A Project Final Report

on

Fake News Detection

Submitted in partial fulfilment of the
requirement for the award of degree of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE & ENGINEERING

Under The Supervisionof
Mr. S.P Ramesh
(Assistant Professor)

Submitted By:

Vaibhav Sahu
Enroll no: 19021011402
Adm no: 19SCSE1010213

Aryan Chaudhary
Enroll no: 19021011366
Adm no: 19SCSE1010171

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING GALGOTIAS UNIVERSITY, GREATER NOIDA, INDIA

December, 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"FAKE NEWS DETECTION"** in partial fulfillment of the requirements for the award of the BACHLOR OF COMPUTER SCIENCE AND ENGINEERING submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of Name… Designation, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

VAIBHAV SAHU(19SCSE1010213)

ARYAN CHAUDHARY(19SCSE1010171)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor Name

Designation

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of 19SCSE1010213-VAIBHAV SAHU ~~and~~ 19SCSE1010171-ARYAN CHAUDHARY ~~has~~ has been held on _____ and work is recommended for the award of BACHLOR OF COMPUTER SCIENCE AND ENGINEERING.

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date:

Place:

ABSTRACT

Most of the smart phone users prefer to read the news via social media over internet. The news websites are publishing the news and provide the source of authentication. The question is how to authenticate the news and articles which are circulated among social media like WhatsApp groups, Facebook Pages, Twitter and other micro blogs & social networking sites. It is harmful for the society to believe on the rumors and pretend to be a news. The need of an hour is to stop the rumors especially in the developing countries like India, and focus on the correct, authenticated news articles. This paper demonstrates a model and the methodology for fake news detection. With the help of Machine learning and natural language processing, it is tried to aggregate the news and later determine whether the news is real or fake using Support Vector Machine. The results of the proposed model is compared with existing models. The proposed model is working well and defining the correctness of results upto 93.6% of accuracy.

In our modern era where the internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via click-baits. In this paper, we aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the websitepublishingthenews.

List of Figures

1. **Flow Chart Diagram** **10**

2. **Architecture Diagram** **11**

Contents

Title			Page No.
Abstract			**I**
Contents			**II**
List of Table			**III**
List of Figures			**IV**
Chapter 1	**Introduction**		**1**
	1.1	Introduction	**2**
	1.2	Formulation of Problem	**5**
		1.2.1 Tool and Technology Used	**9**
Chapter 2	**Literature Survey**		**10-12**
	2.1	Existing System	
	2.2	Problem Defintion	
Chapter 3	**Functionality/Proposed Model**		**13-24**
Chapter 4	**RelatedWork**		**25-27**
Chapter 5	**Conclusion and Future Scope**		**28-30**
	5.1	Conclusion	
	5.2	Future Scope	
	Reference		**31-33**

CHAPTER 1

Introduction

Fake news denotes a type of yellow press which intentionally presents misinformation or hoaxes spreading through both traditional print news media and recent online social media. Fake news has been existing for a long time, since the “Great moon hoax” published in 1835. In recent years, due to the booming developments of online social networks, fake news for various commercial and political purposes has been appearing in large numbers and widespread in the online world. With deceptive words, online social network users can get infected by these online fake news easily, which has brought about tremendous effects on the offline society already. During the 2016 US president election, various kinds of fake news about the candidates widely spread in the online social networks, which may have a significant effect on the election results. According to a post-election statistical report, online social networks account for more than 41.8% of the fake news data traffic in the election, which is much greater than the data traffic shares of both traditional TV/radio/print medium and online search engines respectively. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely, which will be the main tasks studied in this paper.

As an increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to hunt out and consume news from social media instead of traditional news organizations. The explanations for this alteration in consumption behaviors are inherent within the nature of those social media platforms: (i) it's often more timely and fewer expensive to consume news on social media compared with traditional journalism , like newspapers or

television; and (ii) it's easier to further share, discuss , and discuss the news with friends or other readers on social media. For instance, 62 percent of U.S. adults get news on social media in 2016, while in 2012; only 49 percent reported seeing news on social media. It had been also found that social media now outperforms television because the major news source. Despite the benefits provided by social media, the standard of stories on social media is less than traditional news organizations. However, because it's inexpensive to supply news online and far faster and easier to propagate through social media, large volumes of faux news, i.e., those news articles with intentionally false information, are produced online for a spread of purposes, like financial and political gain. it had been estimated that over 1 million tweets are associated with fake news "Pizzagate" by the top of the presidential election. Given the prevalence of this new phenomenon, "Fake news" was even named the word of the year by the Macquarie dictionary in 2016. The extensive spread of faux news can have a significant negative impact on individuals and society. First, fake news can shatter the authenticity equilibrium of the news ecosystem for instance; it's evident that the most popular fake news was even more outspread on Facebook than the most accepted genuine mainstream news during the U.S. 2016 presidential election. Second, fake news intentionally persuades consumers to simply accept biased or false beliefs. Fake news is typically manipulated by propagandists to convey political messages or influence for instance, some report shows that Russia has created fake accounts and social bots to spread false stories. Third, fake news changes the way people interpret and answer real news, for instance, some fake news was just created to trigger people's distrust and make them confused; impeding their abilities to differentiate what's true from what's not. To assist mitigate the negative effects caused by fake news (both to profit the general public and therefore the news ecosystem). It's crucial that we build up methods to automatically detect fake news broadcast on social media.

Internet and social media have made the access to the news information much easier and comfortable. Often Internet users can pursue the events of their concern in online form, and increased number of the mobile devices makes this process even easier. But with great possibilities come great challenges. Mass media have an enormous influence on the society, and because it often happens, there's someone who wants to require advantage of this fact. Sometimes to realize some goals mass-media may manipulate the knowledge in several ways. This result in producing of the news articles that isn't completely true or maybe completely false. There even exist many websites that produce the fake news almost exclusively. The rise of fake news during the 2016 U.S. Presidential Election highlighted not only the dangers of the effects of fake news but also the challenges presented when attempting to separate fake news from real news. Fake news may be a relatively new term but it is not necessarily a new phenomenon. Fake news has technically been around at least since the appearance and popularity of one-sided, partisan newspapers in the 19th century. However, advances in technology and the spread of news through different types of media have increased the spread of fake news today. As such, the effects of fake news have increased exponentially in the recent past and something must be done to prevent this from continuing in the future. I have identified the three most prevalent motivations for writing fake news and chosen only one as the target for this project as a means to narrow the search in a meaningful way. The first motivation for writing fake news, which dates back to the 19th century one-sided party newspapers, is to influence public opinion. The second, which requires more recent advances in technology, is the use of fake headlines as clickbait to raise money. The third motivation for writing fake news, which is equally prominent yet arguably less dangerous, is satirical writing. While all three subsets of fake news, namely, clickbait, influential, and satire, share the common thread of being fictitious, their widespread effects are vastly different. As such, this paper will

focus primarily on fake news as defined by politifact.com, “fabricated content that intentionally masquerades as news coverage of actual events.” This definition excludes satire, which is intended to be humorous and not deceptive to readers. Most satirical articles come from sources like “The Onion“, which specifically distinguish themselves as satire. Satire can already be classified, by machine learning techniques according to. Therefore, our goal is to move beyond these achievements and use machine learning to classify, at least as well as humans, more difficult discrepancies between real and fake news.

The dangerous effects of fake news, as previously defined, are made clear by events such as in which a man attacked a pizzeria due to a widespread fake news article. This story along with analysis from provide evidence that humans are not very good at detecting fake news, possibly not better than chance . As such, the question remains whether or not machines can do a better job. There are two methods by which machines could attempt to solve the fake news problem better than humans. The first is that machines are better at detecting and keeping track of statistics than humans, for example it is easier for a machine to detect that the majority of verbs used are “suggests” and “implies” versus, “states” and “proves.” Additionally, machines may be more efficient in surveying a knowledge base to find all relevant articles and answering based on those many different sources. Either of these methods could prove useful in detecting fake news, but we decided to focus on how a machine can solve the fake news problem using supervised learning that extracts features of the language and content only within the source in question, without utilizing any fact checker or knowledge base. For many fake news detection techniques, a “fake” article published by a trustworthy author through a trustworthy source would not be caught. This approach would combat those “false negative” classifications of fake news. In essence, the task would be equivalent to what a human faces when reading a hard copy of a newspaper article,

without internet access or outside knowledge of the subject (versus reading something online where he can simply look up relevant sources). The machine, like the human in the coffee shop, will have only access to the words in the article and must use strategies that do not rely on blacklists of authors and sources. The current project involves utilizing machine learning and natural language processing techniques to create a model that can expose documents that are, with high probability, fake news articles. Many of the current automated approaches to this problem are centered around a "blacklist" of authors and sources that are known producers of fake news. But, what about when the author is unknown or when fake news is published through a generally reliable source? In these cases it is necessary to rely simply on the content of the news article to make a decision on whether or not it is fake. By collecting examples of both real and fake news and training a model, it should be possible to classify fake news articles with a certain degree of accuracy. The goal of this project is to find the effectiveness and limitations of language-based techniques for detection of fake news through the use of machine learning algorithms including but not limited to convolutional neural networks and recurrent neural networks. The outcome of this project should determine how much can be achieved in this task by analyzing patterns contained in the text and blind to outside information about the world.

This type of solution is not intended to be an end-to end solution for fake news classification. Like the "blacklist" approaches mentioned, there are cases in which it fails and some for which it succeeds. Instead of being an end-to-end solution, this project is intended to be one tool that could be used to aid humans who are trying to classify fake news. Alternatively, it could be one tool used in future applications that intelligently combine multiple tools to create an end-to-end solution to automating the process of fake news classification.

CHAPTER 2

Literature Survey

A look at contemporary scholarly work shows that the issue of fake news has been a major concern amongst scholars from various backgrounds. For instance, some authors have observed that fake news is no longer a preserve of the marketing and public relations departments. In the stead, the problem is increasingly being regarded as part of the responsibilities associated with the information technology (IT) department. Traditionally, it was believed that the two departments mentioned above were the ones to deal with any implications arising from the dissemination of misleading news related to an organization. However, current research indicates that fake news is considered to be a threat to information security. The involvement of the IT department, therefore, is premised on the idea that it would help avert the various risks associated with the problem. Similarly, other authors have noted that the participation of IT professionals in resolving matters concerning fake news is paramount considering the demands of the contemporary corporate environment. Rather than as it was the case a few years ago when perpetrators of such gimmicks were motivated by just attracting web traffic, the practice has evolved into a matter that includes the involvement of hackers.

Specifically, some content publishers have resorted to including material that contains malicious code as part of the content provided on their web pages, leading those who visit such sites to click the links and download the malware without their knowledge. Such developments, according to the scholars, have exposed modern companies to further risk of cyber intrusion as the perpetrators of the fake news tend to target employees of certain organizations with the aim of exploiting the latter's curiosity.

Our project is an web application which gives you the guidance of the day to day rountine of fake news, spam message in daily news chanel , Facebook, Twitter, Instagram and other social media.We have shown some data analysis from our dataset which have retrive from many online social media and display the main source till now fake news and true news are engaged.

Our project is tangled with multiple model trained by our own and also some pretrained model extracted from Felipe Adachi. The accuracy of the model is around 95% for all the selfmade model and 97% for this pretrained model. This model can detect all news and message which are related to covid-19, political news, geology ,etc.

2.1 Existing System

We can get online news from different sources like social media websites, search engine, homepage of news agency websites or the factchecking websites. On the Internet, there are a few publicly available datasets for Fake news classification like Buzzfeed News, LIAR, BS Detector etc. These datasets have been widely used in different research papers for determining the veracity of news. In the following sections, I have discussed in brief about the sources of the dataset used in this work.This Existing system can help us to trained our model using machine learningtechnique.

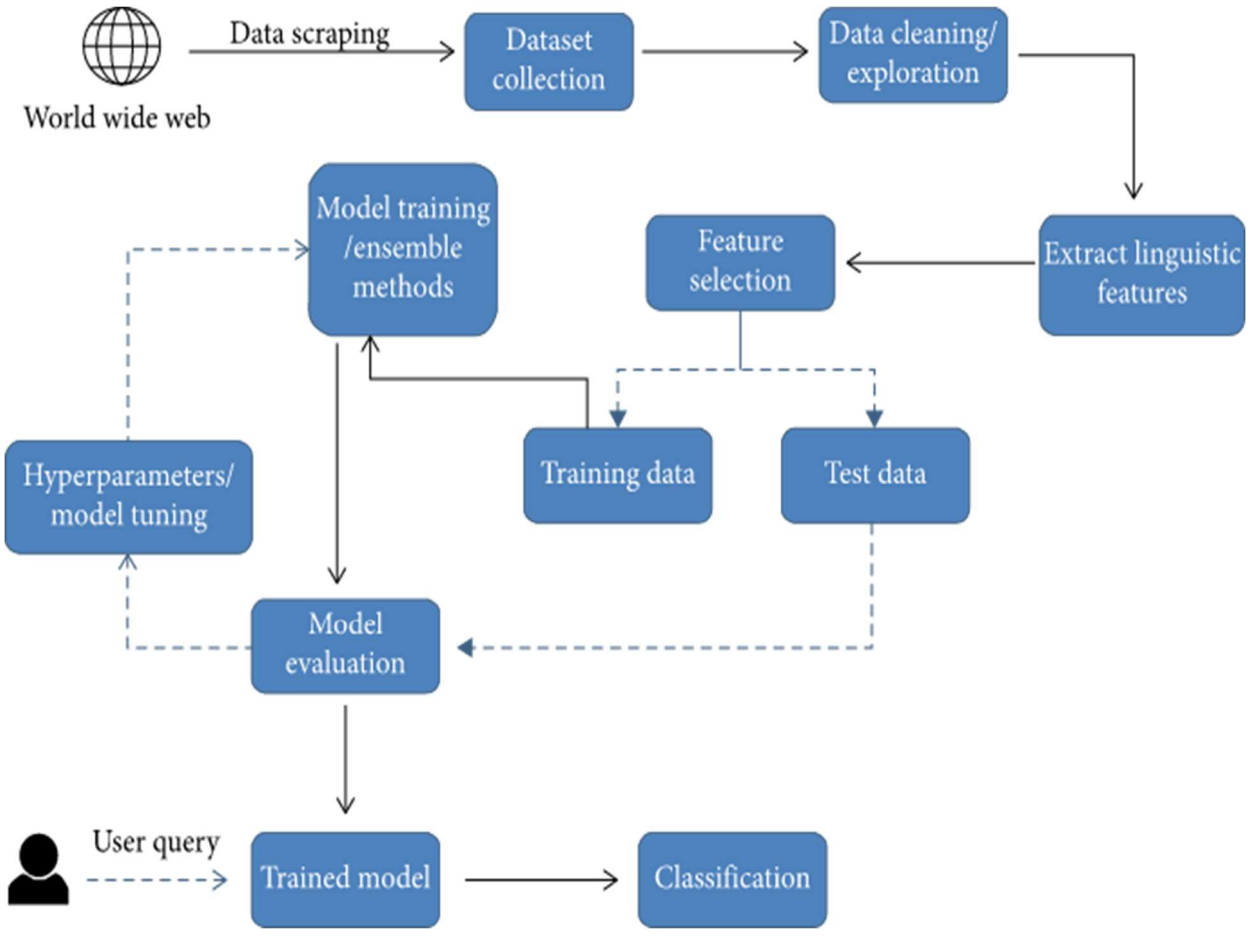
2.2 Problem Definiton

The system is an Web application which help user to detect the fake news. We have given the text box where the user has the option to paste the message or paste the url link of the news and other message link and after that it gives the reality of it. All the user gives data to detector may save for further use in order to update the statue of model, data analysis in future. We also help user by giving some guidance of how to prevent from such false event and how to stop with such event from spreading it.

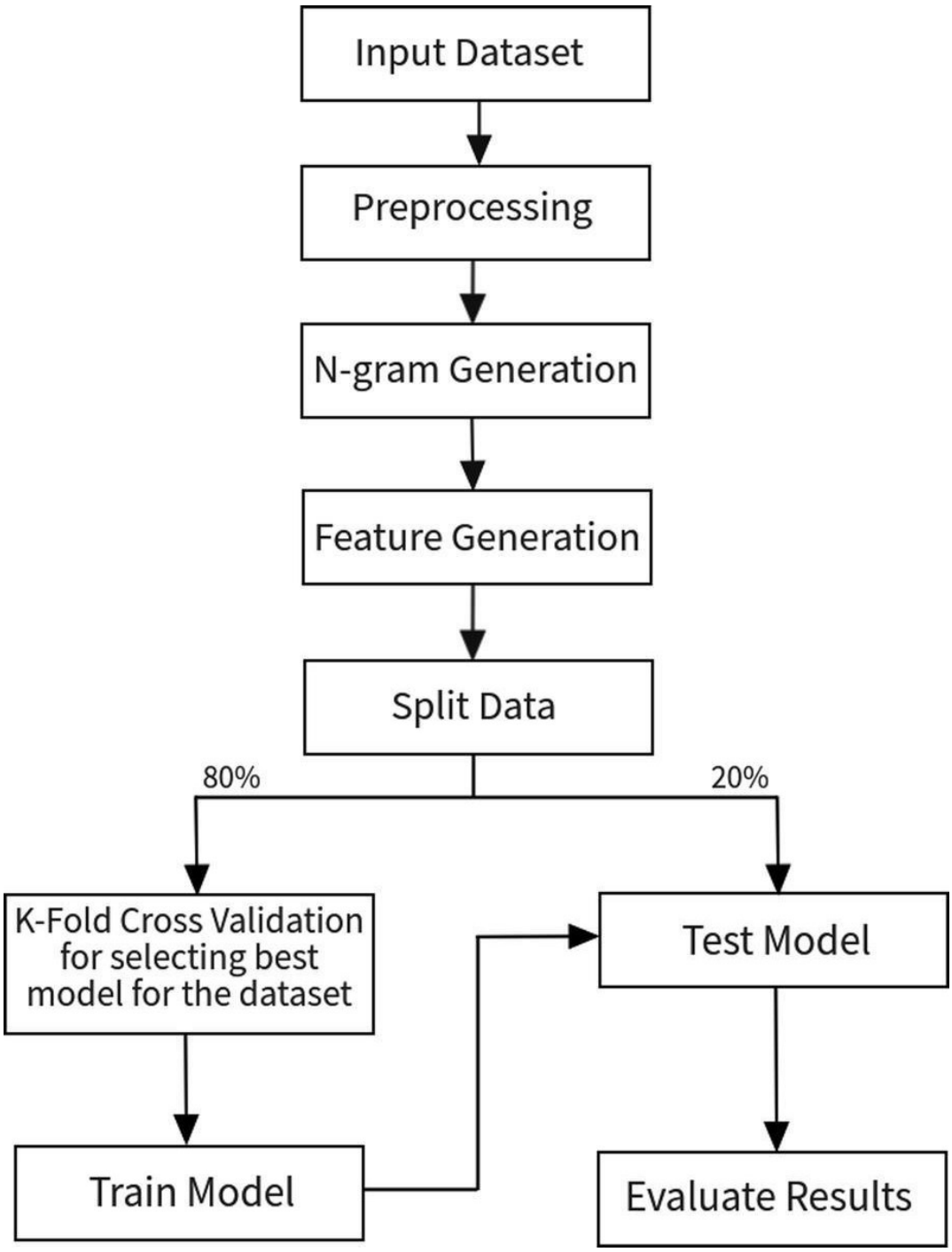
CHAPTER 3

Proposed Model

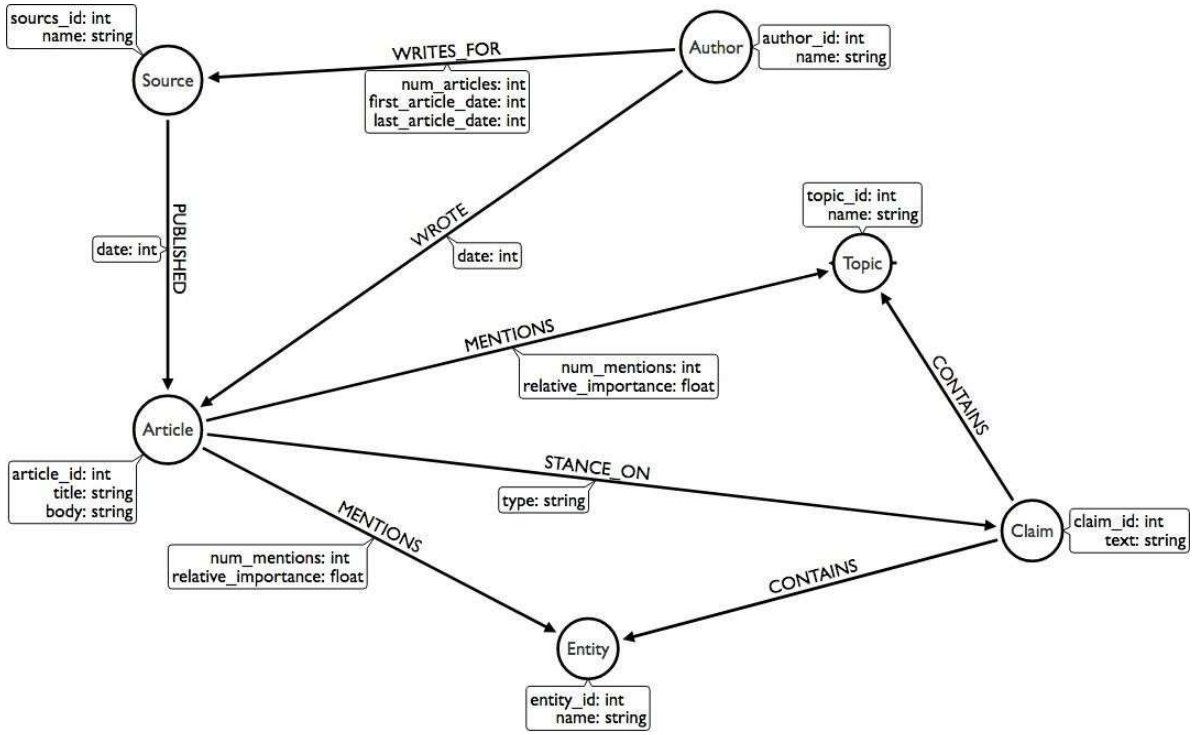
The system is an Web application which help user to detect the fake news. We have given the text box where the user has the option to paste the message or paste the url link of the news and other message link and after that it gives the reality of it. All the user gives data to detector may save for further use in order to update the statue of model, data analysis in future. We also help user by giving some guidance of how to prevent from such false event and how to stop with such event from spreading it.



Flow Chart


Input Dataset
Preprocessing
N-gram Generation
Feature Generation
Split Data
80%
20%
K-Fold Cross Validation for selecting best model for the dataset
Test Model
Train Model
Evaluate Results

Architecture Diagram



3.1 Sentence-Level Baselines

I have run the baselines described in, namely multi-class classification done via logistic regression and support vector machines. The features used were n grams and TF-IDF. N-grams are consecutive groups of words, up to size "n". For example, bi-grams are pairs of words seen next to each other. Features for a sentence or phrase are created from n-grams by having a vector that is the length of the new "vocabulary set," i.e. it has a spot for each unique n-gram that receives a 0 or 1 based on whether or not that n-gram is present in the sentence or phrase in question. TF-IDF stands for term frequency inverse document frequency. It is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. As a feature, TF-IDF can be used for stop-word filtering, i.e. discounting the value of words like "and,", "the", etc. whose counts likely have no effect on the classification of the text. An alternative approach is removing stopwords (as defined in various packages, such as Pythons NLTK).

Table 4.1: Preliminary Baseline Results

Model	Vectorizer	N-gram Range	Penalty, C	Dev Score
Logistic Regression	Bag of Words	1-4	0.01	0.2586
Logistic Regression	TF-IDF	1-4	10	0.2516
SVM w. Linear Kernel	Bag of Words	1	10	0.2508
SVM w. RBF kernel	Bag of Words	1	1000	0.2492

Additionally, we explored some of the characteristic n-grams that may influence Logistic Regression and other classifiers. In calculating the most frequent n-grams for “pants-fire” phrases and those of “true” phrases, we found that the word “wants” more frequently appears in “pants-fire” (i.e. fake news) phrases and the phrase “states” more frequently appears in “true” (i.e. real news) phrases. Intuitively, This makes sense because it is easier to lie about what a politician wants than to lie about what he or she has stated since the former is more difficult to confirm. This observation motivates the experiments in Section 4.2, which aim to find a more full set of similarly intuitive patterns in the body texts of fake news and real news articles.

3.2 Document-Level

Deep neural networks have shown promising results in NLP for other classification tasks such as. CNNs are well suited for picking up multiple patterns, and sentences do not provide enough data for this to be useful. However, a CNN baseline modeled off of the one described for NLP in [15] did not show a large improvement in accuracy on this task using the Liar Dataset. This is due to the lack of context provided in sentences. Not surprisingly, the same CNN performance on the full body text datasets we created was much higher.

3.2.1 Tracking Important Trigrams

The nature of this project was to decide if and how machine learning could be useful in detecting patterns characteristic of real and fake news articles. In accordance with this purpose, we did not attempt to build deeper and better neural nets in order to improve performance, which was already much higher than expected.Instead, we took steps to analyze the most basic neural net. We wanted to learn what patterns it was learning that resulted in such a high accuracy of being able to classify fake and real news. If a human were to take on the task of picking out phrases that indicate fake or real news, they may follow guidelines such as those in. This and similar guidelines often encourage readers to look for evidence supporting claims because fake news claims are often unbacked by evidence. Likewise, these guidelines encourage people to read the full story, looking for details that seem “far-fetched.” Figures 4.1 and 4.2 show examples of the phrases a human might pick up on to decide if an article is fake or real news. We were curious to see if a neural net might pick up on similar patterns.

3.2.2 Topic Dependency

As we suspected from the makeup of the dataset which can be seen from 4.7 which demonstrates a general overview of the makeup of both of the datasets, there is a significant difference in the subjects being written about in fake news and real news, even in the same time range with the same current events going up. More specifically, you can see that the concentration of articles that involve “Hillary”, “Wikileaks”, and “republican” is higher in Fake News than it is in real news. This is not to say that these words did not appear in real news, but they were not some of the “most frequent” words there. Additionally, words like ”football” and “love” 24 appear very frequently in the real news dataset, but these are topics that you can imagine would not be written about, or rarely be written about, in fake news. The “hot topics” of fake news present another issue in this task. We do not want a model that simply chooses a classification based on the probability that a fake or real news article would be written on that topic just like we would never tell a person that every article written about Hillary is fake news or every article written about love is real news. The way we accounted for these differences in the dataset was by separating our training set and tests sets on the presence/absence of certain words. We tried this for a number of topics that were present in both fake news and real news but had different proportions in the two categories. The words we chose were “Trump”, “election”, “war”, and “email.” To create a model that was not biased about the presence of one of these words, we extracted all body texts which did not contain that word. We used this set as the training set. Then, we used the remaining body texts that did contain the target word as the test set. The accuracy of the model on the test set represents transfer learning in the sense that the model was trained on a number of articles about topics other than the target word and had

to use what it learned to classify texts about the target word. The accuracies were still quite high, as demonstrated in section 5. This shows that the model was learning patterns of language other than those specific words. This could mean that it learned similar words because of the word embeddings or it could mean that it learned completely different words to “pay attention” to, or both.

3.2.3 Cleaning

Pre-processing data is a normal first step before training and evaluating the data using a neural network. Machine learning algorithms are only as good as the data you are feeding them. It is crucial that data is formatted properly and meaningful features are included in order to have sufficient consistency that will result in the best possible results. As seen in, for computer vision machine learning algorithms, pre-processing the data involves many steps including normalizing image inputs and dimensionality reduction. The goal of these is to take away some of the unimportant distinguishing features between different images. Features like the darkness or brightness are not beneficial in the task of labeling the image. Similarly, there are portions of text that are not beneficial in the task of labeling the text as real or fake.

The task of pre-processing data is often an iterative task rather than a linear one. This was the case in this project where we used a new and not yet standardized dataset. As we found certain unmeaningful features that the neural net was learning, we learned what more we needed to pre-process from the data.

Non-English Word Removal

Two observations that lead us to more pre-processing were the presence of run-on words and proper nouns in the most important trigrams for classification. An example of a run on word that we saw frequently was in the "most fake" trigram category was "NotMyPresident" that came from a trending "hashtag" on twitter. There were also decisive trigrams that were simply pronouns like "Donald J Trump." Proper nouns could not possibly be helpful in a meaningful way to a machine learning algorithm trying to detect language patterns indicative of real or fake news. We want our algorithm to be agnostic to the subject material and make a decision based on the types of words used to describe whatever the subject is. Another algorithm may aim to fact check statements in news articles. In this situation, it would be important to maintain the proper nouns/subjects because changing the proper noun in the sentence "Donald J. Trump is our current president" to "Hillary Clinton is our current president" changes the classification of true fact to false fact. However, our purpose is not fact checking but rather language pattern checking, so removal of proper nouns should aid in pointing the machine learning algorithms in the right direction as far as finding meaningful features.We removed "non-English" words by using PyEnchants version of the English dictionary. This also accounted for removal of digits, which should not be useful in this classification task, and websites. While links to websites may be useful in classifying the page rank of an article, it is not useful for the specific tool we were trying to create.

Source Pattern Removal

Another observation was that the two real news sources had some specific patterns that were easily learnable by the machine learning algorithms. This was more of an issue with the real news sources than the fake news sources because there were many more fake news sources than real news sources. More specifically, there were 244 fake news sources and only 128 neurons so the algorithm couldnt simply attune one neuron to each of the fake news sources patterns. There were only two 27 real news sources, however. Therefore, the algorithm was able to pick up easily on the presence or absence of these patterns and use that, without much help from other words or phrases, to classify the data. There were a few separate steps in removing patterns from the real news sources. The New York Times articles of a particularly common section often started off with “Good morning. (or evening) Heres what you need to know:” This, along with other repeated sentences were always in italics. To account for the lack of consistency in the exact sentences that were repeated, we had to scrape the data again from the URLs and remove anything that was originally in italics. Another repeated pattern in the New York Times articles was parenthetical questions with links to sign up for emails, for example “Want to get California Today by email? Sign up.)“. Another pattern was in The Guardian, articles almost always ended with “Share on FacebookShare on TwitterShare via EmailShare on LinkedInShare on PinterestShare on Google+Share on WhatsAppShare on MessengerReuse this content” which is the result of links/buttons on the bottom of the webpage to share the article. When

removing the non-English words, we were left with “on on on on on this content” which was enough of a pattern to force the model to learn classification almost solely based on its presence or absence. Note that this was a particularly strong pattern because it was consistent throughout the Guardian articles from all sections of the Guardian. Also, the majority of articles in our real news set are from the Guardian.

3.2.4 Describing Neurons

Although the accuracy was high in the classification task even after extensive pre-processing of the data, we wanted a way to more qualitatively evaluate how and what the neural net was learning the classification. Understanding and visualizing the way a CNN encodes information is an ongoing question. It is an infinitely more challenging pattern when there are more than one convolutional layer, which is why we kept our neural net shallow. For CNNs with one convolutional layer, [19] shows a way to visualize any CNN single neuron as a filter in the first layer, in terms of the image space. We were able to use a similar method to “visualize” the CNN neurons as filters in the first (and only) layer in terms of text space. Instead of finding the location in each image of the window that caused each neuron to fire the most, we find the location in the pre-processed text of the trigram (or length 3 sequence of words) that caused each neuron to fire the most. As the authors of were able to identify patterns of colors/lines in images that caused firing, we were able to identify textual patterns that caused firing. Textual patterns are more

difficult to visualize than image space patterns. While similar but nonidentical RGB pixel values look similar, two words that are mathematically "similar" in their embedding but non-identical do not look similar. They do, however, have similar meanings. In order to get a general grasp of the meaning of words/trigrams that each neuron was firing most highly for, we followed similar steps to those described in the section of 4.2.1. However, instead of finding those neurons that had the highest/lowest weight × activation, we looked at each neuron, and which trigram in each body text resulted in the pooled value for that neuron. Then, we accumulated all of the trigrams for each neuron and summarized them by counting the instances of each word in the trigram. Our algorithm reported the words with the highest counts, excluding stopwords as described by NLTK (i.e. words like "the", "a", "by", "it", which are not meaningful in this circumstance). We were able to observe some clear patterns detected by certain neurons.

CHAPTER 4

Related Work

4.1 Spam Detection

The problem of detecting not-genuine sources of information through content based analysis is considered solvable at least in the domain of spam detection, spam detection utilizes statistical machine learning techniques to classify text (i.e. tweets or emails) as spam or legitimate. These techniques involve pre-processing of the text, feature extraction (i.e. bag of words), and feature selection based on which features lead to the best performance on a test dataset. Once these features are obtained, they can be classified using Nave Bayes, Support Vector Machines, TF-IDF, or K-nearest neighbors classifiers. All of these classifiers are characteristic of supervised machine learning, meaning that they require some labeled data in order to learn the function where, m is the message to be classified and is a vector of parameters and Cspam and Cleg are respectively spam and legitimate messages. The task of detecting fake news is similar and almost analogous to the task of spam detection in that both aim to separate examples of legitimate text from examples of illegitimate, ill-intended texts. The question, then, is how can we apply similar techniques to fake news detection. Instead of filtering like we do with spam, it would be beneficial to be able to flag fake news articles so that readers can be warned that what they are reading is likely to be fake news. The purpose of this project is not to decide for the reader whether or not the document

is fake, but rather to alert them that they need to use extra scrutiny for some documents. Fake news detection, unlike spam detection, has many nuances that arent as easily detected by text analysis. For example, a human actually needs to apply their knowledge of a particular subject in order to decide whether or not the news is true. The “fakeness” of an article could be switched on or off simply by replacing one persons name with another persons name. Therefore, the best we can do from a content-based standpoint is to decide if it is something that requires scrutiny. The idea would be for a reader to do leg work of researching other articles on the topic to decide whether or not the article is actually fake, but a “flagging” would alert them to do so in appropriate circumstances.

4.2 Stance Detection

In December of 2016, a group of volunteers from industry and academia started a contest called the Fake News Challenge. The goal of this contest was to encourage the development of tools that may help human fact checkers identify deliberate misinformation in news stories through the use of machine learning, natural language processing and artificial intelligence. The organizers decided that the first step in this overarching goal was understanding what other news organizations are saying about the topic in question. As such, they decided that stage one of their contest would be a stance detection competition. More specifically, the organizers built a dataset of headlines and bodies of text and challenged competitors to build classifiers that could correctly label the stance of a body text, relative to a given headline, into one of four categories: “agree”, “disagree”, “discusses” or “unrelated.”

The top three teams all reached over 80% accuracy on the test set for this task. The top teams model was based on a weighted average between gradient-boosted decision trees and a deep convolutional neural network.

4.3 Benchmark Dataset

demonstrates previous work on fake news detection that is more directly related to our goal of using a text-only approach to make a classification. The authors not only create a new benchmark dataset of statements (see Section 3.1), but also show that significant improvements can be made in fine-grained fake news detection by using meta-data (i.e. speaker, party, etc) to augment the information provided by the text.

CHAPTER 5

Conclusion

Fake news and Clickbaits interfere with the ability of a user to discern useful information from the Internet services especially when news becomes critical for decision making. Considering the changing landscape of the modern business world, the issue of fake news has become more than just a marketing problem as it warrants serious efforts from security researchers. It is imperative that any attempts to manipulate or troll the Internet through fake news or Clickbaits are countered with absolute effectiveness. We proposed a simple but effective approach to allow users install a simple tool into their personal browser and use it to detect and filter out potential Clickbaits. The preliminary experimental results conducted to assess the method's ability to attain its intended objective, showed outstanding performance in identify possible sources of fake news. Since we started this work, few fake news databases have been made available and we're currently expanding our approach using R to test its effectiveness against the new datasets.

The main contribution of this project is support for the idea that machine learning could be useful in a novel way for the task of classifying fake news. Our findings show that after much pre-processing of relatively small dataset, a simple CNN is able to pick up on a diverse set of potentially subtle language patterns that a human may (or may not) be able to detect. Many of these language patterns are intuitively useful in a humans manner of classifying fake news. Some such intuitive patterns that our model has found to indicate fake news include generalizations, colloquialisms and exaggerations. Likewise, our model looks for indefinite or inconclusive words, referential words, and evidence words as patterns that characterize real news. Even if a

human could detect these patterns, they are not able to store as much information as a CNN model, and therefore, may not understand the complex relationships between the detection of these patterns and the decision for classification. Furthermore, the model seems to be relatively unphased by the exclusion of certain “giveaway” topic words in the training set, as it is able to pick up on trigrams that are less specific to a given topic, if need be. As such, this seems to be a really good start on a tool that would be useful to augment humans ability to detect Fake News. Other contributions of this project is include the creation of a dataset for the task and the creation of an application that aids in the visualization and understanding of the neural nets classification of a given body text. This application could be a tool for humans trying to classify fake news, to get indications of which words might cue them into the correct classification. It could also be useful in researchers trying to develop improved models through the use of improved and enlarged datasets, different parameters, etc. The application also provides a way to see manually how changes in the body text affect the classification.

Future Work

Through the work done in this project, we have shown that machine learning certainly does have the capacity to pick up on sometimes subtle language patterns that may be difficult for humans to pick up on. The next steps involved in this project come in three different aspects. The first of aspect that could be improved in this project is augmenting and increasing the size of the dataset. We feel that more data would be beneficial in ridding the model of any bias based on specific patterns in the source. There is also question as to weather or not the size of our dataset is

sufficient. The second aspect in which this project could be expanded is by comparing it to humans performing the same task. Comparing the accuracies would be beneficial in deciding whether or not the dataset is representative of how difficult the task of separating fake from real news is. If humans are more accurate than the model, it may mean that we need to choose more deceptive fake news examples. Because we acknowledge that this is only one tool in a toolbox that would really be required for an end-to-end system for classifying fake news, we expect that its accuracy will never reach perfect. However, it may be beneficial as a stand-alone application if its accuracy is already higher than human accuracy at the same task. In addition to comparing the accuracy to human accuracy, it would also be interesting to compare the phrases/trigrams that a human would point out if asked what they based their classification decision on. Then, we could quantify how similar these patterns are to those that humans find indicative of fake and real news. Finally, as we have mentioned throughout, this application is only one that would be necessary in a larger toolbox that could function as a highly accurate fakenews classifier. Other tools that would need to be built may include a fact detector and a stance detector. In order to combine all of these "routines," there would need to be some type of model that combines all of the tools and learns how to weight each of them in its final decision.

Reference

- Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, “Fake News Detection on Social Media: A Data Mining Perspective” arXiv:1708.01967v3 [cs.SI], 3 Sep 2017
- M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.
- Fake news websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website. Accessed Feb. 6, 2017
- M. Risdal. (2016, Nov) Getting real about fake news. [Online]. Available: https://www.kaggle.com/mrisdal/fake-news
- J. Soll, T. Rosenstiel, A. D. Miller, R. Sokolsky, and J. Shafer. (2016, Dec) The long and brutal history of fake news. [Online]. Available: https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535
- C. Wardle. (2017, May) Fake news. it’s complicated. [Online]. Available: https://firstdraftnews.com/fake-news-complicated/
- T. Ahmad, H. Akhtar, A. Chopra, and M. Waris Akhtar, “Satire detection from web documents using machine learning methods,” pp. 102–105, 09 2014.
- C. Kang and A. Goldman. (2016, Dec) In washington pizzeria attack, fake news broughtreal guns. [Online]. Available: https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html

- C. Domonoske. (2016, Nov) Students have 'dismaying' inability to tell fake news from real, study finds. [Online]. Available: https://www.npr.org/sections/thetwo-way/2016/11/23/503129818/study-finds-students-have-dismaying-inability-to-tell-fake-news-from-real
- M. T. Banday and T. R. Jan, "Effectiveness and limitations of statistical spam filters," arXiv preprint arXiv:0910.2540, 2009.
- S. Sedhai and A. Sun, "Semi-supervised spam detection in twitter stream," arXiv preprint arXiv:1702.01032, 2017.
- A. Bhowmick and S. M. Hazarika, "Machine learning for e-mail spam filtering: Review, techniques and trends," arXiv preprint arXiv:1606.01042, 2016.
- Fake news challenge stage 1 (fnc-i): Stance detection. [Online]. Available: http://www.fakenewschallenge.org/
- W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," arXiv preprint arXiv:1705.00648, 2017.
- Y. Genes, "Detecting fake news with nlp," May 2017. [Online]. Available: https://medium.com/@Genyunus/detecting-fake-news-with-nlp-c893ec31dee8
- S. Agency. (2016, Dec) Bs detector. [Online]. Available: https://github.com/selfagency/bs-detector
- W. Yin, K. Kann, M. Yu, and H. Schutze, "Comparative study of CNN and RNN for natural language processing," CoRR, vol. abs/1702.01923, 2017. [Online]. Available: http://arxiv.org/abs/1702.01923

- D. Britz. (2016, Feb) Implementing a cnn for text classification in tensorflow. [Online]. Available: http://www.wildml.com/2015/12/implementing-a-cnn-for-text-classification-in-tensorflow/
- E. Kiely and L. Robertson. (2016, Dec) How to spot fake news. [Online]. Available: https://www.factcheck.org/2016/11/how-to-spot-fake-news/
- A. Karpathy, "convolutional neural networks for visual recognition." [Online]. Available: http://cs231n.github.io/understanding-cnn/
- N. B, "Image data pre-processing for neural networks becoming human: Artificial intelligence magazine," Sep 2017. [Online]. Available: https://becominghuman.ai/image-data-pre-processing-for-neural-networks-498289068258
- I. Rafegas and M. Vanrell, "Understanding learned cnn features through filter decoding with substitution," arXiv preprint arXiv:1511.05084, 2015.
- OpenSources. [Online]. Available: http://www.opensources.co/