

# Real and Fake News Classifier

Shobhit Srivastava  
Computer Science and

Engineering  
AITS Technologies

shobhitsrivastava18th@gmail.com  
m

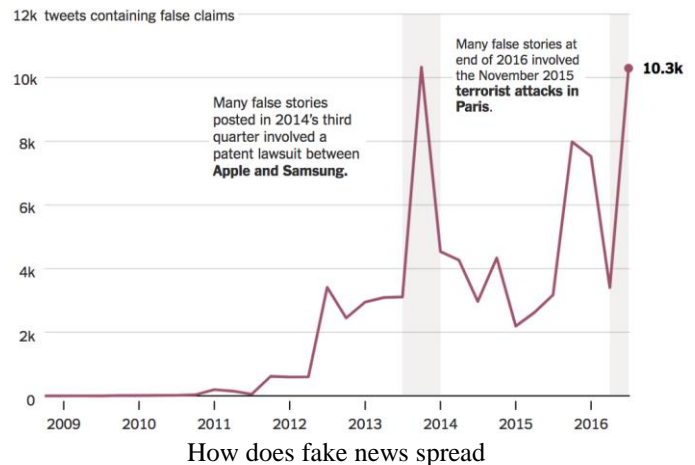
**Abstract -[1]**Recent progress in natural language processing has raised dual-use concerns. While applications like summarization and translation are positive, the underlying technology also might enable adversaries to generate natural fake news: targeted propaganda that closely mimics the style of real news. Modern computer security relied on careful threat modelling. In today's scenario we see a lot of peoples or can say major chunk of population are misguided and mislead by wrong news and outlines. Which in some way affect their decision making and thus taking them out of their consciousness. Also a lot of time we get some viral news and with their alluring content we are misleads and get onto believe about it being reality. So, the main main of this project is to curve the spread of wrong and fake news. Living in the world where everything is related to truth and once again we would be able to revive about our belief.

**Keywords**—Natural Language Processing.

## I. INTRODUCTION

Popularized as a concept in the United States during the 2016 presidential election, fake news is a form of propaganda created to mislead readers, in order to generate views on websites or steer public opinion. Almost as quickly as the issue became mainstream, researchers began developing automated fake news detectors — so-called neural networks that “learn” from scores of data to recognize linguistic cues indicative of false articles. Given new articles to assess, these networks can, with fairly high accuracy, separate fact from fiction, in controlled settings.

One issue, however, is the “black box” problem — meaning there’s no telling what linguistic patterns the networks analyze during training. They’re also trained and tested on the same topics, which may limit their potential to generalize to new topics, a necessity for analyzing news across the internet.



## II. Fake news detecton : An overview

The model is tested on certain news article texts in English language. This approach classifies individual articles based solely on language patterns, which more closely represents a real-world application for news readers. Traditional fake news detectors classify articles based on text combined with source information, such as a Wikipedia page or website.

“In this case, we wanted to understand what was the decision-process of the classifier based only on language, as this can provide insights on what is the language of fake news,” says co-author Xavier Boix, a postdoc in the lab of Tomaso Poggio, the Eugene McDermott Professor in the Department of Brain and Cognitive Sciences (BCS) and director of the Center for Brains, Minds, and Machines (CBMM), a National Science Foundation-funded center housed within the McGovern Institute of Brain Research.

“A key issue with machine learning and artificial intelligence is that you get an answer and don’t know why you got that answer,” says graduate student and first author Nicole O’Brien. “Showing these inner workings takes a first step toward understanding the reliability of deep-learning fake-news detectors.”

The model identifies sets of words that tend to appear more frequently in either real or fake news — some perhaps obvious, others much less so. The findings, the researchers say, points to subtle yet consistent differences in fake news — which favors

exaggerations and superlatives — and real news, which leans more toward conservative word choices.

“Fake news is a threat for democracy,” Boix says. “In our lab, our objective isn’t just to push science forward, but also to use technologies to help society.

... It would be powerful to have tools for users or

companies that could provide an assessment of whether news is fake or not.”



Here's why we have fake news yield curve inversion

### III. Humans are Easily Fooled by Grover-written Propaganda :

We evaluate the quality of disinformation generated by our largest model. We consider four classes of articles: human-written articles from reputable news websites (Human News), Grover-written articles conditioned on the same metadata (Machine News), human-written articles from known propaganda websites (Human Propaganda), and Grover-written articles conditioned on the propaganda metadata. We asked a pool of qualified workers on Amazon Mechanical Turk to rate each article on three dimensions: stylistic consistency, content sensibility, and overall trustworthiness. Results show a striking trend: though the quality of Grover-written news is not as high as human-written news, it is adept at rewriting propaganda.

#### IV. Feature engineering used :

There is a lot of important feature engineering techniques employed in this paper which are normally used in any NLP project at the basic level. I will be making all the points clear.

- All the stopwords are removed from the columns. **Stopwords** contains words like this, that, is, of, off etc.
- All the punctuations are removed from the columns and words are further converted into the lower letter form.
- All maximum and minimum occurred words are removed from the data such as to avoid any noise and those are the words which finds least occurrence in any such news articles and texts.
- Alpha words are filtered out and further they are lemmatized so as to provide each words with base such as to avoid same word occurring many times and increasing redundant.

For crunching the words in number forms we have used tf-idf method which work as assigning some sort of number to the most occurring and common words.

### V. Use of

#### Unsupervised method:

As the labels provided in the dataset were quite imbalance and were not in identified state to describe them.

So, I decided to adopt the method of unsupervised learning. Unsupervised learning is a machine learning paradigm we go on to train the machine learning algorithm without any label data.

For the algorithm, I have used **K- Means Clustering** method for that.

### VI. K Means

#### Clustering :

For finding the cluster, we are using the k means clustering method, which samples the data with small distance of separability in one cluster and the another one in the other cluster. The **centroid** of the cluster changes after each point added into the cluster and thus through this way model learning. While testing with test, the test data point is compared with all the available cluster centroid and then the smaller distance one is assigned the specific group.

### VII. Dealing with Unknown

#### Biases in the Data :

Anyone who's worked on even a slightly complicated **NLP** project knows there's no smooth sailing to building a model. There will be obstacles along the way. You might miss out on a certain point, or an unknown bias might creep in which no one would have thought of in a million years. Also while sorting out or removing the unimportant words from the data may lead to loss of some of the most important words which may cost us later later. Now coming over the bias, by sorting out the most occurring or the least occurring word from the column we can deal with that.

### VIII. How does a model distinguish between human and machine text? :

The generated model is tested with many set of hypothesis, so as to test with thich set of words it is working well. The words which made the model work efficiently as selected through a lot of trial and run. The most important words are kept in the data while rest and rarely occurring words are removed from it. The words which are most common in a lot of fake news documented are sorted out through a lot of research from other papers and documented also, and at last all the relevent ideas got unified with the documents and this is how it is able to distinguish between real and fake aritcles.

The models is still in the improvement stage, a lot of new tactice are still required to to make the model learn.

### IX.

#### Authors and Affiliations :

The present paper is written thoroughly by myself. I have tried to write it in the best possible way, by taking the reference from many Youtube videos and also from some other research papers for the guidance and all.

### X. ACKNOWLEDGEMENT

In the acknowledgement, I want to thanks the AITS team to be always there in the skype group and got me all the possible help with the frequent mails. Also, I want to thank my friends out there in my college hostel who have helped me out with any possible requirement.

## **XI. REFERENCES**

- [1] <https://arxiv.org/pdf/1905.12616v1.pdf>
- [2] Noam Shazeer and Mitchell Stern. Adafactor: Adaptive learning rates with sublinear memory cost. In International Conference on Machine Learning, pages 4603–4611, 2018.