**Analysis of classifiers for fake news detection**

**Vasu Agarwal**

*School of Computing Science and Engineering*, *Vellore Institute of Technology, Vellore, Tamil Nadu*

*vasu.agarwal2014@vit.ac.in*

**Prof. Parveen Sultana H**

*School of Computing Science and Engineering*, *Vellore Institute of Technology, Vellore, Tamil Nadu*

*hparveensultana@vit.ac.in*

**Abstract –** Fake news detection is a difficult task. This paper discusses the approach of NLP and machine learning to try and solve this problem. We have used bag-of-words, n-grams, and have trained the data on 4 classifiers to see which of them works well for this specific dataset of labelled news statements. The f1 scores show that logistic regression and SVM work well for the selected dataset.

# Introduction

The term “fake news” was almost non-existent in the general context and media providers but times have changed. Fake news is a term that has been used to describe very different issues, from satirical articles to completely fabricated news and plain government propaganda in some outlets. Fake news, information bubbles, news manipulation and the lack of trust in the media are growing problems with huge ramifications in our society. However, in order to start addressing this problem, we need to have an understanding on what Fake News is. Only then can we look into the different techniques and fields of machine learning (ML), natural language processing (NLP) and artificial intelligence (AI) that could help us fight this situation.

## “Fake news” has been used in a multitude of ways in the last half a year and multiple definitions have been given. For instance, the New York times defines it as “a made-up story with an intention to deceive”. Measuring fake news or even defining it properly could very quickly become a subjective matter, rather than an objective metric. Despite all these drawbacks, several people and organizations have tried to categorize fake news in different ways.

## Motivation

The widespread problem of fake news is very difficult to tackle in today’s digital world where there are thousands of information sharing platforms through which fake news or misinformation may propagate. It has become a greater issue because of the boom in AI which bring along artificial bots that may be used to create and spread fake news. The situation is dire because many people believe anything they read on the internet and the ones who are amateur or are new to the digital technology may be easily fooled. A similar problem is fraud that may happen due to spam or malicious emails and messages. So, it is compelling enough to try and tackle this challenge to stop the rates of crime, political unrest, grief, and many other problems.

## Outline

## Text is one form which is difficult to process simply because of various linguistic features and styles like sarcasm, metaphors, etc. Also, there are so many languages and every language has its own style, script and words. Natural language processing (NLP) is a branch of artificial intelligence and it deals with techniques that can utilize text and make predictions and inferences. Although a lot of research has been done in NLP, there is still a lot of scope for research and application in different fields and scenarios.

The aim of this work is to create a system or model that can use the data of past news reports and predict the chances of a news report being fake or not.

# Related Work

Many researchers have tried solving this problem in a lot of different ways to test which methods works and get desirable results.

A few studies have discussed fake news detection approaches from a data mining perspective, including feature extraction and model construction.

Attempts to use TF-IDF and LSTMs with SVM have also been made in the past but that was not very promising.

After these, people tried using deep learning to solve the problem.

A paper showed a method of ‘capture, ‘score’, ‘integrate’ and creates a model of recurrent neural networks for stance detection of fake news.

Another study showed the use of LSTM-based Recurrent Neural Network with stacked ReLu layers and creates a state of the art feed forward network.

Even the use of convolutional neural networks has been done while some others tried linguistically infused neural networks.

So, we can see that many attempts have been made but it is safe to assume that a state of the art approach does not exist as of now. There is a lot of room for development and research in this area especially because news statements have so many variables attached to them: sarcasm, abbreviations, metaphors, etc.

However, efforts have been made to arrange reliable and vast data into a quality dataset. One such benchmark dataset has been used in this project.

Fake news problem is growing at an alarming rate and it needs to be addressed more seriously.

# Data

Categorizing a news statement as “fake news” could be a very challenging task. For this reason, we have used an existing [dataset](https://www.kaggle.com/mrisdal/fake-news) that has already collected and classified fake news. The data source used for this project is LIAR dataset which contains 3 files with .tsv format for test, train and validation. Below is some description about the data files used for this project.

The dataset has been cited in the paper by William Yang Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, to appear in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), short paper, Vancouver, BC, Canada, July 30-August 4, ACL.

The original dataset contained 13 variables/columns for train, test and validation sets. For the sake of simplicity, we have chosen only 2 variables from this original dataset for this classification. The other variables can be used later to achieve more detailed analysis.

The two columns that have been used are: ‘Statement’, which is the actual news statement itself, and ‘Label’, which refers to the statement being true/false.

The procedure used for reducing number of classes in the dataset:

* True, mostly-true, half-true become ‘True’.
* Barely-true, false, pants-fire become ‘False’.

# Model

The performance of a classifier working on text depends on the size and quality of the text data (or corpus) and also the features of the text. Common noisy words called ‘stopwords’ are useless words when it comes to text feature extraction and they only contribute towards feature dimensionality and can be discarded for better performance.

This helps in reducing the size of the text corpus and add text context for feature extraction. Also, lemmatization is used to convert words to their core meaning and this results in multiple word conversion into a single discrete representation.

In this step, first the test, train and validation data has been read and then tokenization and stemming has been done. After this, some exploratory data analysis has been performed like response variable distribution and data quality checks like null or missing values.

We have performed feature extraction and selection methods from sci-kit learn python libraries. For feature selection, I have used methods like simple bag-of-words and n-grams and then term frequency like tf-tdf weighting.

N-grams are essentially permutations of word combinations. They help in providing context to the text by combining nearby words and making a single feature out of them.

The extracted features are fed into different classifiers. I have used Naïve Bayes, Logistic Regression, Linear SVM and Random forest classifiers from sklearn. Each of the extracted features were used in all of the classifiers. Once fitting the model, I compared the ‘f1’ score and checked the confusion matrix. After fitting all the classifiers, 2 best performing models were selected as candidate models for fake news classification. I have performed parameter tuning by implementing GridSearchCV methods on these candidate models and chosen best performing parameters for these classifiers. Finally, selected model was used for fake news detection with the probability of truth. In Addition to this, I have also used Precision-Recall and learning curves to see how training and test set performs when we increase the amount of data in our classifiers.

The data is rarely evenly distributed. So, in such cases we use confusion matrix to measure the performance of a classifier. True positives are the correct predictions of the classifier and false positives are the incorrect predictions. Using these numbers, we can find the precision of a classifier.



The finally selected, best performing model is ‘Logistic Regression’, which was saved using the ‘pickle’ library function.

Using this model, predictions can be made on a new news statement.

# Results

# Conclusion, Limitations and Scope for future work

The fake news problem is at large and is spreading rapidly as information becomes easier to access with advancement in technology. It is clearly visible from the recent US presidential elections that fake news can have a huge impact in politics and more so on the lives of masses. With the help of artificial intelligence, we can control and limit the spread of such misinformation more quickly and efficiently than manually. The work in this project proposes a stacked model which fine tunes the informational insight gained from the data at each step and then tries to make a prediction.

Although many attempts have been made to solve the problem of fake news, any significant success has yet to be achieved. With huge amounts of data collected from websites like facebook, twitter, etc, the best models improve every day. With the use of deep neural networks, the future work in this field seems promising.

The use of NLP or deep learning is debatable since both approaches have shown varied results on different datasets. It cannot be firmly decided that which approach is better but maybe a hybrid approach can be experimented.

The limitations that come packaged with this problem is that, the data is erratic and this means that any type of prediction model can have anomalies and can make mistakes. This cannot be afforded at any costs since it can lead to severe conditions.

For future improvements, concepts like POS tagging, word2vec and topic modelling can be done. These will give the model a lot more depth in terms of feature extraction and fine-tuned classification.

***Word2Vec*:** The Word2Vec technique converts text to features while maintaining the original relationships between words in a corpus. It is a combination of techniques and is one of the best feature extraction techniques in NLP. It converts features into vector form.

***Topic Modelling:*** News can contain a vast range of topics. Just the classification based on labels is not enough of we want to achieve realistic results. For this reason, we can use an advanced technique called topic modelling.

Topic modelling categories each piece of text into topics and using this we can make more accurate predictions. The most popular topic modelling technique used in NLP is ‘Latent Dirichlet Allocation’, also known as LDA.

Use of LDA can add another layer of depth to the fake news classification task.

# References

Weblinks:

1. https://miguelmalvarez.com/2017/03/23/how-can-machine-learning-and-ai-help-solving-the-fake-news-problem/
2. https://opendatascience.com/blog/how-to-build-a-fake-news-classification-model/
3. https://nycdatascience.com/blog/student-works/identifying-fake-news-nlp/
4. https://github.com/sumeetkr/AwesomeFakeNews
5. https://towardsdatascience.com/i-trained-fake-news-detection-ai-with-95-accuracy-and-almost-went-crazy-d10589aa57c
6. https://www.fakenewschallenge.com/

Journals:

1. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Haun Liu, “Fake News Detection on Social Media: A Data Mining Perspective”, Computer Science & Engineering, Arizona State University, Tempe, AZ, USA, Charles River Analytics, Cambridge, MA, USA
2. Natali Ruchansky, Sungyong Seo, Yan Liu, “CSI: A Hybrid Deep Model for Fake News Detection”, University of Southern California, Los Angeles, California
3. Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer, “The spread of misinformation by social bots”, Indiana University, Bloomington
4. Sahil Chopra, Saachi Jain, John Merriman Sholar, “Towards Automatic Identification of Fake News: Headline-Article Stance Detection with LSTM Attention Models”, Department of Computer Science Stanford University
5. Stephen Pfohl, Oskar Triebe, Ferdinand Legros, “Stance Detection for the Fake News Challenge with Attention and Conditional Encoding”, Stanford University
6. Neel Rakholia, Shruti Bhargava, “Is it true?” – Deep Learning for Stance Detection in News, Stanford University
7. William Yang Wang, “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection, Department of Computer Science University of California, Santa Barbara, Santa Barbara, CA 93106 USA
8. Samir Bajaj, “The Pope Has a New Baby!” Fake News Detection Using Deep Learning, Stanford University
9. Svitlana Volkova, Kyle Shaffer, Jin Yea Jang and Nathan Hodas, “Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter”, Data Sciences and Analytics Group, National Security Directorate, Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA 99354