Applied Machine Learning Homework 5

Due 2 May, 2022 (Monday) 11:59PM EST

Natural Language Processing

Eshan Kumar

ek3227

We will train a supervised training model to predict if a tweet has a positive or negative sentiment.

Dataset loading & dev/test splits

1.1) Load the twitter dataset from NLTK library

```python
In [1]: import nltk
        nltk.download('twitter_samples')
        from nltk.corpus import twitter_samples
```

```
[nltk_data] Downloading package twitter_samples to
[nltk_data]     /Users/eshankumar/nltk_data...
[nltk_data]   Package twitter_samples is already up-to-date!
```

1.2) Load the positive & negative tweets

```python
In [2]: all_positive_tweets = twitter_samples.strings('positive_tweets.json')
        all_negative_tweets = twitter_samples.strings('negative_tweets.json')
```

```python
In [6]: print("Example of positive tweet:\t" + all_positive_tweets[4])
        print("\nExample of negative tweet:\t" + all_negative_tweets[4])
```

```
Example of positive tweet:      yeaaaah yippppy!!!  my accnt verified rqst has succeed got a blue tick mar
k on my fb profile :) in 15 days

Example of negative tweet:      Dang starting next week I have "work" :(
```

1.3) Create a development & test split (80/20 ratio):

```python
In [39]: from sklearn.model_selection import train_test_split
         import pandas as pd

         posDF = pd.DataFrame(all_positive_tweets)
         posDF['sentiment'] = 1
         negDF = pd.DataFrame(all_negative_tweets)
         negDF['sentiment'] = 0

         tweetDF = posDF.append(negDF, ignore_index=True)

         X_data = tweetDF.drop(columns=['sentiment'])
         y_data = tweetDF['sentiment']

         X_dev_raw, X_test_raw, y_dev, y_test = train_test_split(X_data, y_data, stratify=y_data,
                                                  test_size = 0.2, random_state=42)
```

```
/var/folders/mg/dz4x1g196tg095k9bq0ckpw40000gn/T/ipykernel_5950/3089239367.py:9: FutureWarning: The frame.
append method is deprecated and will be removed from pandas in a future version. Use pandas.concat instea
d.
  tweetDF = posDF.append(negDF, ignore_index=True)
```

Data preprocessing

We will do some data preprocessing before we tokenize the data. We will remove # symbol, hyperlinks, stop words & punctuations from the data. You can use the re package in python to find and replace these strings.

1.4) Replace the # symbol with '' in every tweet

```
In [50]: X_dev_list = X_dev_raw[0].tolist()
         X_test_list = X_test_raw[0].tolist()

         X_dev_poundless = [tweet.replace("#", "") for tweet in X_dev_list]
         X_test_poundless = [tweet.replace("#", "") for tweet in X_test_list]
```

```
1
```

1.5) Replace hyperlinks with '' in every tweet

```
In [53]: import re

         X_dev_no_link = [re.sub(r'http\S+', '', tweet) for tweet in X_dev_poundless]
         X_test_no_link = [re.sub(r'http\S+', '', tweet) for tweet in X_test_poundless]

         print(f"Before removing hyperlink:\t{X_dev_poundless[3]}")
         print(f"\nAfter removing hyperlink:\t{X_dev_no_link[3]}")
```

```
Before removing hyperlink:      Thankyou x :)  https://t.co/Y21tv7APVO (https://t.co/Y21tv7APVO)

After removing hyperlink:       Thankyou x :)
```

1.6) Remove all stop words

```
In [64]: from sklearn.feature_extraction.text import ENGLISH_STOP_WORDS
         print(f"Number of stop words: {len(ENGLISH_STOP_WORDS)}")
         print(f"Some stop words: {list(ENGLISH_STOP_WORDS)[:20]}")

         def remove_stop_words(inputList):
             outputList = []
             for tweet in inputList:
                 newtweet = tweet
                 for word in ENGLISH_STOP_WORDS:
                     newtweet = newtweet.replace(" "+word+" ", " ")
                 outputList.append(newtweet)
             return outputList

         X_dev_no_stop = remove_stop_words(X_dev_no_link)
         X_test_no_stop = remove_stop_words(X_test_no_link)

         print(f"\nBefore removing stop words:\n{X_dev_no_link[16]}")
         print(f"\nAfter removing stop words:\n{X_dev_no_stop[16]}")
```

```
Number of stop words: 318
Some stop words: ['next', 'show', 'some', 'each', 'de', 'nothing', 'almost', 'nine', 'my', 'than', 'they',
'what', 'un', 'eg', 'besides', 'these', 'twenty', 'where', 'once', 'else']

Sentence with stop words included:
@JabongIndia ready and eagerly waiting for the next question. bring it on :) JabongatPumaUrbanStampede Jab
ongatPumaUrbanStampede

Sentence with stop words removed:
@JabongIndia ready eagerly waiting question. bring :) JabongatPumaUrbanStampede JabongatPumaUrbanStampede
```

1.7) Remove all punctuations

```
In [65]: X_dev_no_punc = [re.sub(r'[^\w\s]', '', tweet) for tweet in X_dev_no_stop]
         X_test_no_punc = [re.sub(r'[^\w\s]', '', tweet) for tweet in X_test_no_stop]

         print(f"\nBefore removing punctuation:\n{X_dev_no_stop[16]}")
         print(f"\nAfter removing punctuation:\n{X_dev_no_punc[16]}")
```

```
Before removing punctuation:
@JabongIndia ready eagerly waiting question. bring :) JabongatPumaUrbanStampede JabongatPumaUrbanStampede

After removing punctuation:
JabongIndia ready eagerly waiting question bring  JabongatPumaUrbanStampede JabongatPumaUrbanStampede
```

1.8) Apply stemming on the development & test datasets using Porter algorithm

```python
In [77]:  from nltk.tokenize import word_tokenize, sent_tokenize
          from nltk.stem import *
          import nltk
          nltk.download('punkt')

          def stemSentence(sentence):
              token_words = word_tokenize(sentence)
              stem_sentence=[porter.stem(word) for word in token_words]
              return " ".join(stem_sentence)
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     /Users/eshankumar/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

```python
In [79]:  porter = PorterStemmer()

          X_dev = [stemSentence(tweet) for tweet in X_dev_no_punc]
          X_test = [stemSentence(tweet) for tweet in X_test_no_punc]

          print(f"\nBefore stemming:\n{X_dev_no_punc[16]}")
          print(f"\nAfter stemming:\n{X_dev[16]}")
```

```
Before stemming:
JabongIndia ready eagerly waiting question bring  JabongatPumaUrbanStampede JabongatPumaUrbanStampede

After stemming:
jabongindia readi eagerli wait question bring jabongatpumaurbanstamped jabongatpumaurbanstamped
```

Model training

1.9) Create bag of words features for each tweet in the development dataset

```python
In [90]:  from sklearn.feature_extraction.text import CountVectorizer
          vector = CountVectorizer()

          vector.fit(X_dev)
          print(f"Some feature names in the fitted vector (dictionary):\n\
                  {vector.get_feature_names()[4000:4020]}")
          X_dev_BOW_vector = vector.transform(X_dev)
          X_test_BOW_vector = vector.transform(X_test)

          print(f"\nShape of X_dev:\t\t{X_dev_BOW_vector.shape}")
          print(f"Shape of X_test\t\t{X_test_BOW_vector.shape}")
```

```
Some feature names in the fitted vector (dictionary):
['ebook', 'ebookwoman', 'ecclestech', 'eckoxsoldi', 'eclaire605209', 'ecotr', 'eczema', 'eczemanomor', 'e
d', 'ed_dur', 'eddie_avi', 'edelabayog', 'edenador', 'edgar_trilla', 'edin_gam', 'edinunineuro', 'edit',
'editi', 'editionsdulivr', 'edm']

Shape of X_dev:          (8000, 14735)
Shape of X_test          (2000, 14735)

/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/utils/deprecation.py:87: FutureWarning: F
unction get_feature_names is deprecated; get_feature_names is deprecated in 1.0 and will be removed in 1.
2. Please use get_feature_names_out instead.
  warnings.warn(msg, category=FutureWarning)
```

1.10) Train a supervised learning model of choice on the development dataset

```
In [85]: from sklearn.linear_model import LogisticRegressionCV

         lr_BOW = LogisticRegressionCV().fit(X_dev_BOW_vector, y_dev)
```

/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

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

Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

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

Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

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


/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

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


**1.11) Create TF-IDF features for each tweet in the development dataset**

In [91]:
```python
from sklearn.feature_extraction.text import TfidfVectorizer
vector = TfidfVectorizer()

vector.fit(X_dev)
print(f"Some feature names in the fitted vector (dictionary):\n\
        {vector.get_feature_names()[4000:4020]}")
X_dev_TFIDF_vector = vector.transform(X_dev)
X_test_TFIDF_vector = vector.transform(X_test)

print(f"\nShape of X_dev:\t\t{X_dev_TFIDF_vector.shape}")
print(f"Shape of X_test\t\t{X_test_TFIDF_vector.shape}")
```

```
Some feature names in the fitted vector (dictionary):
['ebook', 'ebookwoman', 'ecclestech', 'eckoxsoldi', 'eclaire605209', 'ecotr', 'eczema', 'eczemanomor', 'e
d', 'ed_dur', 'eddie_avi', 'edelabayog', 'edenador', 'edgar_trilla', 'edin_gam', 'edinunineuro', 'edit',
'editi', 'editionsdulivr', 'edm']

Shape of X_dev:         (8000, 14735)
Shape of X_test         (2000, 14735)
```

1.12) Train the same supervised learning algorithm on the development dataset with TF-IDF features

```
lr_TFIDF = LogisticRegressionCV().fit(X_dev_TFIDF_vector, y_dev)
```

/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/pr
eprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.or
g/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
/Users/eshankumar/mambaforge/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:814: Convergenc
eWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

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

1.13) Compare the performance of the two models on the test dataset

```
In [93]: print(f"\nScore of logistic regression model trained on BOW representation:\t\
             {lr_BOW.score(X_test_BOW_vector, y_test)}")
         print(f"Score of logistic regression model trained on TF-IDF representation\t\
             {lr_TFIDF.score(X_test_TFIDF_vector, y_test)}")
```

```
Score of logistic regression model trained on BOW representation:        0.755
Score of logistic regression model trained on TF-IDF representation      0.754
```