FAKE NEWS DETECTION USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

K.PRASANTH 11191205031

N.PRAVEEN 111916205034

S.VIJAY 111916205053

In partial fulfillment for the award of the degree

Of

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

S.A. ENGINEERING COLLEGE

CHENNAI – 600077.

ANNA UNIVERSITY: CHENNAI 600025

SEPTEMBER 2020

ANNA UNIVERSITY: CHENNAI 600025

BONAFIDE CERTIFICATE

Certified that this project report **"FAKE NEWS DETECTION USING MACHINE LEARNING"**is the bonafide work of **"K.PRASANTH(111916205031),N.PRAVEEN(111916205034),S.VIJAY(111916205053)"** who carried out the project work under my supervision.

SIGNATURE

Dr.AHMED MUDASSAR ALI, M.E.,Ph.D.,
HEAD OF THE DEPARTMENT
PROFESSOR
Dept. of Information Technology
S.A Engineering College
Chennai-600 077.

SIGNATURE

Mrs.V.AUXILIA OSVIN NANCY,M.TECH.,
SUPERVISOR
ASSISTANT PROFESSOR
Dept. of Information Technology
S.A Engineering College
Chennai-600 077.

Submitted to Project and Viva-Voce Examination held on ___22.09.2020_____

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We express our gratitude to beloved founder (**Late)Thiru D.SUDHARSSANAM, Founder** and the honorable chairman **Thiru D.DURAISWAMY,** who gave us the opportunity to do this project.

We express our proud thanks and deep sense of gratitude to **Thiru S.AMARNAATH,** Correspondent and We would like to sincerely thank **Thiru D.SABARINATH.,** Director, for the facilities and support provided by them in the college.

We are thankful to **Dr.G.S.KUMARSAMY, M.E.,Ph.D., Principal,** for providing all necessary facilities for undertaking the project.

We express our deep sense of gratitude and heartfelt thanks to **Dr.AHMED MUDASSAR ALI, M.E.,Ph.D., Head of Department of Information Technology** for giving valuable suggestions, expert guidance and encouragement which paved the way for the successful completion of our project work.

We are deeply obliged to our Project Guide **Mrs.V.AUXILIA OSVIN NANCY,(M.TECH) Assistant Professor,** Department of Information Technology for offering his valuable guidelines and suggestions during course work of project.We also extend our sincere thanks to all faculty members and non-teaching staff members of Department of Information Technology, who have given their co-operation and helped us to complete the project successful.

ABSTRACT

Today Fake News has become a major problem wreaking havoc all over the world. Therefore building an algorithm with the best possible accuracy will be a revelation and it will have a massive impact on the social issues which are prevalent as well as on the current political scenario. Our proposed model will be implemented for any domain news. Recently, fake news has been incurring many problems to our society. As a result, many researchers have been working on identifying fake news. Most of the fake news detection systems utilize the linguistic feature of the news. However, they have difficulty in sensing highly ambiguous fake news which can be detected only after identifying meaning and latest related information. In this paper, to resolve this problem, we shall present a new Korean fake news detection system using fact DB which is built and updated by human's direct judgement after collecting obvious facts. Our system receives a proposition, and search the semantically related articles from Fact DB in order to verify whether the given proposition is true or not by comparing the proposition with the related articles in fact DB. To achieve this, we utilize a deep learning model, Bidirectional Multi-Perspective Matching for Natural Language Sentence(BiMPM), which has demonstrated a good performance for the sentence matching task. However, BiMPM has some limitations in that the longer the length of the input sentence is, the lower its performance is, and it has difficulty in making an accurate judgement when an unlearned word or relation between words appear. In order to overcome the limitations, we shall propose a new matching technique which exploits article abstraction as well as entity matching set in addition to BiMPM. In our experiment, we shall show that our system improves the whole performance for fake news detection

TABLE OF CONTENTS

CHAPTER 1

1 INTRODUCTION

Fake news has been incurring many problems to our society. Fake news is the ones that the writer intends to mislead in order to achieve his/her interests politically or economically on purpose. With the generation of a huge volume of internet news and social media. It becomes much more difficult to identify fake news personally.Recently,manresearchershave worked on fake news detection system which automatically determines if any opinion claimed in the article contains fake content . In a large context, the forms of their research are carried out with the method that connects the linguistic pattern of news to deception, and that verifies deception by utilizing external knowledge . The first approach can quickly verify fake news at a low cost. However, in order to detect clever fake news, it is necessary to grasp the semantic content of the article rather than partial patterns and verify it through external facts updated by human. Therefore, we search the input proposition and related articles from the Fact DB, and develop the Korean fake news detection system to verify if the found articles and proposition are semantically related. Our system utilize the Fact DB, a collection of true articles which humans have directly built to detect fake news. The system receives the proposition as an input, and verifies if an article that supports the content exists in the fact DB. To verify that the given article and proposition are semantically similar, we recognize a input proposition and an article as separate sentences, and utilize sentence matching techniques to measure similarity. Through sentence matching, we judge whether two sentences are semantically consensual. Sentence matching has been rapidly developed as a result of success in Natural Language Processing

(NLP) field. Recently, models using the BiDirectional Long Short-Term Memory (BiLSTM) and the attention mechanism in the machine comprehension studies have shown good performance in the sentence matching data set (SNLI) . In our paper, we use BiMPM model to identify the authenticity of the input proposition. In order to train the model, we build a base data set which reflects various word relations and grammaticality in the Korean news sentence. We shall show that BiMPM model works well in the sentence matching of Korean sentences. However, even though BiMPM has shown good performance in various datasets, it has some limitationssuch thatthe longerthe length ofthe inputsentence is, the lower its performance is, and it has difficulty in making an accurate judgement when an unlearned word or relation between words appear. In order to over come the limitations , we shall propose a new matching technique which exploits article abstraction and entity matching set in addition to BiMPM . In our experiment, we shall show that our system improves the whole performance for fake news detection. The outline of our paper is as follows: Chapter 2 describes about machine learning and deep learning for the sentence matching techniques used in our system. Chapter 3 introduces the architecture of our proposed system and explains how we implement each module of the system. Chapter 4 describes about the data set we use, and the performance evaluation of the system. The last chapter gives a conclusion.

1.1 OBJECTIVE

To detect the fake news spread by social media.To achieve the goal, various combination of algorithm will be trained.Fake News detection will be trained for

each domain. So that higher level of accuracy can be achieved.For example separate model for cinema news, political news, sports news and so on.

1.2 MOTIVATION AND SCOPE

Social Media and online news articles serve as a major source of news and for data for people since it can be approached easily , has a subsidized costing and is readily available-just a click away. Hence, we have proposed a new solution for fake news detection which incorporates counter vector and Term Frequency-Inverse Document Frequency(TF- IDF) with Multinomial and Passive Aggressive Classifier as an important feature to improve the accuracy.

CHAPTER 2

2.LITERATURE SURVEY

2.1 INTRODUCTION

The purpose of literature survey is to give the brief overview and also establish complete information about the reference papers. The goal of Literature Survey is to completely specify the technical details related to the main project in a concise and unambiguous manner.

2.2 Supervised Learning for Fake News Detection.

Author: Julio C. S. Reis, Andr_e Correia, Fabr_ıcio Murai, Adriano Veloso, and Fabr_ıcio Benevenuto

A large body of recent works has focused on understanding and detecting fake news stories that are disseminated on socialmedia. To accomplish this goal, theseworks explore several types of features extracted from news stories, including source and posts from socialmedia. In addition to exploring the main features proposed in the literature for fake news detection,wepresent a newset of features and measure the prediction performance of current approaches and features for automatic detection of fake news. Our results reveal interesting findings on the usefulness and importance of features for detecting false news. Finally,we discuss how fake news detection approaches can be used in the practice, highlighting challenges and opportunities. A key problem today is that social media has become

a place for campaigns of misinformation that affect the credibility of the entire news ecosystem. A unique characteristic of news on social media is that anyone can register as a news publisher without any upfront cost (e.g., anyone can create a Facebook page claiming to be a newspaper or news media organization). Consequently, not only traditional news, corporations are increasingly migrating to social. Along with this transition, not surprisingly, there are growing concerns about fake news publishers posting “fake” news stories,and often disseminating them widely using “fake” followers.1 As the extensive spread of fake news can have a serious negative impact on individuals and society, the lack of scalable fact checking strategies is especially worrisome.Not surprisingly, recent research efforts aredevoted not only to better comprehend this phenomenon1 but also to automatize the detection of fake news.

2.3 Fake News Detection on Social Media: A Data Mining Perspective.

Author: Kai Shuy, Amy Slivaz, Suhang Wangy, Jiliang Tang , and Huan Liu

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 ine_ective or not applicable. First, fake news is intentionally written to mislead readers to believe false information, which makes it di_cult 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 producedata 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 future research directions for fake news detection on social media.

2.4 BI-DIRECTIONAL ATTENTION FLOW FOR MACHINE COMPREHENSION.

Author: Minjoon Aniruddha Kembhavi, Ali Farhadi,Hananneh Hajishirzi1

Machine comprehension (MC), answering a query about a given context paragraph, requires modeling complex interactions between the context and the query. Recently, attention mechanisms have been successfully extended to MC. Typically these methods use attention to focus on a small portion of the context and summarize it with a fixed-size vector, couple attentions temporally, and/or

often form a uni-directional attention. In this paper we introduce the Bi-Directional Attention Flow (BIDAF) network, a multi-stage hierarchical process that represents the context at different levels of granularity and uses bidirectional attention flow mechanism to obtain a query-aware context representation without early summarization. Our experimental evaluations show that our model achieves the state-of-the-art results in Stanford Question Answering Dataset (SQuAD) and CNN/DailyMail cloze test. The tasks of machine comprehension (MC) and question answering (QA) have gained significant popularity over the past few years within the natural language processing and computer vision communities. Systems trained end-to-end now achieve promising results on a variety of tasks in the text and image domains. One of the key factors to the advancement has been the use of neural attention mechanism, which enables the system to focus on a targeted area within a context paragraph (for MC) or within an image (for Visual QA), that is most relevant to answer the question (Weston et al., 2015; Antol et al., 2015; Xiong et al., 2016a). Attention mechanisms in previous works typically have one or more of the following characteristics. First, the computed attention weights are often used to extract the most relevant information from the context for answering the question by summarizing the context into a fixed-size vector. Second, in the text domain, they are often temporally dynamic, whereby the attention weights at the current time step are a function of the attended vector at the previous time step. Third, they are usually uni-directional, wherein the query attends on the context paragraph or the image.

.

2.5 Bilateral Multi-Perspective Matching for Natural Language Sentences.

Author: ZhiguoWang,Wael Hamza, Radu Florian

Natural language sentence matching is a fundamentaltechnology for a variety of tasks. Previous approaches either match sentences from a single direction or only apply single granular (wordby- word or sentence-by-sentence) matching. In this work, we propose a bilateral multi-perspective matching (BiMPM) model. Given two sentences P and Q, our model first encodes them with a BiLSTM encoder. Next, we match the two encoded sentences in two directions P against Q and Q against P. In each matching direction, each time step of one sentence is matched against all timesteps of the other sentence from multiple perspectives. Then, another BiLSTM layer is utilized to aggregate the matching results into a fixed-length matching vector. Finally, based on the matching vector, a decision is made through a fully connected layer. We evaluate our model on three tasks: paraphrase identification, natural language inference and answer sentence selection. Experimental results on standard benchmark datasets show that our model achieves the state-of-the-art performance on all tasks.Natural language sentence matching (NLSM) is the task of comparing two sentences and identifying the relationship between them. It is a fundamental technology for a variety of tasks. For example, in a paraphrase identification task, NLSM is used to determine whether two sentences are paraphrase or not [Yin et al., 2015]. For a natural language inference task, NLSM is utilized to judge whether a hypothesis sentence can be inferred from a premise sentence [Bowman et al., 2015]. For question answering and information retrieval tasks, NLSM is employed to assess the relevance between query-answer pairs and rank all the candidate answers [Wang et al., 2016d]. For machine comprehension tasks, NLSM is used for matching a passage with a question and pointing out the correct answer span [Wang et al., 2016b]. With the renaissance of neural network models [LeCun et al., 2015; Peng

et al., 2015a; Peng et al., 2016], two types of deep learning frameworks were proposed for NLSM.

2.6 Unsupervised Learning of Sentence Embeddings using Compositional n-Gram Features.

Author: Matteo Pagliardini Iprova SA, Prakhar Gupta EPFL, Martin Jaggi EPFL.

The recent tremendous success of unsupervised word embeddings in a multitude of application raises the obvious question if similar methods could be derived to improve embeddings (i.e. semantic representations) of word sequences as well. We present a simple but efficient unsupervised objective to train distributed representations of sentences. Our method outperforms the state-of-the-art unsupervised models on most benchmark tasks, highlighting the robustness of the produced general-purpose sentence embeddings. Improving unsupervised learning is of key importance for advancing machine learning methods, as to unlock access to almost unlimited amounts of data to be used as training resources. The majority of recent success stories of deep learning does not fall into this category but instead relied on supervised training (in particular in the vision domain). A very notable exception comes from the text and natural language processing domain, in the form of semantic word embeddings trained unsupervised (Mikolov et al., 2013b,a; Pennington et al., 2014). Within only a few years from their invention, such word representations – which are based on a simple matrix factorization model as we formalize below – are now routinely trained on very large amounts of raw text data, and have become ubiquitous building blocks of a majority of current state-of-the-art NLP applications. While very useful semantic representations are available

for words, it remains challenging to produce and learn such semantic embeddings for longer pieces of text, such as sentences, paragraphs or entire documents.

2.7 Studying Fake News via Network Analysis:Detection and Mitigation.

Author: Kai Shu1, H. Russell Bernard2 and Huan Liu1

Social media is becoming increasingly popular for news consumption due to its easy access, fast dissemination, and low cost. However, social media also enables the wide propagation of “fake news”, i.e., news with intentionally false information. Fake news on social media can have significant negative societal effects. Identifying and mitigating fake news also presents unique challenges. To tackle these challenges, many existing research efforts exploit various features of the data, including network features. In essence, a news dissemination ecosystem involves three dimensions on social media, i.e., a content dimension, a social dimension, and a temporal dimension. In this chapter, we will review network properties for studying fake news, introduce popular network types and propose how these networks can be used to detect and mitigate fake news on social media. Social media has become an important means of large-scale information sharing and communication in all occupations, including marketing, journalism, public relations, and more. The reasons for this change in consumption behaviors are clear: (i) it is often faster and cheaper to consume news on social media compared to news on traditional media, such as newspapers or television; and (ii) it is easier to share, comment on, and discuss the news with friends or other readers on social

media. However, the low cost, easy access, and rapid dissemination of information of social media draws a large audience and enables the wide propagation of “fake news”, i.e., news with intentionally false information. Fake news on social media is growing fast in volume and can have negative societal impacts. First, people may accept deliberate lies as truths; second, fake news can change the way people respond to legitimate news; and finally, the prevalence of fake news has the potential to break the trustworthiness of the entire news ecosystem. In this chapter, we discuss recent advancements–based on a network perspective–for the detection and mitigation of fake news.

2.8 The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News.

Author: Nguyen Vo and Kyumin Lee

A large body of research work and efforts have been focused on detecting fake news and building online fact-check systems in order to debunk fake news as soon as possible. Despite the existence of these systems, fake news is still wildly shared by online users. It indicates that these systems may not be fully utilized. After detecting fake news, what is the next step to stop people from sharing it? How can we improve the utilization of these fact-check systems? To fill this gap, in this paper, we (i) collect and analyze online users called guardians, who correct misinformation and fake news in online discussions by referring fact-checking URLs; and (ii) propose a novel fact-checking URL recommendation model to encourage the guardians to engage more in fact-checking activities. We found that the guardians usually took less than one day to reply to claims in online

conversations and took another day to spread verified information to hundreds of millions of followers. Our proposed recommendation model outperformed four state-of-the-art models by 11%∼33%. Our source code and dataset are available at ACM Reference Format: Nguyen Vo and Kyumin Lee . 2018. The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News. In SIGIR '18: The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, July 8–12, 2018, Ann Arbor, MI, USA. ACM, New York, NY, USA. Fake news, misinformation, rumor or hoaxes are one of the most concerning problems due to their popularity and negative effects on society. Particularly, social networking sites (e.g., Twitter and Facebook) have become a medium to disseminate fake news. Therefore, companies and government agencies have paid attention to solving fake news. For example, Facebook has a plan to combat fake news1 and the FBI has investigated disinformation spread by Russia and other countries2. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee

2.9 Deception for Detection News:Three Types of Fakes.

Author: Victoria L. Rubin, Yimin Chen and Niall J. Conroy

A *fake news detection* system aims to assist users in detecting and filtering out varieties of potentially deceptive news. The prediction of the chances that a particular news item is intentionally deceptive is based on the analysis of previously seen truthful and deceptive news. A scarcity of deceptive news, available as corpora for predictive modeling, is a major stumbling block in this field of natural language processing (NLP) and deception detection. This paper discusses three types of fake news, each in contrast to genuine serious reporting, and weighs their pros and cons as a corpus for text analytics and predictive modeling. Filtering, vetting, and verifying online information continues to be essential in library and information science (LIS), as the lines between traditional news and online information are blurring.The boundaries between news production and information creation and sharing are gradually blurring in the current online news environments and social media (Chen, Conroy, & Rubin, 2015). Fleeting news becomes part of the static searchable information realm. Broad audiences gain access to news through traditional search engines, digital forms of mainstream news channels and citizen information sharing platforms. While the issue of news verification and fact-checking has traditionally been a matter of journalistic endeavor, LIS offers meaningful insights on credibility assessment and cutting-edge information technologies. As LIS is redefining itself in the age of fast-streaming big data and text analytics, it is turning to automated methods for filtering, vetting, and verifying online information. Deceptive news, such as fake news, phony press releases and hoaxes, may be misleading or even harmful, especially when they are disconnected from their original sources and contexts. “A 2012 report of Pew Internet Research on the future of big data (Anderson and Rainie, 2012) argues that even though by 2020 big data is likely to have a transformational effect on our knowledge and understanding of the world, there is also danger from inaccurate or false information (called “distribution of harms”

(Rumors and Deception in Social Media Workshop, 2015). Occasionally reports of non-existent, surreal, alarming events have been taken seriously. For instance, "Jack Warner, the former FIFA vice president, has apparently been taken in by a spoof article from the satirical website *The Onion*" (Topping, 2015) after *The Onion* had suggested that the FIFA corruption scandal would result in a 2015 Summer Cup in the U.S. *Journalistic deception* is "an act of communicating messages verbally (a lie) or nonverbally through the withholding of information with the intention to initiate or sustain a false belief" (Elliot and Culver, 1992, cited in Lee, 2004, p. 98). Cases of outright fabrications by celebrity hosts or journalists have been recently reported (Compton & Benedetti, 2015; Shingler, 2015). The journalism community observes that the responsibility for knowing what is true shifts to individuals, as news consumers, with two predominant overwrought views: "Utopians have heralded this as the end of journalism and the information monopoly of elites and see a citizen media culture that instantly self-corrects – a kind of pure information democracy. Critics see a world without editors, of unfettered spin, where the loudest or most agreeable voice wins and where truth is the first casualty" (Kovach & Rosentiel, 2010, p. 7). With distinctions between genuine and misleading news eroding, "few news verification mechanisms currently exist, and the sheer volume of the information requires novel automated approaches" (Rubin, Conroy, & Chen, 2015).

2.0.1 Separating Facts from Fiction: Linguistic Models to Classify Suspiciousand Trusted News Posts on Twitter.

Author:Svitlana Volkova, Kyle Shaffer, Jin Yea Jang and Nathan Hodas

Pew research polls report 62 percent of U.S. adults get news on social media (Gottfried and Shearer, 2016). In a December poll, 64 percent of U.S. adults said that "made-up news" has caused a "great deal of confusion" about the facts of current events (Barthel et al., 2016). Fabricated stories in social media, ranging from deliberate propaganda to hoaxes and satire, contributes to this confusion in addition to having serious effects on global stability. In this work we build predictive models to classify 130 thousand news posts as suspicious or verified, and predict four subtypes of suspicious news – satire, hoaxes, clickbait and propaganda. We show that neural network models trained on tweet content and social network interactions outperform lexical models. Unlike previous work on deception detection, we find that adding syntax and grammar features to our models does not improve performance. Incorporating linguistic features improves classification results, however, social interaction features are most informative for finer-grained separation between four types of suspicious news posts. Popular social media platforms such as Twitter and Facebook have proven to be effective channels for disseminating falsified information, unverified claims, and fabricated attention-grabbing stories due to their wide reach and the speed at which this information can be shared. Recently, there has been an increased number of disturbing incidents of fabricated stories proliferated through social media having a serious impact on real-world events (Perrott, 2016; Connolly et al., 2016) False news stories distributed on social media vary depending on the intent behind falsification. Unlike verified news, suspicious news tends to build narratives rather than report facts. On one extreme is disinformation which communicates false facts to deliberately deceive readers or promote a biased agenda.

2.1.1 Fake News Detection Using Sentiment Analysis.

Author: Bhavika Bhutani Neha Rastogi Priyanshu Sehgal ArchanPurwar

Social media is one of the most revolutionary inventions of the present times. With its own set of advantages and disadvantages it is extremely essential for each one of us. Today Fake News has become a major problem wreaking havoc all over the world. Therefore building an algorithm with the best possible accuracy will be a revelation and it will have a massive impact on the social issues which are prevalent as well as on the current political scenario. Social Media and online news articles serve as a major source of news and for data for people since it can be approached easily , has a subsidized costing and is readily available-just a click away. However, it does have several negative impacts too such as no check on the source or authenticity and validity of the views being endorsed. Hence, we have proposed a new solution for fake news detection which incorporates sentiment as an important feature to improve the accuracy. It also investigates the performance of proposed method using three different data sets. Results show that proposed solution performs well. Moreover, the comparison is also made with other methods under this study. The information that we procure from the Internet and the Web cannot be relied on. Over the past few years spreading of rumors and fake information has reached a tipping point so much so that it has begun to affect social issues and political problems as well. The amount of time spent on social media platforms and online news sites has increased at an alarming rate. Thus most of the information that they have is acquired from these sources. Although social media can be accessed from anywhere and at any time and is also free but it provides anonymity while expressing out opinion therefore leading to a lack of accountability which greatly reduces the authenticity of data received from them as compared to a newspaper or any other trusted news source. Lack of constant

supervision and an overseeing authority has allowed the wrongdoers to run amok and spread false information. Fake news is the deliberately falsified news which is sent out to fool people and make them believe in otherwise false information. Misusing the news being broadcasted to various consumers can only result in confusion and chaos because various versions of the truth would be present. This disinformation may be spread in order to re-establish popularity or just for fun. In either case we need to figure out an effective way to identify fake and falsified information and prevent it from spreading further. clickbait ("eye-catching" headlines) accounts. The intent behind propaganda and clickbait varies from opinion manipulation and attention redirection to monetization and traffic attraction. Hoaxes are another type of disinformation that aims to deliberately deceive the reader (Tambuscio et al., 2015; Kumar et al., 2016). On the other extreme is satire, e.g., @TheOnion, where the writer's primary purpose is not to mislead the reader, but rather entertain or criticize (Conroy et al., 2015). However, satirical news and hoaxes may also be harmful, especially when they are shared out of context (Rubin et al., 2015).

CHAPTER 3

3. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Existing system receives a proposition, and search the semantically related articles from Fact DB in order to verify whether the given proposition is true or not by comparing the proposition with the related articles in fact DB.To achieve this, we utilize a deep learning model, Bidirectional Multi Perspective Matching for Natural Language Sentence(BiMPM), which has demonstrated a good performance for the sentence matching task. However, even though BiMPM has shown good performance in various datasets, it has some limitations such that the longer the length of the input sentence is, the lower its performance.

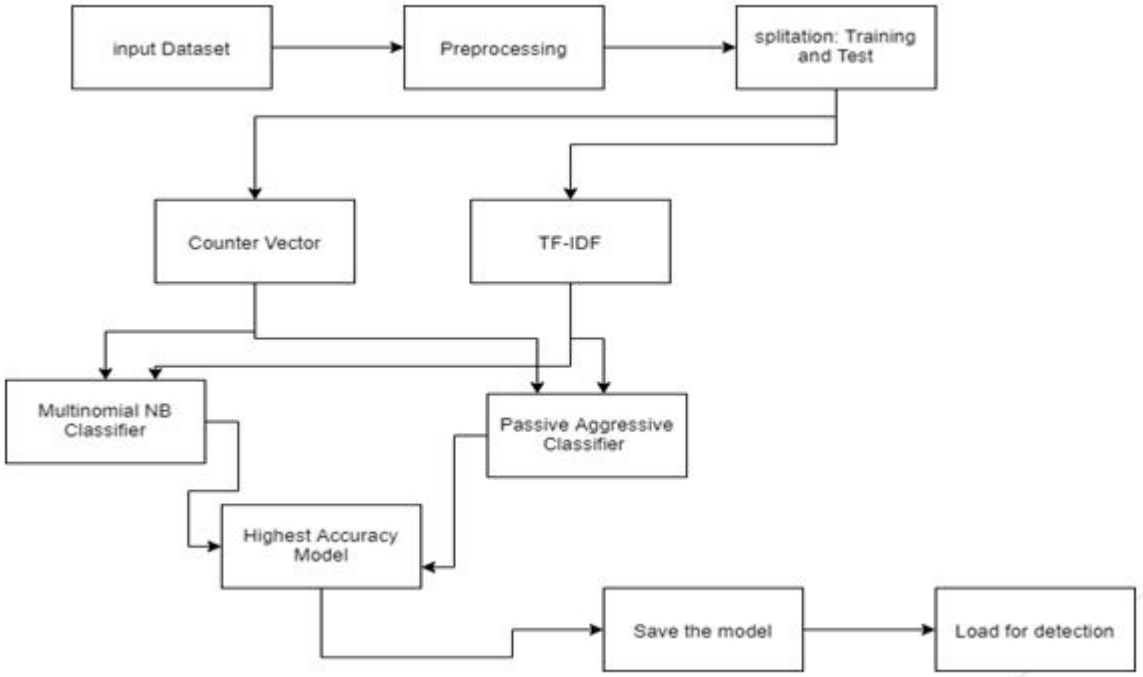
3.2 PROPOSED SYSTEM

Proposed system classify the fake and real news using Machine Learning algorithm.It incorporates counter Vector and Term Frequency-inverse Document Frequency with MultinomialNB and Passive Aggressive Classifier.The Model which gives highest accuracy will implemented in the detecting the fake news.Proposed system which incorporates counter vector and Term Frequency-Inverse Document Frequency(TF- IDF) with Multinomial and Passive Aggressive Classifier.Existing system has a draw back for longer sentence and it showed only 78 percentage of accuracy.Our proposed will work on for longer sentence to classify as well as to improve the accuracy

.

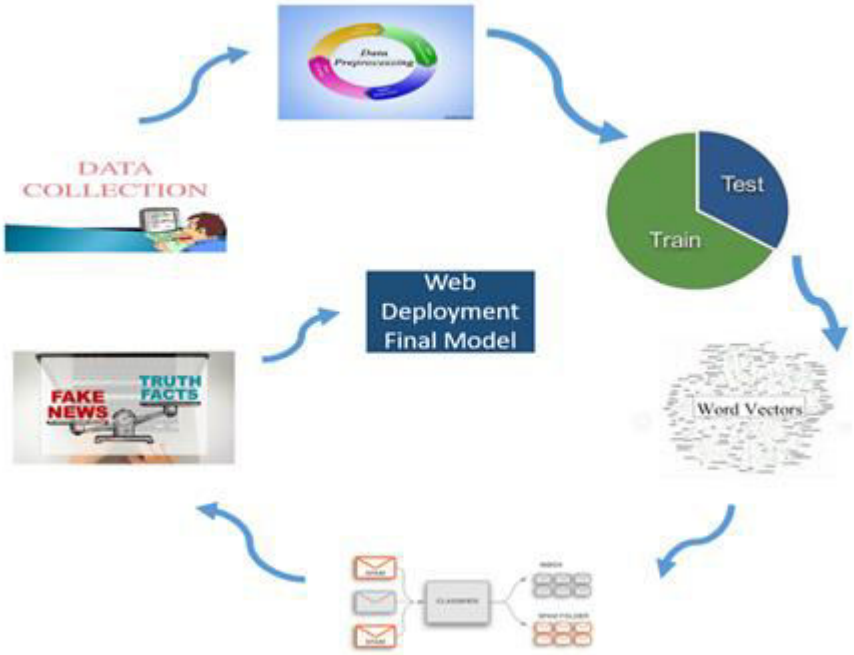
3.3 SYSTEM DESIGN

SYSTEM DESIGN



3.3.1

SYSTEM ARCHITECTURE



3.4 METHODOLOGY

Data Collection , the data-acquisition process, “fake news” and “real news”. Collecting the fake news was easy as Kaggle released a fake news dataset consisting of 13,000 articles published during the 2016 election cycle. Now the later part is very difficult. That is to get the real news for the fake news dataset. It requires huge work around many Sites because it was the only way to do web scraping thousands of articles from numerous websites. With the help of web scraping a total of 5279 articles, real news dataset was generated, mostly from media organizations (New York Times, WSJ, Bloomberg, NPR, and the Guardian) which were published around 2015 – 2016. Collected data will be pre-processed which means encoding the categorical information in the data. Dropping unwanted parameters, scaling the parameter values to achieve normal distribution (Zero mean and Standard Deviation as one), handling missing values and so on. In Machine

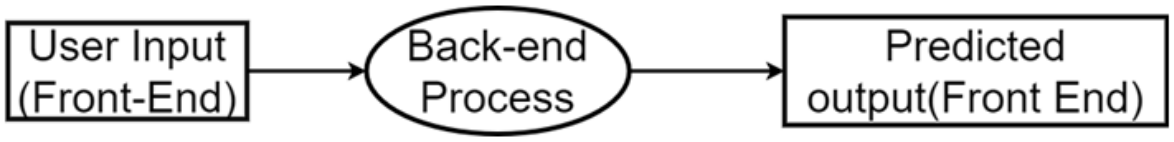
Learning validation confirms the accuracy of the model. To test the model using training dataset value gives the accuracy of the model . so it is necessary to split the Training and Test Dataset. Splitting Ratio may be 8: 2. Word Embedding 's or Word Vectorization is a methodology in NLP to map words or phrases from vocabulary to a corresponding vector of real numbers which used to find word predictions, word similarities/semantics. Compute similar words Text classifications. Document clustering/grouping . Feature extraction for text classifications . Text classification also known as *text tagging* or *text categorization* is the process of categorizing text into organized groups. By using Natural Language Processing (NLP), text classifiers can automatically analyze text and then assign a set of pre-defined tags or categories based on its content. Text classification is becoming an increasingly important part of businesses as it allows to easily get insights from data and automate business processes. Some of the most common examples and use cases for automatic text classification include the following: Topic Modelling , Sentimental Analysis . In the Proposed system have used Multinomial And Passive Aggressive Classifier with different combination of Counter vector and TF-IDF . The model with best accuracy will be implemented for the web deployment . Web development will have the input text field ,in which fake news will be provide . After Submit button backend model takes the input and predict with the model. The result will be thrown back to front end , the news is fake or real . In addition to this planned to display the learned keywords with their respective weight for both fake and real news.

CHAPTER 4

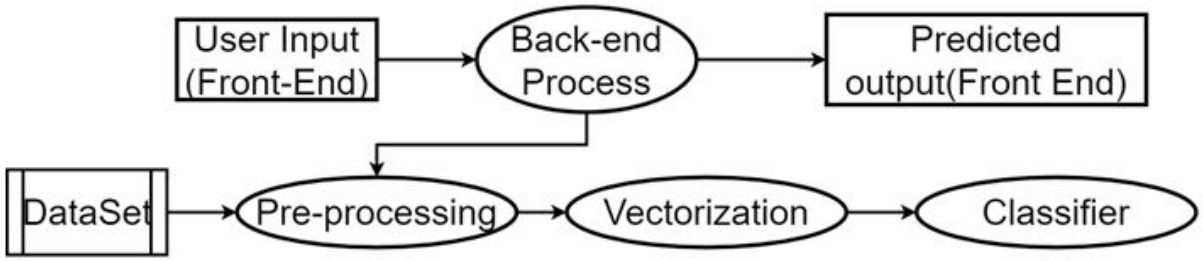
4. UML DIAGRAMS

4.1 DATA FLOW DIAGRAM

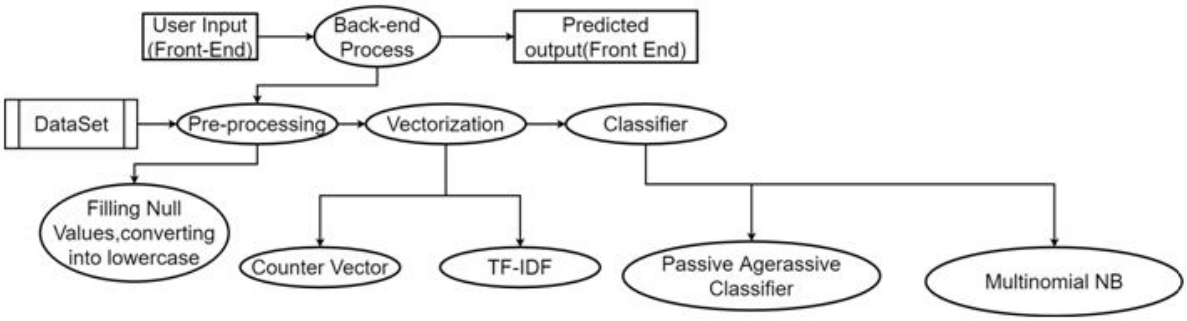
DATA FLOW DIAGRAM:LEVEL 0



DATA FLOW DIAGRAM:LEVEL 1

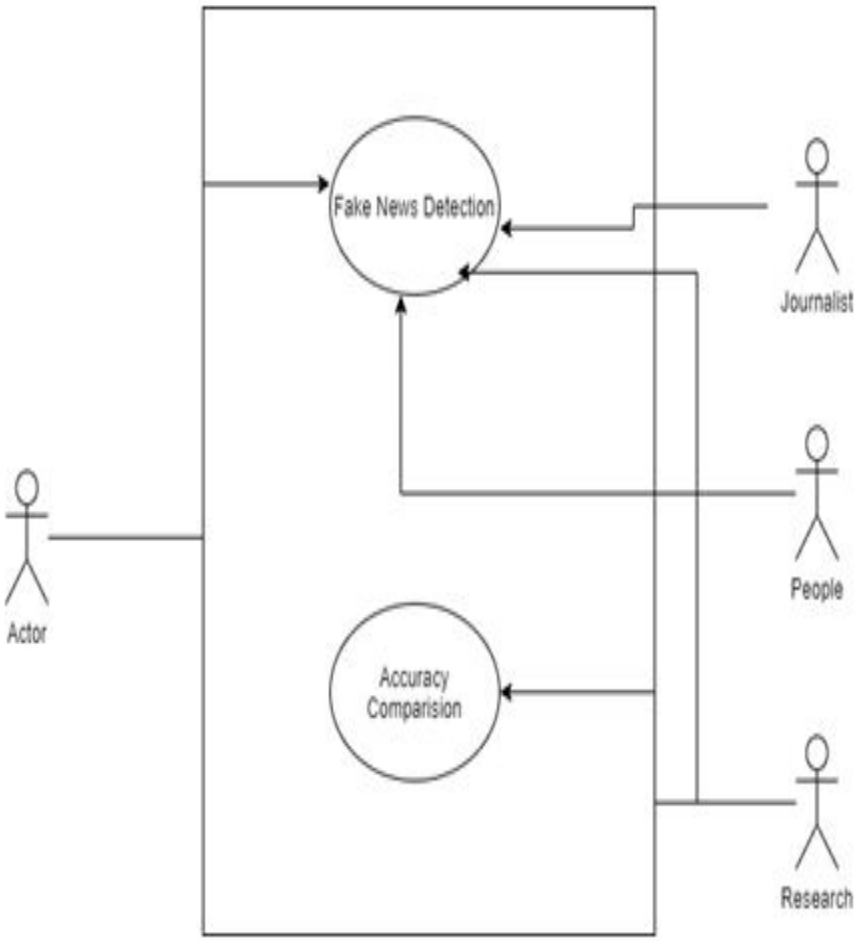


DATA FLOW DIAGRAM:LEVEL 2

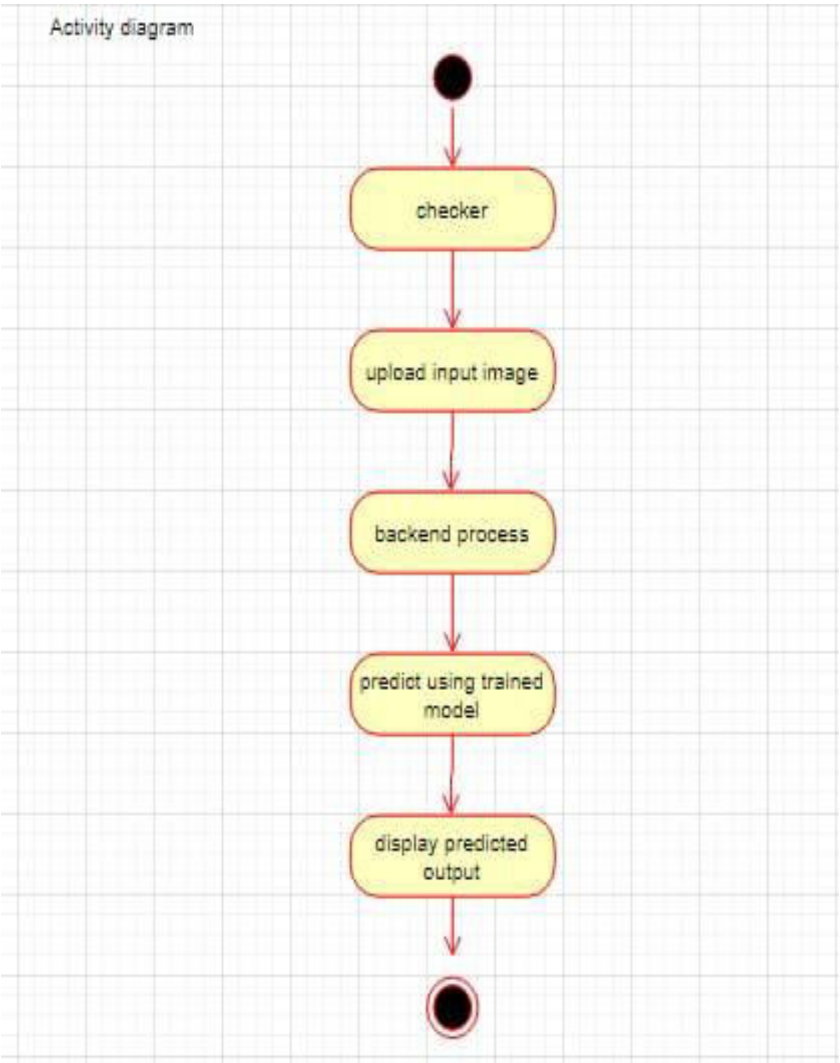


4.2ARCHITECTURE DIAGRAM

USE CASE DIAGRAM



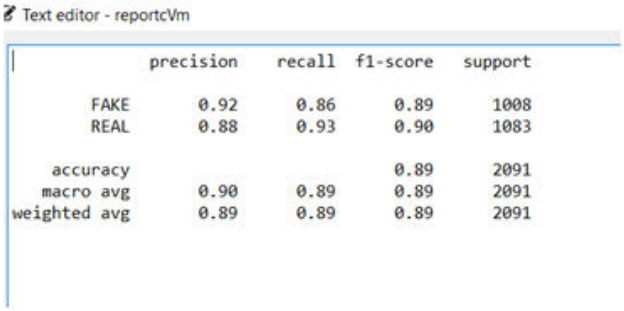
4.3 ACTIVITY DIAGRAM


Activity diagram
checker
upload input image
backend process
predict using trained model
display predicted output

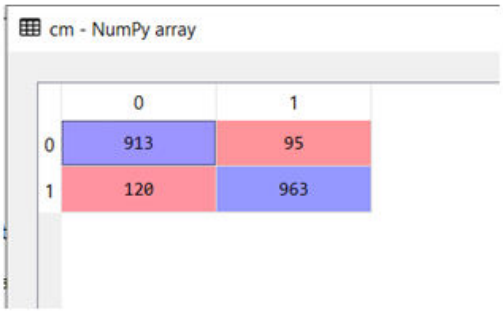
CHAPTER 5

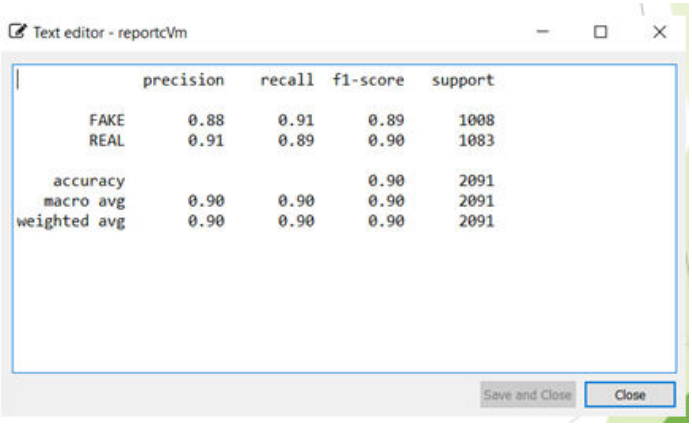
5. RESULT

5.1 Output Screen shot: Counter Vector + MultinomialNB



5.2 Output Screen shot: Counter Vector + Passive Aggressive Classifier

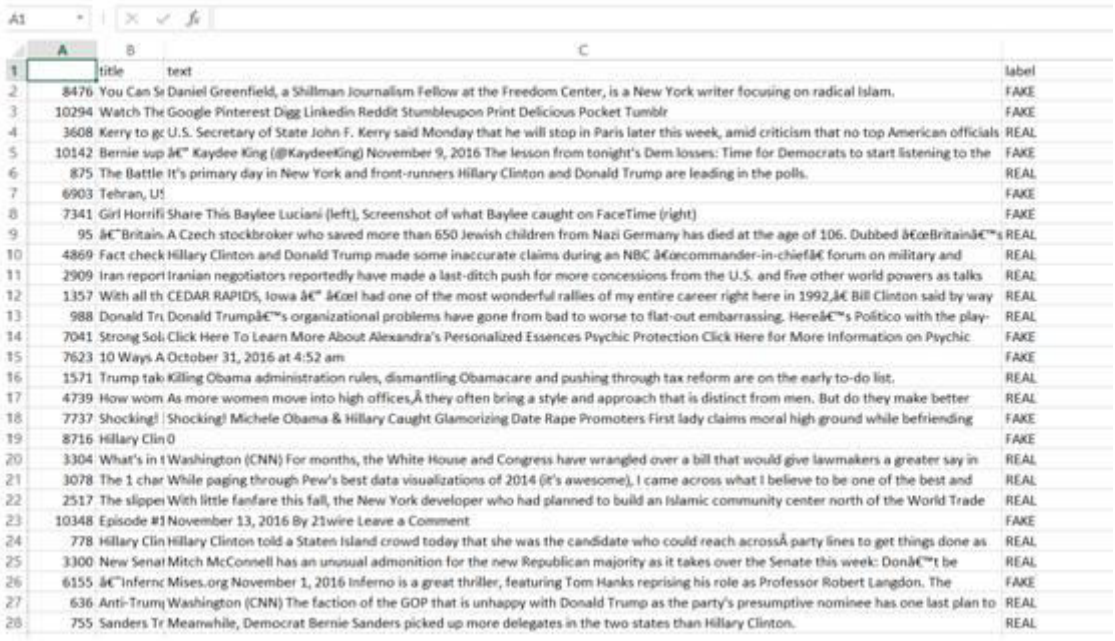




5.3 COMPARISION TABLE

Accuracy	Counter Vector	TF-IDF
MultinomialNB Classifier	0.89	0.85
Passive Aggressive Classifier	0.89	0.93

5.4 RAW DATASET



	A	B	C	
1		title	text	label
2	8476	You Can Si	Daniel Greenfield, a Shillman Journalism Fellow at the Freedom Center, is a New York writer focusing on radical Islam.	FAKE
3	10294	Watch The	Google Pinterest Digg Linkedin Reddit Stumbleupon Print Delicious Pocket Tumblr	FAKE
4	3608	Kerry to go	U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that no top American officials	REAL
5	10142	Bernie sup	â€" Kaydee King (@KaydeeKing) November 9, 2016 The lesson from tonight's Dem losses: Time for Democrats to start listening to the	FAKE
6	875	The Battle	It's primary day in New York and front-runners Hillary Clinton and Donald Trump are leading in the polls.	REAL
7	6903	Tehran, U		FAKE
8	7341	Girl Horrifi	Share This Baylee Luciani (left), Screenshot of what Baylee caught on FaceTime (right)	FAKE
9	95	â€"Britain	A Czech stockbroker who saved more than 650 Jewish children from Nazi Germany has died at the age of 106. Dubbed â€œBritainâ€™s	REAL
10	4869	Fact check	Hillary Clinton and Donald Trump made some inaccurate claims during an NBC â€œcommander-in-chiefâ€ forum on military and	REAL
11	2909	Iran report	Iranian negotiators reportedly have made a last-ditch push for more concessions from the U.S. and five other world powers as talks	REAL
12	1357	With all th	CEDAR RAPIDS, Iowa â€" â€œI had one of the most wonderful rallies of my entire career right here in 1992,â€ Bill Clinton said by way	REAL
13	988	Donald Tr	Donald Trumpâ€™s organizational problems have gone from bad to worse to flat-out embarrassing. Hereâ€™s Politico with the play-	REAL
14	7041	Strong Sol	Click Here To Learn More About Alexandra's Personalized Essences Psychic Protection Click Here for More Information on Psychic	FAKE
15	7623	10 Ways A	October 31, 2016 at 4:52 am	FAKE
16	1571	Trump tak	Killing Obama administration rules, dismantling Obamacare and pushing through tax reform are on the early to-do list.	REAL
17	4739	How wom	As more women move into high offices,Â they often bring a style and approach that is distinct from men. But do they make better	REAL
18	7737	Shocking!	Shocking! Michele Obama & Hillary Caught Glamorizing Date Rape Promoters First lady claims moral high ground while befriending	FAKE
19	8716	Hillary Clin	0	FAKE
20	3304	What's in t	Washington (CNN) For months, the White House and Congress have wrangled over a bill that would give lawmakers a greater say in	REAL
21	3078	The 1 char	While paging through Pew's best data visualizations of 2014 (it's awesome), I came across what I believe to be one of the best and	REAL
22	2517	The slippe	With little fanfare this fall, the New York developer who had planned to build an Islamic community center north of the World Trade	REAL
23	10348	Episode #1	November 13, 2016 By 21wire Leave a Comment	FAKE
24	778	Hillary Clin	Hillary Clinton told a Staten Island crowd today that she was the candidate who could reach acrossÂ party lines to get things done as	REAL
25	3300	New Senat	Mitch McConnell has an unusual admonition for the new Republican majority as it takes over the Senate this week: Donâ€™t be	REAL
26	6155	â€"Inferno	Mises.org November 1, 2016 Inferno is a great thriller, featuring Tom Hanks reprising his role as Professor Robert Langdon. The	FAKE
27	636	Anti-Trum	Washington (CNN) The faction of the GOP that is unhappy with Donald Trump as the party's presumptive nominee has one last plan to	REAL
28	755	Sanders Tr	Meanwhile, Democrat Bernie Sanders picked up more delegates in the two states than Hillary Clinton.	REAL

5.5 INPUT UI


Home
Fake News detection
Dropdown link
Welcome!!
Please fill the respective field
Fakenews*
When President Obama defended the right of developers to build the project, he was â€" surprise, surprise â€" accused of being out of touch, an
submit

CHAPTER 6

6. OUTPUT

6.1 OUTPUT UI:FAKE

.

6.2 OUTPUT:UI REAL

6.3 CODE SCREEN SHOT PRE-PROCESSING

```
import pandas as pd
import numpy as np
import itertools
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer, HashingVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn import metrics
import matplotlib.pyplot as plt
df=pd.read_csv('news.csv', index_col=None)
#Get shape and head
df.shape
df.head()
df = df.set_index('Unnamed: 0')
#DataFlair - Get the labels
labels=df.label
labels.head()


y = df.label

df = df.drop('label', axis=1)

X_train, X_test, y_train, y_test = train_test_split(df['text'], y, test_size=0.33, random_state=53)
```

6.4 VECTORIZATION:COUNTER VECTOR

```
count_vectorizer = CountVectorizer(stop_words='english')
count_train = count_vectorizer.fit_transform(X_train)
print(count_train)
count_test = count_vectorizer.transform(X_test)
```

6.5 CODESCREEN SHOT:COUNTER VECTOR + PASSIVE AGGRESSIVE CLASSIFIER

```
linear_clf_count = PassiveAggressiveClassifier(max_iter=50)
linear_clf_count.fit(count_train, y_train)
pred = linear_clf_count.predict(count_test)
score = metrics.accuracy_score(y_test, pred)
reportcVm=metrics.classification_report(y_test, pred)
print("accuracy:   %0.3f" % score)
cm = metrics.confusion_matrix(y_test, pred, labels=['FAKE', 'REAL'])
```

6.6 LEARNED FEAUTER NAMES FOR FAKE LABEL

Feature Names of the model
For the label FAKE

Sl.No	Feature Value	Feature Name
0	-4.862391226960536	2016
1	-4.304651555598096	october
2	-4.154606815472189	hillary
3	-3.2451802744076557	share
4	-3.0235793912872957	article
5	-2.9325582809937116	november
6	-2.698546986795109	print
7	-2.4923993641533757	oct
8	-2.4332185507951474	email
9	-2.344951348670061	source

6.7 LEARNED FEAUTER NAMES FOR REAL LABEL

For the label REAL

Sl.No	Feature Value	Feature Name
0	1.6772482548292398	saturday
1	1.6822636319685351	conservatives
2	1.6948102442719937	deal
3	1.6989531498836563	sen
4	1.7034403934045221	polarization
5	1.7530839062743413	recounts
6	1.7606928392544126	campaign
7	1.8127241132670309	debate
8	1.8167688314323351	week
9	1.8225036906822334	paris

CHAPTER 7

7.1CONCLUSION:

The proposed system, classify the given news is fake or real. We have used the political news which was published by UCI Machine Learning Respiratory. We have compared with two Vectorization and two classifier, total combination accuracy of four . Tf-idf and Counter Vector are used for Vectorization . Passive Aggressive classifier and Multinomial Naive Bays Classifier . Better accuracy model was deployed in the web development . Same procedure will be used for all the domain wise fake news detection.

REFERNCE:

1) Supervised Learning for Fake News Detection**.**
Author**:**Julio C. S. Reis, Andr_e Correia, Fabr_ıcio Murai, Adriano Veloso, and Fabr_ıcio Benevenuto

2) Fake News Detection on Social Media: A Data Mining Perspective.
Author**:**Kai Shuy, Amy Slivaz, Suhang Wangy, Jiliang Tang , and Huan Liu

3) Bi-directional attentional flow for machine.
Author:Minjoon Aniruddha Kembhavi, Ali Farhadi,Hananneh Hajishirzi1

4) Bilateral Multi-Perspective Matching for Natural Language Sentences.
Author:ZhiguoWang,Wael Hamza, Radu Florian

5) Unsupervised Learning of Sentence Embeddings using Compositional n-Gram Features.
Author**:**Matteo Pagliardini Iprova SA, Prakhar Gupta EPFL, Martin Jaggi EPFL**.**

6) Studying Fake News via Network Analysis:Detection and Mitigation.
Author:Kai Shu1, H. Russell Bernard2 and Huan Liu1

7) The Rise of Guardians: Fact-checking URL Recommendation
to Combat Fake News.
Author**:**Nguyen Vo and Kyumin Lee

8) Deception for DetectionforNews:Three Types of Fakes.

Author:Victoria L. Rubin, Yimin Chen and Niall J. Conroy

9) Separating Facts from Fiction: Linguistic Models to Classify Suspiciousand Trusted News Posts on Twitter.

Author:Svitlana Volkova, Kyle Shaffer, Jin Yea Jang and Nathan Hodas

10) Fake News Detection Using Sentiment Analysis.

Author:Bhavika Bhutani Neha Rastogi Priyanshu Sehgal ArchanPurwar

R.M.K

COLLEGE OF ENGINEERING AND TECHNOLOGY

R.S.M. Nagar, Puduvoyal - 601 206.

EIGHTH NATIONAL CONFERENCE ON COMPUTING TECHNOLOGIES TODAYS AND BEYOND (NCCTTB'2020)

CERTIFICATE

Certified that Mr/~~Ms~~ Prasanth. K

from S. A. Engineering College has presented the paper

titled Fake News Detection Using Machine Learning

in the National Conference on Computing Technologies Todays and Beyond (NCCTTB'20) Organized by the Department of Computer Science and Engineering held on 17th Feb 2020 at RMK College of Engineering and Technology.

COORDINATOR HOD PRINCIPAL

R.M.K

COLLEGE OF ENGINEERING AND TECHNOLOGY

R.S.M. Nagar, Puduvoyal - 601 206

EIGHTH NATIONAL CONFERENCE ON
COMPUTING TECHNOLOGIES TODAYS AND BEYOND
(NCCTTB'2020)

CERTIFICATE

Certified that Mr/~~Ms~~ Praveen N

from S.A. Engineering College has presented the paper titled Fake News Detection Using Machine Learning in the National Conference on Computing Technologies Todays and Beyond (NCCTTB'20) Organized by the Department of Computer Science and Engineering held on 17th Feb 2020 at RMK College of Engineering and Technology.

ORDINATOR

HOD

PRINCIPAL

R.M.K

COLLEGE OF ENGINEERING AND TECHNOLOGY

R.S.M. Nagar, Puduvoyal - 601 206.

EIGHTH NATIONAL CONFERENCE ON COMPUTING TECHNOLOGIES TODAYS AND BEYOND (NCCTTB'2020)

CERTIFICATE

Certified that Mr/~~Ms~~ Vijay.S

from S.A. Engineering College has presented the paper

titled Fake News Detection Using Machine Learning

in the National Conference on Computing Technologies Todays and Beyond (NCCTTB'20) Organized by the Department of Computer Science and Engineering held on 17th Feb 2020 at RMK College of Engineering and Technology.

COORDINATOR HOD PRINCIPAL

International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 2, February 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Fake News Detection using Machine Learning

Prasanth. K[1], Praveen. N[1], Vijay. S[1], Auxilia Osvin Nancy. V[2]

[1]B.Tech Student, IT, SAEC, Chennai, Tamil Nadu, India
[2]Assistant Professor, IT, SAEC, Chennai, Tamil Nadu, India

ABSTRACT

Recently, fake news has been incurring many problems to our society. As a result, many researchers have been working on identifying fake news. Most of the fake news detection systems utilize the linguistic feature of the news. However, they have difficulty in sensing highly ambiguous fake news which can be detected only after identifying meaning and latest related information. In this paper, to resolve this problem, we shall present a new Korean fake news detection system using fact DB which is built and updated by human's direct judgement after collecting obvious facts. Our system receives a proposition, and search the semantically related articles from Fact DB in order to verify whether the given proposition is true or not by comparing the proposition with the related articles in fact DB. To achieve this, we utilize a deep learning model, Bidirectional Multi-Perspective Matching for Natural Language Sentence(BiMPM), which has demonstrated a good performance for the sentence matching task. However, BiMPM has some limitations in that the longer the length of the input sentence is, the lower its performance is, and it has difficulty in making an accurate judgement when an unlearned word or relation between words appear. In order to overcome the limitations, we shall propose a new matching technique which exploits article abstraction as well as entity matching set in addition to BiMPM. In our experiment, we shall show that our system improves the whole performance for fake news detection.

KEYWORDS: *Fake news detection, Sentence matching, Natural Language Processing, Deep learning, BiLSTM model, Machine Learning*

How to cite this paper: Prasanth. K | Praveen. N | Vijay. S | Auxilia Osvin Nancy. V "Fake News Detection using Machine Learning" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-2, February 2020, pp.512-514, URL: www.ijtsrd.com/papers/ijtsrd30014.pdf

IJTSRD30014

INRODUCTION

Fake news has been incurring many problems to our society. Fake news is the ones that the writer intends to mislead in order to achieve his/her interests politically or economically on purpose [1]. With the generation of a huge volume of internet news and social media. It becomes much more difficult to identify fake news personally. Recently, many researchers have worked on fake news detection system which automatically determines if any opinion claimed in the article contains fake content [2]. In a large context, the forms of their research are carried out with the method that connects the linguistic pattern of news to deception, and that verifies deception by utilizing external knowledge [3]. The first approach can quickly verify fake news at a low cost. However, in order to detect clever fake news, it is necessary to grasp the semantic content of the article rather than partial patterns and verify it through external facts updated by human. Therefore, we search the in put proposition and related articles from the Fact DB, and develop the fake news detection system to verify if the found articles and proposition are semantically related.

Related Works:

Recently, as the deep learning in the NLP field has been developed, various types of the sentence matching techniques have been introduced. We introduce the related research of the sentence matching techniques as we divide the works into the unsupervised learning, and supervised learning based works.

A. Unsupervised Learning:

One of the most important elements in the sentence matching is the way of expressing a word into a data structure. The existing method of expressing words is one-hot encoding vector. However, this method requires lots of dimension to express a single word, and cannot express the relation between words. Overcoming these shortcomings, the word-to-vector (word2vec) [5] method was proposed which maps significant information into the vector of fixed dimensions. The word2vec is enabled to learn the weight to increase the probability that the nearby words will appear for the main word, and uses the corresponding weight as a vector. As an extended research of word2vec, sentence-to-vector (sent2vec).

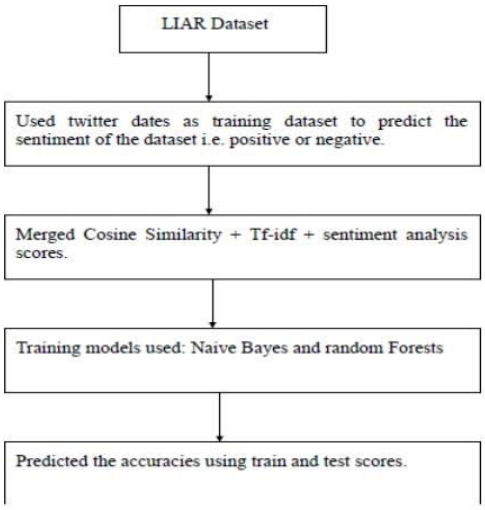
B. Supervised Learning:

Recently, the research of the machine comprehension is developed with attention mechanism and BiLSTM. LSTM resolved the vanishing gradient problem of Recursive Neural Network (RNN) by adding the layer that forgets the past information, and remembers the current one to the cell. Since LSTM handles sequential inputs, it is often used for encoding and decoding of sentences. However, as the length of LSTM becomes longer, the model loses the information, and it shows a tendency to remember the latter information. Therefore, scholars worked on improving the performance of the existing LSTM through attention mechanism which reminds important information selectively. They also showed using BiLSTM together can improve performance. The Bi-Directional Attention Flow (BiDAF) [7] minimizes the

loss of information by applying the attention mechanism at each time stamp of LSTM. In particular, BiMPM [8] applied BiLSTM and attention mechanism to sentence matching.

Proposed System:

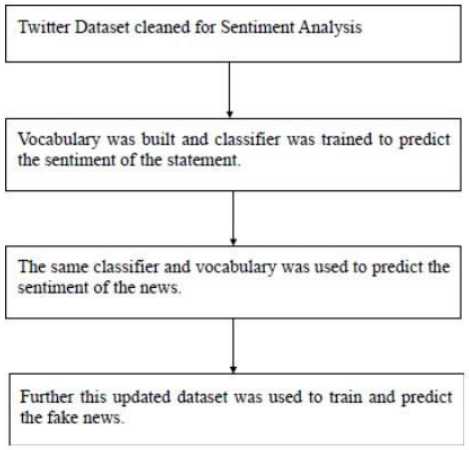
This section discusses the proposed solution for fake news detection by combining Fake News with Sentiment Analysis. Proposed solution is shown in Fig. 1. It consists of various steps as below:



Step 1 : Merged data set was prepared using from different data sets namely Politifact, Kaggle and Emergent datasets.

Step 2 : The different text preprocessing techniques like bigrams (series of two words taken from a given text) ,trigrams (continuous series of three words taken from example text), CountVectorizer (count of terms in vector/ text , term frequency-inverse document frequency (tf-idf) vectorizer.

Step 3 : We have used tf-idf vectorizer on twitter dataset along with cosine similarity to build our vocabulary. Then Naive Bayes classifier was used to predict the sentiment of news statement of test data set (Merged data set) as shown in Fig. 2.



Step 4 : We added additional columns: tf-idf scores, sentiments and Cosine similarity scores in Merged data set.

Step 5 : Training model was built using Naive Bayes and Random Forest (train-test ratio: 3:1)

Step 6 : Performance is evaluated and compared using accuracy.

Proposed solution consists of important steps 2 to 4 as preprocessing. It uses tf-idf Vectorizer with cosine similarities method for tokenizing a collection of text documents along with building a vocabulary of pre-existing words. Further we encoded the novel documents using that vocabulary. The encoded vector is returned with length of the entire vocabulary (bag of words) and an integer count for the number of times each word appeared had in the document.

System Evaluation:

We evaluate the performance of proposed system in this section. Given the relevant article on the input proposition, the evaluation verifies the ability to determine whether the semantic content of the input proposition can be found in the relevant article. We train the BiMPM which is the foundation of our system, and the experiments identify how much the performance improves by adding modules proposed previously. We first build the data set directly to train the BiMPM to output true or false when given a short article consisting of three or four sentences and propositions. In the datasets construction, the following policy is set up to proceed with the learning.

1. Extract one sentence from a short article, and use it into an input proposition.
2. Generate the data, which is true through variations of thesaurus, a change of word orders, and omission of some contents.
3. Distort some information such as numbers, nouns, and verbs or omit words to generate false data.

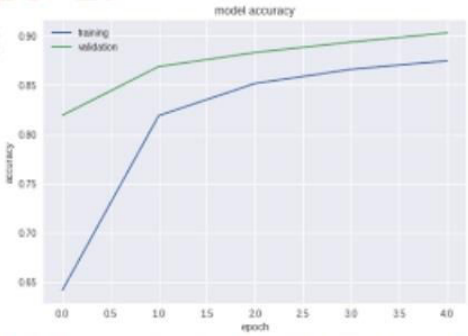


Fig.2. Training accuracy of test set for each epoch.

Proved the sentences used in the new test set are longer, and consist of new words that are not in the previous dataset. In terms of the using True Positive Rate (TPR) as the y-axis and False Positive Rate (FPR) as the x-axis.

Conclusion:

In this paper, we have proposed the fake news detection system using Machine learning which is built and updated by human's direct judgement. Our system receives a

proposition as an input to verify, and search the related articles so that it verifies if the article found by the entered proposition can be semantically concurred with the proposition. To achieve this, we utilized model which is a deep learning model for sentences matching and machine learning. However, even though has shown good performance in various datasets, it has some limitations such that the longer the length of the input sentence is, the lower its performance is, and it has difficulty in making an accurate judgement when an unlearned word or relation between words appear. In order to overcome the limitations, we have presented the new matching technique which makes use of article abstraction as well as entity matching set besides BiMPM. In our experiment, we have shown that our system improved the whole performance for fake news detection.

Reference:

[1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of economic perspectives, vol. 31, no. 2, pp. 211–36, 2017.

[2] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," in Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, p. 82, American Society for Information Science, 2015.

[3] V. L. Rubin, Y. Chen, and N. J. Conroy, "Deception detection for news:three types of fakes," in Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, p. 83, American Society for Information Science, 2015.

[4] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, "A large annotated corpus for learning natural language inference," arXiv preprintarXiv: 1508.05326, 2015.

[5] M. Potthast, S. K¨opsel, B. Stein, and M. Hagen, "Click bait detection," in Advances in Information Retrieval. Springer, 2016.

[6] M. Pagliardini, P. Gupta, and M. Jaggi, "Unsupervised learning of sentence embeddings using compositional n-gram features," arXiv preprintarXiv: 1703.02507, 2017.

[7] V. Piek et al., "Newsreader: How semantic web helps natural language Recognizing click bait as false news," in ACM MDD, 2015.

[8] W. W. Yang. " liar, liar pants on fire: A new benchmark dataset for fake news detection," arXiv preprint arXiv: 1705.00648, 2017.

[9] Z. Wang, W. Hamza, and R. Florian, "Bilateral multi-perspective matching for natural language sentences," arXiv preprint arXiv: 1702.03814,2017.

[10] K. Shu, D. Mahudeswaran, and H. Liu. "Fake News Tracker: a tool for fake news collection, detection, and visualization," Computational and Mathematical Organization Theory, pp.1-2, 2018.