269494 Final Exam

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

: The system that is used to detect the message (news) that related with sentiment for the reader who want to know the news is positive, natural or negative, so that we will be able to use our model to predict the news message and the output is sentiment

• Number of classes, features, size of training and testing data. (2 points)
: My work, there are 3 classes the first class is positive ,second is Neutral and the third is Negative

 Feature: There are two features are 'Sentiment' and 'Message'

Size of training and testing data:

For the training and testing size I choose 40% for training and 60% for testing as the instruction.

Results (5 points)

```
Logistic Regression Classification Report:
              precision    recall  f1-score   support

    negative       0.54      0.56      0.55       230
     neutral       0.77      0.80      0.79      1153
    positive       0.62      0.55      0.58       555

    accuracy                           0.70      1938
   macro avg       0.64      0.64      0.64      1938
weighted avg       0.70      0.70      0.70      1938

Logistic Regression Confusion Matrix:
[[128  78  24]
 [ 62 927 164]
 [ 49 201 305]]
```

Precision of each class are:

Negative: 54%
Natural: 77%
Positive: 62%

Recall of each class are:

Negative: 56%
Natural: 80%
Positive: 55%

In my opinion we should focus on "Recall" because we are want to make the model correctly to detected about negative sentiment.
Problems that may arise are: The model predicts negative news as positive news. The model predicts positive news as negative news. Or the model predicts natural news whether it is positive or negative news. Those problems make the model inaccurate, which is an inefficient model. Because those problems may come from not processing data well enough.