Fake News Detection

Zaure Zhussipbekova, Shakshi Sharma and **William Mukose**
University of Tartu, Estonia

Abstract

Fake News are widely spread across the Internet especially after the 2016 U.S Presidential Elections. Therefore, in this paper, we study the detection of Fake News through news articles. We utilize two types of analysis while detecting Fake news: first, descriptive analysis and second, predictive analysis. This way, we are able to differentiate the structure between fake and true news in a better way than just using classification models. We tried both traditional as well as neural network approaches. The best performing model is, neural network in our case in terms of accuracy, precision, recall, and F1 Score on both test and validation dataset. We have provided the code repository also in order to replicate our results.

1 Introduction

We work on a Kaggle project about fake news detection. We have a labelled dataset with several articles that provide examples of both fake and real news. We build several models to try to learn features of fake vs real news, then apply them to the test dataset, also on Kaggle. Initially, we take the baseline as a logistic regression / Naive bag-of-words model, and then we apply neural network. In addition, we also try some descriptive analysis. Our code is available at GitHub[1].

Previous work deals with the fake news detection problems as classification tasks using traditional as well as neural network techniques. However, the task of looking at the structure of fake and real news is still an unexplored area. Thus, apart from performing classification tasks, we would like to explore descriptive analysis such as Topic Modelling, Sentiment analysis, etc.

Our results shows that using both descriptive as well as predictive analysis helps us in identifying the fake news in a better way than only using classification models. This way, we are able to differentiate the structure between fake and true news. We tried both traditional as well as neural network approaches. The best performing model is, neural network in our case in terms of accuracy, precision, recall, and F1 Score on both test and validation dataset.

2 Related Work

Previous work deals with non-neural network approaches. In paper Granik and Mesyura (2017), deal with a Naive Bayes approach to fake news detection. This is in line with our original approach to this task, which was to start with simple methods and go on to more complex ones. This paper simply treated each word as independent for the Naive Bayes method and achieved an accuracy of 74%. The paper also suggested that improvements can be made by having a larger dataset (which we have) and by removing stop words, stemming and using n-grams instead of simple words (which we intend to do). Paper Ahmed et al. (2017) compares linear regression, linear support vector machine, SVM with non-linear kernel, stochastic gradient descent and a decision tree. This paper is also useful for our approach as this introduces machine learning methods more complex than Naive Bayes. The authors found that the highest accuracy was achieved with linear separators. We shall see how these results compare with a neural network approach, described in our other papers. Neural network models are also widely used in fake news detection. In Shu et al. (2019), authors have detected fake news using their proposed neural network framework “dEFEND” on the FakeNewsNet dataset containing the news articles as well as the social context information. Their main contribution is to detect why a particular piece of information is detected as fake/true by the model. In this regard, they utilized the “attention” mechanism to explain the comments captured by their framework. In Wang et al. (2018), authors

[1] https://github.com/92dollarbill/NLP_2021

have focused on detecting newly arrived fake news using multi-modal features, that is, images and text. Their dataset is from Twitter and Weibo social networking sites. They detected fake news by proposing the Event Adversarial Neural Network (EANN) framework which extracts the event-invariant features thus, helps in detecting newly arrived fake news. In the paper Oshikawa et al. (2018), the authors compare empirical results of three datasets: LIAR, FEVER and FAKENEWSNET, using various machine learning techniques. Studies on the two first datasets: LIAR and FEVER, showed that LSTM based models achieve the highest accuracy than other methods. Whereas on the FAKENEWSNET dataset, models HC-CB-3 and GCN achieved very high accuracy by the successful utilization of extra data. HC-CB-3 combines content-based and social engagement-based methods. GCN uses graph-like data on the news and publishers' relationships. As a conclusion, authors recommend investigating whether the hand-crafted features can be combined with neural network models. Paper Girgis et al. (2018) covers the implementation of Recurrent Neural Networks models (Vanilla, Gated Recurrent Unit) and LSTM that have been proposed for the detection of online fake news on the prepared LAIR dataset. LSTM showed inefficiency in comparison to GRU and Convolutional Neural Networks, due to expensive calculation and application of backpropagation. Authors found that CNN is the best of other models due to its speed and its best performance.

In the previous literature, fake news detection problems have been widely used as classification tasks using traditional as well as neural network techniques. However, the task of looking at the structure of fake and real news is still an unexplored area. Thus, apart from performing classification tasks, we would like to explore descriptive analysis such as Topic Modelling, Sentiment analysis, etc.

3 Method

The source of the data: The dataset is obtained from the kaggle competition. There are two csv files: test and train. A full training train.csv dataset has the following attributes: id: unique id for a news article title: the title of a news article author: author of the news article text: the text of the article; could be incomplete label: a label that marks the article as potentially unreliable A testing training dataset test.csv has all the same attributes at train.csv without the label. General statistics about the data: Train dataset consists of 20800 rows, the test dataset - 5200, so that the ratio of splitting is 801: unreliable - it is fake news 0: reliable - it's not fake news

Labels are distributed almost evenly: the number of 0 labels is 10387, the number of 1 labels - 10413 In the train set there are 20800 rows and 5 columns In the testing set- 5200 rows and 4 columns

Non neural network methods

Data was uploaded and preprocessed, with all sentences tokenized into words, and converted to lower case. The data passed to the machine learning algorithm was a combination of the news item title and author. As a simplest first step, each sentence was encoded as a 1-hot vector of its words, taking the 10000 most commonly used words in the training data set as a vocabulary size. The data was fitted to several non-neural machine learning models: Logistic Regression using L1 regularization, Logistic Regression using L2 regularization, MultiNomial Naive Bayes, Random Forest and a Gradient Boosted forest model. Hyperparameters were tuned using GridSearch and cross-validation. All these models were implemented through the sklearn library in python.

Neural networks

In addition to the machine learning methods mentioned above, two neural networks were used to predict whether a news item belongs to True or Fake news; a Long-Short Term memory recurring neural network (LSTM), and a Convolutional Neural Network (CNN).These were implemented using tensorflow in Python. For both neural networks, initially, data was preprocessed using the Natural Language Toolkit (nltk) library in Python: sentences were tokenized into words, words were then converted into lower case and converted to their stems: e.g. the words running and ran would be converted into "run". Further, each sentence was converted into an embedding using the tensorflow "one_hot" function with a vocabulary size of 5000, and padded with zeros to a maximum sentence length of 20 characters.

The LSTM consisted of 5 layers in the following order: Embedding, Dropout(0.3), LSTM(100), Dropout (0.3) and Dense(sigmoid activation), with a total of 256,000 trainable parameters. This network was trained over 10 epochs.

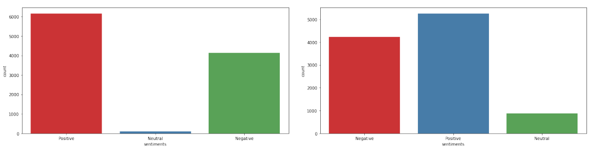
The CNN consisted of 4 layers in the following order: Embedding, Conv1D (activation = 'relu'), GlobalMaxPooling1D, and Dense(sigmoid activation), with a total of 226,000 trainable parameters. This network was also trained over 10 epochs.

4 Results and Discussions

4.1 Descriptive Analysis

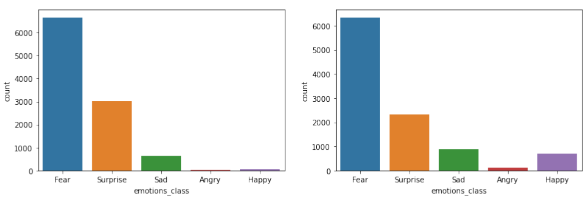
We have analyzed the ‘text’ column here. In particular, we have used the ‘text’ column of the train set in order to perform descriptive analysis only. We have performed pre-processing such as lowercase of the words, tokenization, remove stopwords, stemming, etc.

1. **Sentiments and Emotions Detection:** As shown in Figure 1, 2, we compute the sentiments to check if the sentiments and emotions differ in the Fake News and the True News. We used Vader Lexicon API for sentiments and text2emotion for emotion. It has been observed that True News contains more “Positive” sentiments than Fake News. However, “Negative” sentiments are equal in both types of news. “Neutral” sentiments are less in True News as compared to Fake News. In emotion analysis, Fear and Surprise emotions are equal in both. But Sad, Angry and Happy emotions are more in case of Fake News. Overall, other 3 emotions are more in Fake news.



(a) True News (b) Fake News

Figure 1: Sentiment Analysis



(a) True News (b) Fake News

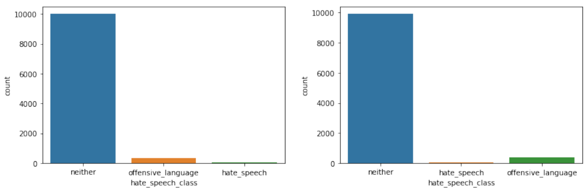
Figure 2: Emotion Analysis

2. **Frequent Words:** We used WordClouds to see if there is any difference between the most frequent words in Fake and True News. We can see that from Figure 3, there are few words that are different in both such as “time”, “call” used more in True News than Fake News.

(a) True News (b) Fake News

Figure 3: Word Clouds

3. **HateSpeech Detection:** We detect if the article is related to hate speech or not using the ‘SONAR’ API. It can be seen in Figure 4 that the “hate speech” and “offensive language” is same in both types of news. So, there is no distinction in terms of Hate Speech.



(a) True News (b) Fake News

Figure 4: Hate Speech

4. **Topic Modeling:** We see which topics are the focus of Fake and True News. Are there any different topics in Fake and True News? We use the LDA approach to find out the first top ten topics for both True and Fake News. It is clear from Figure 5, 6, the differentiating factor is that the True News contains all the topics related to Donald Trump only whereas Fake News topics are a combination of Donald Trump and Hillary Clinton.

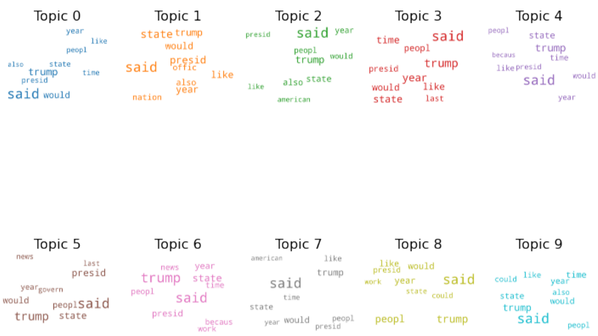


Figure 5: True News LDA

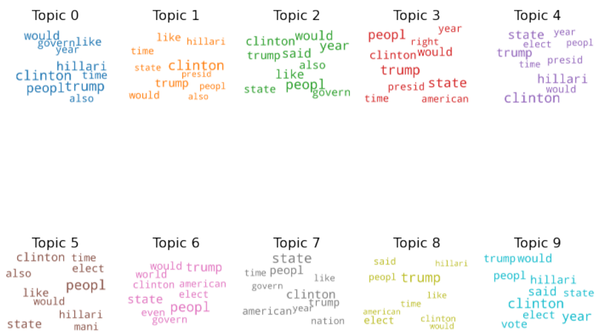
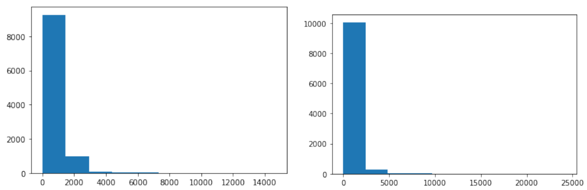


Figure 6: Fake News LDA

5. **Length of the 'text' column:** We also calculated the length of the 'text' column and plot the histogram as displayed in Figure 7. The length of both News differ. Specifically, Fake News contains less length as compared to True News. In addition, we also performed "median" of the length. True News contains 773 median length while Fake News has 382. This confirms that Fake News content has low length compared to True News.



(a) True News (b) Fake News

Figure 7: Length of the news articles

4.2 Predictive Analysis

Predictions were submitted to a Kaggle competition and graded, and the test accuracy displayed is the Kaggle score that this team received. Test data was not used in developing the original one-hot matrix of most common words to avoid data leakege, and this explains why validation scores (which were calculated on a part of the train set) are consistently higher than test scores, which were calculated on a test set which did not take part in any stage of the training process.

Non-neural network machine learning accuracies achieved were in the 88-92% range. The results are presented in more detail in table 1 below.

Neural networks: were significantly more accurate, with both achieving accuracies above 99% on the test set, and the Convolutional Neural Network performed slightly better. Their results are in table 2 below:

Table 1: Non Neural Network Prediction Results

Models	Train acc	Val acc	F1 score	Test acc
Logistic Regression, L1	0.99	0.99	0.99	0.93
Logistic Regression, L2	0.99	0.97	0.98	0.92
Naive Bayes	0.98	0.96	0.96	0.89
Random Forest	1.0	0.98	0.99	0.92
Gradient Booster	0.99	0.99	0.99	0.92

Table 2: Neural Network Prediction Results

Models	Train acc	Val acc	F1 score	Test acc
LSTM	1.0	0.99	0.99	0.99
CNN	0.99	0.99	0.99	0.99

5 Conclusion

We analyze the Fake news using descriptive and predictive analysis. We found that we were able to differentiate between Fake and True news using both analyses. Furthermore, we achieve good predictive results. Thus, we can say that we are able to detect Fake news quite well.

Our neural network results are very good, possibly even too good. We suspect that the reason may be that the news source is often within the text, and certain news sources are much more trustworthy than others. While we were unable to extract the news source from this data (it is part of the text itself, not in a separate column), an interesting line of investigation would be doing this classification task on data where news sources are unknown.

References

Hadeer Ahmed, Issa Traore, and Sherif Saad. 2017. Detection of online fake news using n-gram analysis and machine learning techniques. In *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*, pages 127–138. Springer.

Sherry Girgis, Eslam Amer, and Mahmoud Gadallah. 2018. Deep learning algorithms for detecting fake news in online text. In *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, pages 93–97. IEEE.

Mykhailo Granik and Volodymyr Mesyura. 2017. Fake news detection using naive bayes classifier. In *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pages 900–903. IEEE.

Ray Oshikawa, Jing Qian, and William Yang Wang. 2018. A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.

Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019. defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 395–405.

Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pages 849–857.