**Title of Project**

Submitted in partial fulfillment of the requirements

of the degree of

**Bachelor of Technology**

by

**Name of the Student**

**(Reg. No.\_\_\_\_\_\_\_)**

Supervisor (s):

**Name of Supervisor (s)**

****

**Department of Electronics and Telecommunication Engineering,**

**Shri Guru Gobind Singhji Institute of Engineering & Technology, Vishnupuri, Nanded, Maharashtra, India, 431606.**

**2018-19**

**DECLARATION**

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

**Rutik Jayram Torambe**

**Reg. No.:2020BEC014**

Date: \_\_\_\_\_\_\_\_\_\_

**CERTIFICATE**

This is to certify that the report entitled “**Title of the project”** being submitted by  **Rutik Jayram Torambe** (2020BEC014) to **Shri Guru Gobind Singhji Institute of Engineering and Technology, Vishnupuri, Nanded (M.S.), India,** as partial fulfillment for the award of the degree of  **Bachelor of Technology** in **Electronics and Telecommunication Engineering**, is a record of bonafide work carried out by him under our supervision and guidance. The matter contained in this report has not been submitted to any other university for the award of any degree or diploma.

**Name of supervisor Dr. M. B. Kokare**

**Supervisor H.O.D.**

**Elect. and Telecom. Engg. Dept. Elect. and Telecom. Engg. Dept.**

**APPROVAL SHEET**

This report entitled "Title of project" by Rutik Jayram Torambe is approved for the degree of Bachelor of Technology.

**Examiners** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Supervisor (s)** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date : \_\_\_\_\_\_\_\_\_\_\_\_

Place : \_\_\_\_\_\_\_\_\_\_\_\_

**ACKNOWLEDGEMENT**

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crowned our efforts with success.

I have great pleasure in expressing my deep sense of gratitude to the respected chairman Mr. Sunil Raithatta for having provided me with great infrastructure and well-furnished labs.

I take this opportunity to express my profound gratitude to our respected Director Dr. Manesh B Kokare for his constant support and encouragement.

I am grateful to the Head of the Department Dr. M.V. Bhalerao, Department of Electronics & Telecommunication Engineering, for her unfailing encouragement and suggestion given to me in the course of my Final Year Project work.

Guidance and deadlines play a very important role in successful completion of the seminar report on time. I also convey my gratitude to Dr. Kamble Usha, Professor, Department of Electronics & Telecommunication Engineering, for having constantly monitored the development of the Final Year Project work and setting up precise deadlines.

Finally, a note of thanks to, Department of Electronics & Telecommunication Engineering, both teaching and non-teaching staff for their co-operation extended to me.

**Rutik Jayram Torambe**

**Abstract**

In recent years, the rapid growth of online social networks has led to a surge in the dissemination of fake news, driven by various commercial and political motives, posing significant challenges to trustworthiness and integrity in the online sphere. Deceptive content on social media platforms can swiftly infiltrate users' feeds, exerting profound impacts on offline society. This study delves into the principles, methodologies, and algorithms aimed at promptly detecting fake news articles, their creators, and associated subjects within online social networks, along with assessing their efficacy. The precision of information dissemination, particularly on social media platforms, has emerged as a pressing concern. However, the sheer scale of web-based data complicates the task of identifying, evaluating, and rectifying such misleading content, commonly referred to as "fake news." In this paper, we propose a method for detecting fake news and explore its application on Facebook, a prominent online social media platform. Leveraging a Naive Bayes classification model, our approach predicts whether a Facebook post is genuine or fabricated. Furthermore, we discuss various techniques to enhance the accuracy of these predictions. Our findings underscore the viability of employing machine learning methodologies to address the challenge of fake news detection.

**Keywords:** Fake news, Social media, Detection algorithms, Information trustworthiness, Machine learning, Online disinformation.

**CONTENTS**

**List of Figures i**

**List of Tables ii**

**Abbreviations, Notations and Nomenclature iii**

|  |  |  |
| --- | --- | --- |
| **1.** | **Introduction**  1.1 MOTIVATION  1.2 OBJECTIVE  1.3 OVERVIEW OF PROJECT | **1**  3  4 |
| **2.** | **Literature Survey**   * 1. MEDIA RICH FAKE NEWS DETECTION   2.1.1 WEAKLY SUPERVISED LEARNING FOR FAKE NEWS   * 1. FAKE NEWS DETECTION IN SOCIAL MEDIA   2. THE SPREA OF FAKE NEWS BY SOCIAL BOTS   3. MISLEADING ONLINE CONTENT | **5** |
| **3.** | **Pre-Requirements of Projects**  3.1  3.2  3.3 |  |
| **4.** | **Proposed Method**  4.1  4.2  4.3 |  |
| **5.** | **Experimental Results**  5.1  5.2  5.3 |  |
| **6.** | **Conclusions and Future Scope** |  |

**References**

**Appendix (if any)**

**CHAPTER 1**

**Introduction**

These days‟ fake news is creating different issues from sarcastic articles to a fabricated news and plan government propaganda in some outlets. Fake news and lack of trust in the media are growing problems with huge ramifications in our society. Obviously, a purposely misleading story is “fake news” but lately blathering social media’s discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred viewpoints. The importance of disinformation within American political discourse was the subject of weighty attention, particularly following the American president election. The term 'fake news' became common parlance for the issue, particularly to describe factually incorrect and misleading articles published mostly for the purpose of making money through page views. In this paper, it is seemed to produce a model that can accurately predict the likelihood that a given article is fake news. Facebook has been at the epicenter of much critique following media attention. They have already implemented a feature to flag fake news on the site when a user sees’ it; they have also said publicly they are working on to distinguish these articles in an automated way. Certainly, it is not an easy task. A given algorithm must be politically unbiased – since fake news exists on both ends of the spectrum – and also give equal balance to legitimate news sources on either end of the spectrum. In addition, the question of legitimacy is a difficult one. However, in order to solve this problem, it is necessary to have an understanding on what Fake News.

**1.1 MOTIVATION**

We will be training and testing the data, when we use supervised learning, it means we are labeling the data. By getting the testing and training data and labels we can perform different machine learning algorithms but before performing the predictions and accuracies, the data is need to be preprocessing i.e. the null values which are not readable are required to be removed from the data set and the data is required to be converted into vectors by normalizing and tokening the data so that it could be understood by the machine. Next step is by using this data, getting the visual reports, which we will get by using the Mat Plot Library of Python and Sickit-Learn. This library helps us in getting the results in the form of histograms, pie charts or bar charts.

**1.2 OBJECTIVE**

The objective of this project is to examine the problems and possible significances related with the spread of fake news. We will be working on different fake news data set in which we will apply different machine learning algorithms to train the data and test it to find which news is the real news or which one is the fake news. As the fake news is a problem that is heavily affecting society and our perception of not only the media but also facts and opinions themselves. By using the artificial intelligence and the machine learning, the problem can be solved as we will be able to mine the patterns from the data to maximize well defined objectives. So, our focus is to find which machine learning algorithm is best suitable for what kind of text dataset. Also, which dataset is better for finding the accuracies as the accuracies directly depends on the type of data and the amount of data. The more the data, more are your chances of getting correct accuracy as you can test and train more data to find out your results.

**1.3 OVERVIEW OF PROJECT**

With the advancement of technology, digital news is more widely exposed to users globally and contributes to the increment of spreading and disinformation online. Fake news can be found through popular platforms such as social media and the Internet. There have been multiple solutions and efforts in the detection of fake news where it even works with tools. However, fake news intends to convince the reader to believe false information which deems these articles difficult to perceive. The rate of producing digital news is large and quick, running daily at every second, thus it is challenging for machine learning to effectively detect fake news.

**CHAPTER 2**

**Literature Survey**

The available literature has described many automatic detection techniques of fake news and deception posts. Since there are multidimensional aspects of fake news detection ranging from using chatbots for spread of misinformation to use of click baits for the rumor spreading. There are many click baits available in social media networks including Facebook which enhance sharing and liking Proceedings of posts which in turn spreads falsified information. Lot of work has been done to detect falsified information.

**2.1 MEDIA RICH FAKE NEWS DETECTION: A SURVEY**

In general, the goal is profiting through click baits. Click baits 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.

**2.1.1 WEAKLY SUPERVISED LEARNING FOR FAKE NEWS DETECTION ON TWITTER**

The problem of automatic detection of fake news in social media, e.g., on Twitter, has recently drawn some attention. Although, from a technical perspective, it can be regarded as a straight-forward, binary classification problem, the major challenge is the collection of large enough training corpora, since manual annotation of tweets as fake or non-fake news is an expensive and tedious endeavor. In this paper, we discuss a weakly supervised approach, which automatically collects a large-scale, 4 but very noisy training dataset comprising hundreds of thousands of tweets. During collection, we automatically label tweets by their source, i.e., trustworthy or untrustworthy source, and train a classifier on this dataset. We then use that classifier for a different classification target, i.e., the classification of fake and non-fake tweets. Although the labels are not accurate according to the new classification target (not all tweets by an untrustworthy source need to be fake news, and vice versa), we show that despite this unclean inaccurate dataset, it is possible to detect fake news with an F1 score of up to 0.9.

**2.2 FAKE NEWS DETECTION IN SOCIAL MEDIA**

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 click baits. Click baits 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. Automatic Online Fake News Detection Combining Content and Social Signals The proliferation and rapid diffusion of fake news on the Internet highlight the need of automatic hoax detection systems. In the context of social networks, machine learning (ML) methods can be used for this purpose. Fake news detection strategies are traditionally either based on content analysis (i.e. analyzing the content of the news) or - more recently - on social context models, such as mapping the news‟ diffusion pattern. In this paper, we first propose a novel ML fake news detection method which, by combining news content and social context features, outperforms 5 existing methods in the literature, increasing their already high accuracy by up to 4.8%. Second, we implement our method within a Facebook Messenger chatbot and validate it with a real-world application, obtaining a fake news detection accuracy of 81.7%. In recent years, the reliability of information on the Internet has emerged as a crucial issue of modern society. Social network sites (SNSs) have revolutionized the way in which information is spread by allowing users to freely share content. As a consequence, SNSs are also increasingly used as vectors for the diffusion of misinformation and hoaxes. The amount of disseminated information and the rapidity of its diffusion make it practically impossible to assess reliability in a timely manner, highlighting the need for automatic hoax detection systems. As a contribution towards this objective, we show that Facebook posts can be classified with high accuracy as hoaxes or non-hoaxes on the basis of the users who "liked" them. We present two classification techniques, one based on logistic regression, the other on a novel adaptation of Boolean crowdsourcing algorithms. On a dataset consisting of 15,500 Facebook posts and 909,236 users, we obtain classification accuracies exceeding 99% even when the training set contains less than 1% of the posts. We further show that our techniques are robust: they work even when we restrict our attention to the users who like both hoax and non-hoax posts. These results suggest that mapping the diffusion pattern of information can be a useful component of automatic hoax detection systems.

**2.3 THE SPREAD OF FAKE NEWS BY SOCIAL BOTS**

The massive spread of fake news has been identified as a major global risk and has been alleged to influence elections and threaten democracies. Communication, cognitive, social, and computer scientists are engaged in efforts to study the complex causes for the viral diffusion of digital misinformation and to develop solutions, while search and social media platforms are beginning to deploy countermeasures. However, to date, these efforts have been mainly informed by anecdotal evidence rather than systematic data. Here we analyze 14 million messages spreading 400 6 thousand claims on Twitter during and following the 2016 U.S. presidential campaign and election. We find evidence that social bots play a key role in the spread of fake news. Accounts that actively spread misinformation are significantly more likely to be bots. Automated accounts are particularly active in the early spreading phases of viral claims, and tend to target influential users. Humans are vulnerable to this manipulation, retweeting bots who post false news. Successful sources of false and biased claims are heavily supported by social bots. These results suggests that curbing social bots may be an effective strategy for mitigating the spread of online misinformation.

**2.4 MISLEADING ONLINE CONTENT**

Tabloid journalism is often criticized for its propensity for exaggeration, sensationalization, scare-mongering, and otherwise producing misleading and low-quality news. As the news has moved online, a new form of tabloidization has emerged: “click baiting., Clickbait” refers to “content whose main purpose is to attract attention and encourage visitors to click on a link to a particular web page” [“clickbait”, n.d.] and has been implicated in the rapid spread of rumor and misinformation online. This paper examines potential methods for the automatic detection of clickbait as a form of deception. Methods for recognizing both textual and non-textual click baiting cues are surveyed, leading to the suggestion that a hybrid approach may yield best results. Big Data Analytics and Deep Learning are two high-focus of data science. Big Data has become important as many organizations both public and private have been collecting massive amounts of domain-specific information, which can contain useful information about problems such as national intelligence, cyber security, fraud detection, marketing, and medical informatics. Companies such as Google and Microsoft are analyzing large volumes of data for business analysis and decisions, impacting existing and future technology. Deep Learning algorithms extract high level, complex abstractions as data representations through a hierarchical learning 7 process. Complex abstractions are learnt at a given level based on relatively simpler abstractions formulated in the preceding level in the hierarchy. A key benefit of Deep Learning is the analysis and learning of massive amounts of unsupervised data, making it a valuable tool for Big Data Analytics where raw data is largely unlabeled and un-categorized. In the present study, we explore how Deep Learning can be utilized for addressing some important problems in Big Data Analytics, including extracting complex patterns from massive volumes of data, semantic indexing, data tagging, fast information retrieval, and simplifying discriminative tasks. We also investigate some aspects of Deep Learning research that need further exploration to incorporate specific challenges introduced by Big Data Analytics, including streaming data, high-dimensional data, scalability of models, and distributed computing. We conclude by presenting insights into relevant future works by posing some questions, including defining data sampling criteria, domain adaptation modeling, defining criteria for obtaining useful data abstractions, improving semantic indexing, semi - supervised learning, and active learning.