```
C:\Users\mkhan\PycharmProjects\SentimentAnalysis\venv\Scripts\python.exe C:/Users/mkhan/AppData/Local/JetBrains/Toolbox/apps/PyCharm-C/ch-0/222.4345.23/plugins/python-ce/helpers/pydev/pydevconsole.py --mode=client --host=127.0.0.1 --port=62260

import sys; print('Python %s on %s' % (sys.version, sys.platform))
sys.path.extend(['C:\\Users\\mkhan\\PycharmProjects\\SentimentAnalysis'])

PyDev console: starting.

Python 3.9.13 (tags/v3.9.13:6de2ca5, May 17 2022, 16:36:42) [MSC v.1929 64 bit (AMD64)] on win32
>>> # importing libraries
... import constants
... import definitions
... import os
... from src import load_data, data_cleaning, text_preprocessing, embedding, train_model, train_test_data_prep
...
... import pandas as pd
... pd.set_option('display.max_columns', None)
...
... # Model Training
... ----------------------------------------------------------------------
... # load data
... data = load_data.LoadData(os.path.join(definitions.ROOT_DIR, definitions.DATA_DIR, definitions.TRAIN_FILE))
... data.load_data()
...
... # clean data
... clean = data_cleaning.CleanData(data.df)
... clean.clean_data()
...
... # text preprocessing
... preprocessed = text_preprocessing.TextPreprocessing(clean.df)
... preprocessed.preprocess()
...
... # train-test split
... train_test_data = train_test_data_prep.TrainTestPrep(preprocessed.df)
... X_train, X_test, y_train, y_test = train_test_data.train_test()
...
```

```
... # bert embeddings
... encode_data = embedding.Embedding(X_train, X_test)
... encode_data.encode()
...
... # train model
... model = train_model.TrainModel()
... model.train_model(X_train=encode_data.X_train,
...                   y_train=y_train,
...                   X_test=encode_data.X_test,
...                   y_test=y_test)
...
... print(model.clf)
...
Data loaded with shape:  (31962, 3)
Snapshot of Data:
    id  label
   tweet
0   1      0   @user when a father is dysfunctional and is s...
1   2      0  @user @user thanks for #lyft credit i can't us...
2   3      0                                   bihday your majesty
3   4      0  #model   i love u take with u all the time in ...
4   5      0             factsguide: society now    #motivation
Data shape before cleaning:  (31962, 3)
Sentiment Distribution:
 0    29720
1     2242
Name: label, dtype: int64
 0    0.929854
1    0.070146
Name: label, dtype: float64
Data shape after cleaning:  (31962, 3)
Sentiment Distribution:
 0    29720
1     2242
Name: label, dtype: int64
 0    0.929854
1    0.070146
Name: label, dtype: float64
Converting to lower case..
Removing Punctuations..
Tokenization..
Lemmatization..
Joining words to text..

 Shape of Data:  (31962, 4)

Snapshot of Data:
    id  label
```

```
 81 tweet  \
 82 0   1      0  user when a father is dysfunctional and is so
     ...
 83 1   2      0  user user thanks for lyft credit i cant use ca
    ...
 84 2   3      0                                  bihday your
    majesty
 85 3   4      0  model i love u take with u all the time in urð
    ...
 86 4   5      0                  factsguide society now
    motivation
 87
 88                                             raw_text
 89 0   @user when a father is dysfunctional and is s...
 90 1  @user @user thanks for #lyft credit i can't us...
 91 2                               bihday your majesty
 92 3  #model   i love u take with u all the time in ...
 93 4             factsguide: society now    #motivation
 94 Shape of input data:  (31962, 4)
 95 Shape of training data:  (22373,)
 96 Shape of training data:  (9589,)
 97
 98 Converting to list..
 99
100 Complete!..
101 Encoding using BERT..
102 Encoding Train data..
103 Batches: 100%|██████████| 700/700 [48:52<00:00,  4.19s/it]
104 Complete!..
105 Encoding Test data..
106 Batches: 100%|██████████| 300/300 [07:56<00:00,  1.59s/it]
107 Complete!..
108 Fitting SVM Classifier
109 Complete!..
110 Predicting Train
111 Predicting Test
112 Evaluation on Train..
113 Confusion Matrix -
114  [[20611   166]
115  [  675   921]]
116 Classification report -
117                precision    recall  f1-score   support
118
119             0       0.97      0.99      0.98     20777
120             1       0.85      0.58      0.69      1596
121
122     accuracy                           0.96     22373
123    macro avg       0.91      0.78      0.83     22373
```

```
weighted avg       0.96      0.96      0.96     22373

AUC score -
 0.9560257854531887
Evaluation on Test..
Confusion Matrix -
 [[8862   81]
 [ 293  353]]
Classification report -
               precision    recall  f1-score   support

           0       0.97      0.99      0.98      8943
           1       0.81      0.55      0.65       646

    accuracy                           0.96      9589
   macro avg       0.89      0.77      0.82      9589
weighted avg       0.96      0.96      0.96      9589

AUC score -
 0.9484907683301432
SVC(kernel='linear', probability=True)
>>> import joblib
>>> joblib.dump(model.clf, os.path.join(definitions.ROOT_DIR,
    definitions.DATA_DIR, 'model.pkl'))
['C:\\Users\\mkhan\\PycharmProjects\\SentimentAnalysis\\data\\
    model.pkl']

```