

# **AUTOMATIC IDENTIFICATION OF FAKE NEWS**

Enrollment No. - 14103010, 14103078, 14103108  
Name of Students - Nikita Jain, Saumya Pandey, Aditi Bhardwaj  
Name of Supervisor - Dr. Gagandeep Kaur



**May -2018**

**Submitted in partial fulfilment of the Degree of**

**Bachelor of Technology**

**In**

**Computer Science Engineering**

**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING & INFORMATION  
TECHNOLOGY**

**JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA**

**TABLE OF CONTENTS**

<b>Chapter No.</b>	<b>Topics</b>
	Student Declaration
	Certificate from the Supervisor
	Acknowledgement
	Summary
	List of Figures
	List of Tables
	List of Symbols and Acronyms
<b>Chapter-1</b>	<b>Introduction</b>
	1.1 General Introduction
	1.2 Problem Statement
	1.3 Empirical Study
	1.4 Solution Approach
	1.5 Support for Novelty/ significance of problem
	1.6 Tabular comparison of other existing approaches
<b>Chapter-2</b>	<b>Literature Survey</b>
	2.1 Summary of papers studied
	2.2 Integrated summary of the literature studied
<b>Chapter 3:</b>	<b>Analysis, Design and Modelling</b>
	3.1 Requirements Specifications
	3.2 Functional and Non-Functional requirements
	3.3 Overall architecture with component description and dependency details
	3.4 Design Documentation
<b>Chapter-4:</b>	<b>Implementation details and issues</b>
	4.1 Implementation details and issues
	4.1.1 Implementation Issues
	4.1.2 Algorithms
	4.2 Risk Analysis and Mitigation
<b>Chapter-5:</b>	<b>Testing</b>

5.1 Testing Plan

5.2 Component decomposition and type of testing required

5.3 Error and Exception Handling

## **Chapter-6 Findings & Conclusion**

6.1 Findings

6.2 Conclusion

6.3 Future Work

References

Appendices

(II)

## DECLARATION

We hereby declare that this submission is our own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Place:JIIT,Noida

Signature:

Date: 24-05-2018

Name: Nikita Jain

Enrollment No: 14103010

Signature:

Name: Saumya Pandey

Enrollment No: 14103078

Signature:

Name: Aditi Bhardwaj

Enrollment No:14103108

(III)

**CERTIFICATE**

This is to certify that the work titled “**Automatic Identification of Fake News**” submitted by “**Nikita Jain, Saumya Pandey, Aditi Bhardwaj**” of **B.Tech (Computer Science and Engineering)** of Jaypee Institute of Information Technology University, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of any other degree or diploma.

Signature of Supervisor

Name of Supervisor      Dr. Gagandeep Kaur

Designation

Date                              24-05-2018

(IV)

**ACKNOWLEDGEMENT**

Signature of the Student

Name of Student Nikita Jain

Enrollment Number 14103010

Date 24-05-2018

Signature of the Student

Name of Student Saumya Pandey

Enrollment Number 14103078

Date 24-05-2018

Signature of the Student

Name of Student Aditi Bhardwaj

Enrollment Number 14103108

Date 24-05-2018

## SUMMARY

The spread of fake news articles has generated noticeable concern recently, as false or misleading stories can spread faster and reach a wider audience over social media. Fake news are online stories that appear to be factual but are not. From 2016 US Presidential elections to GPS chips in currency notes- were among the many incidences where fabricated news was used to mislead, influence, promote prejudices and manipulate the beliefs of vast majority of audience. Fabricated news is capable of not only tarnishing reputation of person/organization, but also causing huge revenue losses. This area of study is garnering a lot of attention from the researchers. Our application Fake Busters (FBS) tends to explore the relationship between the headline article pair in the news firstly through hand crafted feature modelling and secondly, through Long Short Term Memory (LSTM) model and Bidirectional Long Short Term Memory (BiLSTM) model thereby leveraging Artificial Intelligence (AI) to solve this problem [3]. Further, we evaluate the efficiency of our models by calculating accuracy, precision, support and F1 score. After carefully, studying about the menace of clickbait for monetary purposes or mining cryptocurrency illegally by enticing the user to click a headline we identified that clickbait is also a subcategory or a type of fake news that is generated or circulated with the motive of engaging the user into their content whether fake or true through catchy headlines. It is extremely difficult to spot them in the ginormous amounts of data generated each day on the digital media thus we leverage machine learning models like Logistic Regression, Multi-Layer perceptron, Random Forest, Support Vector Machine (SVM) media to detect the percentage of clickbait in any headline.

---

Signature of Student

Name: Nikita Jain

Date: 24-05-2018

---

Signature of Student

Name: Saumya Pandey

Date: 24-05-2018

---

Signature of Student

Name: Aditi Bhardwaj

Date: 24-05-2018

---

Signature of Supervisor

Name: Dr. Gagandeep Kaur

Date: 24-05-2018

## LIST OF SYMBOLS AND ACRONYMS

1. NLP: Natural Language Processing
2. SVM: Support Vector Machine
3. RF: Random Forest
4. MLP: Multi-Layer Perceptron
5. LSTMs: Long-Short-Term-Memory Models
6. BiLSTM: Bidirectional LSTM

## LIST OF FIGURES

1.	Fake Busters- Web Application Workflow
2.	Stance Detection Model Flowchart
3.	Clickbait Detection Model Flowchart
4.	Dataset Characteristics
5.	Fake Busters Screenshots
6.	Clickbait Screenshots

## LIST OF TABLES

1.	Hyperparameter Tuning
2.	Baseline Model Accuracy



# CHAPTER 1: INTRODUCTION

## 1.1 General Introduction

With advent of digital age, vast majority of people spend their time online, as a result the consumption of the news from the online sources has exponentially increased as compared to the conventional news media [1]. On one side, online media provides easy accessibility of information and is also cost efficient but on the flip side it has become a source of propagation of large scale poor quality news. The tectonic shift in the news consumption pattern poses a serious threat as social media has become hub for deliberately spreading fake news. From 2016 US Presidential elections to GPS chips in currency notes- were among the many incidences where fabricated news was used to mislead, influence, promote prejudices and manipulate the beliefs of vast majority of audience. Fabricated news is capable of not only tarnishing reputation of person/organization, but also causing huge revenue losses. In 2013 false news regarding bomb explosion in White House by AP (Associated Press) resulted in 1% decline in Standard & Poor 500 index along with causing US\$136 Billion loss which highlights how fake news dissemination has many far reaching consequences [29]. Thus, fake news can be defined as are stories that may appear to be true or legitimate but are designed to intentionally disseminate rumors or falsified information at large scale for promoting biases, financial or political gain and other propagandas. The most disturbing part is that the fake news stories have become so prevalent due the level of anonymity that online media provides that it is sometimes published by reliable media sources. Therefore, it inhibits the ability of a daily news consumer to segregate fake news and real news [1].

Social media platforms have gained massive popularity over the years. The amount of information being shared on the microblogging sites is unimaginable. People use it to stay connected, stay updated about latest happenings and present their opinions on various subjects. Due to lack of proper filtering mechanisms, vetting of bogus sources the cases of fake news on Twitter have also become a prevalent phenomenon. Cybercriminals hacking profiles to tweet false information leading to diffusion of fake news is also one of the major problem [2].

Human fact checking is heavily prone to biases so it is important to come up with solutions that aid the journalists to flag and detect suspicious false news effectively [3]. Thus, it is clearly evident that technological advancements have a key role to play in large scale

information diffusion in these times, while there are numerous ways in which the problem of deceptive news can be handled namely-statistical, theoretical, crowdsourcing, machine learning etc. It becomes imperative to utilize machine learning techniques to analyze and appropriately handle the problem of spread of fake news as it has the potential to provide more accurate results.

This area of study is garnering a lot of attention from the researchers. Our application Fake Busters (FBS) tends to explore the relationship between the headline article pair in the news firstly through hand crafted feature modelling and secondly, through Long Short Term Memory (LSTM) model and Bidirectional Long Short Term Memory (BiLSTM) model thereby leveraging Artificial Intelligence (AI) to solve this problem [3]. Further, we evaluate the efficiency of our models by calculating accuracy, precision, support and F1 score. After carefully, studying about the menace of clickbait for monetary purposes or mining cryptocurrency illegally by enticing the user to click a headline we identified that clickbait is also a subcategory or a type of fake news that is generated or circulated with the motive of engaging the user into their content whether fake or true through catchy headlines. It is extremely difficult to spot them in the ginormous amounts of data generated each day on the digital media thus we leverage machine learning models like Logistic Regression, Multi-Layer perceptron, Random Forest, Support Vector Machine (SVM) media to detect the percentage of clickbait in any headline.

## **1.2 Problem Statement**

From 2016 US Presidential elections to GPS chips in currency notes and the recent 2018 Karnataka election goofups- were among the many incidences where fabricated news was used to mislead, influence, promote prejudices and manipulate the beliefs of vast majority of audience. Fabricated news is capable of not only tarnishing reputation of person/organization, but also causing huge revenue losses.

After studying about the menace of clickbait for monetary purposes or mining cryptocurrency illegally by enticing the user to click a headline we identified that clickbait is also a subcategory or a type of fake news that is generated or circulated with the motive of engaging the user into their content whether fake or true through catchy headlines. It is extremely difficult to spot them in the ginormous amounts of data generated each day on the digital media we leverage a web application-Fake Busters (FBS) that leverages machine learning models-LSTM, BiLSTM, Random Forest, Multi-layer perceptron, Logistic Regression and SVM to predict the stance between the headline and corresponding text and

classify the article into four classes-“agrees, discusses, disagree and unrelated”. The same text is then supplied into the clickbait detection model which works on two models-the data is supplied to ML models and the URL is sent to Mozscape and Alexa API to get the credibility and reputation score. The results from both the

## 1.3 Empirical Study

### 1.3.1 Mozscape

Moz crawls the web continuously, searching for new content and re-crawling existing content to help us detect the ranking of the URLs. We can save each URL and other interesting details about that page: HTTP status code, page title, links, and other information. For calculating Mozscape values, they try and compute a link graph of the web and uses that data to generate metrics which further help us to determine how important each page is to rankings[12]. We call url-metrics with a request being sent to the api with an url to analyze and return the rankings to refine Mozscape's search. The *Cols* or the column parameter indicates the data which Mozscape returns.

The Mozscape index contains a lot of information and this information can help us identify a lot. To help you get the data you want, Mozscape also includes query parameters to limit, offset, filter, sort, and scope our results and to view the results as per our convenience. Query parameters are those which help restrict the amount of data returned; this may seem less important to us when we sent for a call analyzing only one URL, but the ability to receive only the data we specify becomes increasingly important when we are making multiple calls per second[13].

#### Limitations:

FREE ACCESS	PAID ACCESS
Free access allows for one request every ten seconds, up to 25,000 rows per month.	Paid access has no such restrictions and limits.
Free access provides relatively limited information about many different websites	Paid access provides information about all the websites and gives access to all the parameters.

<p>Free access lets you make calls to the url-metrics and links endpoints.</p> <ul style="list-style-type: none"> <li>• Free access allows access to only some <i>Cols</i> parameter values.</li> <li>• Each Mozscape call lists which parameter values are available to free users.</li> </ul>	<p>Paid access lets you use every call and parameter Mozscape offers.</p> <ul style="list-style-type: none"> <li>• Paid access includes the anchor-text and top-pages endpoints.</li> </ul>
---	---

### 1.3.2 Alexa Website Ranking

FREE ACCESS	PAID ACCESS
Free access allows checking of rank for only limited number of websites	Paid access has no such restrictions and limits.
Free access provides relatively limited information about many different websites	Paid access provides information about all the websites and gives access to all the parameters.

Alexa is a global traffic rank given to a website on the basis of how well it is performing relative to all other sites over a period of last 3 months. As per Alexa's official website, this rank is calculated using a "combination" of the estimated average daily unique visitors to the site and the estimated number of page views on the site over the past 3 months.

Alexa ranking system audits and calculate the frequency of visits on websites. This means if the same user is visiting the website more than once on a same day, it is counted as a single visit. The algorithm they use is pretty simple. The parameters used to measure the traffic are 'reach' (number of Alexa users visiting a site in a day) and 'page views' (number of times a particular page or URL is viewed by Alexa users).

### 1.3.3 Floyd Hub-Cloud Platform

FloydHub is a Platform-as-a-Service for training and deploying your deep learning models in the cloud. FloydHub takes care of the engineering ground work so you can focus on the core of your problem. It provide 22 hours' worth of free CPU/GPU credits for training of the model and thus proves to be a great option for beginners to train their

model on this platform. This platform works via command line and is much simpler to use as compared to AWS P2 instance.

### **1.3.4 Google Collaboratory-Cloud Platform**

Google Collaboratory is an initiative by Google to provide Free 7 hours of GPU access to machine learning enthusiasts and young researchers who are trying their hands on deep learning for the first time to train their models online, as they might have the resources and the machine capacity to train their models locally.

## **1.4 Solution Approach**

Recently, the issue of fake news has inspired many scientific studies to spot falsified news in the mammoth amount of content generated every day. There are several techniques that have been applied to solve the problem of fake news like sentiment analysis, building a bias classifier, developing crowdsourcing application which all have major flaw in which totally relying on public or absolute labelling of true and false by the classifier is also prone to bias based on the dataset on which the algorithm was trained on. Thus, stance detection is one solution to solve this problem.

Stance detection refers to identifying the correlation of the any given claim or headline with its corresponding piece of text and then identifying the relationship between the two and labelling the text into different categories. Therefore, in order to perform stance detection and also combine clickbait identification models we developed a web-application using Flask named Fake Busters that has two modules-first module is used for testing the percentage of clickbait using the neural network models and Alexa traffic rank, second module deploys both the stance detection as well as the clickbait detection models for predicting the credibility of the news available by pasting the headline or the url in the given web application.

## **1.5 Significance of problem**

Digital media platforms have gained massive popularity over the years. The amount of information being shared online every day is unimaginable. People use it to stay connected, stay updated about latest happenings and present their opinions on various subjects. Due to lack of proper filtering mechanisms, vetting of bogus sources the cases of fake news have now become a prevalent phenomenon. Cybercriminals, propagandists and people with

nefarious intent have led to large scale diffusion of fake news making it impossible for humans to manually spot fake news.

Human fact checking is heavily prone to biases so it is important to come up with solutions that aid the journalists to flag and detect suspicious false news effectively [3]. Thus, it is clearly evident that technological advancements have a key role to play in large scale information diffusion in these times, while there are numerous ways in which the problem of deceptive news can be handled namely-statistical, theoretical, crowdsourcing, machine learning etc. It becomes imperative to utilize machine learning techniques to analyze and appropriately handle the menace of spread of fake news as it has the potential to provide more accurate results

## 1.6 Tabular Comparison with existing Approaches

S. NO.	EXISTING APPROACH	FLAWS
1.	Sentimental analysis of the data as positive, negative and neutral	Sentimental analysis is not a complete parameter for judging the credibility of the news
2.	Analysis based on public votes- crowd sourcing application	It can be based on individual's opinion so the results may be ambiguous. Cyberbots can be deployed to sway the sentiments of news towards one side.
3.	Use of static registry of URL's to determine if the article is clickbait and fake	The registry may or may not be updated regularly which exposes a serious flaw in technique for prediction
4.	Use of only baseline models like SVM, XGBoost, Logical Regression etc. for identification	These models don't take into account the long term dependencies that might exist in the article or the news and thus provide less accurate results when compared to neural networks.

## **CHAPTER 2: LITERATURE SURVEY**

### **2.1 Summary of Research papers**

#### **Click bait: Forward-reference as lure in online news headlines [16]**

##### **Journal of Pragmatics, Elsevier**

The widespread outreach of clickbait headlines used to lure the masses into reading the news on social media is increasing at an alarming rate. It surveys 10 online Danish news websites to identify how forward-referencing works and study its impact on news consumption. Based on author's preliminary analyses, research identified the five most frequently used discourse deictic and cataphoric lexemes in the dataset. Searching for these words in the expanded data set and afterwards checking the headlines for discourse deictic forward-reference or cataphora, it proved possible to conduct a large scale comparative analysis of forward-reference in headlines of Danish news websites. The results are depicted in this paper- support the hypotheses on a correlation between forward-reference in headlines and tabloidization and commercialization. However, the results could not be regarded as fully comprehensive, but rather as part of a tendency assessment.

#### **“8 Amazing Secrets for Getting More Clicks”: Detecting Clickbaits in News Streams Using Article Informality [17]**

##### **Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)**

Clickbaits are articles with misleading titles, exaggerating the content on the landing page. Their goal is to entice users to click on the title in order to monetize the landing page.

The content on the landing page is usually of low quality. Their presence in user homepage stream of news aggregator sites may adversely impact user experience. Hence, it is important to identify and demote or block them on homepages. In this paper, author presents a machine-learning model to detect clickbaits. He uses a variety of features and show that the degree of informality of a webpage (as measured by different metrics) is a strong indicator of it being a clickbait [17]. Formality measure, no. of swear words, no. of POS tags of different categories, forward reference etc. are some of the features extracted for identifying the clickbaits. The author conducts extensive experiments to evaluate the approach and analyze properties of clickbait and non-clickbait articles [17]. Their model achieves high performance (74.9% F-1 score) in predicting clickbaits.

## **Diving Deep into Clickbaits: Who Use Them to What Extents in Which Topics with What Effects? [18]**

### **IEEE Transactions on Secure & Dependable Computing**

To attract the readers to click on an article and subsequently visit the media site, the outlets often come up with catchy headlines accompanying the article links, which lure the readers to click on the link. Such headlines are known as Clickbaits [18]. It creates a set of supervised clickbait classification models to identify clickbait headlines and used distributed subword embedding technique to transform the words in the corpus to 300 dimensional embeddings [18]. Skip-Gram Model was used to generate word embeddings. These embeddings were used to map sentences to a vector space over which a softmax function was applied to the classifier. 10-fold cross validation along with the evaluation metrics like precision, recall and F-1 score were calculated and the model thereby achieved an accuracy of 98.3% accuracy on a labelled dataset. Moreover, the author in addition to identifying the clickbait headlines in the corpus, used the embeddings to measure the distance between the headline and the first paragraph, known as intro, of a news article. Finally, close scrutiny of the social media posts also reveals that broadcast type media has higher percentage of usage of clickbait practice than the print media and non-news type broadcast media mostly contributes to it.

## **Malicious URL detection using convolutional neural network [19]**

### **IEEE Transactions on Innovative computing**

A Malicious URL or a website can be defined as a network which hosts a variety of unrequested content in the form of spam, which can be prone to launch attacks. Users unknowingly visit such web sites and become victims of various types of scams, including monetary loss, leak of private information. Popular types of attacks using malicious URLs include: Phishing and Social Engineering, and Spam[20]. Specifically, Black-lists contains a list of URLs that have been confirmed to be malicious in the past. Such a technique is extremely fast in detecting such type of URLs as query overhead is less, and hence is very simpler to implement[21]. Additionally, such a technique would have a very low false-positive rate that is probability of detecting URLs that are not malicious is very low. Attackers nowadays use creative techniques to attack the blacklists and fool users by modifying the URL making it appear genuine via obfuscation. All of these possibilities try to hide the malicious intentions of the website by masking the malicious URL and making it appear genuine[22].



A little crawler has been setup by them and then they have crawled a lot of malicious links from various websites. The second task they did was finding out clear URLs this dataset was readily available so this time there was no need of setting up a crawler to fetch the data. So, a total of 420464 URLs were fetched out of which around 75643 were malicious and others were clean. Finally, they applied convolution neural network algorithm was used to detect malicious URL because this algorithm takes less time and it is fast as compared to other algorithms. Input given to the network was: URL of the webpage and output obtained from the algorithm was: Bad or Good[19].

### **Identifying Malicious Web Domains Using Machine Learning Techniques with Online Credibility and Performance Data [23]**

**Evolutionary Computation, 2016 IEEE Congress (CEC) [23]**

Invalidated and unwanted redirects and forwards are possible when a web application accepts the input that is not trusted can cause the unwanted redirects redirect the request to a URL contained with the malicious URL. The objective of this paper was to improve the performance of machine learning which will further help us in accurately identifying malicious web domains based on their popularity and performance. For the purpose of identifying and detecting phishing web domains, they have collected malicious web domains. Specifically, two datasets, one with malware only (Malware) and the other with both malware and phishing (Malware Phishing) domains were collected to build and evaluate the classifiers for detecting phishing web domains.

In this paper, they have presented a study to basically investigate the performance of machine learning techniques for identifying malware and phishing web domains using online popularity and performance data. Sixteen data features have been studied and analysed which represent the popularity and performance of web domains from where the data was crawled, and nine well-known machine learning techniques were examined.

### **Towards Detecting Compromised Accounts on Social Networks [24]**

**IEEE Transactions on Dependable and Secure Computing [24]**

Compromising social network accounts has become a profitable course of action for cybercriminals. Attackers can distribute their malicious messages or disseminate fake information to a large user base by hijacking control of a popular media or business account. These incidents often result in serious threats ranging from a tarnished reputation to multi-billion dollar monetary losses on financial markets. In this research paper, we

have made use of various techniques to identify whether the individual high-profile accounts are compromised or not [24]. We often see that high-profile accounts frequently have one characteristic that makes the detection reliable and even more predictable —they show consistent behaviour over time.

In the research we present COMPA, the first detection system designed to identify compromised social network accounts. Basically COMPA is a very simple based framework: that often the social network account holders develop habits over time, and these habits are fairly stable. If the message significantly deviates from the learned behavioural profile, COMPA flags it as a possible compromise. As part of our research we first show that high profile accounts often have well-defined behavioural profiles that allow COMPA to detect compromises with very low false positives.

### **Detecting Malicious URLs using Machine Learning Techniques [25]**

#### ***Computational Intelligence (SSCI), 2016 IEEE Symposium Series [25]***

With the undeniable prominence of the World Wide Web as the paramount platform supporting knowledge dissemination and increased, economic activity, the security aspect continues to be at the forefront of many companies and governments' research efforts. Symantec's 2016 Internet Security Report [25] elaborates on an ample array of global threats that includes corporate data breaches, attacks on browsers and websites, spear phishing attempts, ransomware and other types of fraudulent, cyber activities. Another well-known and surprisingly quite effective strategy, is baiting the users to click on a malicious Uniform Resource Locator (URL), which leads to the system being somehow compromised. In this paper, the detection of malicious URLs as a binary classification problem and study the performance of several well-known classifiers, namely Naive Bayes, Support Vector Machines, Multi-Layer Perceptron, Decision Trees, Random Forest and k-Nearest Neighbours has been addressed. Furthermore, we adopted a public dataset comprising 2.4 million URLs (instances) and 3.2 million features [25]. The data for this research, presented by Ma et al., consists of 121 sets of URLs that have been collected for 121 different days. For Day0, 16,000 URLs have been collected, Day45 contains only 130 entries. For all other days, 20,000 URLs are registered. All URLs are characterized by multiple attributes and each is classified as either malicious or benign. The entire dataset consists of over 2.3 million URLs, each having over 3.2 million features. The overwhelming majority of these features can be identified as binary attributes.

### **Evaluating Weightless Neural Networks for Bias Identification on News [26]**

### ***Networking, Sensing and Control (ICNSC), 2017 IEEE***

In this paper, the analysis has actually compared the WiSARD classifier, a lightweight efficient weightless neural network architecture, against Logistic Regression, Gradient Tree Boosting, SVM and Naive Bayes for identification of polarity in news. Further knowing and motivated by the fast pace at which news feeds are being published, we envision the increasing need for efficient and accurate mechanisms for bias detection. WiSARD has presented itself as a good choice for the task of bias identification in fake news section, specially in dynamic contexts, due to its online learning ability and comparable accuracy when contrasted against the considered alternatives. In this paper, the main goal was to automatically classify articles of news based on their political bias. Their goal was to find the greatest number of pages of interest with the least computational cost. When driven by the page contents, they are classified as specific or focused. These actually gave us the correct data for analysis.

### **Incentivizing the Dissemination of Truth Versus Fake News in Social Networks [27]**

#### **In System of Systems Engineering Conference (SoSE), 2017 IEEE**

The online social networks, cause emergent properties, thus making authentication processes difficult, given availability of multiple sources. In this study, we show how this conflict, can be modelled as a volunteer's dilemma. We also show how the public contribution through news subscription (shared rewards) can impact the dominance of truth over fake news in the network. Further moving on to the model we start with, we showcase the model variables and parameters. In the model a number of regular, agents ( $N$ ) are interacting in social system [27]. These regular agents can either participate ( $M$ ) in validating and authenticating the truthful news or consume news ( $N-M$ ). The minimum group size of volunteers is required to achieve the public good ( $k$ ). The cost of volunteering and the cost of failing to achieve, the public good occurs in case. Having an rewarding mechanism,  $s$  is distributed among the volunteering agents. As the model is a symmetric model, the equilibrium, shows the volunteering ratio or the probability of volunteering among users ( $p$ ) [27].

To run the model and show the results, we define two comparing factors. The first factor calculates the difference between averaged payoff of volunteering and defecting agents. Apparently, the measure shows whether the agents benefit from either volunteering or defecting depending on the difference being positive or negative. For the second factor,

we consider the fake news agents, we calculate the expected difference between the payoff of volunteering and defecting among fake news agents.

### **Stakeholder Mining and Its Application to News Comparison [28]**

#### **Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference**

Due to intentions of news agencies and their sponsors, in a sense, news is never free from bias. Bias often causes readers to misunderstand the facts of actual events and even the whole story. Although a large number of studies have been made on analyzing bias by means of comparing related news articles, conventional methods present related articles and ask users to compare them. To the best of our knowledge, models and criteria for bias analysis have not yet been well studied [28]. In this paper, we propose a novel stakeholder mining mechanism for analyzing bias in news articles by comparing descriptions of stakeholders. The mechanism is based on the presumption that interests often induce bias of news agencies. As we use the term, a “stakeholder” is a participant in an event described in a news article who should have some relationships with other participants in the article [28]. The approach attempts to elucidate bias of articles from three aspects: stakeholders, interests of stakeholders, and the descriptive polarity of each stakeholder. This paper presents a user study to validate the proposed methods. To detect the participants sharing interests in an event, we build a lexical resource named RelationshipWordNet and construct a sentence structure named relationship structure. A synset that express a better (resp. worse) relationship has a larger positive (resp. negative) score. We proposed a stakeholder mining mechanism and de-scribed how it can be applied to bias analysis. The experimental results obtained indicate that the stakeholder mining mechanism is helpful for discovering news articles’ bias. In future work we will attempt to identify identical entities that were referred to using different words. It will also be necessary to use lexical resources appropriately to the meanings of words according to context.

## **2.2 Integrated Summary**

In order to develop an approach for effectively handling the task of detecting fake news we explored several research papers. The first paper discusses the fundamental problem in NLP (Natural Language Processing) related to textual entailment which is sequence based task similar to the stance detection [4]. The author utilized *Stanford Natural Language Inference* (SNLI) [5] dataset which contained labeled sentence pairs [5]. The task involved determining how the sentences may contradict each other, may be unrelated or one may be

a logical consequence of another sentence [4]. He used feature based classifier along with neural network models-LSTM for sequence to sequence based models and classify the hypothesis-premise pairs as an entailment, contradiction or unrelated. In another research paper involving sequence modelling, we studied different types of recurrent units on music modelling and speech signal modelling [7]. Further analysis revealed that advanced Recurrent Neural Networks (RNN) like LSTM performed much better than the conventional *tanh* units thereby making RNN the fundamental unit in the task of sequence modelling [7][9].

Previous research in the area of stance detection, bias detection proved to be of utmost importance. The paper analyzed “Emergent”- a dataset which had been labelled by the journalists as true, false or unverified for detecting rumor in the given set [8]. Author utilized Logistic Regression classifier after condensing each article into corresponding headline. The classifier was built on features that categorized the given headline-claim into one of the three stances-for, against or observing with an accuracy of 73% [8] [9]. Other paper explored sentiment bias detection in news articles and websites by visualizing and detecting the sentiment tendencies that are mapped into four dimension rather than simply positive and negative [10]. It involved providing differences in sentiment among the subtopics and the various websites by extracting relevant information thereby generating sentiment graph for assessing the credibility of the piece of news in the process [10]. The task of assigning probabilistic scores for classifying the news articles as optimistic and pessimistic was explored in another paper [11]. It utilized semantic analysis in which weights were assigned to the themes based on the scores computed after parsing the HTML contents of the page. Predefined weights were used to assign total score to positive and negative themes that leveraged Slant Engine a semantic analysis system. These papers successfully provided the foundation for beginning the stance detection tasks.

Wide range of research has been done when it comes to social media platforms like Twitter but in order to proceed with fake news identification it was integral to go through few papers that outlined similar tasks. In order to map the fake news detection on Twitter a paper was looked into which explained COMPA a detection system aimed at identifying compromise of the high profile twitter accounts by their behavioural profiling. System utilized several features of the tweets being sent like the *time*, *message source*, *message text*, *message topic* and the *links* in order to identify if the accounts has been hijacked by a cybercriminal [12]. Author then calculated the global anomaly score to detect anomaly in

the usual behaviour thereby by successfully determining compromised accounts on Twitter. In yet another paper decision tree classifier, Naive Bayes were used as classification models to detect fake images on Twitter during crisis event. It elucidates how user based features for detection performed poorly as compared to tweet based features through construction of retweet graphs and temporal analysis of the tweets shared per hour [13]. In addition to this the paper about Fake Tweet Buster [14] aimed to detect fake images on twitter using reverse image search, user based features like *no. of followers* in the network along with crowdsourcing for identification. All the papers regarding identification of fake images failed to utilize the statistical features of the images.

Thus, these papers helped in developing and deriving the models from them after thorough evaluation of various parameters investigated in each paper. The following section provides in depth view of the approach undertaken to detect news fabrication.

## **CHAPTER 3: ANALYSIS, DESIGN AND MODELLING**

### **3.1 Requirements Specifications**

#### **3.1.1 Functional Requirements**

##### 3.1.1.1 Model Features

- 3.1.1.1.1 No. of stopwords
- 3.1.1.1.2 No. of POS tags
- 3.1.1.1.3 N-gram
- 3.1.1.1.4 Catchy words
- 3.1.1.1.5 Length of Headlines

##### 3.1.1.2 Neural Networks and Baseline Models

- 3.1.1.2.1 LSTM
- 3.1.1.2.2 Bidirectional LSTM
- 3.1.1.2.3 SVM
- 3.1.1.2.4 MLP
- 3.1.1.2.5 Random Forest

##### 3.1.1.3 Mozscape and Alexa Ranking

- 3.1.1.3.1 Node Js
- 3.1.1.3.2 API access keys
- 3.1.1.3.3 GET and POST Requests
- 3.1.1.3.4 Store Results and analyse

### **3.2 Non-Functional Requirements**

#### **3.2.1 Hardware Requirements**

- **Operating System:** Microsoft Windows 7 Professional/ Windows 8.1/ Windows 10 or Linux
- **Processor:** 2.6 GHz Intel Core i5 or greater
- **Memory:** Minimum 4 GB RAM
- **Disk space:** 2 GB of free disk space
- **Graphics Card:** NVIDIA GeForce GTX860 and GTX870

### 3.2.2 Software Requirements:

- Python 2.7 or 3.5
- Gensim
- Jupyter Notebook
- NLTK, Scikitlearn
- Matplotlib, Numpy, Pandas
- Node Js
- Anaconda
- Flask

### 3.2.3 Performance Requirements:

System becomes slow on training the model for greater than 100 epochs on 8GB RAM and needs GPU acceleration by using CuDA and CuDNN on the system. The Mozscape API Free version allows sending one request per 10 seconds which makes the system extremely slow and thus reduces the overall performance of the system.

## 3.3 Overall architecture with component description and dependency details

### 3.3.1 Keras

Keras is an open source software library used for the purpose of conducting machine learning and deep neural networks research and for numerical computation using data flow graphs. The flexible architecture allows you to deploy computation to one or more CPUs or GPUs in a desktop.

### 3.3.2 Node Js

Node.js is a JavaScript runtime built on Chrome's JavaScript engine. Node.js uses an event-driven, non-blocking I/O model that makes it lightweight and efficient. Node.js' package ecosystem, npm, is the largest ecosystem of open source libraries in the world

### 3.3.3 Libraries

- NLTK, Scikitlearn
- Matplotlib, Numpy, Pandas

- Pyscape
- Keras
- Urllib3, requests, bs4
- Simplejson
- Tqdm
- Gensim
- Zip
- Gzip
- Textblob
- Seaborn

### 3.4 Design Documentation

#### 3.4.1 Control Flow Diagram

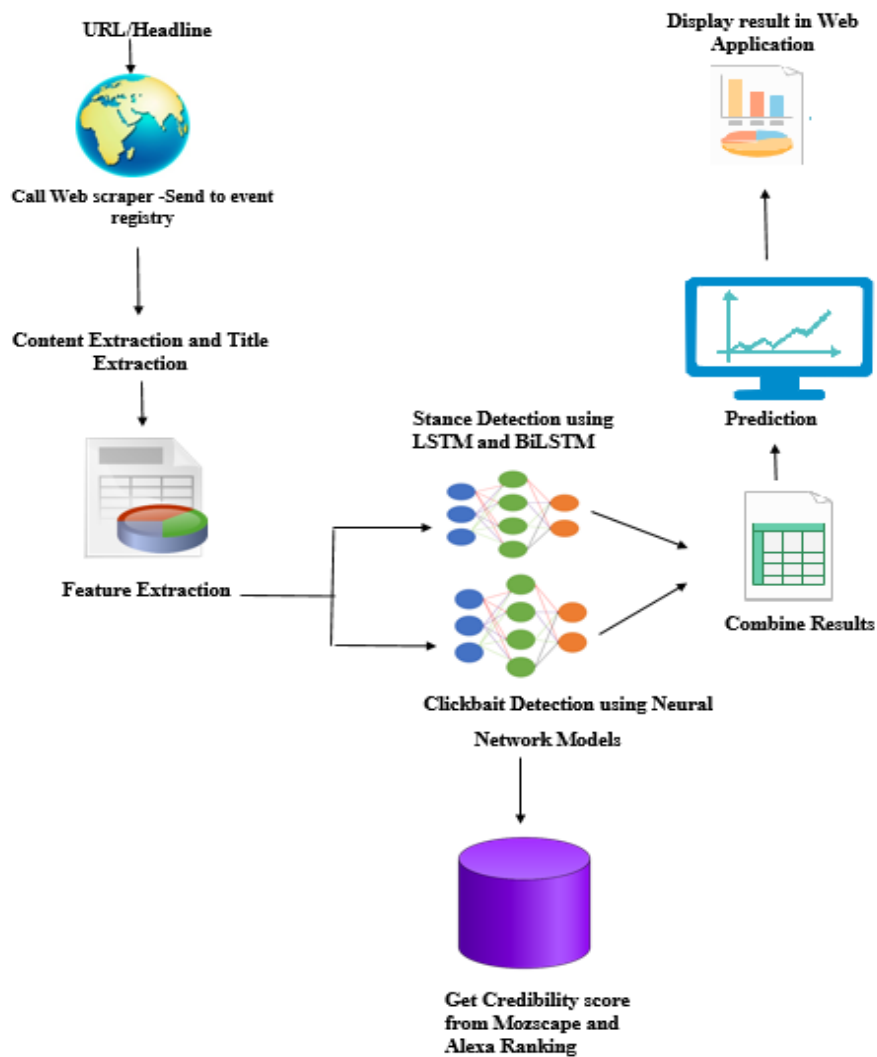
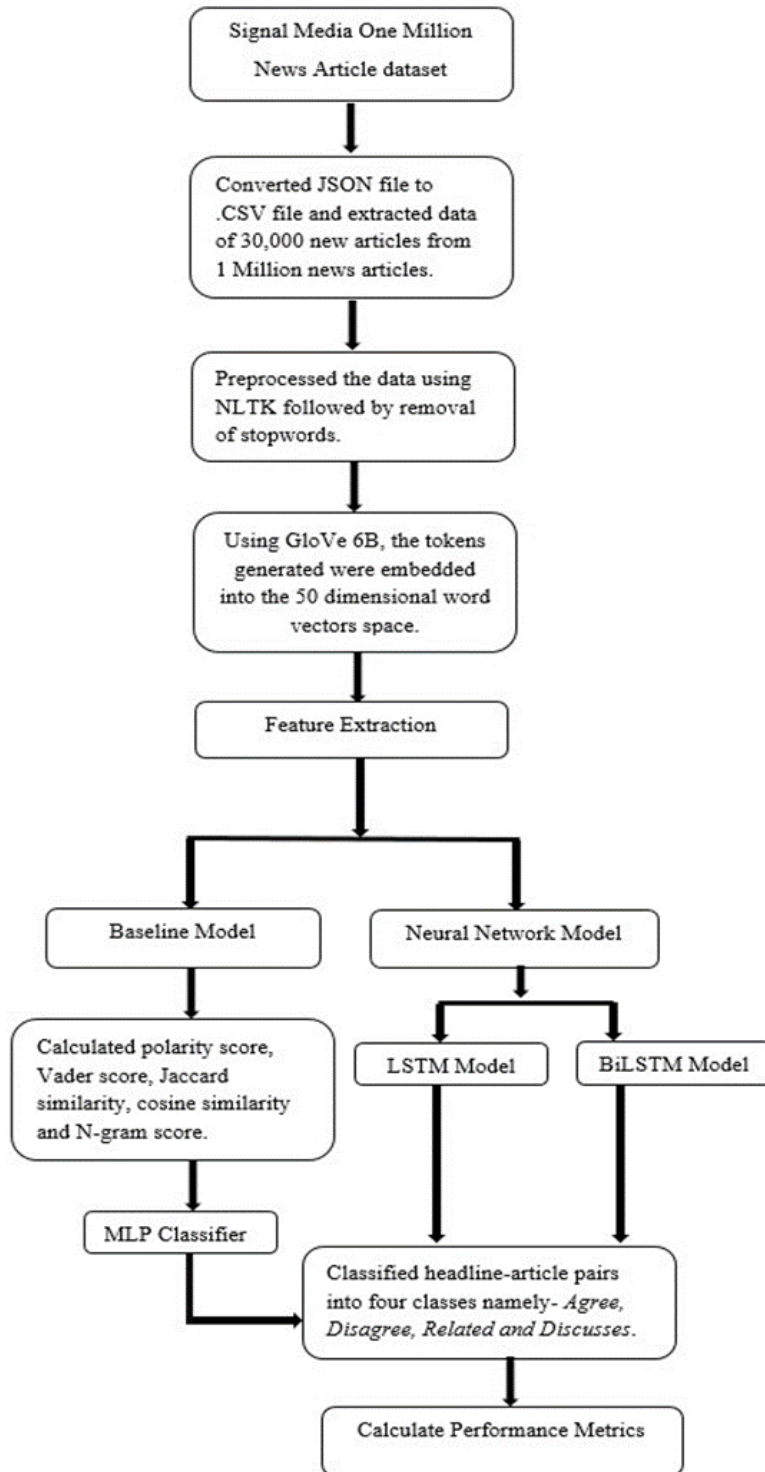
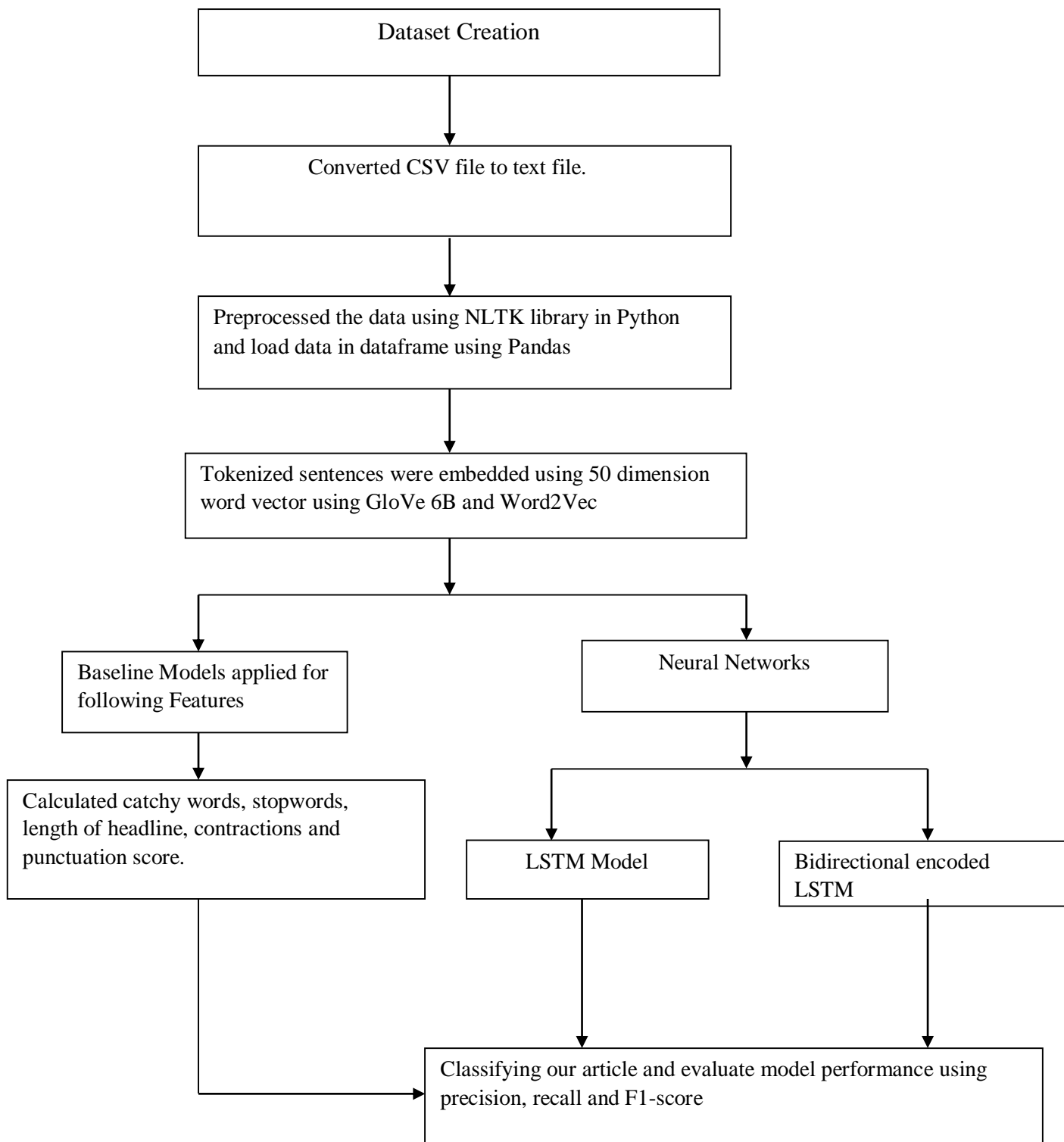


Figure 1:Fake Busters-Web Application Workflow





**Figure 2: Stance Detection Model Flowchart**



**Figure 3:Clickbait Detection Model Flowchart**

# CHAPTER 4: IMPLEMENTATION

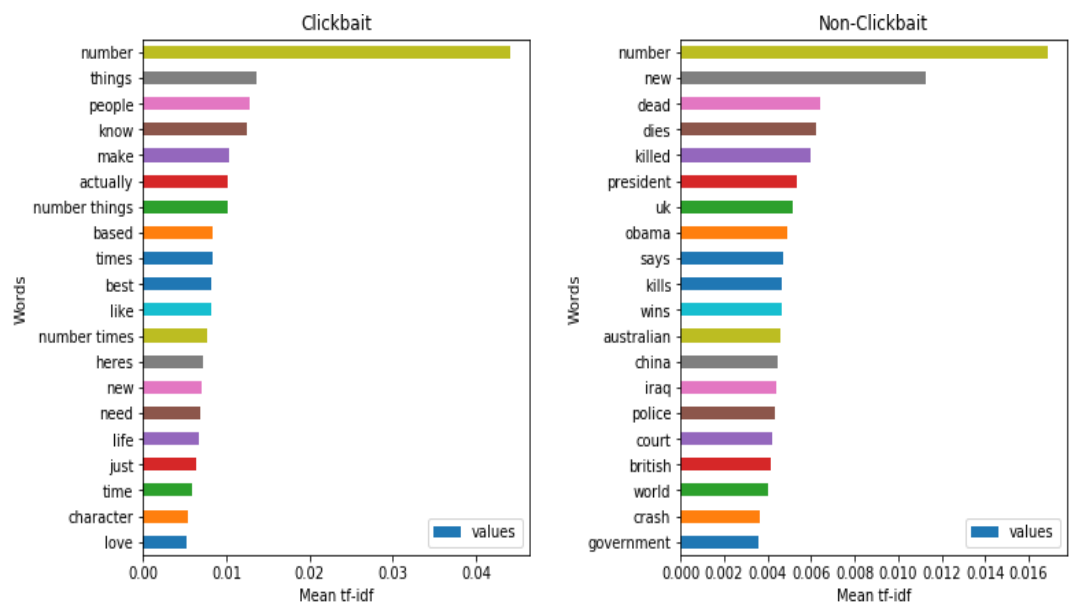
## 4.1 Implementation Details

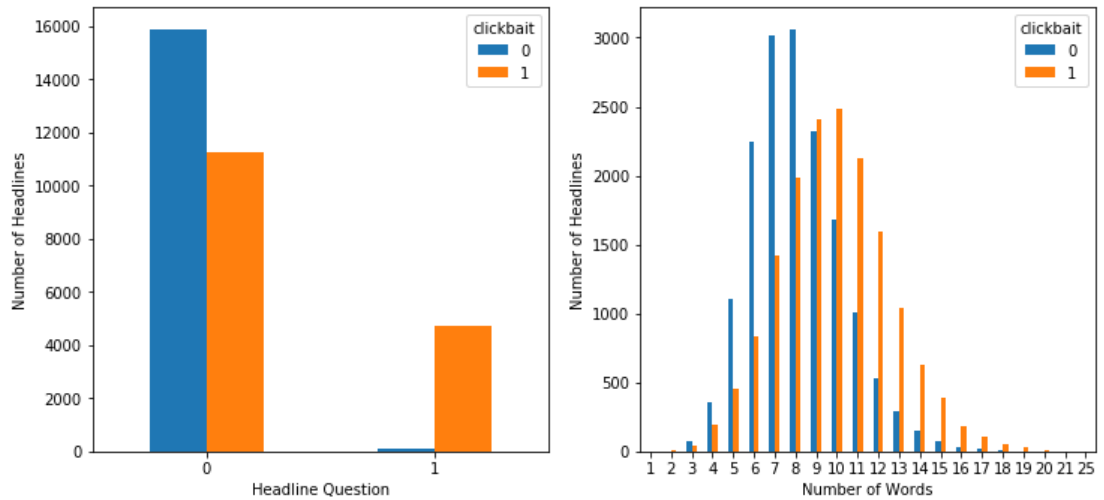
### 4.1.1 Dataset for stance detection

The Signal Media One Million News Article [30] (SMOM) - is the dataset that was used for the stance detection of headline-article pair in this research work. The articles in the dataset [15], were gathered from heterogeneous sources including BBC, The Hindu, The New York Times etc. It contains one million headline-article pair in JSON format from various news magazines, blogs and local outlets which can be downloaded from official signal media website. The average article length ranges from 400 to 1000 words for the news articles while for blogs the average article length is 1200 words. The main features of the dataset are: The JSON file had the following format: Title, Content, Media-Type, Source, and Published.

### 4.1.2 Clickbait Dataset

The dataset was created by crawling the data from various websites like The Hindu, The NYT, The Guardian, Buzzfeed and The ViralNova, Scoopwhoop and Times of India which was stored in txt file. The dataset was manually labelled into clickbait and non-clickbait and consists of following characteristics depicted through the graphs below. A total of 80,000 headlines have been collected from various sources.





**Figure 4: Dataset Characteristics**

#### 4.1.3 Issues

There were several limitations that were identified in the SMOM dataset, for e.g. lack of proper access right leads to getting warning message in the dataset instead of the actual content, syndication involves same article being published by multiple media houses and, missing links i.e. the dataset only consists of textual data and doesn't contain hyperlinks to original article [15].

## 4.2 Algorithms Used

### • LSTM

Long Short Term Memory networks —called —LSTMs are used to solve the problem of persistence by taking into account the previous knowledge i.e. neural networks that take long term dependencies into account. Each LSTM consists of three gates namely — input gate, output gate and the forget gate that helps in maintaining the flow for information to go through the net-work [17]. The input gate maintains the flow of the new value into the memory, while the forget gate is used to manage the amount of time the value stays in the memory and the output gate manages the degree to which the value in the memory is used for computation of the output activation of the given block of LSTM [17]. Fig.4 shows LSTM algorithm used in our work.

- Bidirectional-LSTM (BiLSTM)
- As shown in Fig.6, Bidirectional LSTMs is a modified version of conventional LSTM that is used to improve the accuracy of the given classification problem. In problems where all timesteps of the input sequence are available, BiLSTM is

preferred. It is a combination of 2 LSTM trained on the input sequence. The first LSTM takes the input sequence and the second LSTM takes the reversed input sequence as the input. This can provide more contextual information to the network thereby increasing the efficiency many folds.

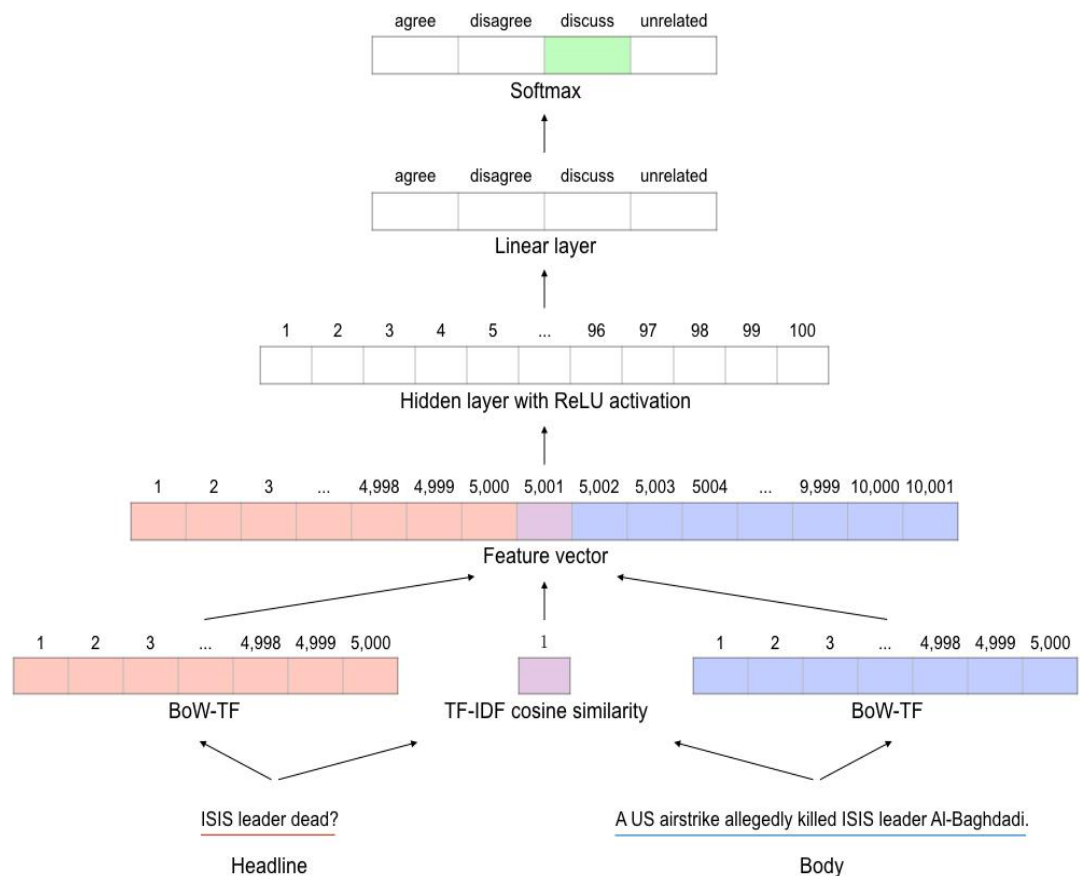
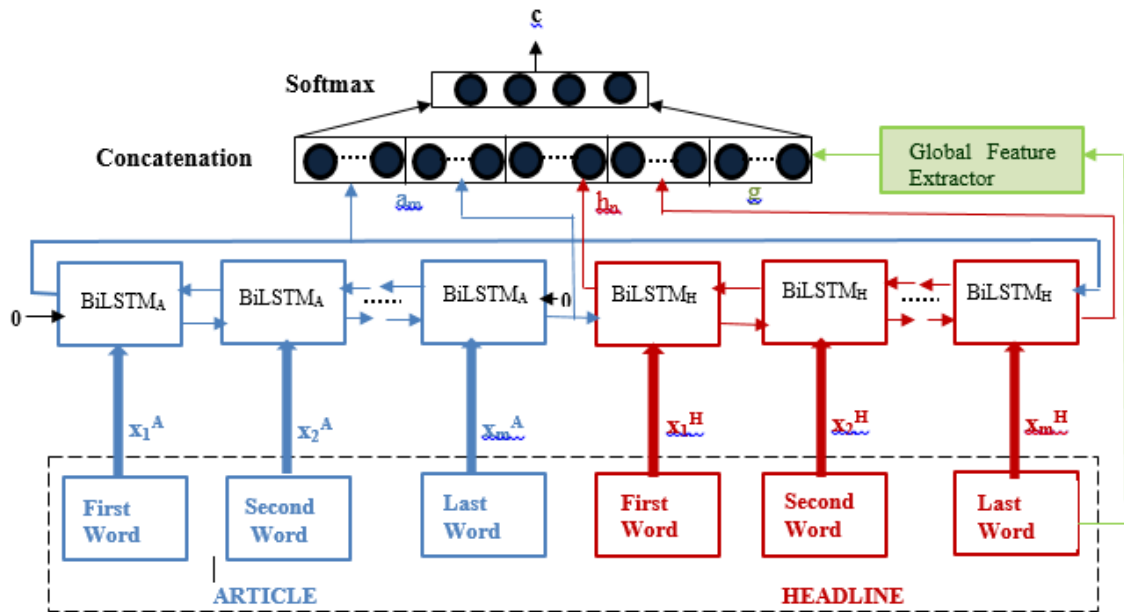


Figure 5: Working of stance detection Model

- **Bidirectional-LSTM (BiLSTM)**

As shown in Fig.6, Bidirectional LSTMs is a modified version of conventional LSTM that is used to improve the accuracy of the given classification problem. In problems where all timesteps of the input sequence are available, BiLSTM is preferred. It is a combination of 2 LSTM trained on the input sequence. The first LSTM takes the input sequence and the second LSTM takes the reversed input sequence as the input. This can provide more contextual information to the network thereby increasing the efficiency many folds.



- **SVM-Support Vector Machine**

It represents the points in space separated by a hyper-plane which classifies the data into defined classes. SVM is generally used for linear classification but can effectively classify non-linear data as well. This implies given a training set that is labelled, SVM defines an optimal hyper-plane which classifies the data into respective classes [18]. Fig. 5 shows the working of SVM where input space represents the news data, which are mapped into the feature space using various parameters mentioned above. This is followed by binary classification of the input space into clickbait and not clickbait respectively.

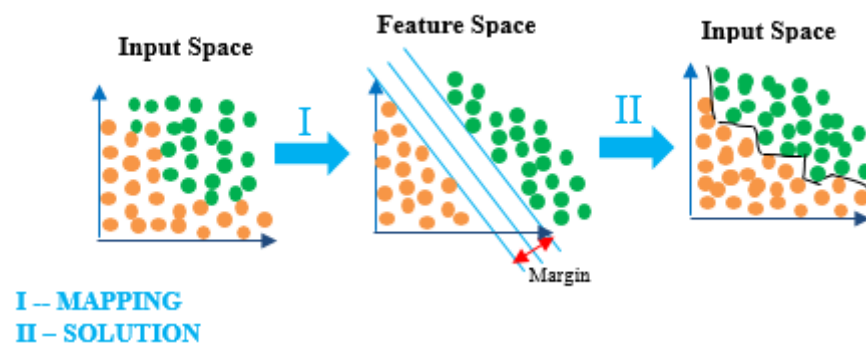


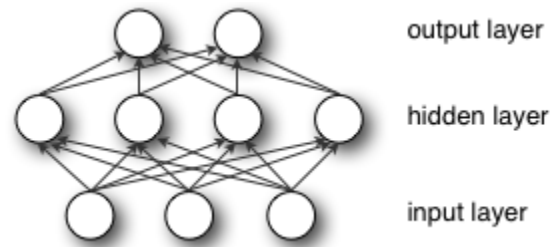
Fig. 5. Working of Support Vector Machine

- **Random Forest**

Random Forest is an ensemble learning algorithm which is used for both classification and regression. It is known as an ensemble algorithm because it involves combining more than one algorithms for same or different kind for

classifying the data [19] [20]. Random forest classifier during training of the dataset constructs multiple decision trees and then outputs the most frequently occurring class of the individual trees for the test object [19]. One of the main advantage of using random forest in this work is that it can handle large datasets and solves the problem of over fitting.

- **Multi-Layer Perceptron-MLP**



A multilayer perceptron (MLP) is a class of feedforward artificial neural network. An MLP consists of at least three layers of nodes. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training.[1][2] Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

#### **Cross-Validation:**

In order to solve the problem of overfitting k-fold CV is applied. This procedure involves splitting the dataset into training and test set respectively [31]. When performing k-fold CV, the dataset is randomly segregated into k sets, where k-1 sets are used for training the model while the rest along with test set is utilized for testing the model [31]. This technique therefore helps in achieving higher accuracy.

#### **Hyperparameter Tuning:**

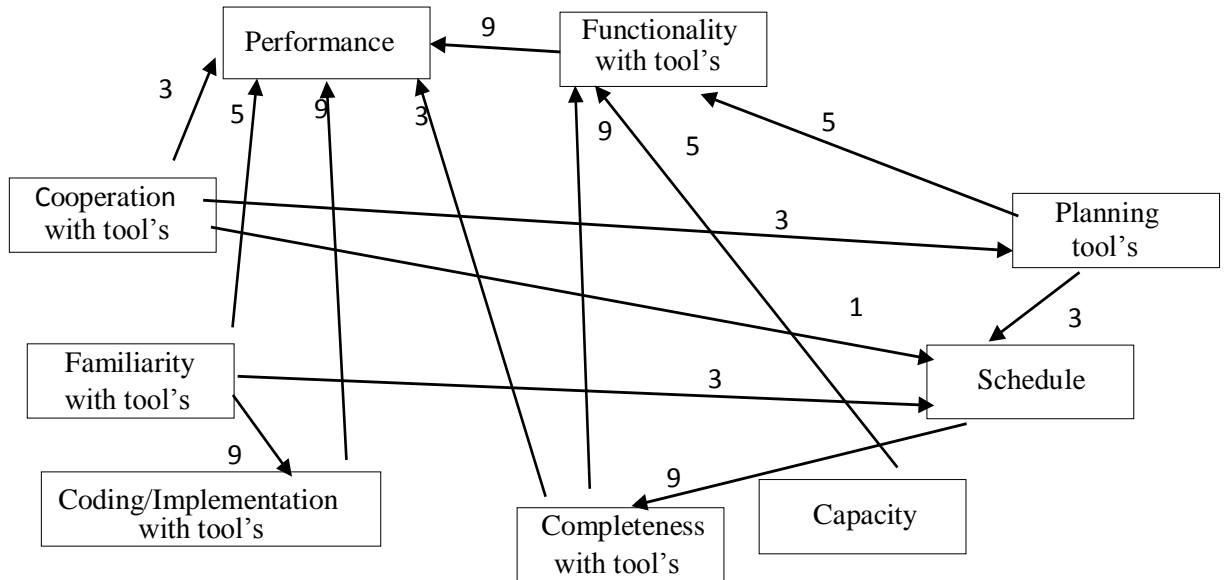
The hyperparameters selection was independent of the implemented neural network model. Due to constraints in the computation power of the system on which the models were implemented it was impractical and impossible to tune the neural networks on various hyperparameters and thus, we choose no. of epochs ( $e$ ) as the hyperparameter for analyzing its impact on the models. We performed hyperparameter tuning by changing the no. of epochs in the different models and checking their accuracy.

### 4.3 Risk Analysis & Mitigation

Risk Id	Classification	Description of Risk	Risk Area	Prob.	Impact	RE (P*I)
A1	Product Engineering	Design and selection of algorithms used	Functionality	Medium (3)	High (5)	15
A1	Product Engineering	Requirements of complete data to be worked on	Completeness	High (5)	Medium (3)	15
A3	Product Engineering	Inadequate systems properties, System properties do not match with tool's Requirements	Coding/Implementation	Low (1)	High (5)	5
A2	Product Engineering	Efficient Design Model runs for the test data in a go	Performance	Medium (3)	High (5)	15
B1	Development Environment	Development Process that is Familiarity with languages, IDEs used	Familiarity	Low (1)	Medium (3)	3
B5	Development Environment	Work Environment should be stable and well cooperated	Cooperation	Low (1)	Medium (3)	3
B3	Development Environment	Management Process. Proper planning of phases	Planning	Low (1)	Medium (3)	3
B2	Development Environment	Development System. The csv files were too large that they caused system to hang multiple times	Capacity	Medium (3)	High (5)	15



C1	Program Constraints	Resources should be available at appropriate time	Schedule	Low (1)	Medium (3)	3
----	---------------------	---	----------	---------	------------	---



**Figure6: Inter-relationship Graph**

Table 2: Risk wise total weighting factor

S.No.	Risk Area	# of Risk statements	Weights (In+ Out)	Total Weight	Priority
1	Performance	6	3+5+9+3+9	29	1
2	Functionality	5	9+9+5+5	28	2
3	Completeness	2	9+9+3	21	3
4	Coding/Implementation	2	9+9	18	4
5	Familiarity	1	9+5+3	17	5
6	Schedule	2	1+3+3+9	16	6
7	Planning	4	3+5+3	11	8
8	Capacity	1	9	9	9
9	Cooperation	2	3+3+1	7	7

Table 3: Risk Statement

Risk Statement	Risk Area	Priority of Risk Area in IG
Risk of availability of data	Completeness	3

1. Parameter Tampering—Changing information in a site’s URL parameter. Because many applications fail to confirm the correctness of CGI parameters embedded inside a hyperlink, parameters can be easily altered and redirected to malicious sites
2. Buffer Overflow—Closure of business. By exploiting a flaw in a form to overload a server with excess information, hackers can often cause the server to crash and shut down the web site.
3. Cross-Site Scripting—Hijacking/Breach of Trust. When hackers inject malicious code into a site, the false scripts might be executed in a context that appears to have originated from Fake Busters.
4. Backdoor and Debug Options—Often, programmers leave in debug options in order to test the site before it goes live. Sometimes, in haste, we might forget to close the holes, giving hackers free access to sensitive information.

#### **Mitigation:**

- Words and sentence tokenization and stop words removal using NLTK libraries.
- Effectiveness: Ease in breaking of sentences into words and removal of unnecessary words so that data could be read and processed correctly.
- Use of Pandas library to read and pre-process the csv data
- Effectiveness: Easy and relevant in data processing

Avoid, accept, reduce/control, or transfer. For each risk we encounter during our project, we as a team will have to deal with it. A little thought about the work and a proper pre-planning of work and doing the work more systematically enables more options than just a major product recall or recovery which makes the probability of arising of risks should be low. Risk mitigation plans should

- Identify the root causes of risks that have been identified and the causes which have led to these risks & Prioritize mitigation alternatives so that all the risks can be avoided.
- Evaluate the common causes that have led to these risks.
- Identify mitigation strategies, methods, and tools for each major risk and analyse how each risk could be avoided.

## CHAPTER 5: TESTING

### 5.1 Testing Plan

Types of Test	Will test be performed?	Comments/Explanations	Software Component
Requirement Testing	Yes	Input expected from user	To be tested on corpus implementation
Unit Testing	Yes	Python found compatible with algorithms	To be tested on Python
Integration	Yes	Successful software integration on all operating systems	The client program tested on different inputs
Performance	Yes	The software responds to any query in minimal time	To be tested on the software
Stress	Yes	Able to summarize sufficient amount of queries	To be tested on software
Compliance	Yes	In compliance with the existing competition needs	Suite to be tested and compared with existing technologies
Security	Yes	To ensure that candidates data are not leaked	Resume uploader
Load	No	Implementation does not allow multiple users to request at same time	To be tested on user's component
Stress	No	It is applied to check the response of the system when the system is given a load beyond its specified limits. This situation will not occur in our software, so it is not performed.	

## 5.2 Component Decomposition and Type of Testing Required

Sr. No	Components that require testing	Type of testing Required	Technique for writing test cases
1	Running of algorithm	Unit, performance Stress	White box
2	Load Balancing	Unit, integration	Black box
3	Connectivity	Performance	White box
4	Results	Unit	Black box

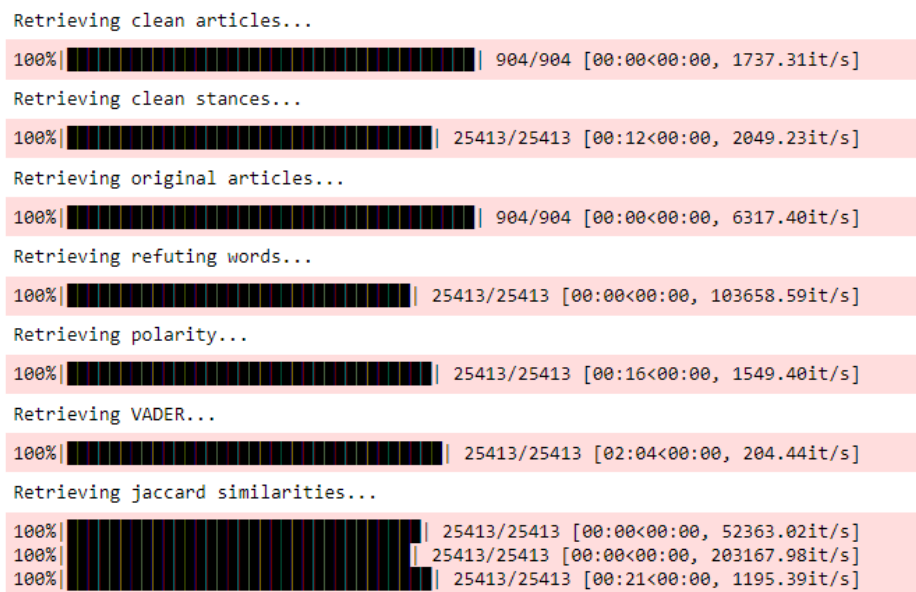
## 5.3 Error and Exception Handling

Test case id	Test case	Debugging Technique
1.	Dimensional Error	Padding input sequences
2.	Unicode Decode Error	Opening of file in utf-8 encoded format
3.	Shape mismatch Error	Converting the input sequences to appropriate dimensions using reshape function.

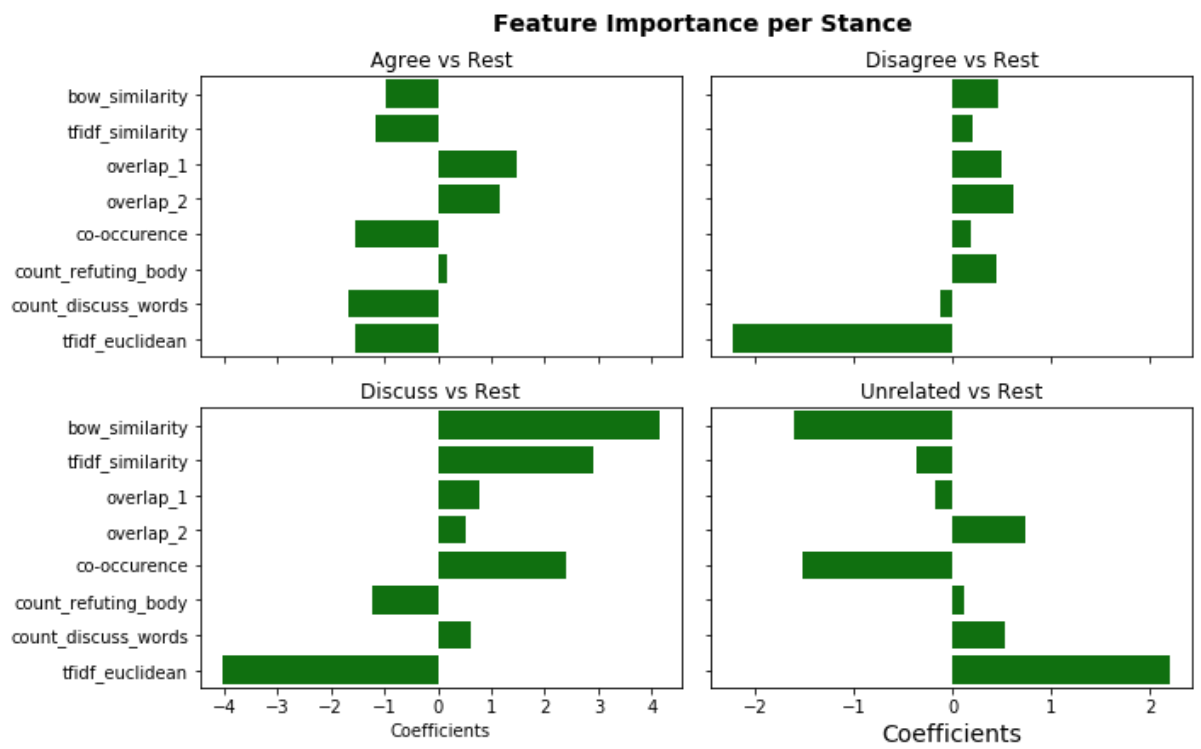
# CHAPTER 6: FINDINGS AND CONCLUSION

## 6.1 Findings

Based on the analyses of the features for stance detection model, the following figure shows the feature importance for each category of class. The figure below also shows the training process of the baseline model of stance detection.



**Fig.7.** Process of training the baseline model on various features.



Classifier	Set	Accuracy	Weighted F1-score
Random Forest	Validation	81.37	86.71
Random Forest	Test	77.07	83.60
SVM	Validation	78.13	84.05
SVM	Test	75.22	81.79
MLP	Validation	79.32	85.57
MLP	Test	75.50	83.27
Random Forest	Validation	81.98	87.36
Random Forest	Test	76.53	83.53
SVM	Validation	79.30	85.33
SVM	Test	76.07	82.91
MLP	Validation	80.23	86.98
MLP	Test	75.64	83.84

**Figure 8. Baseline Model Accuracy Stance Detection**

## Clickbait Detection:

The collinearity matrix given below defines which features can be dropped from model training process thereby defining their relative priority.

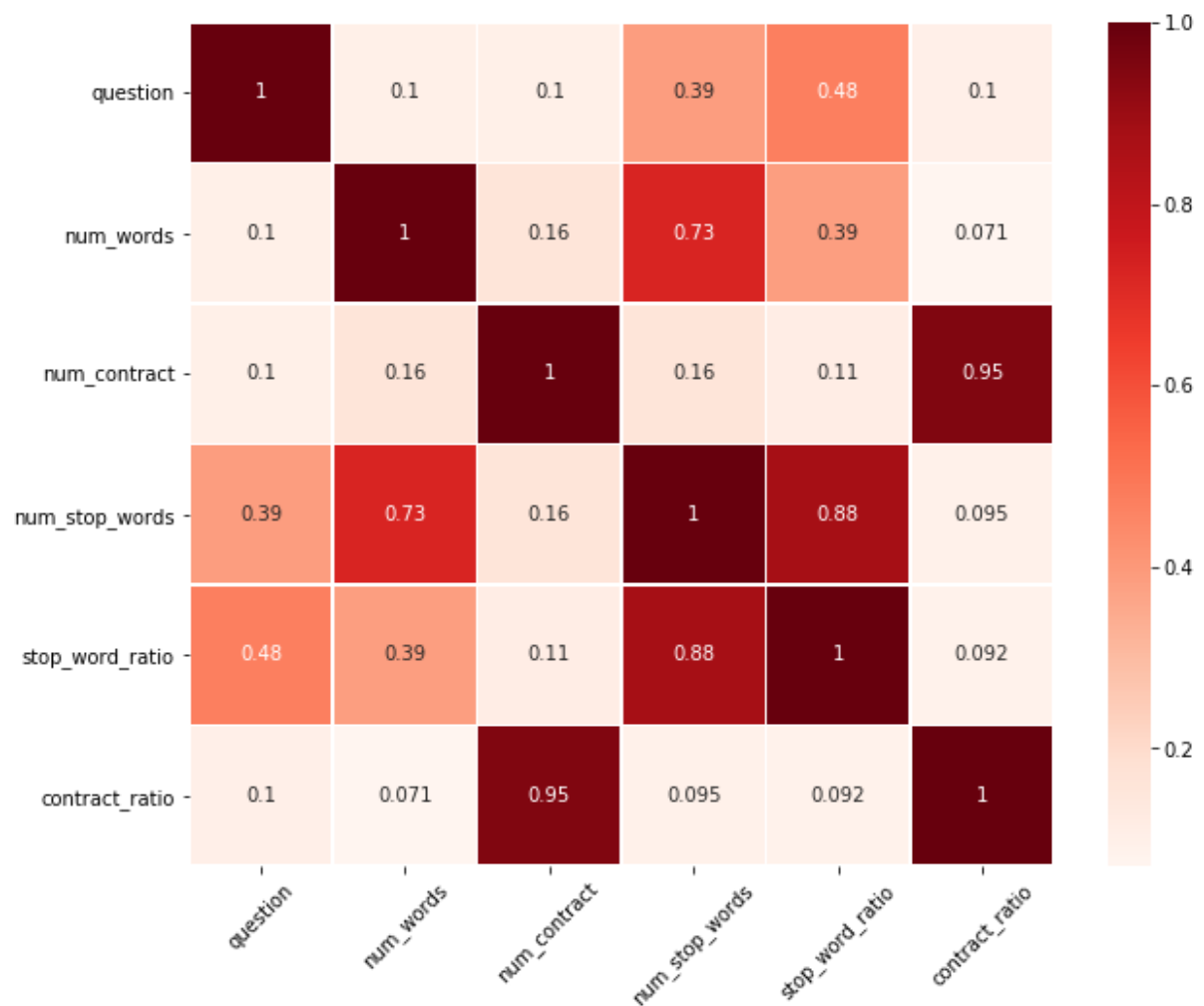


Figure 9. Collinearity Matrix for features of clickbait models

#### Baseline Evaluation Metric-SVM

	precision	recall	f1-score	support
0	0.79	0.89	0.83	3204
1	0.87	0.76	0.81	3196
avg / total	0.83	0.82	0.82	6400

#### Baseline Evaluation Metric-Random Forest

	precision	recall	f1-score	support
0	0.93	0.97	0.95	3204
1	0.97	0.93	0.95	3196
avg / total	0.95	0.95	0.95	6400

**Baseline Evaluation Metric-Logistic Regression**

	precision	recall	f1-score	support
0	0.97	0.98	0.98	3204
1	0.98	0.97	0.98	3196
avg / total	0.98	0.98	0.98	6400

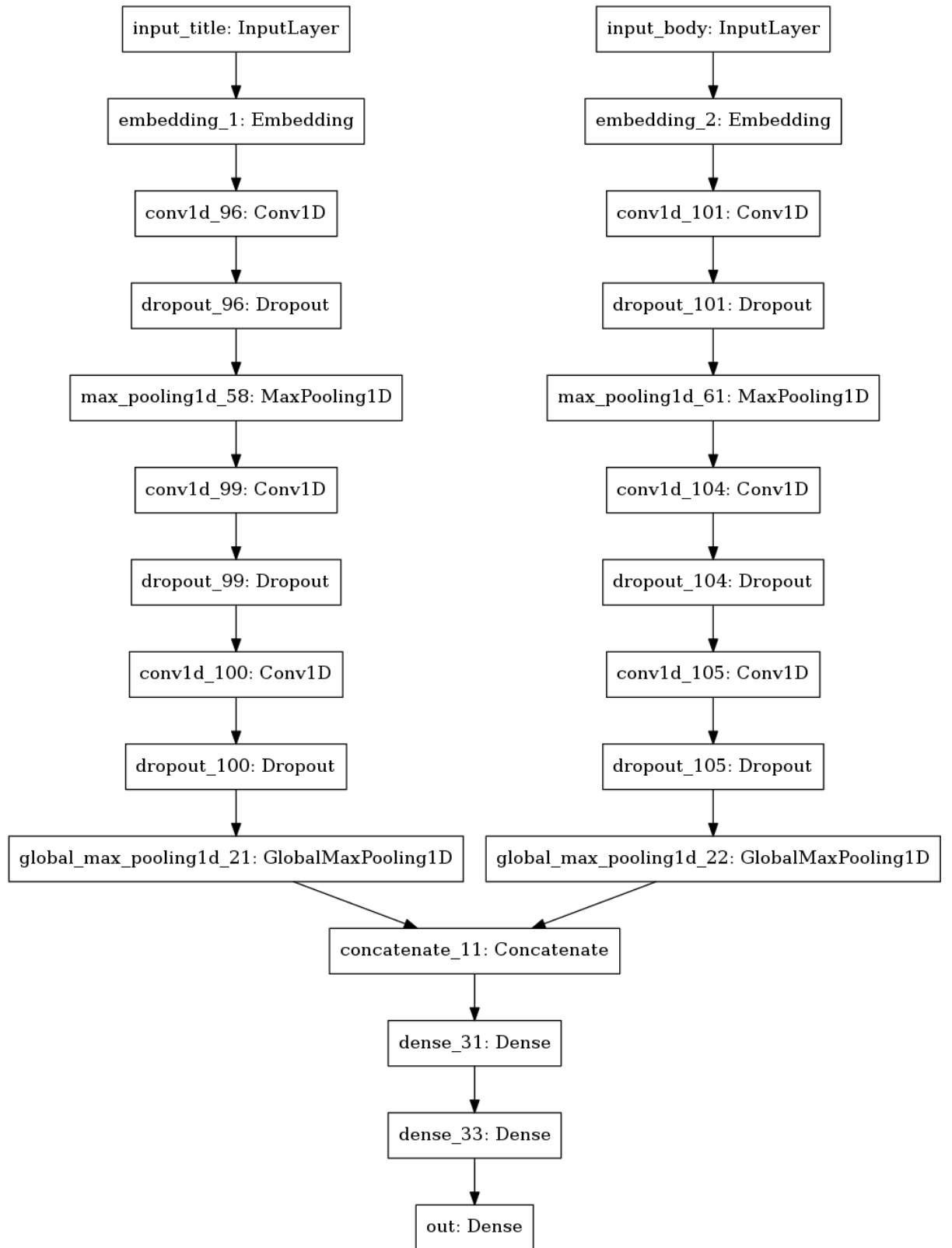
**Baseline Evaluation Metric-MLP**

	precision	recall	f1-score	support
0	0.97	0.98	0.98	3204
1	0.98	0.97	0.98	3196
avg / total	0.98	0.98	0.98	6400

**Table 3.** Results of hyperparameter Tuning (Epoch Loss) for LSTM, BiLSTM Clickbait detection Models

Model Type	Activation Function	e=20(%)	e=50(%)	e=100(%)
LSTM	ReLU	91.25	92.45	94.56
BiLSTM	ReLU	89.03	93.90	95.01
LSTM	Tanh	94.07	95.00	95.35
BiLSTM	Tanh	95.89	96.46	96.82
LSTM	Softmax	96.95	97.05	98.04
BiLSTM	Softmax	98.68	99.23	99.08






**Figure 8 showing the structure of BiLSTM with highest accuracy**


## Fake Busters-Web Application and Clickbait module

The given sequence of screenshots of the web application show the working of the app. The first screen prompts the user to enter the headline or the url which is followed the processing of the entered articles. The final screen shows a popup for either fake or not fake news along with stances for the referred sources.


 Fake Busters

Check your facts before you slip on them  
Validate your article claims against our machine learning system to predict its credibility

Sources ▼


 Search

\*\*\* We'll use this claim to find similar articles and see where they stand against this claim.

 Fake Busters

Check your facts before you slip on them  
Validate your article claims against our machine learning system to predict its credibility

Sources ▼


 Search

Checking your article...

\*\*\* We'll use this claim to find similar articles and see where they stand against this claim.

Check your facts before you slip on them  
Validate your article claims against our machine learning system to predict its credibility

<https://timesofindia.indiatimes.com> Sources Search



**Not fake news!**

Based on the sources we checked and referenced this article against, it is most likely credible!


OK

Sources:	Stance:	Article:
ABC News	Discuss	<a href="http://www.abc.net.au/news/2017-09-05/sydney-resid...">http://www.abc.net.au/news/2017-09-05/sydney-resid...</a>
military-technologies.net	Unrelated	<a href="http://www.military-technologies.net/2017/08/18/uav-...">http://www.military-technologies.net/2017/08/18/uav-...</a>

Your source: [indiatimes.com](http://indiatimes.com)


Sources we used to predict your article:

Sources:	Stance:	Article:
ABC News	Discuss	<a href="http://www.abc.net.au/news/2017-09-05/sydney-resid...">http://www.abc.net.au/news/2017-09-05/sydney-resid...</a>
military-technologies.net	Unrelated	<a href="http://www.military-technologies.net/2017/08/18/uav-...">http://www.military-technologies.net/2017/08/18/uav-...</a>
Federal News Radio	Unrelated	<a href="https://federalnewsradio.com/u-s-news/2017/09/what-...">https://federalnewsradio.com/u-s-news/2017/09/what-...</a>
Anderson Observer	Unrelated	<a href="http://andersonobserver.com/news/2017/9/16/hurrica...">http://andersonobserver.com/news/2017/9/16/hurrica...</a>
USA Today	Discuss	<a href="https://www.usatoday.com/story/money/2017/09/08/h...">https://www.usatoday.com/story/money/2017/09/08/h...</a>
Los Angeles Times	Discuss	<a href="http://www.latimes.com/nation/la-updates-hurricane-ir...">http://www.latimes.com/nation/la-updates-hurricane-ir...</a>
Global News	Discuss	<a href="http://globalnews.ca/news/3682835/republican-doubts-...">http://globalnews.ca/news/3682835/republican-doubts-...</a>
Business Insider	Discuss	<a href="http://www.businessinsider.com/ap-gop-doubts-and-an...">http://www.businessinsider.com/ap-gop-doubts-and-an...</a>
Bloomberg Business	Discuss	<a href="https://www.bloomberg.com/news/articles/2017-09-07...">https://www.bloomberg.com/news/articles/2017-09-07...</a>
Economic Times	Discuss	<a href="http://economictimes.indiatimes.com/news/internation...">http://economictimes.indiatimes.com/news/internation...</a>
KHOU	Discuss	<a href="http://www.khou.com/news/eclipse/coolest-thing-ever-...">http://www.khou.com/news/eclipse/coolest-thing-ever-...</a>

 **Fake Busters**

Validate your article claims against our machine learning system to predict its credibility

<http://www.fakingnews.com> Sources Search



**Fake news!**

Based on the sources we checked and referenced this article against, it is most likely not credible!

OK

Your source: [fakingnews.com](http://fakingnews.com)

Sources we used to predict your article:

## Implementation Guide

### Clickbait Detection Approach

- Dataset creation using custom web scrapper
- Labelling the dataset
- Load the dataset in the dataframe & apply feature selection
- Build model and deploy the web app
- For each url scrape data & feed to model
- Based on features in the text, show % of clickbait

### Clickbait Headlines Examples

We Can Tell You Who You Should friend With Just Four Questions

23 Things You Probably Shouldn't Say To Someone

17 Thoughts We All Had While taking the Exams

How Well Do You Remember Season 3 Of "Silicon Valley"

Adele Said A Mic Dropped In The Piano And That's Why Her Grammy Performance Was Off

# **Headline:**

Things you don't know about Kate Middleton

# **Result: Baity**

Probability of Clickbait = 99.8%

Keep Calm and **Check Clickbait!!**

## 6.2 Conclusion

Social media and the rapidly increasing menace of fake news dissemination are two sides of the same coin that are highly correlated. Our research work conducted on the dataset created is the proof of concept that deep learning models can provide an essential pathway towards successful flagging of fabricated news in the near future. In this research, BiLSTM with attention achieved an accuracy of 98.3% thereby proving to be fruitful in detecting the false news.

## 6.3 Future Work

In the future, we wish to implement synthetic gradients for training the neural networks to greatly reduce the training time for the models. In addition, to this we want to deploy the application on web servers that can be accessed publically, and pitch the idea in a start-up submit to receive adequate funding, thereby fighting the battle against fake news on all platforms. Currently, the scope of study is restricted to detecting fake news for news media and doesn't cover all the social media platforms to lack of adequate system for processing massive amount of data in real time on current 8GB RAM systems.

# REFERENCES

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.
2. Egele, M., Stringhini, G., Kruegel, C., & Vigna, G. (2017). Towards detecting compromised accounts on social networks. *IEEE Transactions on Dependable and Secure Computing*, 14(4), 447-460.
3. Chopra, S., Jain, S., & Sholar, J. M. (2017). Towards Automatic Identification of Fake News: Headline-Article Stance Detection with LSTM Attention Models.
4. Pfohl, S., Triebe, O., & Legros, F. (2017). Stance Detection for the Fake News Challenge with Attention and Conditional Encoding.
5. Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*.
6. Rocktäschel, T., Grefenstette, E., Hermann, K. M., Kočiský, T., & Blunsom, P. (2015). Reasoning about entailment with neural attention. *arXiv preprint arXiv:1509.06664*.
7. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
8. Ferreira, W., & Vlachos, A. (2016). Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 1163-1168).
9. Chaudhry, A. K., Baker, D., & Thun-Hohenstein, P. Stance Detection for the Fake News Challenge: Identifying Textual Relationships with Deep Neural Nets.
10. Zhang, J., Kawai, Y., Nakajima, S., Matsumoto, Y., & Tanaka, K. (2011, January). Sentiment bias detection in support of news credibility judgment. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on* (pp. 1-10). IEEE.
11. Musgrove, T., Walsh, R., & Ridge, P. (2011, September). Automated profiling of the balance of optimism and pessimism in online news content. In *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on* (pp. 1-6). IEEE.
12. Egele, M., Stringhini, G., Kruegel, C., & Vigna, G. (2017). Towards detecting compromised accounts on social networks. *IEEE Transactions on Dependable and Secure Computing*, 14(4), 447-460.
13. Gupta, A., Lamba, H., Kumaraguru, P., & Joshi, A. (2013, May). Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 729-736). ACM.
14. Saez-Trumper, D. (2014, September). Fake tweet buster: a webtool to identify users promoting fake news on twitter. In *Proceedings of the 25th ACM conference on Hypertext and social media* (pp. 316-317). ACM.
15. Corney, D., Albakour, D., Martinez-Alvarez, M., & Moussa, S. (2016, March). What do a million news articles look like?. In *NewsIR@ ECIR* (pp. 42-47)
16. Blom, J. N., & Hansen, K. R. (2015). Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76, 87-100.
17. Biyani, P., Tsioutsoulis, K., & Blackmer, J. (2016, February). "8 Amazing Secrets for Getting More Clicks": Detecting Clickbaits in News Streams Using Article Informality. In *AAAI* (pp. 94-100).
18. Rony, M. M. U., Hassan, N., & Yousuf, M. (2017, July). Diving Deep into Clickbaits: Who Use Them to What Extents in Which Topics with What Effects?. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 232-239). ACM.
19. Abdi, F. D., & Wenjuan, L. MALICIOUS URL DETECTION USING CONVOLUTIONAL NEURAL NETWORK.
20. *Survey on malicious web pages detection techniques* D. R. Patil and J. Patil [available]- <https://pdfs.semanticscholar.org/1f67/b724614a50f90688d1db450eed7916d6dad8.pdf>
21. *An empirical analysis of phishing blacklists*. S. Sheng, B. Wardman, G. Warner, L. F. Cranor, J. Hong, and C. Zhang [available]- <http://repository.cmu.edu/cgi/viewcontent.cgi?article=1286&context=hcii>
22. Sinha, S., Bailey, M., & Jahanian, F. (2008, October). Shades of Grey: On the effectiveness of reputation-based "blacklists". In *Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference on* (pp. 57-64). IEEE.
23. Hu, Z., Chiong, R., Pranata, I., Susilo, W., & Bao, Y. (2016, July). Identifying malicious web domains using machine learning techniques with online credibility and performance data. In *Evolutionary Computation (CEC), 2016 IEEE Congress on* (pp. 5186-5194). IEEE.
24. Egele, M., Stringhini, G., Kruegel, C., & Vigna, G. (2017). Towards detecting compromised accounts on social networks. *IEEE Transactions on Dependable and Secure Computing*, 14(4), 447-460.
25. Vanhoenshoven, F., Nápoles, G., Falcon, R., Vanhoof, K., & Köppen, M. (2016, December). Detecting malicious urls using machine learning techniques. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on* (pp. 1-8). IEEE.
26. Cavalcanti, R. D., Lima, P. M., De Gregorio, M., & Menasche, D. S. (2017, May). Evaluating weightless neural networks for bias identification on news. In *Networking, Sensing and Control (ICNSC), 2017 IEEE 14th International Conference on* (pp. 257-262). IEEE.
27. Ehsanfar, A., & Mansouri, M. (2017, June). Incentivizing the dissemination of truth versus fake news in social networks. In *System of Systems Engineering Conference (SoSE), 2017 12th* (pp. 1-6). IEEE.

28. Ogawa, T., Ma, Q., & Yoshikawa, M. (2010, August). Stakeholder Mining and Its Application to News Comparison. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 440-443). IEEE.
29. E. Lee. (2013). Associated press Twitter account hacked in market-moving attack [Online]. Available: <http://www.bloomberg.com/news/2013-04-23/dow-jones-drops-recovers-after-false-report-on-ap-twitter-page.html>
30. <http://research.signalmedia.co/newsir16/signal-dataset.html>
31. 'Cross-validation: evaluating estimator performance' Accessed on 2 December 2017 [Online] Available: [http://scikit-learn.org/stable/modules/cross\\_validation.html](http://scikit-learn.org/stable/modules/cross_validation.html)
32. Ogawa, T., Ma, Q., & Yoshikawa, M. (2010, August). Stakeholder Mining and Its Application to News Comparison. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 440-443). IEEE.
33. <https://nlp.stanford.edu/projects/glove/>
34. 'Long short-term memory' 2017, Wikipedia, wiki article, Accessed on 14 November 2017 [Online] Available: [https://en.wikipedia.org/wiki/Long\\_short-term\\_memory](https://en.wikipedia.org/wiki/Long_short-term_memory)
35. 'Support vector machine' 2017, Wikipedia, wiki article, Accessed on 20 November 2017 [Online] Available: [https://en.wikibooks.org/wiki/Support\\_Vector\\_Machines](https://en.wikibooks.org/wiki/Support_Vector_Machines)
36. 'Random Forest' 2017, Wikipedia, wiki article, Accessed on 1 December 2017 [Online] Available: [https://en.wikipedia.org/wiki/Random\\_forest](https://en.wikipedia.org/wiki/Random_forest)
37. Analytics Vidhya Content Team 2016, 'A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python)' Accessed on 23 November 2017 [Online] Available: <https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/#nine>
38. E. Lee. (2013). Associated press Twitter account hacked in market-moving attack [Online]. Available: <http://www.bloomberg.com/news/2013-04-23/dow-jones-drops-recovers-after-false-report-on-ap-twitter-page.html>
39. <https://developer.twitter.com/en/docs/basics/rate-limiting>
40. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014 [Online]- <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf>
41. 'Cross-validation: evaluating estimator performance' Accessed on 2 December 2017 [Online] Available: [http://scikit-learn.org/stable/modules/cross\\_validation.html](http://scikit-learn.org/stable/modules/cross_validation.html)
42. <http://research.signalmedia.co/newsir16/signal-dataset.html>
43. Mozscape API overview[accessed on 12 March 2018]online- <https://moz.com/help/guides/moz-api/mozscape/overview>
44. TextMining, "Text Mining Online", [Online] Available :<http://textminingonline.com/dive-into-nltk-part-ii-sentence-tokenize-and-word-tokenize>
45. N-Grams [Online] Available: <https://lagunita.stanford.edu/c4x/Engineering/CS-224N/asset/slp4.pdf>