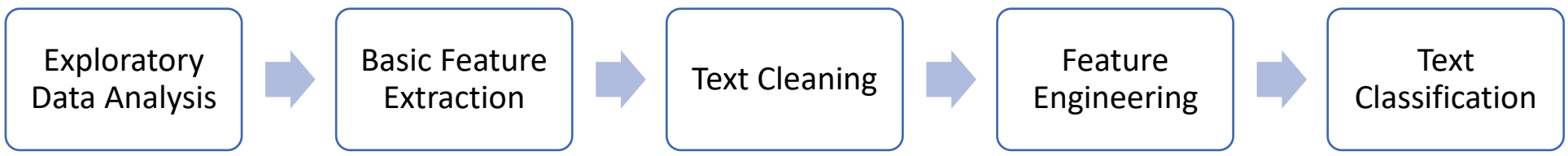
Real or Fake News? Using NLP with Disaster Tweets

Lim Zheng Wei, Institute of Data Capstone Project

Introduction

More people are spending time on the internet and social media which leads to more reliance getting news online. The evolution of communication also gave rise to fake news. In addition, people can post easily, there is nobody to verify the authenticity of the news. Because people shape their opinions based on information that they receive, having fake news can impact lives financially, socially, politically, and more. In this experiment, my aim is detecting real and fake news by classifying Twitter texts based on Natural Language Processing (NLP) and classifiers.

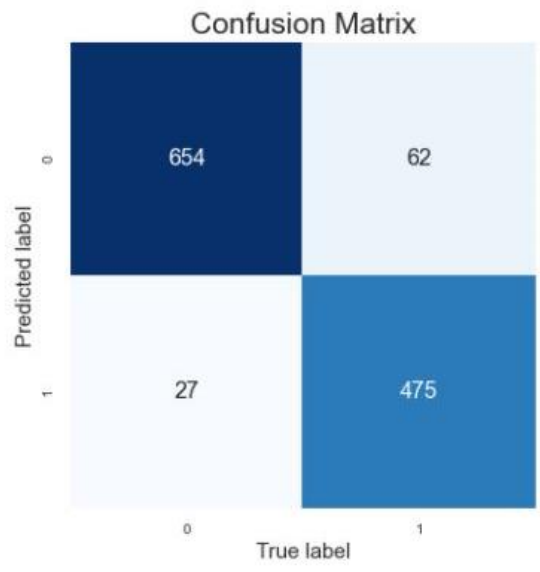
Methodology

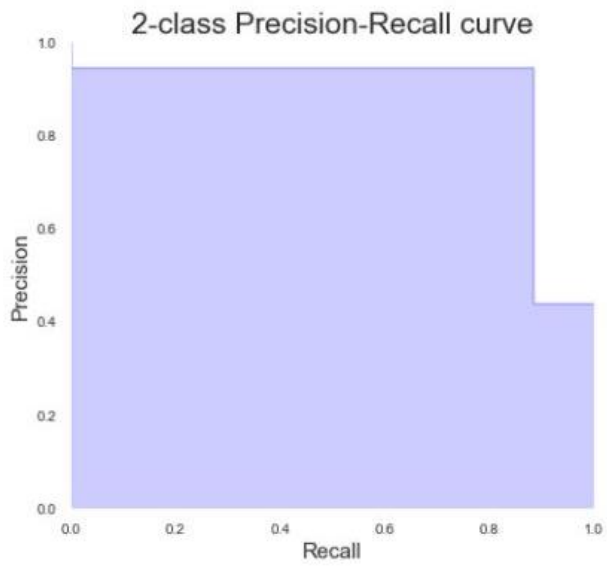


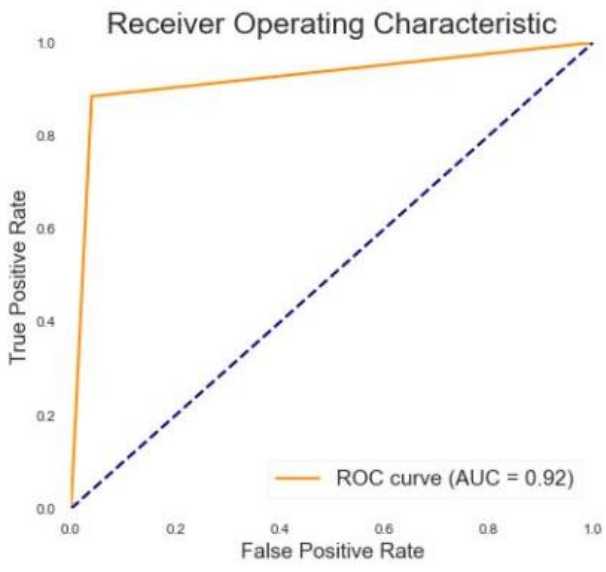
Results

Using Regularised Model from Keras produced the best classification.

```
Accuracy : 0.9269 [(TP + TN) / N] Proportion of predicted labels that match the true labels. Best: 1, Worst: 0
Precision: 0.9462 [TP / (TP + FP)] Not to label a negative sample as positive.        Best: 1, Worst: 0
Recall   : 0.8845 [TP / (TP + FN)] Find all the positive samples.                     Best: 1, Worst: 0
ROC AUC  : 0.9224                                                                     Best: 1, Worst: < 0.5
----------------------------------------------------------------------------------------------------------
TP: True Positives, FP: False Positives, TN: True Negatives, FN: False Negatives, N: Number of samples
```







Word Importance

	Words Associated
Fake News	Think, One, Say, Go, Love, See, Day, Make, Want, Time, Come, Know
Real News	Evacuate, Suicide, Bomber, Flood, Storm, Riot, Kill, Emergency, Northern California, Wildfire, Shoot, Damage, Severe, Thunderstorm

Conclusion & Future Works

My experiment has shown that deep learning is the most reliable method for classifying real and fake news. Considering there are many more texts on social media, for future work, other deep learning methods such as Bert can be used to differentiate the news.