
[learning_curve] Training set sizes: [ 22649  45298 113246 226492]

[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent worke
rs.
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.983, test=0.810) total time=
3.8s

[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    3.9s remaining:    0.
0s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.948, test=0.816) total time=
7.1s

[Parallel(n_jobs=1)]: Done   2 out of   2 | elapsed:   11.0s remaining:    0.
0s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
```

ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...., score=(train=0.897, test=0.823) total time= 1
7.3s

[Parallel(n_jobs=1)]: Done   3 out of   3 | elapsed: 28.5s remaining: 0.
0s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...., score=(train=0.871, test=0.826) total time= 3
8.7s

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...., score=(train=0.982, test=0.812) total time=
5.7s

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...., score=(train=0.949, test=0.815) total time=
7.2s

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...., score=(train=0.896, test=0.822) total time= 1
9.0s

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...., score=(train=0.872, test=0.826) total time= 3
8.2s

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the

```
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
[CV] END ...................., score=(train=0.982, test=0.813) total time=
3.7s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.947, test=0.817) total time=
7.4s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.897, test=0.821) total time=  1
8.7s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.872, test=0.824) total time=  3
8.4s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.982, test=0.809) total time=
3.7s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ...................., score=(train=0.946, test=0.815) total time=
7.2s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
```

```
[CV] END ..................., score=(train=0.898, test=0.822) total time=  1
7.5s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.872, test=0.824) total time=  3
7.5s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.979, test=0.811) total time=
3.8s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.946, test=0.818) total time=
7.6s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.898, test=0.824) total time=  1
7.3s
/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle
arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec
ated and will be removed in a future release. To remove this warning, do the
following: 1) Pass option use_label_encoder=False when constructing XGBClassi
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[CV] END ..................., score=(train=0.872, test=0.829) total time=  3
7.6s
CPU times: user 1h 27min 20s, sys: 3.66 s, total: 1h 27min 24s
Wall time: 5min 38s

[Parallel(n_jobs=1)]: Done  20 out of  20 | elapsed:  5.6min finished
```

In [14]:
```python
myExp.showBaseModelReport(axis_labels,
                          upperValue=0.025)
```

```
Base Model Stats:
Accuracy: 0.83
```

```
Precision: 0.83
Recalll: 0.83
F1 Score: 0.83
Cohen kappa:: 0.67
              precision    recall  f1-score   support

           0       0.83      0.84      0.83     35390
           1       0.84      0.83      0.83     35390

    accuracy                           0.83     70780
   macro avg       0.83      0.83      0.83     70780
weighted avg       0.83      0.83      0.83     70780
```
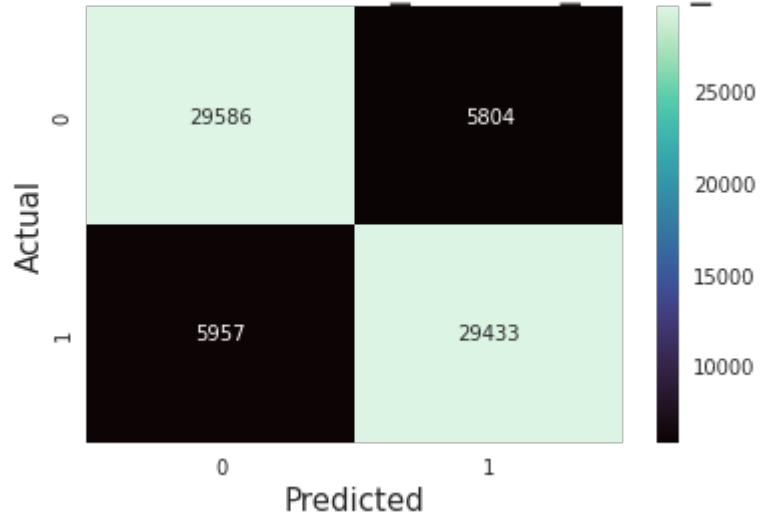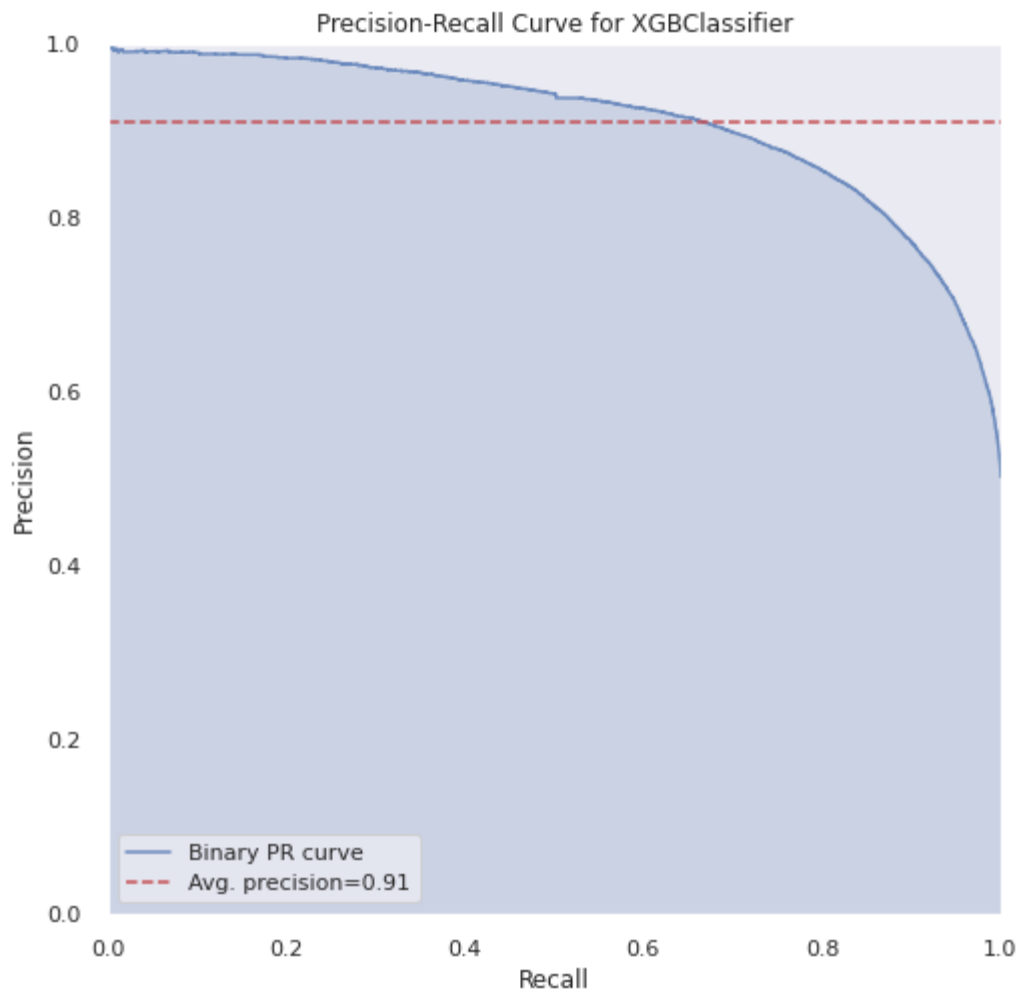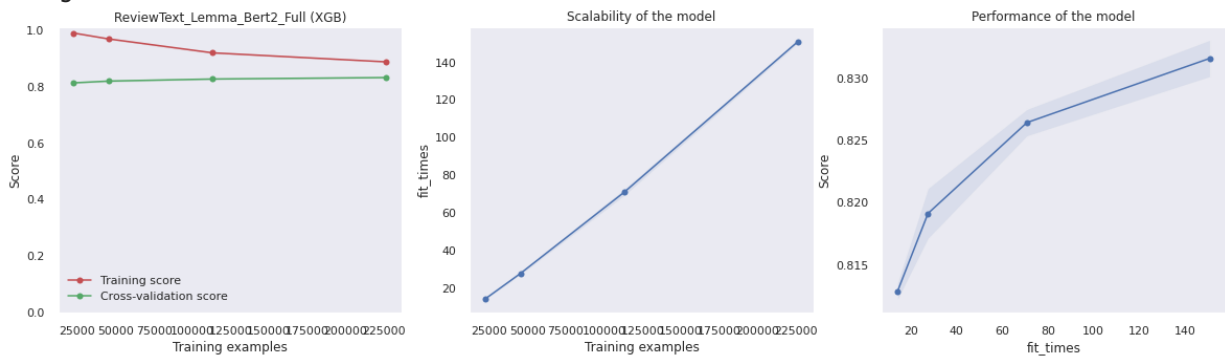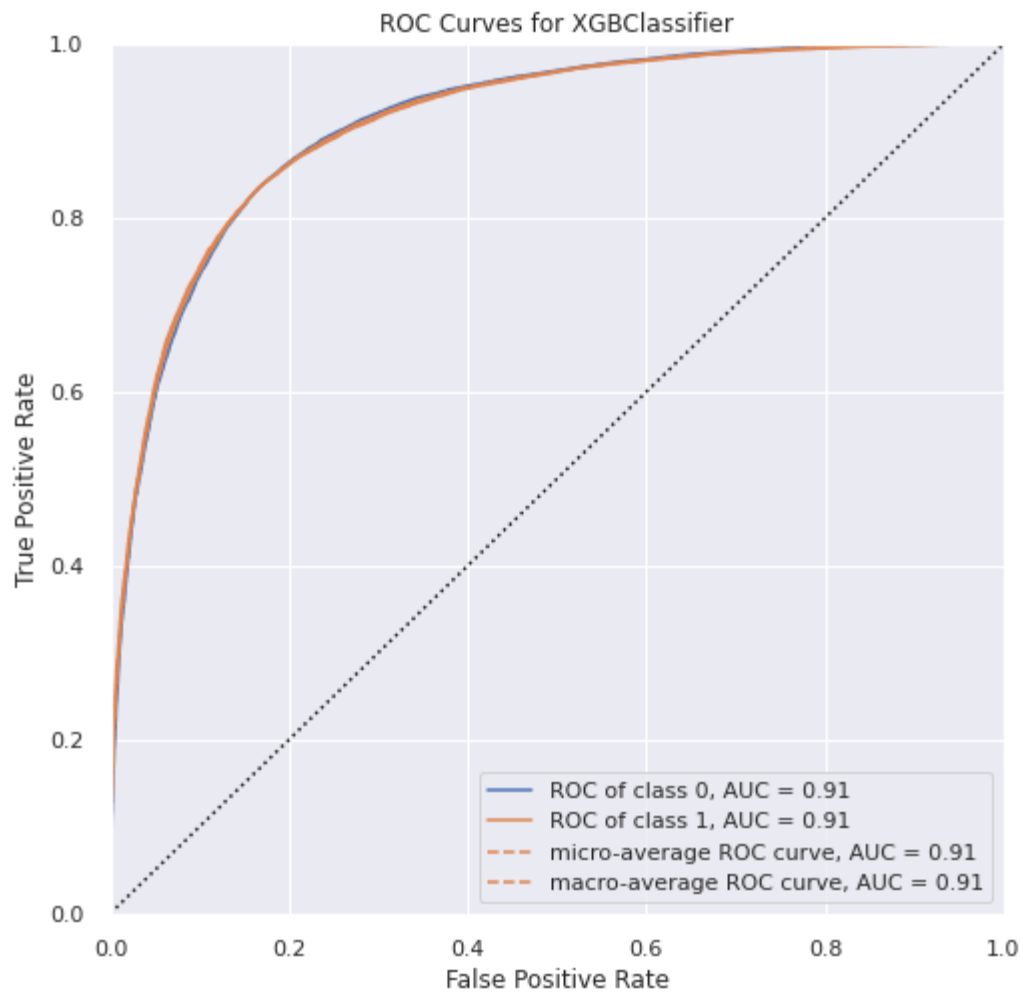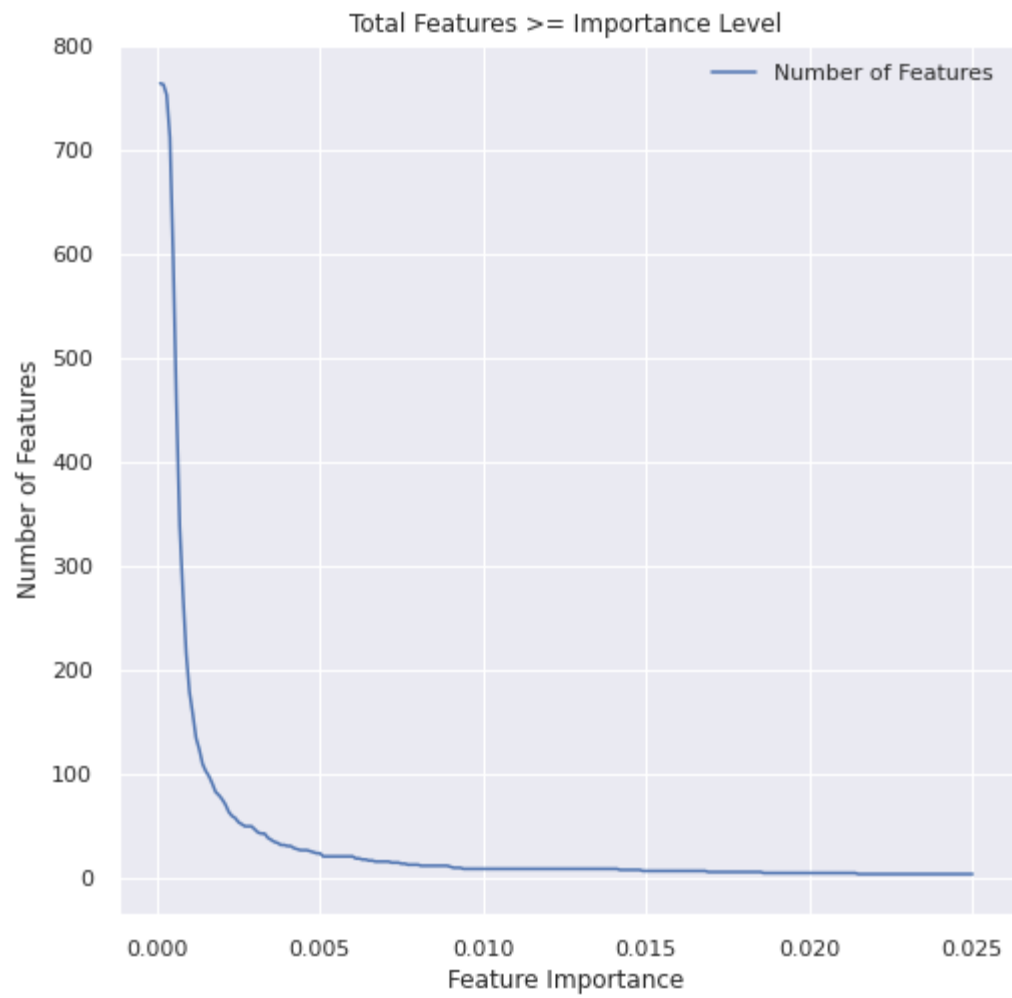
Confusion Matrix: ReviewText_Lemma_Bert2_Full (XGB)

Precision-Recall Curve for XGBClassifier

<Figure size 576x576 with 0 Axes>

Base model ROCAUC not calculated. Starting now

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/skle arn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprec ated and will be removed in a future release. To remove this warning, do the following: 1) Pass option use_label_encoder=False when constructing XGBClassi fier object; and 2) Encode your labels (y) as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
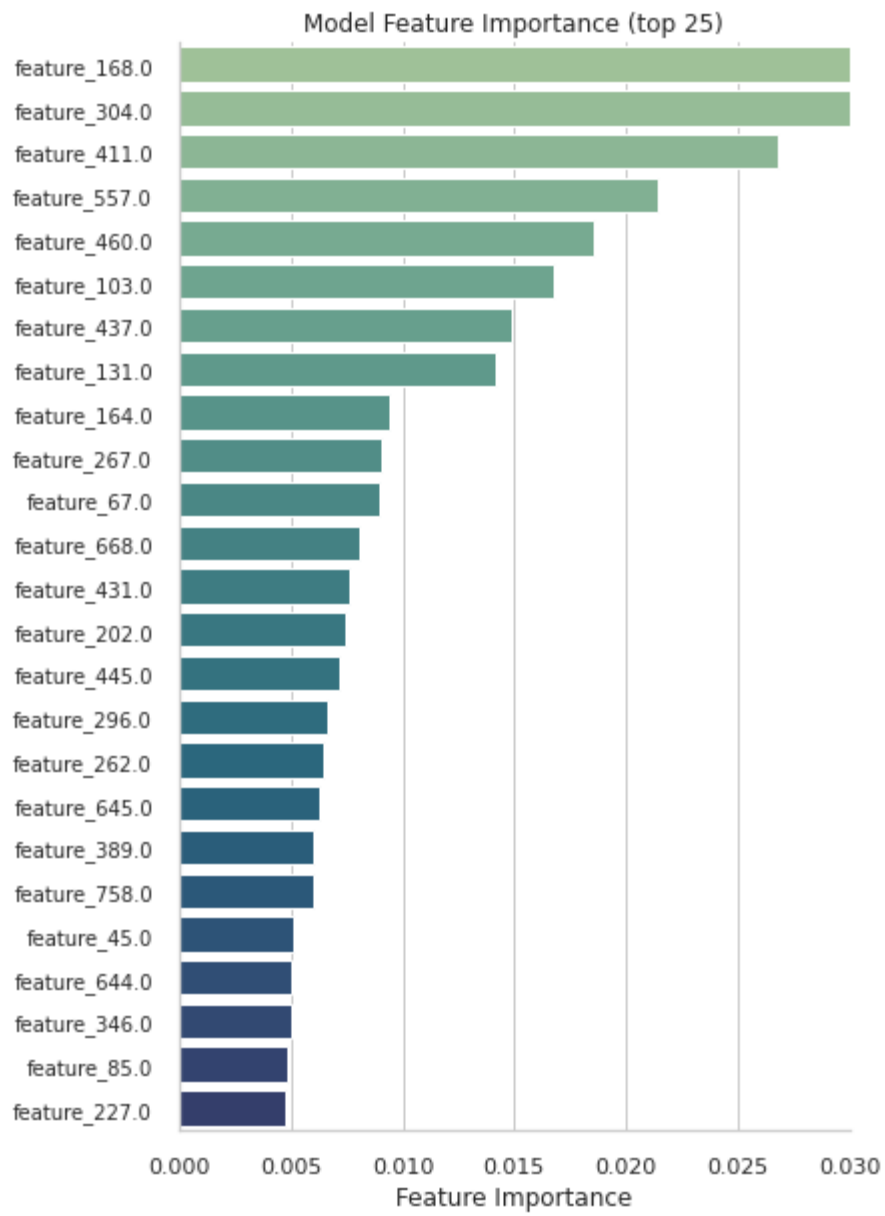 warnings.warn(label_encoder_deprecation_msg, UserWarning)

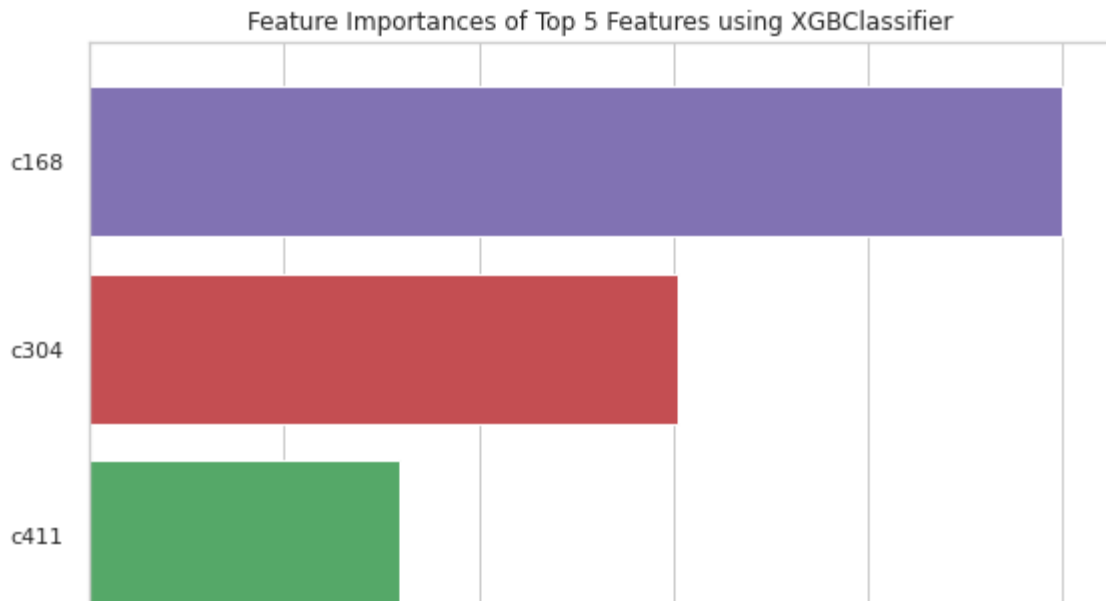ROC Curves for XGBClassifier

ROC of class 0, AUC = 0.91	
ROC of class 1, AUC = 0.91	
micro-average ROC curve, AUC = 0.91	
macro-average ROC curve, AUC = 0.91	

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

Model Feature Importance (top 25)

Feature Importances of Top 5 Features using XGBClassifier

c168

c304

c411

In [15]:
```
myExp.showFinalModelReport(axis_labels,
                          startValue=0.01,
                          increment=0.001,
                          upperValue=0.02)
```
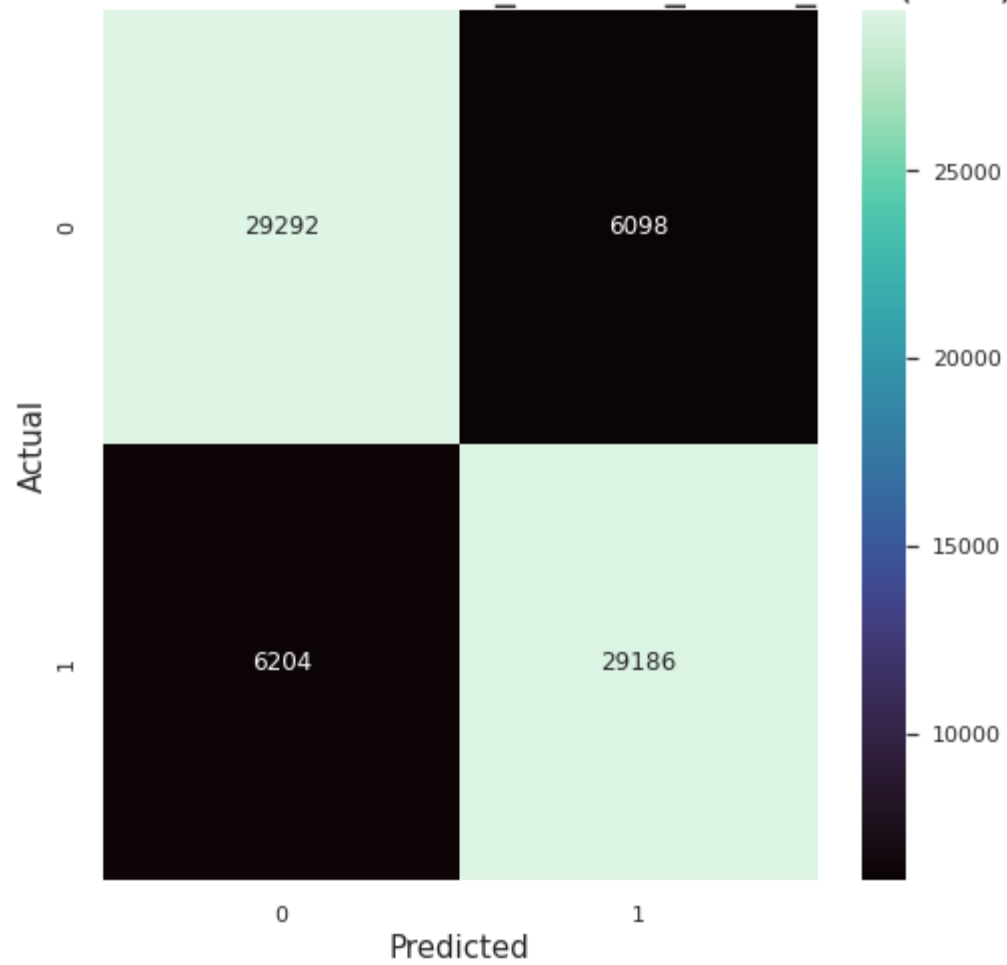
```
Final Model Stats:
Accuracy: 0.83
Precision: 0.83
Recalll: 0.83
F1 Score: 0.83
Cohen kappa:: 0.65
              precision    recall  f1-score   support

           0       0.83      0.83      0.83     35390
           1       0.83      0.82      0.83     35390

    accuracy                           0.83     70780
   macro avg       0.83      0.83      0.83     70780
weighted avg       0.83      0.83      0.83     70780
```
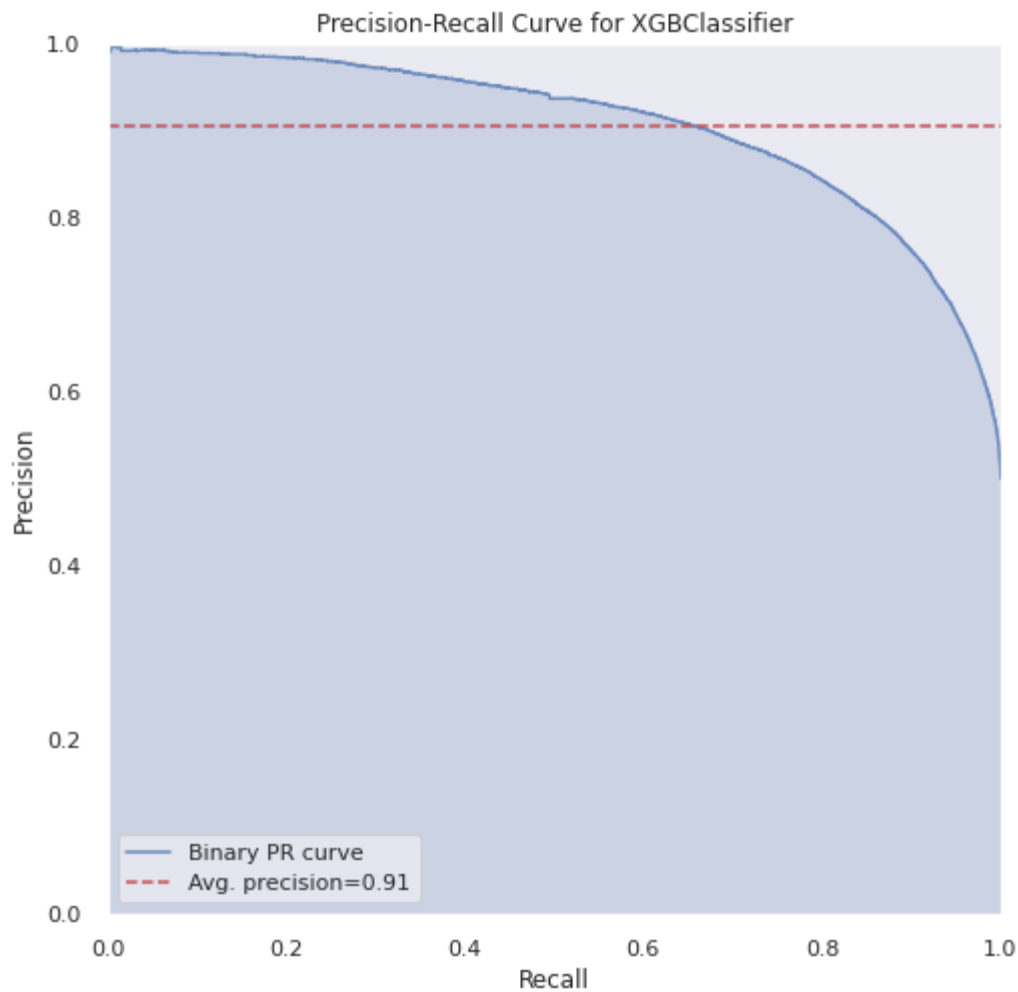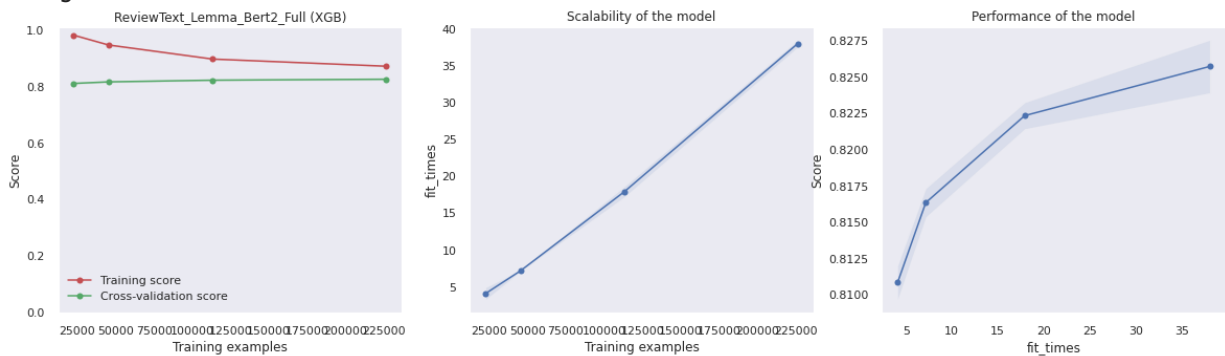
Confusion Matrix: ReviewText_Lemma_Bert2_Full (XGB)

Precision-Recall Curve for XGBClassifier

<Figure size 576x576 with 0 Axes>
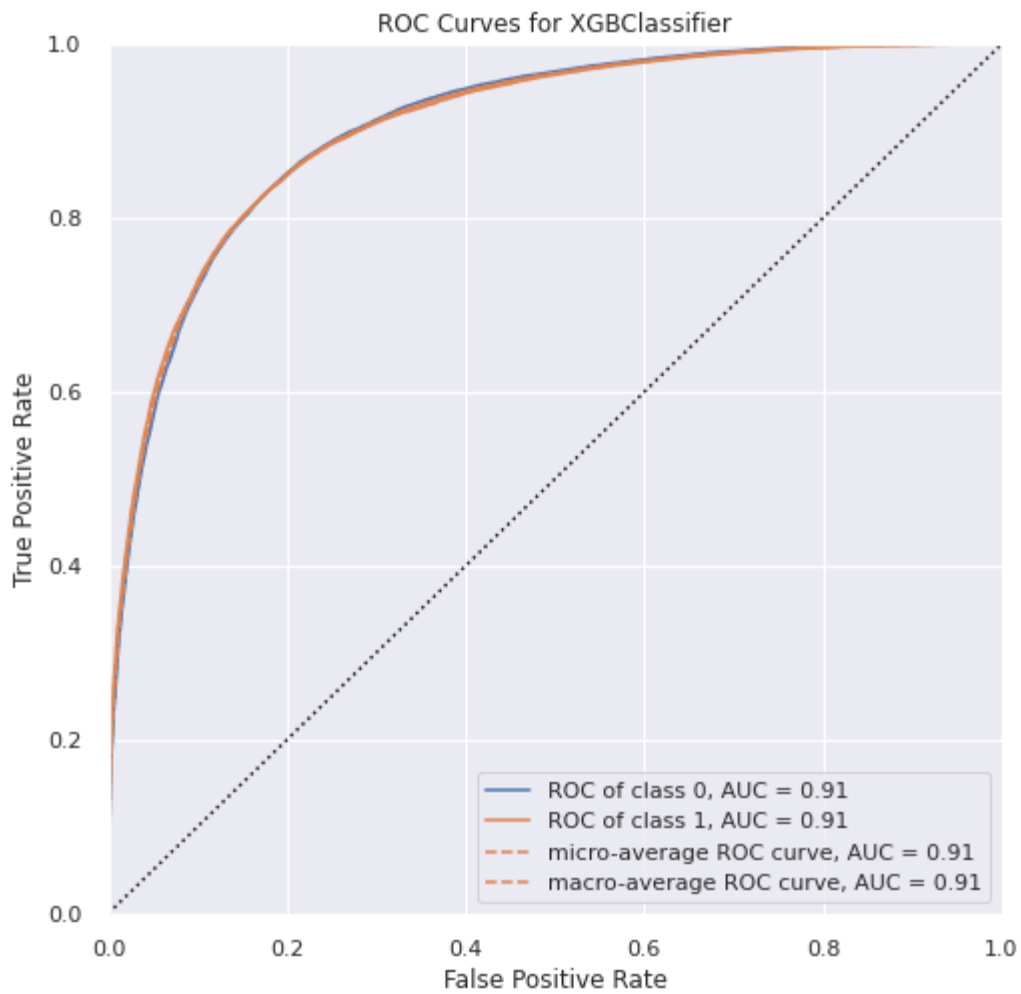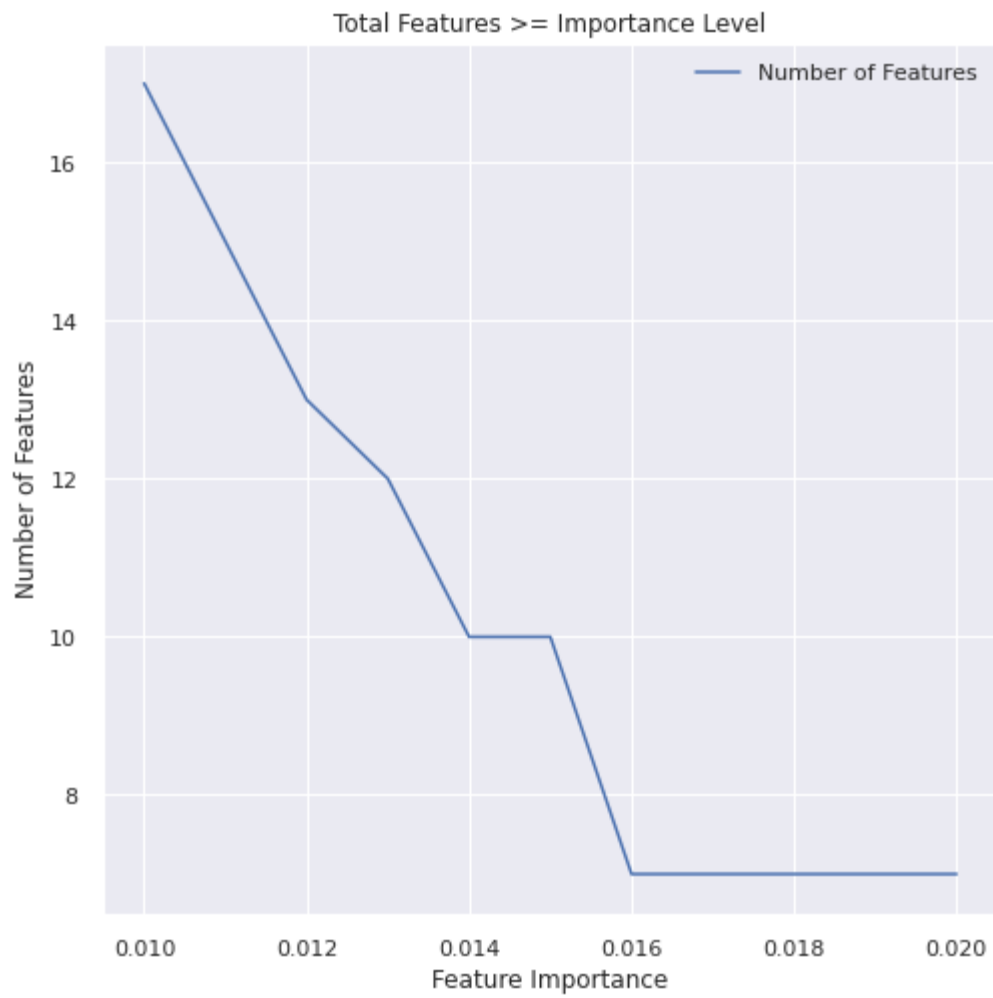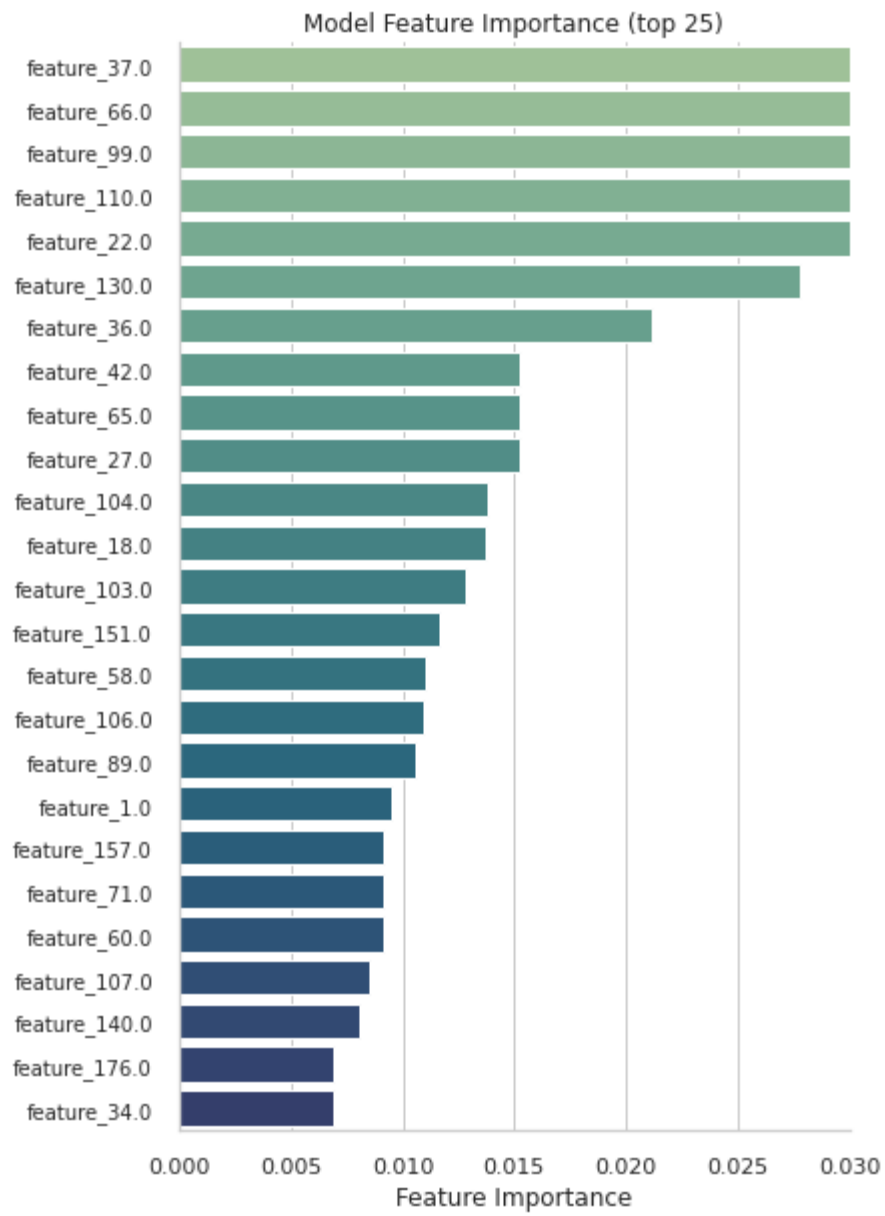

Final model ROCAUC not calculated. Starting now

/home/magni/python_env/ML1010_env2/lib64/python3.7/site-packages/xgboost/sklearn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprecated and will be removed in a future release. To remove this warning, do the following: 1) Pass option use_label_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y) as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)

ROC Curves for XGBClassifier

0%| | 0/12 [00:00<?, ?it/s]

Total Features >= Importance Level

Model Feature Importance (top 25)

Feature Importances of Top 5 Features using XGBClassifier

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

```
DataExperiment summary:
---> projectName: ML1010-Group-Project
---> experimentName: ReviewText_Lemma_Bert2_Full (XGB)
---> isDataPackageLoaded: True
---> isBaseModelLoaded: True
---> isBaseModelPredicted: True
---> isBaseModelLearningCurveCreated: True
---> isFinalModelLoaded: True
---> isFinalModelPredicted: True
---> isFinalModelLearningCurveCreated: True
---> isClassifierLoaded: True
XGBClassifier(base_score=None, booster=None, colsample_bylevel=None,
              colsample_bynode=None, colsample_bytree=None,
              enable_categorical=False, eval_metric='mlogloss', gamma=None,
              gpu_id=None, importance_type=None, interaction_constraints=Non
e,
              learning_rate=None, max_delta_step=None, max_depth=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, reg_alpha=None,
              reg_lambda=None, scale_pos_weight=None, subsample=None,
              tree_method=None, validate_parameters=None, verbosity=None)

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

Save Experiment

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

Scratchpad

In [18]:
```python
from google_colab import data_table
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
/tmp/ipykernel_297554/3560860020.py in <module>
----> 1 from google_colab import data_table
```