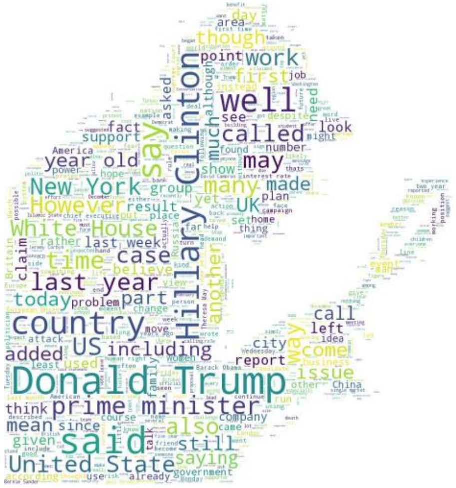
You are Fake News

Fake News Detection
Implement Report



Motivation & Background

The media play a facilitating role – in the easing through of policy action by repetition and reinforcement of media messages, and the absence of proposed alternatives – and also a possible role in shaping behaviour, especially where these are linked to other types of structural support.

Given the impact of media on our lives, Fake News has the potential to disrupt society and important decisions made by individuals and organisations alike, potentially causing havoc and undesired outcomes such as the Washington Pizzeria Attack.

Problem Statement

It is thought that Machine Learning Techniques could be used to evaluate the truthfulness of a given news article.

Therefore, our project aims to utilise Machine Learning Techniques on the detection of Fake News and create a working prototype that is able to identify fake news during the US Presidential Election with reasonable accuracy.

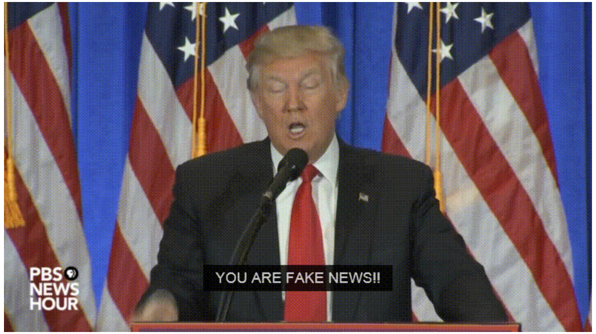
Data collection

We combined 3 different open sourced Fake News Datasets, the 3 of them being : Liar dataset from UCSB, fake news dataset from Kaggle and dataset published by Signal Media in conjunction with the Recent Trends in News Information Retrieval 2016 conference. To crawl true news data, we collect news from *New York Times* and *Guardians*. In different category of news, there will be different kind of fake news, due to the usage of terms. Because we only gather news of U.S president election and focus on mid-2016 to late-2016, in the limited category of news, we have 10305 true news, 8840 fake news, and 12980 other category of news. We split 10% of data for validation.

From this combined dataset, we selected the features in common amongst the dataset and used those features for evaluation. That is, date, author, title and content.

Our Solution

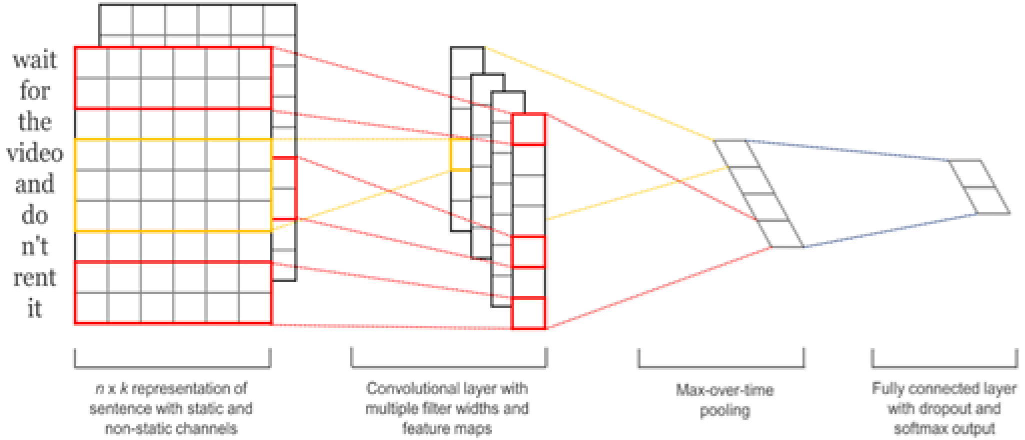
- **Natural Language Processing**
  - Natural Language Toolkit
  - Word Embedding (GloVe)
- **Nerual Networks**
  - Convultional Neural Network
  - Recurrent Neural Network (Bidirectional LSTM)



Our solution utilises Natural Language Processing Tools such as the Natural Language Toolkit in Python that is used to remove stop words, and a word embedding model, Global Vectors for word Representation (GloVe), and a Convolutional Neural Network, as well as a Bidirectional Long Short Term Memory (LSTM) Recurrent Neural Network.

The details of the implementation are shown in the next pages.

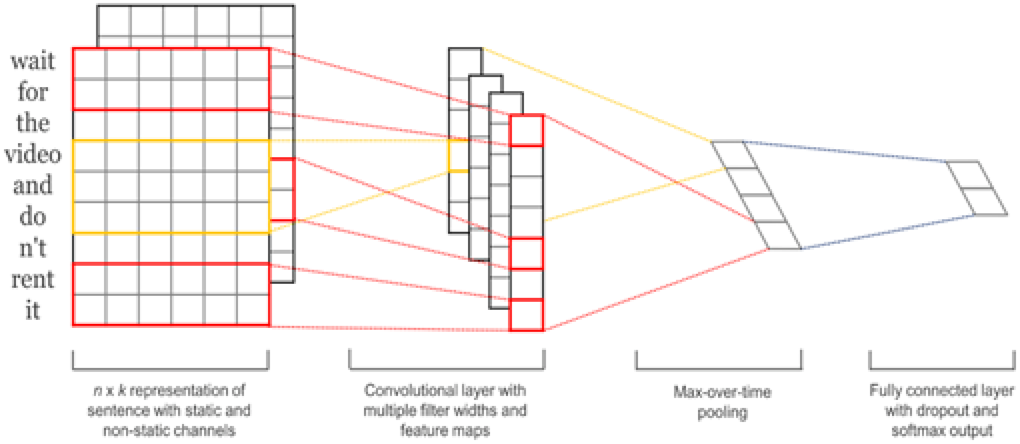
GloVe



▲Diagram showing how GloVe converts words into vectors

GloVe is a pre-trained Word Embedding model which converts words into a vector using the probability of co-occurrence between words to encode the meaning of a given text to determine the relation between words. We encode our metadata (author, title) and content to vectors.

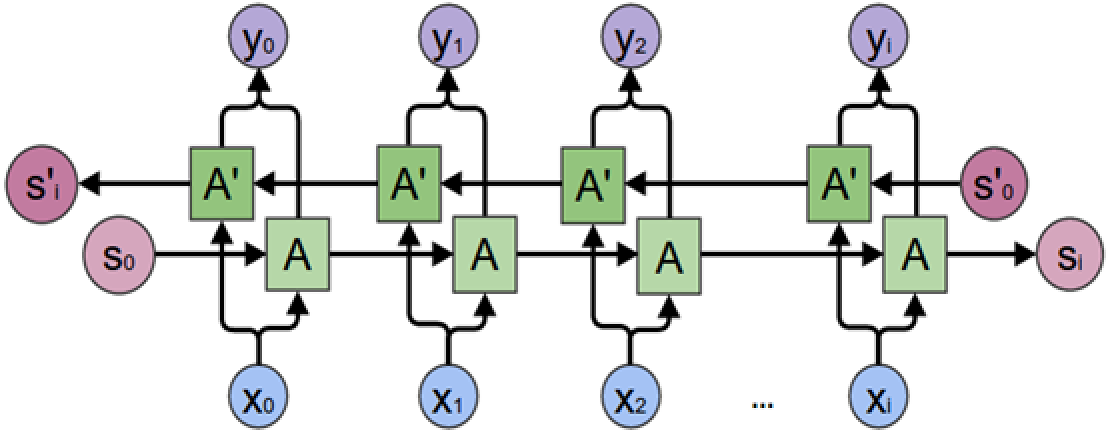
Convolutional Neural Network



▲Diagram of a Convolutional Neural Network

Convolutional neural networks (or convnets for short) are used in situations where data can be expressed as a "map" wherein the proximity between two data points indicates how related they are. Neurons don't have to recognize the whole sentences to discover the pattern. Our metadata will pass through it, and hope the model can learn the critical pattern that affects the news is true or not.

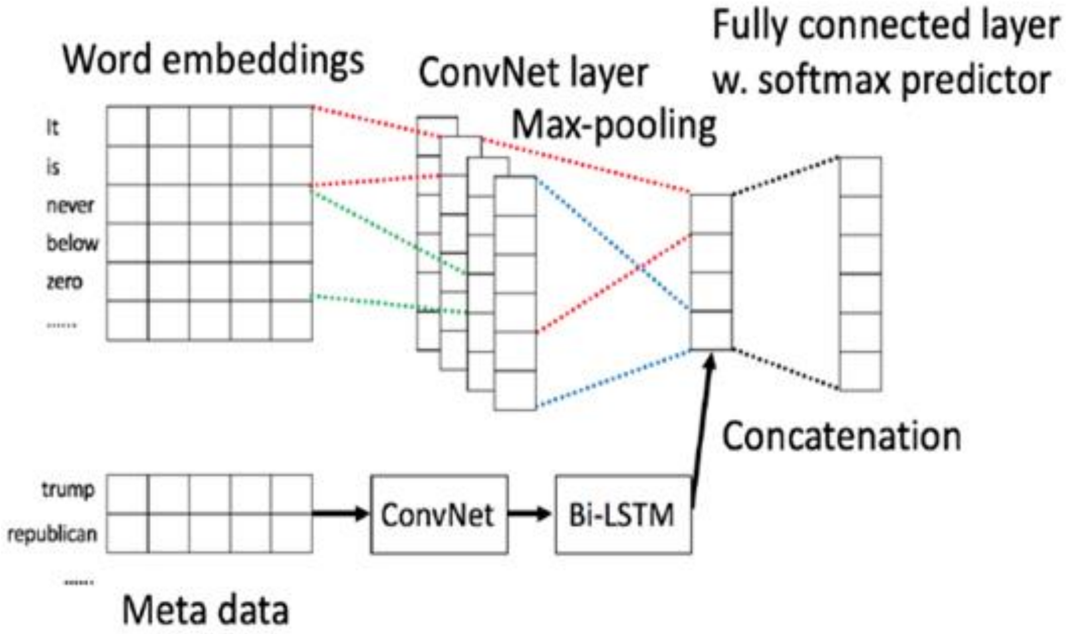
Recurrent Neural Network



▲Diagram of a Bi-directional LSTM Model

In this project, we utilise a Bidirectional Recurrent Network to split the neurons of a regular RNN into two directions, one for positive time direction (forward states), and another for negative time direction(backward states). Those two states' output are not connected to inputs of the opposite direction states. By using two time directions, input information from the past and future of the current time frame can be used unlike standard RNN which requires the delays for including future information. Due to most fake news have template, we hope the model can learn the relations between sentences in order to find out the fake news.

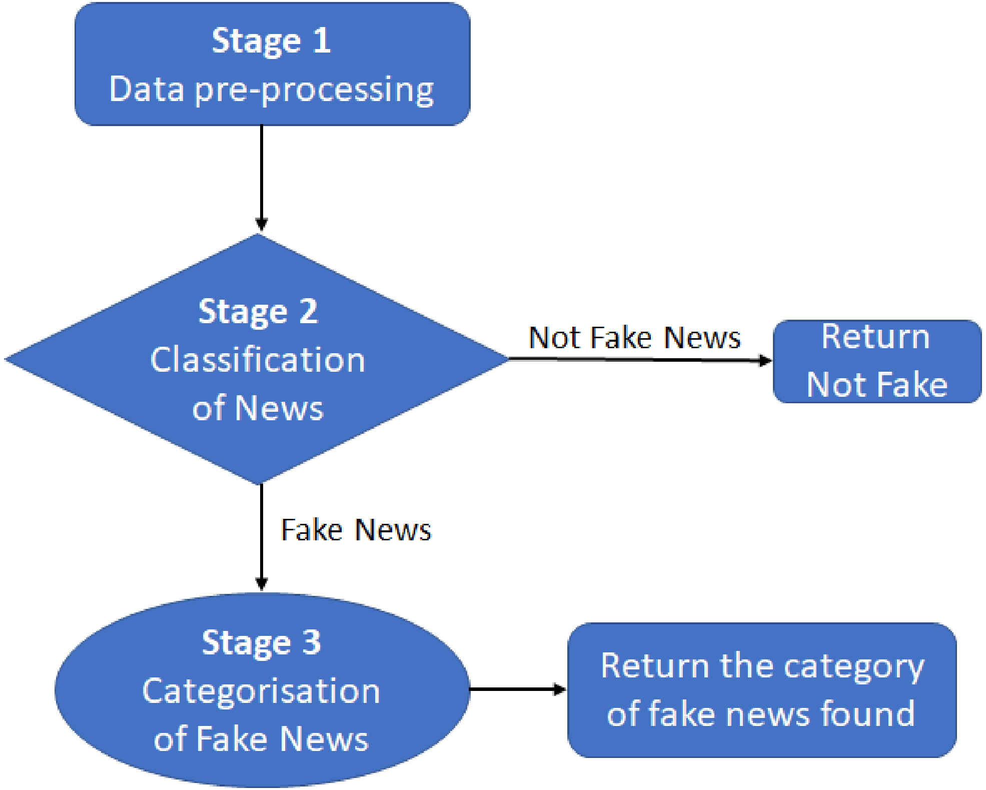
Bi-directional LSTM



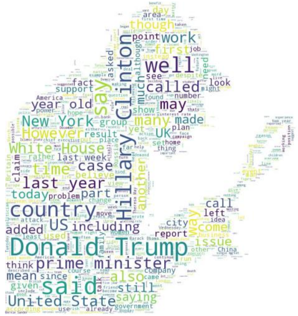
▲Diagram showing our implementation of the Bi-directional LSTM Model for classification of fake news

After the CNN and RNN model, we concatenate them into fully connected layer. Use both of information to predict how fake the news is and what kind of category the news might be.

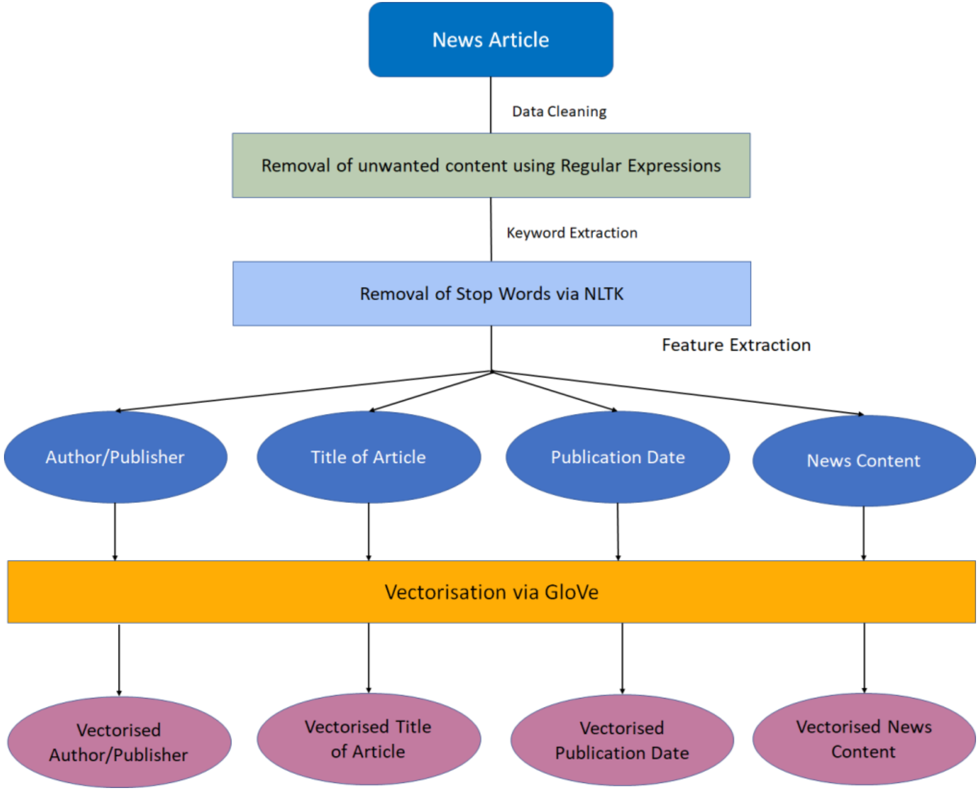
Overview of Implementation



Our Implementation consists of a 3-stage process as shown in the diagram on the left.

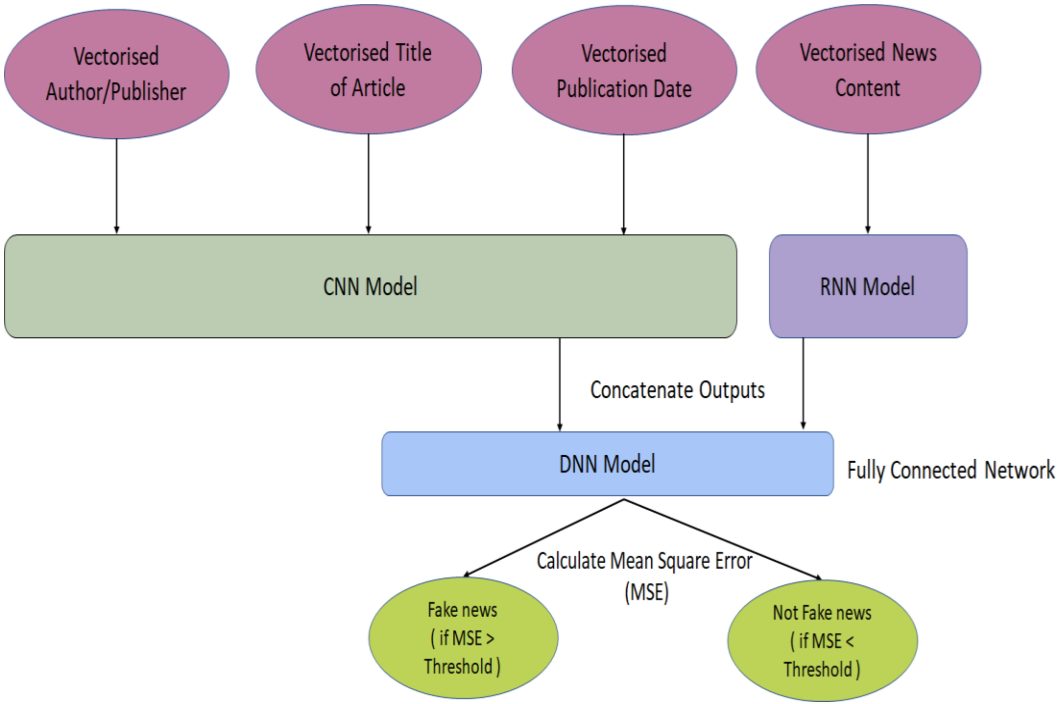


Stage 1 -Data pre-processing



1. We clean up the data provided by the dataset by first removing unwanted content such as “ ? “, “@name” etc.
2. We then remove the stop words in the News Article to identify key points made within the article.
3. Based on the common features amongst our combined datasets, we extract the 4 features as shown in the diagram and vectorise them for further processing via Global Vectors for word representation (GloVe)

Stage 2 - Identification of Fake News

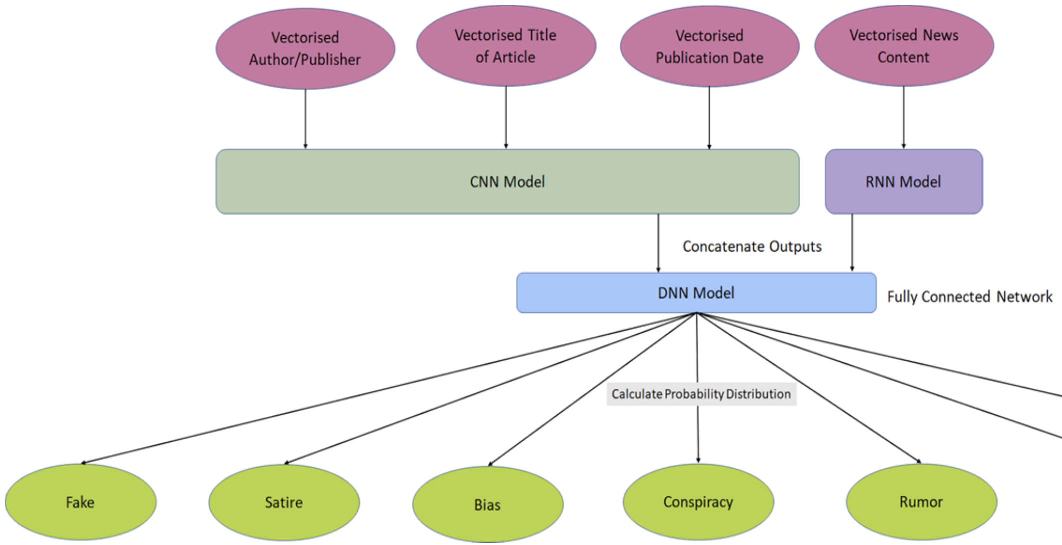


The vectorised features are then passed to either a CNN or RNN model as shown in the diagram.

News Content possesses time-series information, hence it is passed to a RNN model which preserves such information, while Author, Title and Publication Date are passed to a CNN Model.

The results of the CNN and RNN Model will then be concatenated to form a Fully Connected Network, which will output a Mean Square Error Value that will be used to determine the news classification.

Stage 3 - Categorization of Fake News



If the news article is deemed to be fake in stage 2, we then go on to identify the category of fake news that the news article belongs to.
The process is similar to Stage 2, except for the last layer in the DNN Model, where it will output a probability distribution that will be used to identify the category of fake news.

Results

We faced difficulties in testing our model's performance on more datasets as we were unable to collect more testing as most fake news during the recent U.S. Presidential Election has been removed. For the purpose of our testing, we manually selected 23 news articles from BBC (http://www.bbc.com/news/election/us2016) and Twitter (https://twitter.com/president). After which, we categorized the news article manually and used these articles to test the accuracy of our model. We tested our model as seen in the demo video in the next slide.

In stage 2 of our model, we managed to get an accuracy of 47.83% with 11 news articles classified properly, and an accuracy of 78.26% at stage 3 of our model, with 18 news articles categorized properly.

```
Enter the date of the news: 2016-10-27
Enter the author of the news or leave it if unknown: Andrea Lawson Gray
Enter th title of the suspicious news: Comment on HALLOWEEN IN THE CASTRO RETURNS IN 2014!
Enter the article text: It will be recalled that the Halloween bash attracted as many as 500,000 each year and has become a maj
or tourist attraction, second and third only to the Pride parade and Folsom Street Fair. Stabbing and shooting incidents, belie
ved to be perpetrated by straight revelers, prompted city officials to permanently ban the event beginning in 2010.

Well, it's a half half :O However, it's trying to tell the true~
```

DEMO VIDEO: https://drive.google.com/open?id=1ZI4jts3Osnl9R_i-ukDRX5F6D5AWf5qb

Conclusion

With the news we could access, and using the common features amongst the different datasets. We are able to utilise CNN and RNN to make a judgement of the authenticity of a news article, but also the category of fake news a news article might belong to, if it is deemed as fake.

Sometimes when we are looking a news, our thoughts may be led by the author and neglect some parts of news may not be objective and fair. With our model, we hope users can notice some deviations of the news and rethink the view of the news is truly justice.

References

[1] Goldman, C. K. (2016, December 05). In Washington Pizzeria Attack, Fake News Brought Real Guns. Retrieved January 03, 2018, from https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html

[2] Bidirectional recurrent neural networks. (2017, December 17). Retrieved January 03, 2018, from https://en.wikipedia.org/wiki/Bidirectional_recurrent_neural_networks

[3] 22, 2. I. (n.d.). Hashrocket. Retrieved January 03, 2018, from https://hashrocket.com/blog/posts/a-friendly-introduction-to-convolutional-neural-networks

[4] Wang, W. Y. (2017, May 01). "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. Retrieved January 03, 2018, from https://arxiv.org/abs/1705.00648