269494 Final Exam

- What is the news sentiment analysis system? (1 points)

  : The system that is used to detect the news content or message that related with sentiment of the reader whether the news is positive, negative, or natural so that we will be able to use our model to predict the news content and the output as sentiment.

- Number of classes, features, size of training and testing data. (2 points)

  : According to my work, there are 2 classes the first class is sentiment class which define the sentiment of each news as positive, negative, or natural and the second class is message which are the content in the news. For the training and testing size I choose 40% for training and 60% for testing as the instruction is recommend.

- Data Preprocessing (if any)

  : For my data preprocessing I have did alphabet pattern, sequence pattern, and emojis replacement by using WordNetLemmatizer() to help in preprocessing.

```
[[124  74  32]
 [ 43 972 138]
 [ 52 215 288]]
              precision    recall  f1-score   support

    negative       0.57      0.54      0.55       230
     neutral       0.77      0.84      0.81      1153
    positive       0.63      0.52      0.57       555

    accuracy                           0.71      1938
   macro avg       0.66      0.63      0.64      1938
weighted avg       0.71      0.71      0.71      1938
```

As the result:

Precision of each class are

Negative: 57%

Natural: 77%

Positive: 63%

Recall of each class are

Negative: 54%

Natural: 84%

Positive: 52%

In my opinion, the value that should be considered is “Recall” because we are more focus to make the model correctly detected negative news sentiment.

The potential problem they could occur is that, the model has predicted negative news as positive news, the model has predicted positive news as negative news, or the model predict natural news as positive or negative news. Those problem make the model inaccurate which is the model is not efficient enough. The cause of those problem might come from not preprocessing the data well or not tuning the hyperparameter well and so on.

Link to GitHub

https://github.com/komcharn018/FinalExam_269494