

FAKE NEWS DETECTION USING DATA SCIENCE

A Project report submitted in partial fulfilment of the requirements of the award of the degree of

Bachelor of Technology

in

Computer Engineering

by

DIKSHANT MATHUR, PCE19CS052DISHA

SARASWAT, PCE19CS053 HARDIK SINGH,

PCE19CS071 ABHISHEK SHARMA,

PCE19CS011

under the guidance of

Ms. Neha Shrotriya,

Assistant Professor



(Session 2020-21)

Department of Computer EngineeringPoornima

College of Engineering

ISI-6, RIICO Institutional Area, Sitapura, Jaipur – 302022

November-2021

Acknowledgement

We would like to convey our profound sense of reverence and admiration to my supervisor **Ms. Neha Shrotriya**, Associate Professor, **Department of Computer Engineering, Poornima College of Engineering**, for his intense concern, attention, priceless direction, guidance and encouragement throughout this research work.

We are grateful to **Dr. Mahesh Bunde**, Director of Poornima College of Engineering for his helping attitude with a keen interest in completing this dissertation in time.

We extend my heartiest gratitude to all the teachers, who extended their cooperation to steer the topic towards its successful completion. We are also thankful to non-teaching staff of the department to support in preparation of this dissertation work.

Our special heartfelt gratitude goes to **Dr. Surendra Kumar Yadav, HOD, Department of Computer Engineering, Dr. Praveen Gupta, Project Coordinator, Department of Computer Engineering, Poornima College of Engineering**, for unvarying support, guidance and motivation during the course of this research.

We would like to express our deep sense of gratitude towards management of Poornima College of Engineering including **Dr. S. M. Seth**, Chairman Emeritus, Poornima Group and former Director NIH, Roorkee, **Shri Shashikant Singhi**, Chairman, Poornima Group, **Mr. M. K. M. Shah**, Director Admin & Finance, Poornima Group and **Ar. Rahul Singhi**, Director Poornima Group for establishment of institute and providing facilities for my studies.

We would like to take the opportunity of expressing my thanks to all faculty members of the Department, for their kind support, technical guidance, and inspiration throughout the course.

Department Certificate

This is to certify that **Mr. Dikshant Mathur**, Registration No. **PCE19CS052**, of the Computer Science Department of Computer Engineering, has submitted this project report entitled **“FAKE NEWS DETECTION USING DATA SCIENCE”** under the supervision of Ms. Neha Shrotriya, working as Associate Professor in department of Computer Engineering as per the requirements of the bachelor of technology program of Poornima College of Engineering, Jaipur.

Dr. Surendra Kumar Yadav

Head, Dept. of Computer Engineering

Ms. Neha Shrotriya

Project-Guide

Department Certificate

This is to certify that **Ms. Disha Saraswat**, Registration No. **PCE19CS053**, of the Computer Science Department of Computer Engineering, has submitted this project report entitled “**FAKE NEWS DETECTION USING DATA SCIENCE**” under the supervision of Ms. Neha Shrotriya, working as Associate Professor in department of Computer Engineering as per the requirements of the bachelor of technology program of Poornima College of Engineering, Jaipur.

Dr. Surendra Kumar Yadav

Head, Dept. of Computer Engineering

Ms. Neha Shrotriya

Project-Guide

Department Certificate

This is to certify that **Mr. Hardik Singh**, Registration No. **PCE19CS071**, of the Computer Science Department of Computer Engineering, has submitted this project report entitled “**FAKE NEWS DETECTION USING DATA SCIENCE**” under the supervision of Ms. Neha Shrotriya, working as Associate Professor in department of Computer Engineering as per the requirements of the bachelor of technology program of Poornima College of Engineering, Jaipur.

Dr. Surendra Kumar Yadav

Head, Dept. of Computer Engineering

Ms. Neha Shrotriya

Project-Guide

Department Certificate

This is to certify that **Mr. Abhishek Sharma**, Registration No. **PCE19CS011**, of the Computer Science Department of Computer Engineering, has submitted this project report entitled **“FAKE NEWS DETECTION USING DATA SCIENCE”** under the supervision of Ms. Neha Shrotriya, working as Associate Professor in department of Computer Engineering as per the requirements of the bachelor of technology program of Poornima College of Engineering, Jaipur.

Dr. Surendra Kumar Yadav
Head, Dept. of Computer Engineering

Ms. Neha Shrotriya
Project- Guide

Candidate's Declaration

I hereby declare that the work which is being presented in this project report entitled “**FAKE NEWS DETECTION USING DATA SCIENCE**” in the partial fulfilment for the award of the Degree of Bachelor of Technology in (Computer Engineering), submitted in the Department of Computer Engineering, Poornima College of Engineering, Jaipur, is an authentic record of my own work done during the period from August 2021 to December 2021 under the supervision and guidance of Ms. NehaShrotriya.

I have not submitted the matter embodied in this project report for the award of any other degree.

Name of Candidate: **Dikshant Mathur**

Registration no.: **PCE19CS052**

Dated: 29/11/2021

Place: Jaipur

Candidate's Declaration

I hereby declare that the work which is being presented in this project report entitled “**FAKE NEWS DETECTION USING DATA SCIENCE**” in the partial fulfilment for the award of the Degree of Bachelor of Technology in (Computer Engineering), submitted in the Department of Computer Engineering, Poornima College of Engineering, Jaipur, is an authentic record of my own work done during the period from August 2021 to December 2021 under the supervision and guidance of Ms. NehaShrotriya.

I have not submitted the matter embodied in this project report for the award of any other degree.

Name of Candidate: **Disha Saraswat**

Registration no.: **PCE19CS053**

Dated: 29/11/2021

Place: Jaipur

Candidate's Declaration

I hereby declare that the work which is being presented in this project report entitled “**FAKE NEWS DETECTION USING DATA SCIENCE**” in the partial fulfilment for the award of the Degree of Bachelor of Technology in (Computer Engineering), submitted in the Department of Computer Engineering, Poornima College of Engineering, Jaipur, is an authentic record of my own work done during the period from August 2021 to December 2021 under the supervision and guidance of Ms. NehaShrotriya.

I have not submitted the matter embodied in this project report for the award of any other degree.

Name of Candidate: **Hardik Singh**

Registration no.: **PCE19CS071**

Dated: 29/11/2021

Place: Jaipur

Candidate's Declaration

I hereby declare that the work which is being presented in this project report entitled “**FAKE NEWS DETECTION USING DATA SCIENCE**” in the partial fulfilment for the award of the Degree of Bachelor of Technology in (Computer Engineering), submitted in the Department of Computer Engineering, Poornima College of Engineering, Jaipur, is an authentic record of my own work done during the period from August 2021 to December 2021 under the supervision and guidance of Ms. NehaShrotriya

I have not submitted the matter embodied in this project report for the award of any other degree.

Name of Candidate: **Abhishek Sharma**

Registration no.: **PCE19CS011**

Dated: 29/11/2021

Place: Jaipur

TABLE OF CONTENTS

CHAPTER NO.	PARTICULARS	PAGE NO.
	Title Page	2
	Acknowledgment	3
	Certificate by the Department	4-7
	Candidate Declaration	8-11
	Table of Contents	
1	Abstract	13
2	Chapter 1: Introduction	14
3	Chapter 2: Literature Review	15-20
4	Chapter 3: System/Project Development	21
	3.1 Project Description	22
5	Chapter 4: Conclusion	23
6	Chapter 5: Future Scope	24
6	Chapter 6: References	25

ABSTRACT

Due to the large increase in the proliferation of “fake news” in recent years, it has become a widely discussed menace in the online world. In conjunction with this popularity, research of ways to limit the spread has also increased. This paper aims to look at the current research of this area in order to see what automatic fake news detection methods exist and are being developed, which can help online users in protecting themselves against fake news. A systematic literature review is conducted in order to answer this question, with different detection methods discussed in the literature being divided into categories. The consensus which appears from the collective research between categories is also used to identify common elements between categories which are important to fake news detection; notably the relation of headlines and article content, the importance of high-quality datasets, the use of emotional words, and the circulation of fake news in social media groups.

Chapter 1: Introduction

There are numerous reputed websites that post legitimate news contents, and a few other websites such as PolitiFact and Snopes which are used for fact checking. In addition, there are open repositories which are maintained by researchers to keep an up-to-date list of currently available datasets and hyperlinks to potential fact checking sites that may help in countering false news spread. However, we selected three datasets for our experiments which contain news from multiple domains (such as politics, entertainment, technology, and sports) and contain a mix of both truthful and fake articles. The datasets are available online and are extracted from the World Wide Web.

Fake news and hoaxes have been there since before the advent of the Internet. The widely accepted definition of Internet fake news is: fictitious articles deliberately fabricated to deceive readers”. Social media and news outlets publish fake news to increase readership or as part of psychological warfare. In general, the goal is profiting through clickbaits. Clickbaits lure users and entice curiosity with flashy headlines or designs to click links to increase advertisements revenues. This exposition analyzes the prevalence of fake news in light of the advances in communication made possible by the emergence of social networking sites. The purpose of the work is to come up with a solution that can be utilized by users to detect and filter out sites containing false and misleading information. We use simple and carefully selected features of the title and post to accurately identify fake posts. The experimental results show a 99.4% accuracy using logistic classifier

CHAPTER 2: Literature Review

3.1 Paper Review

Paper 1:

TITLE: Fake News Detection Using Machine Learning Algorithms

By -Uma Sharma

Summary:

In the 21st century, the majority of the tasks are done online. Newspapers that were earlier preferred as hard-copies are now being substituted by applications like

Facebook, Twitter, and news articles to be read online. Whatsapp's forwards are also a major source. The growing problem of fake news only makes things more complicated and tries to change or hamper the opinion and attitude of people towards use of digital technology. When a person is deceived by the real news two possible things happen- People start believing that their perceptions about a particular topic are true as assumed.

Thus, in order to curb the phenomenon, we have developed our Fake news Detection system that takes input from the user and classify it to be true or fake. To implement this, various NLP and Machine Learning Techniques have to be used. The model is trained using an appropriate dataset and performance evaluation is also done using various performance measures. The best model, i.e. the model with highest accuracy is used to classify the news headlines or articles. As evident above for static search, our best model came out to be Logistic Regression with an accuracy of 65% . .

Paper 2:

TITLE: Fake News Detection on Social Media: A Data Mining

Perspective

By - Suhang Wang, Kai Shu

Summary:

Social media for news consumption is a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information lead people to seek out and consume news from social media. On the other hand, it enables the wide spread of "fake news", i.e., low quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society.

Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention. Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media ineffective or not applicable. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content; therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination. Second, exploiting this auxiliary information is challenging in and of itself as users' social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. Because the issue of fake news detection on social media is both challenging and relevant, we conducted this survey to further facilitate research on the problem. In this survey, we present a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets. We also discuss related research areas, open problems, and

Paper 3:**TITLE: Analysis of Classifiers for Fake News Detection****By - Vasu Agarwal****Summary:**

As time flows, the amount of data, especially text data increases exponentially. Along with the data, our understanding of AI also increases and the computing power enables us to train very complex and large models faster. Fake news has been gathering a lot of attention worldwide recently. The effects can be political, economic, organizational, or even personal. This paper discusses the approach of natural language processing and machine learning in order to solve this problem. Use of bag-of-words, n-grams, count vectorizer has been made, TF-IDF, and trained the data on five classifiers to investigate which of them works well for this specific dataset of labelled news statements. The precision, recall and f1 scores help us determine which model works best.

Paper 4:**TITLE: A Tool for Fake News Detection****By - Madalina Erascu****Summary:**

The word post-truth was considered by Oxford Dictionaries Word of the Year 2016. The word is an adjective relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief. This leads to misinformation and problems in society. Hence, it is important to make effort to detect these facts and prevent them from spreading. In this paper, we propose machine learning techniques, in particular supervised learning, for fake news detection. More precisely, we used a dataset of fake and real news to train a machine learning model using Scikit-learn library in Python. We extracted features from the dataset using text representation models like Bag-of-Words, Term Frequency-Inverse Document Frequency (TF-IDF) and Bi-gram frequency. We tested two classification approaches, namely probabilistic classification and linear classification on the title and the content, checking if it is clickbait/nonclickbait, respectively fake/real. The outcome of our experiments was that the linear classification works the best with the TF-IDF model in the process of content classification. The Bi-gram frequency model gave the lowest accuracy for title classification in comparison with Bag-of- Words and TF-IDF.

Paper 5:**TITLE: Using data science to detect fake news****By - Eliza Shoemaker****Summary:**

The purpose of this thesis is to assist in automating the detection of Fake News by identifying which features are more useful for different classifiers. The effectiveness of different extracted features for Fake News detection are going to be examined. When classifying text with machine learning algorithms features have to be extracted from the articles for the classifiers to be trained on. In this thesis, several different features are extracted: word counts, ngram counts, term frequency-inverse document frequency, sentiment analysis, lemmatization, and named entity recognition to train the classifiers. Two classifiers are used, a Random Forest classifier and a Naïve Bayes classifier. Training on different features combined with different machine learning algorithms yields different accuracies. By testing the different features on different classifiers, it can be determined which features are the best for Fake News detection. Classifying news articles as either Fake News or as not Fake News is explored using three datasets, which in total contains over 40,000 articles. One of the datasets is used to partly to train the classifiers and partly to test the classifiers. The remaining two datasets are used purely for testing the classifiers

Paper 6:**TITLE: Fake News Detection: A Deep Learning Approach****By - Aswini Thota****Summary:**

Fake news is defined as a made-up story with an intention to deceive or to mislead. In this paper we present the solution to the task of fake news detection by using Deep Learning architectures. Gartner research predicts that “By 2022, most people in mature economies will consume more false information than true information”. The exponential increase in production and distribution of inaccurate news presents an immediate need for automatically tagging and detecting such twisted news articles. However, automated detection of fake news is a hard task to accomplish as it requires the model to understand nuances in natural language. Moreover, majority of the existing fake news detection models treat the problem at hand as a binary classification task, which limits model’s ability to understand how related or unrelated the reported news is when compared to the real news. To address these gaps, we present neural network architecture to accurately predict the stance between a given pair of headline and article body. Our model outperforms existing model architectures by 2.5% and we are able to achieve an accuracy of 94.21% on test data.

Paper 7:**TITLE: Fake News Detection Using Machine Learning Approaches****By - Z Khanam****Summary:**

This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python scikit-learn, NLP for textual analysis. This process will result in feature extraction and vectorization; we propose using Python scikit-learn library to perform tokenization and feature extraction of text data, because this library contains useful tools like Count Vectorizer and Tfidf Vectorizer. Then, we will perform feature selection methods, to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

The research in this paper focuses on detecting the fake news by reviewing it in two stages: characterization and disclosure. In the first stage, the basic concepts and principles of fake news are highlighted in social media. During the discovery stage, the current methods are reviewed for detection of fake news using different supervised learning algorithms.

Paper 8:**TITLE: Fake News Data Exploration and Analytics****By - Awais Yasin****Summary:**

This paper proposes models of machine learning that can successfully detect fake news. These models identify which news is real or fake and specify the accuracy of said news, even in a complex environment. After data-preprocessing and exploration, we applied three machine learning models; random forest classifier, logistic regression, and term frequency-inverse document frequency (TF-IDF) vectorizer. The accuracy of the TFIDF vectorizer, logistic regression, random forest classifier, and decision tree classifier models was approximately 99.52%, 98.63%, 99.63%, and 99.68%, respectively. Machine learning models can be considered a great choice to find reality-based results and applied to other unstructured data for various sentiment analysis applications.

Paper 9:**TITLE: FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network****By - Jiawei Zhang****Summary:**

This paper aims at investigating the principles, methodologies and algorithms for detecting fake news articles, creators and subjects from online social networks and evaluating the corresponding performance. This paper addresses the challenges introduced by the unknown characteristics of fake news and diverse connections among news articles, creators and subjects. This paper introduces a novel automatic fake news credibility inference model, namely FAKEDETECTOR. Based on a set of explicit and latent features extracted from the textual information, FAKEDETECTOR builds a deep diffusive network model to learn the representations of news articles, creators and subjects simultaneously. Extensive experiments have been done on a real-world fake news dataset to compare FAKEDETECTOR with several state-of-the-art models, and the experimental results have demonstrated the effectiveness of the proposed model.

Paper 10:**TITLE: Fake News Detection: A long way to go****By - Sunidhi Sharma****Summary:**

This news needs to be in authentic form which is usually found in adulterated version. Leading us to have a dire need for an identification of real news from any possible fake news. News, being a form of information can be subjective to the proofs and source for its authenticity. As a human, one can easily identify real news from fake news with the help of one's innate capability to deduce logic and outlandish source of the information piece. Just that one needs few trusted sources to check for the facts and myths. But on a real time basis, there is a dire need for some software which can nip such 'false news' in its bud.

Leading it to be one of the most researched area nowadays. Primarily being a part of Information Retrieval, this area is taking up a lot of attention from researchers worldwide to come up with a real-time solution for such an issue.

In this article we have checked and analysed many research articles along with many survey articles and summed up this paper so as to provide the readers with a short idea of what fake news is, its different flavours in the news spectrum, its characteristics and identification basic. We also included the different methods used by prior researchers in the same field. Using few researches as examples we learned about the basics of those methods used in fake news identification. The future aspects are also included in this article along with the challenges one faces while doing.

Paper 11:**TITLE: Fake news detection in social media****By - Kelly Stahl****Summary:**

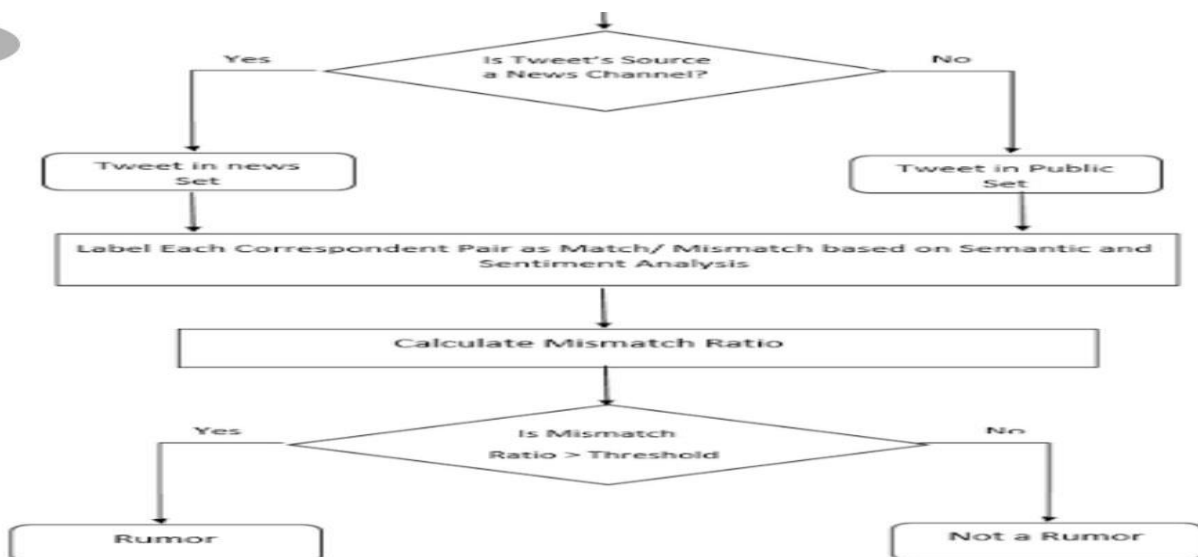
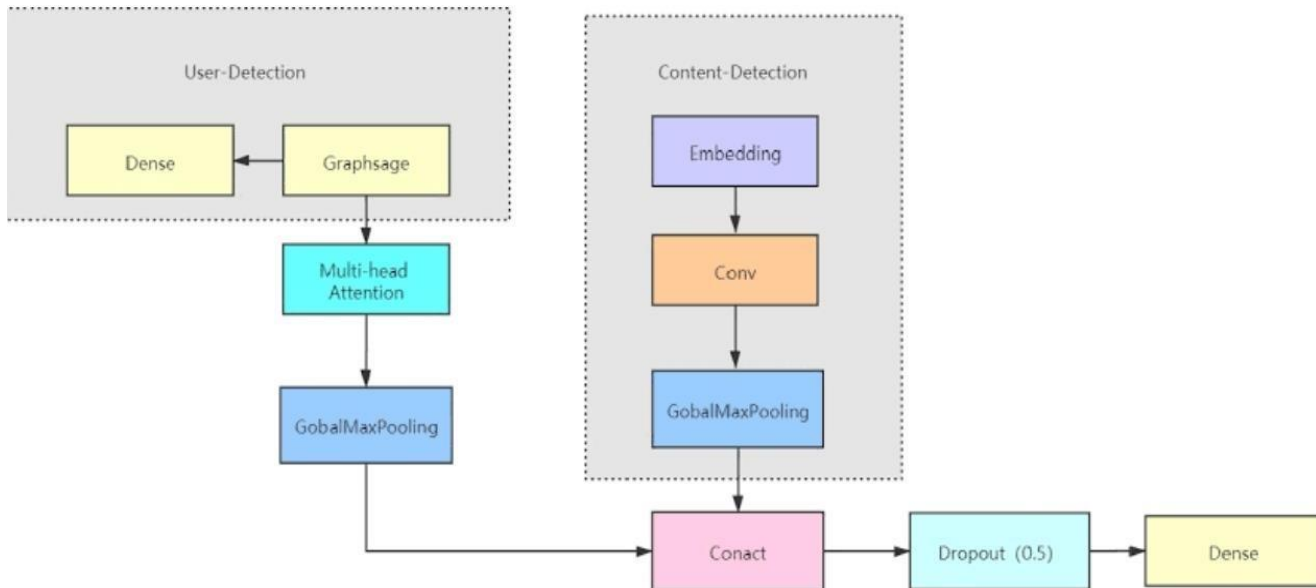
Due to the exponential growth of information online, it is becoming impossible to decipher the true from the false. Thus, this leads to the problem of fake news. This research considers previous and current methods for fake news detection in textual formats while detailing how and why fake news exists in the first place. This paper includes a discussion on Linguistic Cue and Network Analysis approaches, and proposes a three-part method .

Paper 12:**TITLE: Fake News Detection Using Machine Learning Approaches****By - B N Alwase****Summary:**

The fake news on social media and various other media is wide spreading and is a matter of serious concern due to its ability to cause a lot of social and national damage with destructive impacts. A lot of research is already focused on detecting it. This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python scikit-learn, NLP for textual analysis. This process will result in feature extraction and vectorization; we propose using Python scikit-learn library to perform tokenization and feature extraction of text data, because this library contains useful tools like Count Vectorizer and Tfidf Vectorizer. Then, we will perform feature selection methods, to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

Chapter 3: System/Project Development

Attributes involved in the model of the project: -



Project Description

To find patterns in Fake News, first news needs to be collected and labeled. Both Fake News and legitimate news needs to be represented in roughly equal amounts. This is to avoid the frequency of Fake News in the dataset being used as a determining factor in classifying. Having good data is essential producing valid results. Good data in this context is data that is representative of the real world and is generalizable. The dataset used to train the classifiers is the ISOT Fake News Dataset, the largest available dataset of full length Fake News articles .

The ISOT dataset contains 21,417 articles labeled Real and 23,481 that are labeled Fake, totaling 44,898. FakeNewsNet is another data set containing full length articles, however there are only 422 labeled articles in it [6]. And lastly there is a set of 180 articles, 90 Fake and 90 Real, collected by the author which, will be referred to as the Original Data. These two additional datasets will be used to test the accuracy of the trained classifiers. Each model will initially be trained with 80% of the ISOT data. The remaining 20% of the ISOT data will be used to test the accuracy of the trained classifiers. As mentioned, FakeNewsNet and the Original Data will be used for testing as well. The reasoning behind using these additional tests is to make sure we are detecting Fake News and not some other pattern of the ISOT dataset, such as a style of a particular news organization. Each article labeled as Real in the ISOT dataset was collected from Reuters; all articles their started with the word “Reuters”. This pattern could easily be picked up by humans and machines alike. To avoid this issue the beginning “Reuters” phrase was removed from each article.

To find patterns, several different features should be tested. Features are numeric values that describe the text. Examples of these numeric values are word count or the number of times a particular punctuation mark is used. Some features will be more helpful than others, for instance the number of verbs is more likely to be useful compared to the number of times a particular word is used, such as ‘kitten’. The goal is to find the features that are most helpful in detection of Fake News. Next, each extracted feature will be discussed in detail. Word counts are among the most easily obtained features that can be extracted from raw text. It is simply a count of all the terms in a body of text. Word counts are also called a ‘bag of words’, however, to keep names descriptive, we shall call this type of feature a count. To get the word count in texts, scikit-learn’s CountVectorizer is used; the CountVectorizer tokenizes the data and then counts each term [14]. The data can be tokenized by word or by n-gram. N-grams are series of n items, such as words or characters. In this thesis n-grams refers to groupings of two and three characters. For instance, the n-grams of the word ‘feature’ would be as follows: ‘fe’, ‘ea’, ‘at’, ‘tu’, ‘ur’, ‘re’, ‘fea’, ‘eat’, ‘atu’, ‘tur’, and ‘ure’. These features will be referred to as countword and count-ngram respectively. Term frequency-inverse document frequency, or TF-IDF, is calculated as follows: term frequency times the inverse document frequency. Where term frequency is the number of times a term is in a document divided by the number of terms in a document. The inverse document frequency is the logarithm of the number of text (or articles) in the collection divided by the number of texts or articles where the term appears.

Chapter 4: Conclusion

The problems of fake news and disinformation play an important role on nowadays life. This is because the advanced level of technology and communication methods we have enabled information spreading among people without any verification. This is a reason why researchers started searching for solutions to stop fake news and disinformation from spreading easily. However, it is well known that controlling the flow of information online is impossible. In this paper, we performed an attempt to verify the news articles credibility depending on their characteristics. At this aim, we implemented an algorithm combining several classification methods with text models. It performed well, and the accuracy results were relatively satisfying.

With the nature of the Internet as it is, Fake News is easily created and distributed. Fact checking is tedious and time consuming, so automating Fake News detection is critical. Thus Fake News classifiers should be created. However, a classifier does not come out of thin air, it must be trained on already existing data. The quality and quantity of the data is important. Three datasets were used for the research in this thesis. ISOT, a huge dataset of over 40,000 articles. FakeNewsNet is another, much smaller dataset containing 422 articles. Lastly, the Original dataset, containing 180 articles, that was gathered specifically for this research. However, a classifier cannot read, so it must have features extracted for the articles. A feature is a numeric value extracted from the article. Such as a word count, or a count of parts of speech, or more complicated features. Such as a count of the named entities, like businesses or organizations. However, which features work best? Two different classifiers, Random Forests and Naive Bayes, were trained on the 80% of the ISOT dataset reserved for testing using each of the ten different features: Count-word, Countngram, TFIDF-word, TFIDF-ngram, PoS, ER, Lemma, VADER, NLTKStop, and spaCyStop. Then each classifier was tested on the remaining 20% from ISOT, all of FakeNewsNet, and all of the Original dataset. The accuracy results where then examined and conclusions were drawn. The ISOT dataset did not generalize well to the other two datasets used for testing. Making the test results for the 20% testing portion of ISOT get way higher results than the other two datasets. This could be found the fact that ISOT got all of its real news from Reuters and the classifiers ended up being a Reuters vs not-Reuters classifier. Next it was discovered that Count/TFIDF are better standalone features than PoS, ER, and VADER. However, these features still have potential to be used in conjunction with other 23 features. Although Lemma was one of the better features, it was outperformed by Count-word, suggesting that some of the removed data was improving the classification.

Chapter 5: Future scope

We're living in the age of fake news. Fake news consist of deliberate misinformation under the guise of being authentic news, spread via some communication channel, and produced with a particular objective like generating revenue, promoting or discrediting a public figure, a political movement, an organization, etc.

The problems of fake news and disinformation play an important role on nowadays life. This is because the advanced level of technology and communication methods we have enabled information spreading among people without any verification.

This is a reason why researchers started searching for solutions to stop fake news and disinformation from spreading easily. However, it is well known that controlling the flow of information online is impossible. In this paper, we performed an attempt to verify the news articles credibility depending on their characteristics. At this aim, we implemented an algorithm combining several classification methods with text models.

It performed well, and the accuracy results were relatively satisfying. As future work, we plan to better study the combination between the feature extraction methods and the classifiers as we will be able to choose the text representation model that performs best with the classifier.

Moreover, to achieve a higher accuracy, we will have to implement a more sophisticated algorithm which may use data mining technologies with big data, because creating a big dataset including more types of news articles with more class variables (labels) will help raising the accuracy score.

Chapter 6: References

- ❖ A. Gelfert, “Fake news: A definition,” *Informal Log.*, vol. 38, no. 1, pp. 84–117, 2018. [2] K. Rogers and J. Bromwich, “The Hoaxes, Fake News and Misinformation We Saw on Election Day,” *NY Times*, New York, 08-Nov-2016.
- ❖ I. Witten, E. Frank, and M. Hall, *Data Mining* 4th Edition. 2016.
- ❖ W. Y. Wang, “‘Liar, Liar Pants on Fire’: A New Benchmark Dataset for Fake News Detection,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2017, pp. 422–426.
- ❖ D. Byrd, “The science of fake news gets a boost,” 2018. [Online]. Available: <http://earthsky.org/human-world/fake-news-mar-2018-article-science-calling-for-studies>.
- ❖ K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, “FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media,” 2018.
- ❖ “BS DETECTOR.” [Online]. Available: <https://github.com/selfagency/bs-detector>.
- ❖ H. Ahmed, I. Traore, and S. Saad, “Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques,” in *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, 2017, pp. 127–138.
- ❖ H. Ahmed, I. Traore, and S. Saad, “Detecting opinion spams and fake news using text classification,” *Secur. Priv.*, vol. 1, no. 1, p. e9, 2018.
- ❖ Y. Kim, “Convolutional Neural Networks for Sentence Classification,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751.