

Configuration

In [1]:

```
# Parameters
ENABLE_COLAB = False

PROJECT_NAME = 'ML1010-Group-Project'
EXPERIMENT_NAME = 'ReviewText_Lemma_Bert2_Full (XGB_Tune)'
FILE_NAME = '01_ML1010_GP_XGB_Bert2_Full_Tune'
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 18:50
 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
```

```
2022-01-26 18:50:18.597488: 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
2022-01-26 18:50:18.597548: 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]: #axis_labels=[1,2,3,4,5]
axis_labels=[0,1]
#classifier = RandomForestClassifier()
#classifier = XGBClassifier(eval_metric='mlogloss')
classifier = XGBClassifier(silent=True,
                           scale_pos_weight=1,
                           learning_rate=0.01,
                           colsample_bytree = 0.4,
                           subsample = 0.8,
                           objective='binary:logistic',
                           n_estimators=1000,
                           reg_alpha = 0.3,
                           max_depth=4,
                           gamma=10,
                           eval_metric='mlogloss')

ANALYSISIS_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_Cellphone_full.pkl.gz")

    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_Full (XGB_Tune)
---> 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=0.4,
              enable_categorical=False, eval_metric='mlogloss', gamma=10,
              gpu_id=None, importance_type=None, interaction_constraints=None,
              learning_rate=0.01, max_delta_step=None, max_depth=4,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=1000, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, reg_alpha=0.3, reg_lambda=None,
              scale_pos_weight=1, silent=True, subsample=0.8, 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

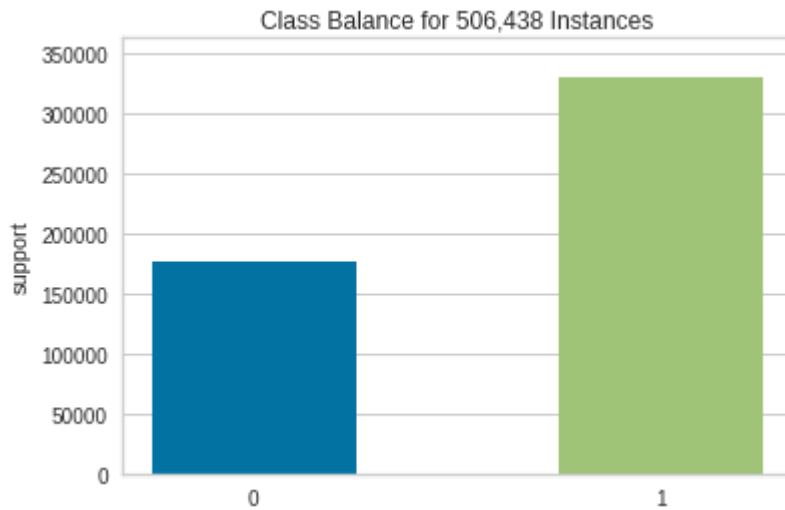
```

In [8]:

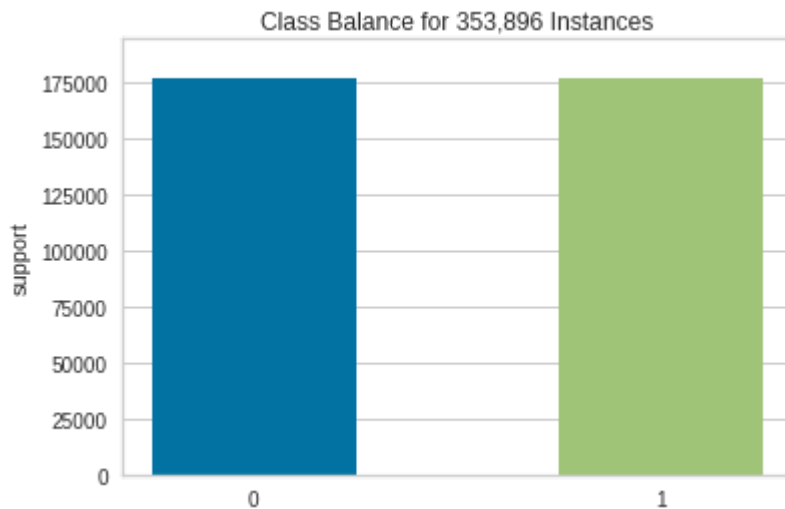
```

myExp.dataPackage.classBalanceUndersample()
myExp.dataPackage.splitTrainTest()

```



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



	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/sklearn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprecated

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 `XGBClassifier` object; and 2) Encode your labels (`y`) as integers starting with 0, i.e. 0, 1, 2, ..., `[num_class - 1]`.

[18:50:45] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

Base Model Stats:

Accuracy: 0.82

Precision: 0.82

Recall: 0.82

F1 Score: 0.82

Cohen kappa: 0.64

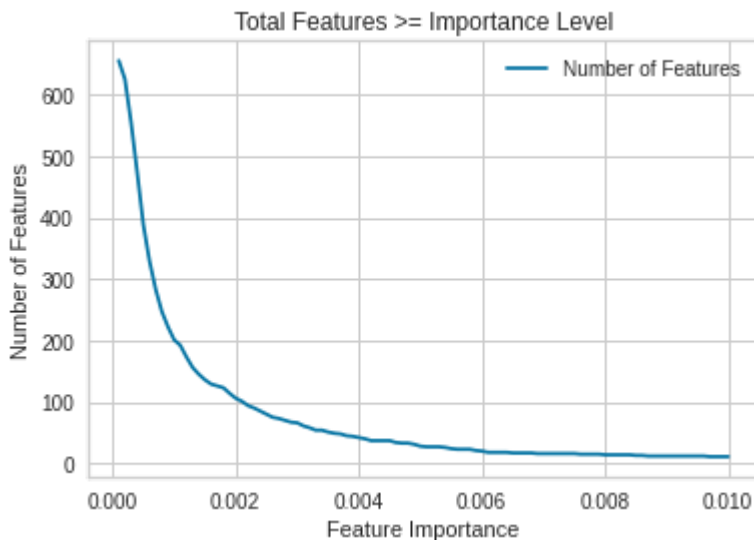
CPU times: user 1h 52min 11s, sys: 5.6 s, total: 1h 52min 16s

Wall time: 11min 50s

In [10]:

```
impDf = myExp.analyzeBaseModelFeatureImportance(returnAbove=0.0010)
```

```
0%|          | 0/101 [00:00<?, ?it/s]
Feature Importance Summary:
--> Original feature count: 768
--> Returned feature count: 201
--> Removed feature count: 567
--> Return items above (including): 0.001
```



In [11]:

```
%%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 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].
[19:02:35] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.

```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```

Final Model Stats:
Accuracy: 0.82
Precision: 0.82
Recall: 0.82
F1 Score: 0.82
Cohen kappa: 0.64
CPU times: user 29min 14s, sys: 1.28 s, total: 29min 16s
Wall time: 3min

```

In [12]:

```

%%time
myExp.createBaseModelLearningCurve()

```

```

[learning_curve] Training set sizes: [ 22649  45298 113246 226492]
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
/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)
[19:05:35] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.

```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```

[CV] END ....., score=(train=0.862, test=0.817) total time= 47.9s
[Parallel(n_jobs=1)]: Done 1 out of 1 | elapsed: 48.0s remaining: 0.0s
/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].

```

```
fier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
[19:06:23] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.843, test=0.818) total time= 1.5
min
[Parallel(n_jobs=1)]: Done 2 out of 2 | elapsed: 2.3min 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)
[19:07:55] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.830, test=0.817) total time= 3.8
min
[Parallel(n_jobs=1)]: Done 3 out of 3 | elapsed: 6.2min 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)
[19:11:46] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.824, test=0.817) total time= 8.0
```


min

```
/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)
```

```
[19:19:47] WARNING: ../src/learner.cc:576:
```

```
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but

then being mistakenly passed down to XGBoost core, or some parameter actually being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.861, test=0.817) total time= 5 2.7s
```

```
/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)
```

```
[19:20:39] WARNING: ../src/learner.cc:576:
```

```
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but

then being mistakenly passed down to XGBoost core, or some parameter actually being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.844, test=0.818) total time= 1.7 min
```

```
/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)
```

```
[19:22:22] WARNING: ../src/learner.cc:576:
```

```
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but

then being mistakenly passed down to XGBoost core, or some parameter actually being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.830, test=0.817) total time= 4.1
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)
[19:26:29] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.824, test=0.817) total time= 8.1
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)
[19:34:38] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.860, test=0.816) total time= 5
1.2s
/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)
[19:35:29] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such

cases.

```
[CV] END ..... , score=(train=0.844, test=0.816) total time= 1.7
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)
[19:37:14] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.829, test=0.816) total time= 3.9
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)
[19:41:06] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.825, test=0.817) total time= 8.0
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)
[19:49:09] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actual

ly being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.861, test=0.819) total time= 4
0.3~
/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)
[19:49:57] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but

then being mistakenly passed down to XGBoost core, or some parameter actual ly being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.843, test=0.820) total time= 1.6
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)
[19:51:32] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but

then being mistakenly passed down to XGBoost core, or some parameter actual ly being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.829, test=0.821) total time= 4.0
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)
[19:55:32] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language

bindings but
 then being mistakenly passed down to XGBoost core, or some parameter actual
 ly being used
 but getting flagged wrongly here. Please open an issue if you find any such
 cases.

```
[CV] END ....., score=(train=0.823, test=0.820) total time= 8.1
/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)
[20:03:36] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language
 bindings but
 then being mistakenly passed down to XGBoost core, or some parameter actual
 ly being used
 but getting flagged wrongly here. Please open an issue if you find any such
 cases.

```
[CV] END ....., score=(train=0.861, test=0.815) total time= 4
9.9s
/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)
[20:04:26] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language
 bindings but
 then being mistakenly passed down to XGBoost core, or some parameter actual
 ly being used
 but getting flagged wrongly here. Please open an issue if you find any such
 cases.

```
[CV] END ....., score=(train=0.843, test=0.817) total time= 1.8
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)
[20:06:13] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.829, test=0.816) total time= 3.9
/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)
[20:10:08] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.824, test=0.816) total time= 8.5
min
CPU times: user 11h 54min 7s, sys: 30 s, total: 11h 54min 37s
Wall time: 1h 13min 1s
[Parallel(n_jobs=1)]: Done 20 out of 20 | elapsed: 73.0min finished
```

In [13]:

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

```
[learning_curve] Training set sizes: [ 22649  45298 113246 226492]
[20:18:36] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
/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)
[CV] END ....., score=(train=0.850, test=0.812) total time= 2
5.7s
[20:19:02] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[Parallel(n_jobs=1)]: Done 1 out of 1 | elapsed: 25.8s 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=0.836, test=0.814) total time= 3
1.6s
[Parallel(n_jobs=1)]: Done 2 out of 2 | elapsed: 57.5s 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)
[20:19:34] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.825, test=0.814) total time= 1.3
min
[Parallel(n_jobs=1)]: Done 3 out of 3 | elapsed: 2.2min 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)
[20:20:51] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.821, test=0.814) total time= 2.6 min
[20:23:25] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
/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)
[CV] END ....., score=(train=0.847, test=0.814) total time= 18.3s
[20:23:43] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
/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)
[CV] END ....., score=(train=0.837, test=0.814) total time= 32.4s
[20:24:16] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
/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)
[CV] END ....., score=(train=0.825, test=0.815) total time= 1.3 min
```

```
/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)
[20:25:32] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.821, test=0.814) total time= 2.6 min
[20:28:09] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
/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)
[CV] END ....., score=(train=0.848, test=0.811) total time= 2 0.0s
[20:28:30] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actual

ly being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
/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)
[CV] END ..... , score=(train=0.837, test=0.813) total time= 3
3.9s
```

```
/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)
[20:29:04] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but

then being mistakenly passed down to XGBoost core, or some parameter actually being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.825, test=0.813) total time= 1.3
min
```

```
/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)
[20:30:21] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but

then being mistakenly passed down to XGBoost core, or some parameter actually being used

but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.821, test=0.815) total time= 2.8
min
```

```
[20:33:07] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language

bindings but
 then being mistakenly passed down to XGBoost core, or some parameter actual
 ly being used
 but getting flagged wrongly here. Please open an issue if you find any such
 cases.

```
/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=0.850, test=0.814) total time= 1
9.5s
[20:33:26] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language
 bindings but
 then being mistakenly passed down to XGBoost core, or some parameter actual
 ly being used
 but getting flagged wrongly here. Please open an issue if you find any such
 cases.

```
/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=0.837, test=0.817) total time= 3
5.6s
/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)
[20:34:02] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language
 bindings but
 then being mistakenly passed down to XGBoost core, or some parameter actual
 ly being used
 but getting flagged wrongly here. Please open an issue if you find any such
 cases.

```
[CV] END ....., score=(train=0.825, test=0.819) total time= 1.4
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 `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)
[20:35:26] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ....., score=(train=0.820, test=0.818) total time= 2.8
min
[20:38:14] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
/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=0.850, test=0.812) total time= 1
7.7s
[20:38:32] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
/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=0.837, test=0.812) total time= 3
4.5s
```

```
/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)
[20:39:07] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.825, test=0.813) total time= 1.4 min
```

```
/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)
[20:40:28] WARNING: ../src/learner.cc:576:
Parameters: { "silent" } might not be used.
```

This could be a false alarm, with some parameters getting used by language bindings but then being mistakenly passed down to XGBoost core, or some parameter actually being used but getting flagged wrongly here. Please open an issue if you find any such cases.

```
[CV] END ..... , score=(train=0.821, test=0.814) total time= 2.8 min
```

```
CPU times: user 3h 59min 15s, sys: 11.8 s, total: 3h 59min 27s
```

```
Wall time: 24min 40s
```

```
[Done: 11.1 / 100 jobs] Done: 20 out of 20. 1.1 loaded. 24.7min finished
```

In [14]:

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

Base Model Stats:

Accuracy: 0.82

Precision: 0.82

Recall: 0.82

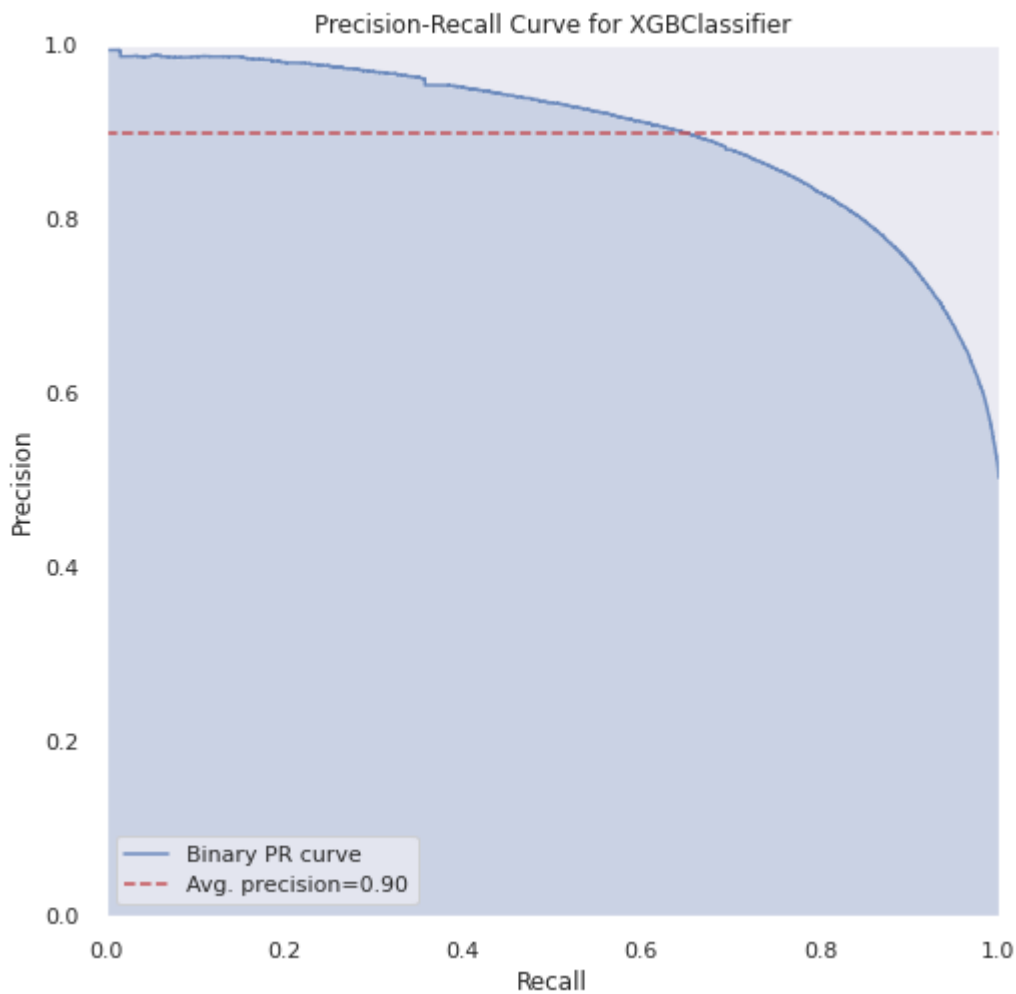
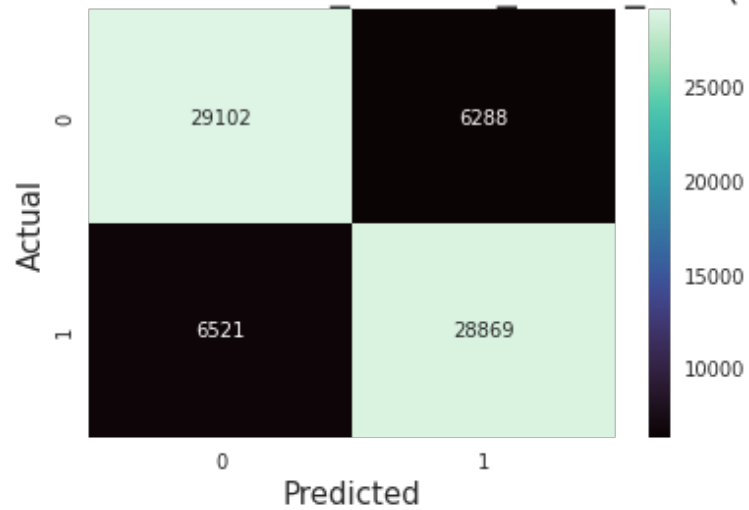
F1 Score: 0.82

Cohen kappa: 0.64

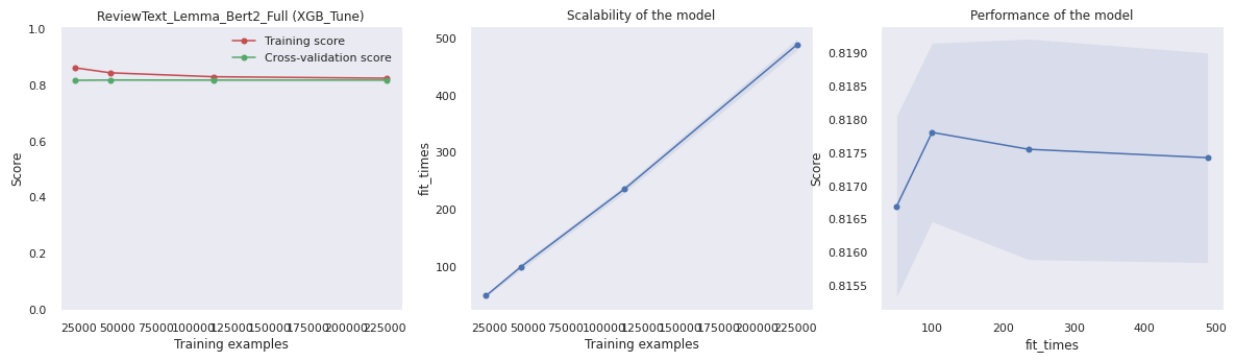
	precision	recall	f1-score	support
0	0.82	0.82	0.82	35390
1	0.82	0.82	0.82	35390

accuracy			0.82	70780
macro avg	0.82	0.82	0.82	70780
weighted avg	0.82	0.82	0.82	70780

Confusion Matrix: ReviewText_Lemma_Bert2_Full (XGB_Tune)



<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/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)
```

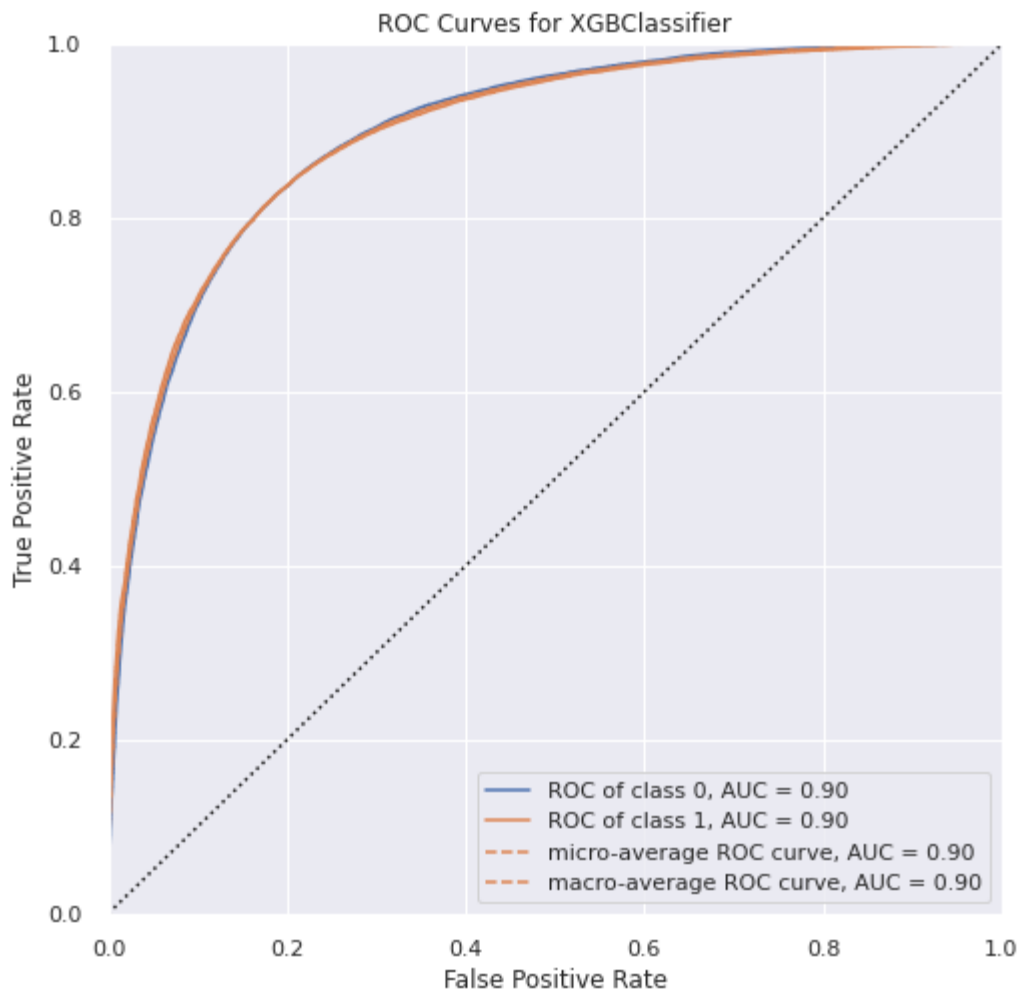
```
[20:43:20] WARNING: ../src/learner.cc:576:
```

```
Parameters: { "silent" } might not be used.
```

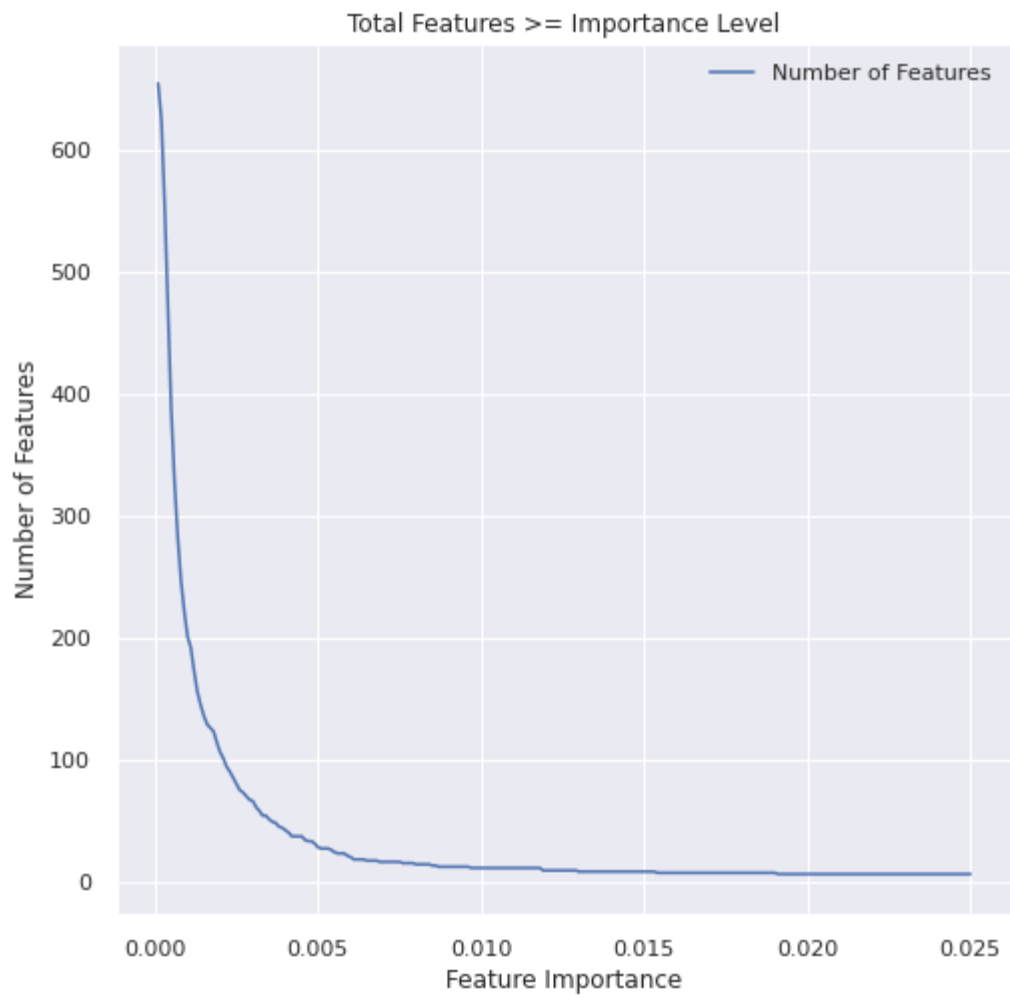
This could be a false alarm, with some parameters getting used by language bindings but

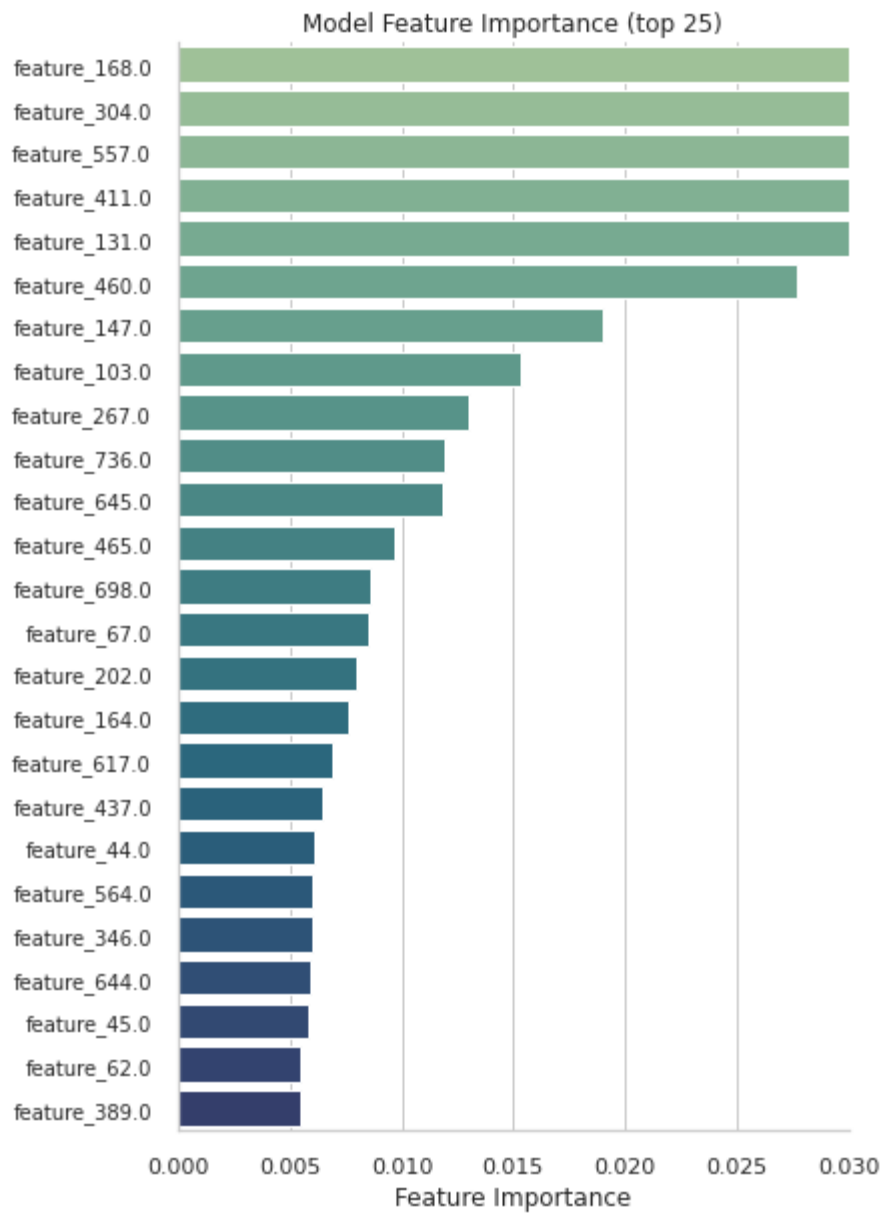
then being mistakenly passed down to XGBoost core, or some parameter actually being used

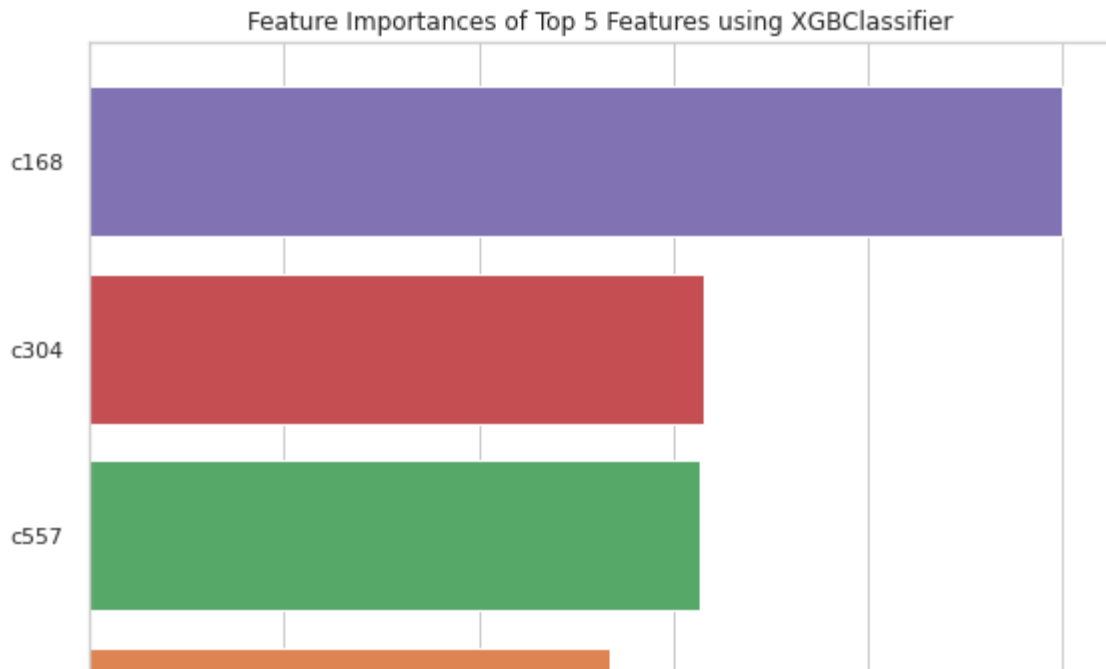
but getting flagged wrongly here. Please open an issue if you find any such cases.



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







```
In [15]: myExp.showFinalModelReport(axis_labels,
                                     startValue=0.01,
                                     increment=0.001,
                                     upperValue=0.02)
```

Final Model Stats:

Accuracy: 0.82

Precision: 0.82

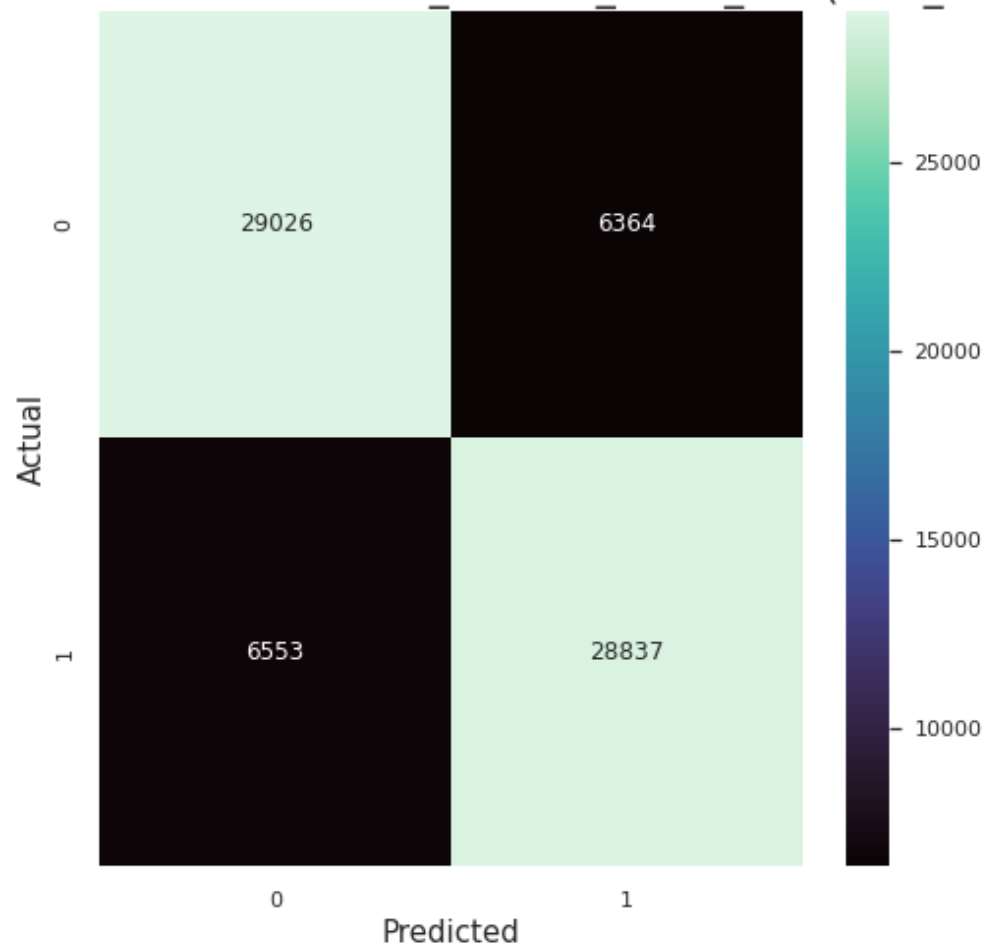
Recall: 0.82

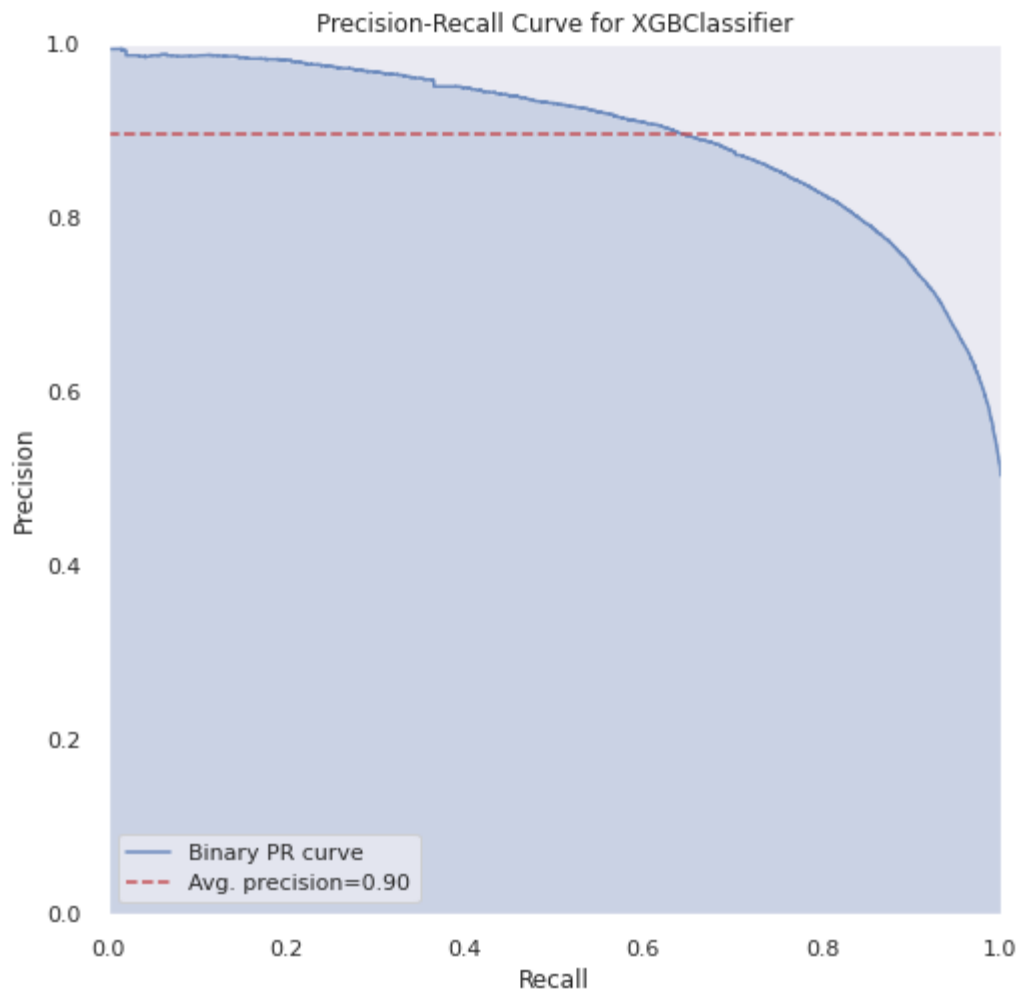
F1 Score: 0.82

Cohen kappa: 0.64

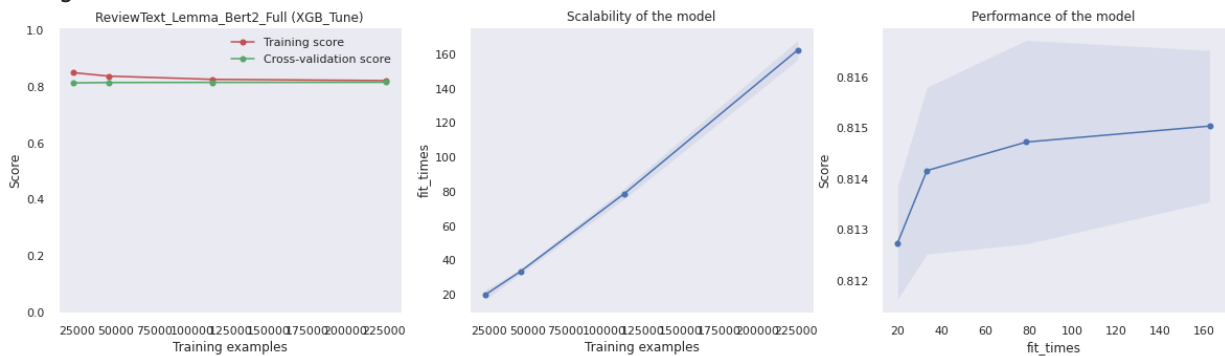
	precision	recall	f1-score	support
0	0.82	0.82	0.82	35390
1	0.82	0.81	0.82	35390
accuracy			0.82	70780
macro avg	0.82	0.82	0.82	70780
weighted avg	0.82	0.82	0.82	70780

Confusion Matrix: ReviewText_Lemma_Bert2_Full (XGB_Tune)





<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)
```

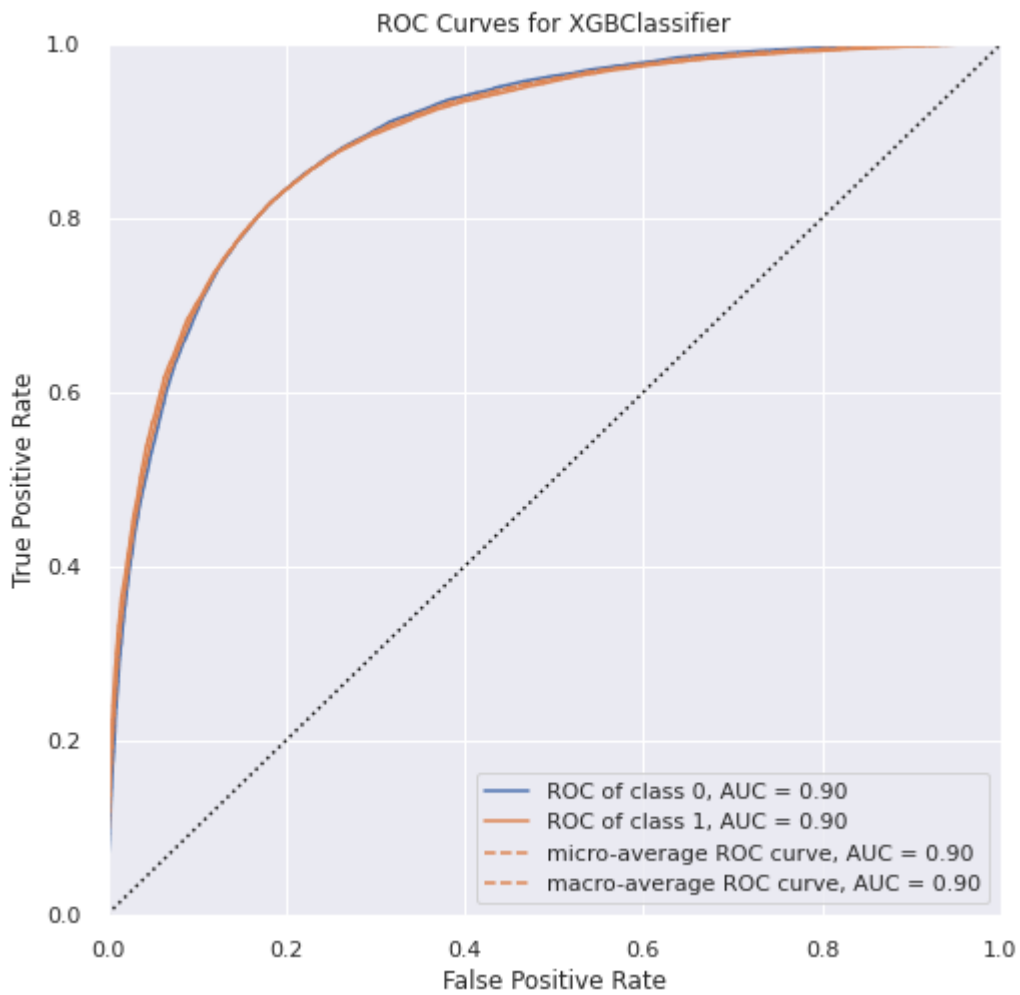
```
[20:59:01] WARNING: ../src/learner.cc:576:
```

```
Parameters: { "silent" } might not be used.
```

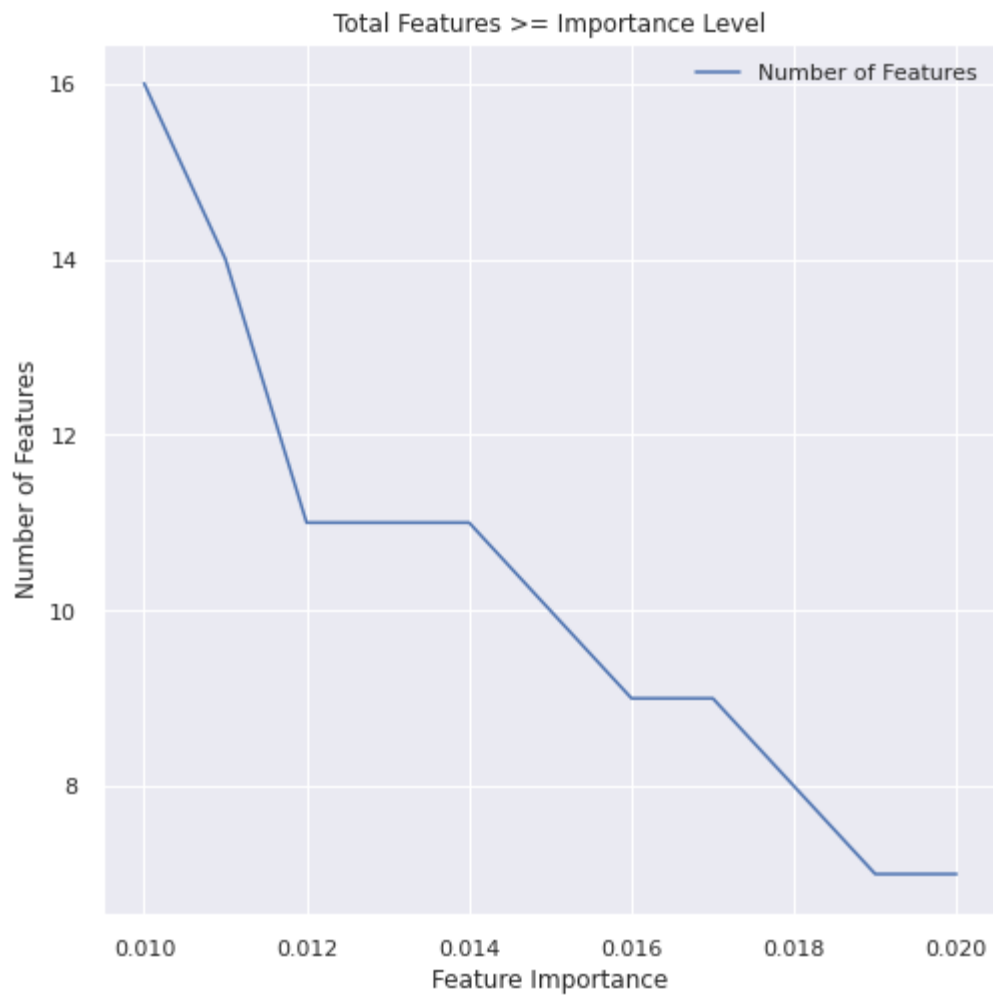
This could be a false alarm, with some parameters getting used by language bindings but

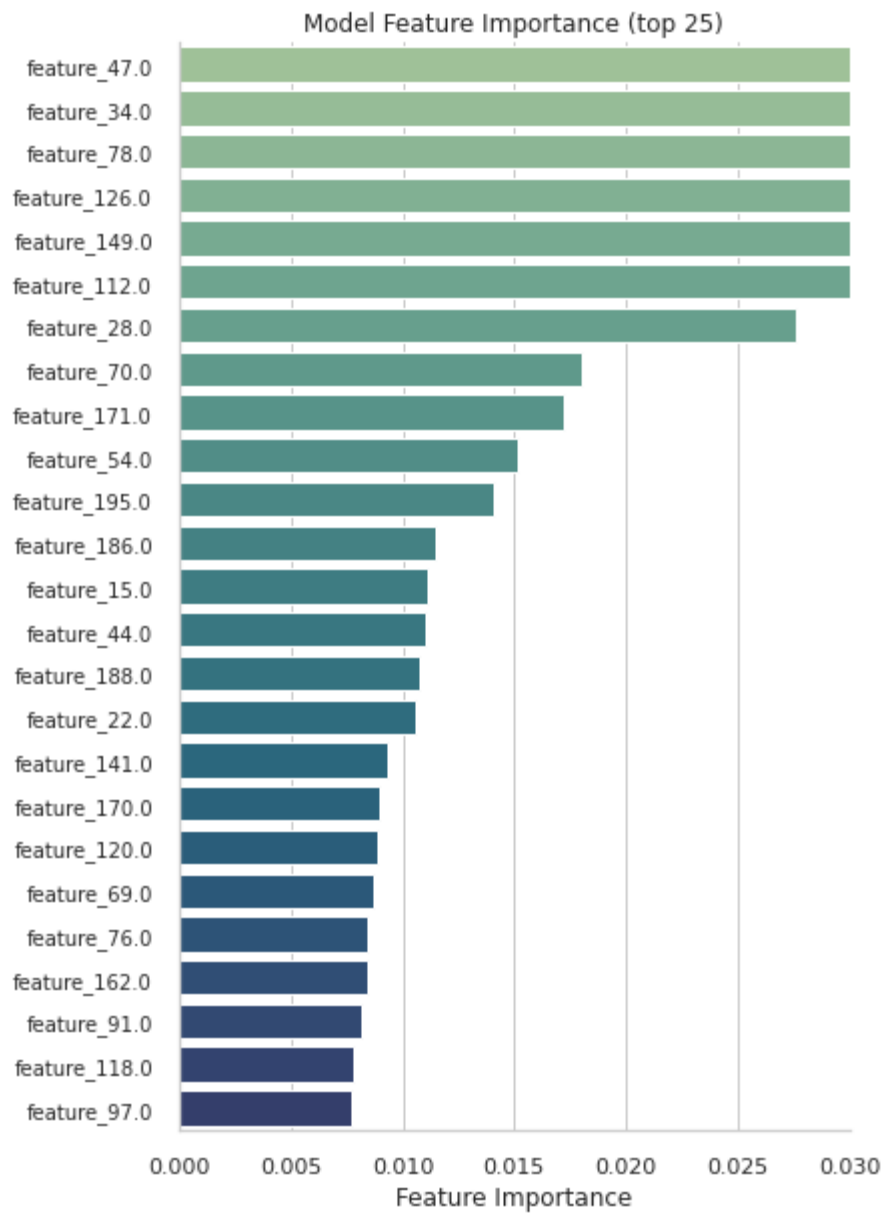
then being mistakenly passed down to XGBoost core, or some parameter actually being used

but getting flagged wrongly here. Please open an issue if you find any such cases.



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





Feature Importances of Top 5 Features using XGBClassifier

In [16]:

`myExp.display()`

DataExperiment summary:

```

---> projectName: ML1010-Group-Project
---> experimentName: ReviewText_Lemma_Bert2_Full (XGB_Tune)
---> 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=0.4,
              enable_categorical=False, eval_metric='mlogloss', gamma=10,
              gpu_id=None, importance_type=None, interaction_constraints=None,
              learning_rate=0.01, max_delta_step=None, max_depth=4,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=1000, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, reg_alpha=0.3, reg_lambda=None,
              scale_pos_weight=1, silent=True, subsample=0.8, 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]:

`jarvis.saveExperiment(myExp, FILE_NAME)`

Scratchpad

In []: