FAKE NEWS DETECTION USING MACHINE LEARNING

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
SOFTWARE ENGINEERING

Submitted by:

Aayush Ranjan

2K16/SWE/01

Under the supervision of

ASSOCIATE PROF. MANOJ KUMAR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi- 110042

JULY, 2018

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi- 110042

CANDIDATE'S DECLARATION

I, Aayush Ranjan, Roll No. 2K16/SWE/01 of M.Tech. Software Engineering, hereby declare that the project Dissertation titled "Fake News Detection Using Machine Learning" which is submitted by me to the Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or similar title or recognition.

Place: Delhi **Aayush Ranjan**

Date: July 29, 2018

COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi- 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Fake News Detection Using Machine Learning", which is submitted by Aayush Ranjan, Roll No. 2K16/SWE/01, Software Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi **ASSOCIATE PROF. MANOJ KUMAR**

Date: July 29, 2018 **SUPERVISOR**

ACKNOWLEDGEMENT

I would like to express my deep gratitude to the Almighty, who bestowed ability and strength in me to complete this work. I also thank my parents and friends for their unceasing encouragement and support. They have always guided me to work on right path of life.

I owe a profound gratitude to my project guide Associate Prof. Manoj Kumar for his expert, sincere and valuable guidance that helped me to develop a new insight into my project. His calm, collected and professionally impeccable style of handling situations not only steered me through every problem, but also helped me to grow as a matured person.

I play on record, my sincere gratitude to the people of Compute Centre, DTU for providing me with all necessary facilities.

Finally, I would like to express my gratitude to all the people who directly, or indirectly, have lent their helping hand in this venture.

Date: July 29, 2018 **AAYUSH RANJAN**

ABSTRACT

The problem of Fake news has evolved much faster in the recent years. Social media has dramatically changed its reach and impact as a whole. On one hand, it's low cost, and easy accessibility with rapid share of information draws more attention of people to read news from it. On the other hand, it enables wide spread of Fake news, which are nothing but false information to mislead people. As a result, automating Fake news detection has become crucial in order to maintain robust online and social media. Artificial Intelligence and Machine learning are the recent technologies to recognize and eliminate the Fake news with the help of Algorithms.

In this work, Machine-learning methods are employed to detect the credibility of news based on the text content and responses given by users. A comparison is made to show that the latter is more reliable and effective in terms of determining all kinds of news. *The method applied in this work is highest posterior probability of tokens in the response of two classes*. It uses frequency-based features to train the Algorithms including Support Vector Machine, Passive Aggressive Classifier, Multinomial Naïve Bayes, Logistic Regression and Stochastic Gradient Classifier. This work also highlights a wide-range of features established recently in this area that gives a clearer picture for the automation of this problem. I have conducted an experiment in this work to match the lists of Fake related words in the text of responses, to find out whether the response based detection is a good measure to determine the credibility or not. The results were found to be very promising and have scope for more research in the area. Linear SVM and Stochastic Gradient Classifier algorithm with Tf-Idf vector achieved Accuracy and ROC Area under curve above 90% and 95% respectively. This work can be used as a significant building block for determining the veracity of Fake news.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TFIDF – Term Freqency Inverse Document Frequency

SVM – Support Vector Machine

SGD- Stochastic Gradient Descent

PAC: Passive Aggressive Classifier

TP: True Positive

FP: False Positive

TN: True Negative

FN: False Negative

CHAPTER 1

INTRODUCTION

With the advancement of technology, information is freely accessible to everyone. Internet provides a huge amount of information but the credibility of information depends upon many factors. Enormous amount of information is published daily via online and print media, but it is not easy to tell whether the information is a true or false. It requires a deep study and analysis of the story, which includes checking the facts by assessing the supporting sources, by finding original source of the information or by checking the credibility of authors etc. The fabricated information is a deliberate attempt with the intent in order to damage/favor an organization, entity or individual's reputation or it can be simply with the motive to gain financially or politically []. "Fake News" is the term coined for this kind of fabricated information, which misleads people. During the Indian election campaigns, we find many such fabricated posts, news articles and morphed pictures circulating on the social media.

In the recent years, a considerable amount of research has been conducted in this area with satisfactory results. With the success and growth of Artificial Intelligence and Machine Learning, technology has relieved human from extraneous efforts. Fake news detection using these technologies can save the society from unnecessary chaos and social unrest.

"***The objective of this project is to build a classifier that is able to predict whether the users claim is fake or real.”*** This project "Fake News Detection System" uses machine learning algorithms and natural language processing techniques. Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to learn with data, without being explicitly programmed [3]. Natural –language processing is an area of computer science and artificial intelligence concerned with interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data [4].

One of the Earlier works [5] was based on text classification on article's body and headlines. The drawback of this approach is that tokens, which are determined with higher posterior probability in two classes, does not necessarily be categorized as important words of those classes because Fake news can be well written with tokens that appeared as important ones in Real class. Hence, a more effective approach is if higher posterior probability is used on responses given by the users rather than body's article.

Social media is used for rapidly spreading false news these days. A famous quote from Wiston Churchill goes by "A lie gets halfway around the world before the truth has a chance to get its pants on." With a large size of active users on social media, the rumors/fake stories spread like a wildfire. Response on such kind of news can prove to be a decisive factor to term the news as 'fake' or 'real'. User provides evidences in the form of multimedia or web links to support or deny the claim. Classification based on this approach would be significant step in this direction. To support this argument, I performed an experiment related to the occurrence of Fake related words in the collection of responses. Section 6.2.2 discusses about this experiment.

CHAPTER 2

LITERATURE SURVEY

Research on fake news detection is a recent phenomenon and is gaining importance everyday due of its huge negative impact on social and civic engagement. In this section, I have reviewed some of the published works in this area.

2.1 Impact of Fake News

Wang et al. [] in his journal says that the plague of fake news not only creates lack of trust in news media but also turbulence in political world. Fake news influences people's decisions regarding whom to vote for during elections. According to the researchers at the Oxford Internet Institute, in the run up to 2016 US Presidential election, Fake news was prevalent and spread rapidly with the help of social media bots [16]. A social bot refers to an account on social media that is programmed to produce content and interact with humans or other malicious bots [6]. Studies reveal that these bots influenced the election online discussions largely [1]. Fake news hinders serious media coverage and makes it more difficult for journalists to cover important news stories [7]. An analysis done by Buzzfeed revealed that the top 20 Fake news stories about the 2016 US Presidential election received more attention on Facebook than the top 20 election stories from 19 major media outlets [8].

Deaths are frequently caused by Fake news. People have been physically attacked over fabricated stories spread on the social media. In Myanmar, the people of Rohingya were arrested, jailed, and in some cases even raped and killed because of Fake news [9]. These attempts seem to have created real life fears and have affected the civic engagement and community conversations.

2.2 Combating Fake News through Machine Learning

Combating fake news is a difficult task. To accomplish whether a news article is a fake by checking the truth of each fact manually is no cakewalk because the truth of the facts exists in continuum and depends heavily upon the nuances of human language, which are difficult to parse in true/false dichotomies [10]. Sloppy written material with grammatical mistakes, may suggest the article is not written by any journalist and can probably be false. The news published/ broadcasted by a unknown media house or newspaper can possibly be Fake news but these factors do not give assurance and therefore definitions and types of Fake news must be properly understood and categorized.

2.2.1 Definitions and its types

Fake news is news that is intentionally and verifiably false and has the potential to mislead viewers/readers. There are two important dimensions of this definition: "intention" and "authenticity". First, fake news propagates misinformation that can be verified. Second, Fake news is created with dishonest intention to mislead public. This definition is widely adopted in recent research analysis [11; 12; 13; 14]. In general, Fake news can be categorized into three groups. In first group - "**Actual Fake News**", we can put those types of news, which are false and made up by the author of the article. The second group –"**Fake news that is actually satire**" is created purely to amuse rather than mislead its audience. Therefore, intentionally misleading and deceptive fake news is different from obvious satire or parody. The third group is "**Poorly reported news that fits an agenda**". This type of news has some real content but is not entirely correct and is designed especially for some political propaganda.

Many researchers have streamlined the types of Fake news to simplify their research. For instance, According to definitions given by [1], there are a few types of news that cannot be called as "Fake"- (1) Satire news having proper context. (2) Misinformation that is created unintentionally. (3) Conspiracy theories those are difficult to put in true/false dichotomies. This paper [1] has presented two main aspects of fake news detection problem: "characterization" and "detection".

2.2.2 Fake news foundations

People who tend to believe their perceptions of reality as only accurate view can believe fake news as true. They think that those who disagree with them are biased and irrational [15]. Also, people who prefer to receive news that confirm their existing belief and views are mostly biased [16], while others are people who are socially conscious and choose a safer side while consuming and discriminating news following the norms of the community, even if the news shared is Fake. These psychological and social human behavioral patterns are the two main foundations of Fake news in the Traditional media. Along with these two factors, malicious twitter bots serves as the foundations of Fake news in Social media [1].

2.2.3 Related Work

According to various researches conducted in this area, Fake news detection methods comprise of four basic types – *Knowledge Based, Style based, Stance based and Visual based*. This section elucidates research in all these types of detection methods and a few other important researches that have received higher recognition. It also presents some of the important features that were used recently in various research papers to determine the credibility of news. The feature extraction is the crucial phase of Machine learning. Table 2.1 shows all these features categorized based on different context.

2.2.3.1 Fake News Detection Methods

Knowledge Based Detection: It aims to use external sources to fact-check the claims made in the news content. Two typical external sources are open web and knowledge graph. Open web sources are compared to the claims in terms of consistency and frequency [18, 19], whereas Knowledge graph is used to check whether the claims can be inferred from existing facts in graph or not [20, 21, 22].Many fact-checking websites (For eg. AltNews, Snopes, Smhoaxslayer, Boomlive) are using domain experts to determine manually the news veracity. Facebook has recently partnered with Indian fact-checking agency Boomlive to spot false news circulation on its website [23]. A problem pertaining to this method is *automated fact-*

checking which is associated with classification of sentences into non-factual, unimportant factual and check-worthy factual statements [24, 27].

Table 2.1 Features applied in previous works to detect fake news [25, 26, 27, 1, 37, 38, 35]

	Linguistic based features	**Visual based features (Images and Videos)**		
News Context	total words, frequency of large words, frequency of unique words, ngrams, bag of word approaches(count, tf-idf, word2vec), no of punctuations(question mark, exclamation mark), no of quoted words , no of external links, no of graphs, average length of graphs, PCFG.	*Visual features*		*Statistical features*
		Clarity score, coherence score, diversity and clustering score, similarity distribution histogram		count, image ratio, multi-image ratio, hot image ratio and long image ratio
Social Context	**User based features (to detect Twitter bots / Fake profiles)**	**Post based features**		**Network based features**
	Individual Level	*Post level*	No. of smiling emoticon, 1st/2nd/3rd pronouns, slangs, readability, WOT score and all the linguistic and embedding features can be applied here.	similarity features between the relevant tweets, following/follower of user who posted tweets, trajectory of spread of news, degree and clustering coefficient, no of users who write posts relevant to same news article etc.
	registration age, no of followers/following, no of tweets authored, difference b/w account creation and relevant tweet authored, has profile image, has a URL, has bio description, has location			
	Group Level	*Group Level*	wisdom of crowd – aggregates feature values of all the relevant posts for the specific news article, average credibility score, amount of disagreement present in conversations, stance features	
	percentage of verified users, average no of followers, author of first tweet in the thread is verified or not			
		Temporal Level	Temporal features (over time) for account age, difference between account age and tweet publication time, author followers/friends/statuses, and the number of tweets per minute.	
Source context	reputation of the source (publisher/author), website registration behavior, internet site age of the publishers			
Similarity based	Jaccard similarity, Cosine similarity			
Other features	probability of news disappearance			

Style Based Detection: Style based detection focuses on the way the content has been presented to the users. Fake news is generally not written by journalist, that being said the style of writing might differ [9]. In [35] the author has implemented deep syntax models using PCFG (Probabilistic Context Free Grammars) to transform sentences into rules like lexicalized/unlexicalized production rules and grandparent rules, which describes syntax structure for deception detection. Another paper [32] implemented deep network models - Convolutional neural networks (CNN) to check the veracity of news. Fake articles sometimes show extreme behavior in favor of a political party. This type of writing style is called as hyper-partisan styles [39]. Linguistic based features can be applied to check this kind of writing style. In some of article's headlines, there is just enough information to make readers curious to go to a certain webpage or video. This type of eye-catching headlines or web links is called as click-bait headlines [1], which can be a source of Fake news.

Style based methods also covers methods which finds out tokens with higher posterior probability in two classes, using word embedding features.[5]used Naïve Bayes algorithm to obtain tokens that were found to be most indicative on the classification and used it for deep learning and logistic regression. They combined the hypothesis obtained from Naïve Bayes, SVM and Logistic regression and observed the average accuracy of 83% on their training set. Although writing styles can largely contribute to detecting fake news but it seem to be less efficient because, Fake news can be written in a style similar to that of real news [10].

Stance based detection: This method compares how a series of posts on social media or a group of reputable sources feels about the claim -Agree, Disagree, Neutral or is Unrelated. In [10], the authors used lexical as well as similarity features fed through a multi-layer perceptron (MLP) with one hidden layer to detect the stance of the articles. They hard-coded reputation score feature (Table 2.1) of various sources based on nationwide research studies. Their model achieved 82% accuracy for pure stance detection on their dataset. Another paper [13] used "wisdom of crowd" feature to improve news verification by discovering conflicting viewpoints in micro blogs with the help of topic model method - Latent Dirichlet Allocation (LDA). Their overall news veracity accuracy reached up to 84%.

Visual Based Detection on Social Media: Digitally altered images are everywhere circulating on social media like a wildfire. Photoshop can be used freely these days to modify images adequately enough to fool people into thinking they are seeing the real picture. The field of multimedia forensics has produced a considerable number of methods for tampering detection in videos [40] and images However,[40]mentions several reasons as to why these methods are not likely to work on social media images. There are also few basic techniques on the web for general people to spot photo-shopped images for e.g. Google's reverse image search, Get image metadata etc. [15] has extracted many visual and statistical based features (shown in Table 2.1) that can be used in detecting the authenticity of the multimedia.

Other related works: [3] implemented Document similarity analysis, that calculates the Jaccard similarity, a widely used similarity measure, between a news 'n' in test set with every news in Fake news training set 'F' and real news training set 'R'. The results obtained were very promising. In [16] the authors have exploited the diffusion patterns of information to detect the hoaxes. Many research papers have used different linguistic and word embedding features. The most common ones are tf-idf, word2vec, punctuations, ngrams, PCFG.

CHAPTER 3

METHODOLOGY

3.1. Response based detection

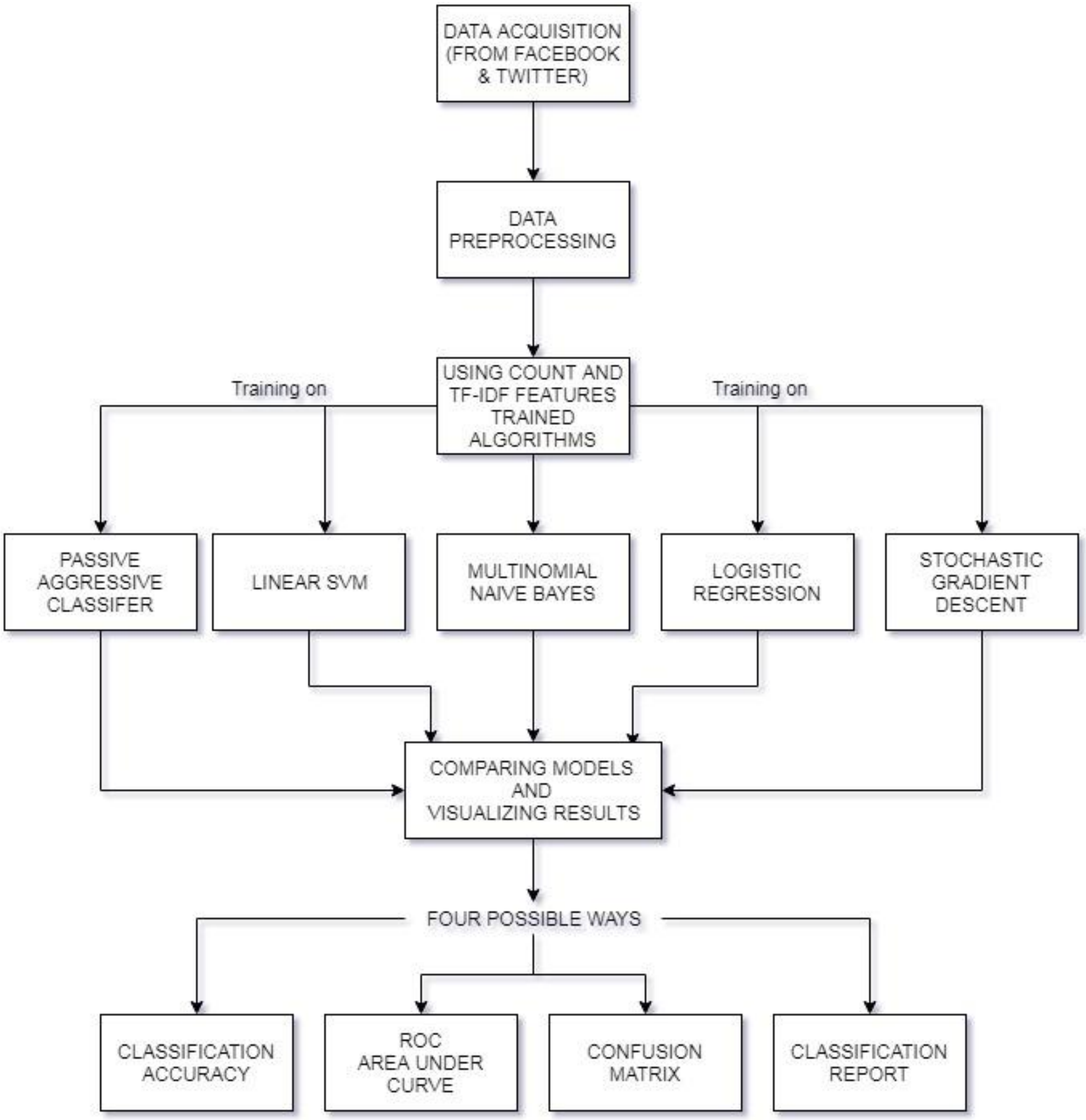
Fake news generally carries strong sentiments and thus circulates in no time on social media. Response based technique takes into consideration the collected responses on tweets/posts to determine the credibility of the news. This project has progressed in two phases. In the First Phase, I implemented the *higher posterior probability* method on the article's body and headlines. Although, I observed higher accuracy results, I found this method to be not very efficient because there is possibility that Fake news can appear in a well-written article. In the second phase, I proposed approach to classify fake news more accurately by *analyzing the response on such news articles"*. Implementation of the same was carried out in five sub phases:

1) Collection of data from social media platform, Facebook and Twitter
2) Choosing relevant features for classification and Training the Model
3) Evaluation of different model performance based on extracted features
4) Improving performance
5) Discussion and Presentation of results

This project was developed in Python using Sci-kit libraries. Python has a huge set of libraries and extensions, which can be easily used in Machine Learning. Sci-Kit Learn [6] library is the best source for machine learning algorithms where nearly all types of machine learning algorithms are readily available for Python, thus easy and quick evaluation of ML algorithms is possible.

3.1.2 Flowchart

Figure 3.1 Flowchart of the method.



CHAPTER 4

DATA COLLECTION AND ANALYSIS

We can get online news from different sources like social media websites, search engine, homepage of news agency websites or the fact-checking websites. On the Internet, there are a few publicly available datasets for Fake news classification like BuzzfeedNews, LIAR, BS Detector, CREDBANK [1] etc. These datasets have been widely used in different research papers for determining the veracity of news. In the following sections, I have discussed in brief about the sources of the dataset used in this work.

4.1 Analysis of two publicly available dataset

In the first phase of this project, which was style based detection on the content/body of the news article; I used two different datasets of varying length and trained the model on each of them. Below are two datasets, which I used.

LIAR: A Benchmark dataset for Fake news detection [32,34]

The original dataset contained 13 columns for train, test and validation files. The training set included 12,386 human-label short statements, sampled from news releases, TV or radio interviews, campaign speeches etc. The data was collected from a Fact-checking website PolitiFact through its API.

For implementation of First phase of the project, I chose only training set file and 2 columns from this file for classification. The other columns can be added later to enhance the performance.

Below were the columns that were used:

Column1: Statement

Column2: Label (True/False)

Classes (labels) were grouped such that in the newly created dataset you find only two labels (True/False) as compared to six present in the original. They were grouped as below:

True	--	True
Mostly-true	--	True
Half-true	--	True
Barely-true	--	False
False	--	False
Pants-fire	--	False

Another dataset obtained from Github [33]

The dataset contained four columns:

i) URL

ii) Headline

iii) Body

iv) Label

This dataset contained 4335 news articles with long body text as compared to short texts in the previous dataset. Average word count of the body in the dataset was 576 words per article. Label was mentioned as 0/1; 0 for Fake news and 1 for Real News. Classification models were trained on this dataset and the performance of the models were compared and best model was chosen. After analyzing this dataset, I found that, fake news were mainly taken from few international fake news websites such as beforeitsnews.com, dailybuzzlive.com, activistpost.com etc., similarly the real news were covered from few main lead newspapers like reuters.com etc.

4.2 Data collection and Analysis for Proposed method

For the proposed method, which is based on Response of the users, I found that none of the publicly available datasets contained Responses. I assembled the required data from Social media websites Twitter and Facebook. There were two main steps of this Data acquisition process:

1) Gathering the Fake and Real news
2) Extracting the Comments and other attributes

4.2.1 Gathering the Fake and Real News:

Fake news collection: I used fact-checking websites in India for this purpose. AltNews.com, Smhoaxslayer.com, Boomlive.com are some of the agencies, which are authentic and recognized for busting Fake news [bbc]. I analyzed the articles posted by them debunking that Fake news. I looked only for the relevant data needed for the construction of the dataset. The relevant data were especially Tweets and Facebook post by different users, which were busted by Fact-checking agencies as Fake. All the Fake Twitter and Facebook posts url were collected in the initial phase.

Real news collection: This was the easier task. I gathered posts/tweets of few reputed news agencies, media news journalists and even some verified users and groups. I picked the news, which carried strong sentiments (negative as well as positive), seeking higher attention but were real. Thus, the dataset created, held resemblance between Fake & Real news in term of gathering attention. This was of course a significant step to measure the performance of the model, because responses to the news with negative sentiment can make users believe that it is Fake.

Total 132 news items were collected for the dataset, out of which 69 were classified as Fake news and 63 as Real news. I intentionally chose to keep the number of news items less but gathered large number response on that news. I picked only those posts on which considerable amount of responses were given. The dataset consisted of 5 columns - ‘users claim’, ‘post/tweet’, ‘url’, ‘comments’ and ‘label’.

4.2.2 Extracting the responses

For each urls of the posts collected, I extracted the comments for the respective posts using Web Scrapping tools in Python – Selenium and Beautiful Soup. With Selenium, we can extract the server version of the page content. Beautiful Soup library on the other hand, cannot do it as it scrapes data from client version of the page. Therefore, Selenium along with Beautiful soup was used to scrape the required data.

I chose first five to six pages of loaded comments to keep the text neither too long nor too short. For convenience, the language of the responses collected was made restricted to English. Facebook has a function called as “Translate all” that converts all the comments to English in one go. In twitter, any Non-English comments have to be translated one by one. Thus, I scrapped comments that were in English or those sentences constructed using English alphabets.

CHAPTER 5

STEPS OF METHOD IMPLEMENTATION

5.1 Text preparation

Social media data is highly unstructured – majority of them are informal communication with typos, slangs and bad-grammar etc. To achieve better insights, it is necessary to clean the data before it can be used for predictive modeling. For this purpose, basic pre-processing was done on the News training data. This step was comprised of

1. **Conversion to Lower case:** First step was to transform the text into lower case, just to avoid multiple copies of the same words. For e.g. while finding the word count, "Response" and "response" is taken as different words.

2. **Removal of Punctuations:** Punctuations does not have much significance while treating the text data. Therefore, removing them helps to reduce the size of overall text.

3. **Stop-words removal:** Stop-words are the most commonly occurring used words in a corpus. These are for e.g. a, the, of, on, at etc. They usually define the structure of a text and not the context. If treated as feature, they would result in poor performance. Therefore, Stop-words were removed from the training data as the part of text cleaning process.

4. **Tokenization**: It refers to dividing the text into a sequence of words or group of words like bigram, trigram etc. Tokenization was done so that frequency-based vectors values could be obtained for these tokens.

5. **Lemmatization:** It converts the words into its word root. With the help of a vocabulary, it does morphological analysis to pick up the root word. In this work, Lemmatization was performed to improve the values of frequency-based vectors.

Text pre-processing was an essential step before the data was ready for analysis. A noise free corpus has a reduced size of the sample space for features thereby resulting in increased accuracy.

5.2 Feature generation

We can use text data to generate a number of features like word count, frequency of large words, frequency of unique words, n-grams etc. By creating a representation of words that capture their meanings, semantic relationships, and numerous types of context they are used in, we can enable computer to understand text and perform Clustering, Classification etc. For this purpose, Word Embedding techniques are used to convert text into numbers or vectors, so that computer can process them.

Word Embedding: A word-embedding format generally tries to map a word to a vector using a dictionary. The following frequency based word embedding vectors was used for training the data. They are also categorized into Linguistic based features.

Count Vector as a feature

Count Vector is a matrix notation of the dataset, in which rows represent the documents in the corpus, columns represent a term from the corpus, and cells represent the count of that particular term in a particular document. The dictionary is created using the list of unique tokens or words in the corpus.

Example: Let us consider three documents in a corpus C, i.e. D1, D2 and D3 containing the text as below:

D1: It was raining heavily yesterday.
D2: Bad weather caused heavy rainfall in London.
D3: Yesterday, London newspapers warned of heavy rainfall.

The dictionary can be created with unique words. The unique words identified are:

[Rain, Heavy, Yesterday, Bad, Weather, London, Newspapers, Warned]

No of Documents D = 3

No of Unique words N = 8

Count Matrix represents the occurrence of every term in every document.

The Count matrix M = 3 X 8 is represented below:

Table 5.1 Showing word document matrix

	Rain	Heavy	Yesterday	Bad	Weather	London	Newspaper	Warned
D1	1	1	1	0	0	0	0	0
D2	1	1	0	1	1	1	0	0
D3	1	1	1	0	0	1	1	1

A column can be called a word vector for the corresponding word in the Matrix M. Word vector for “Yesterday” is [1,0,1]. Count vector outputs all those words or tokens from the highest frequency in the Corpus to the lowest frequency. For e.g. Rain, Heavy has the highest occurrence in the Corpus, so they lead the word list in the dictionary. This feature was used for the proposed method to give the machine learning models idea that which words do social media users often use when they see a Fake or Real news.

TF-IDF vectors as a feature:

TF-IDF weight represents the relative importance of a term in the document and entire corpus.

TF stands for Term Frequency: It calculates how frequently a term appears in a document. Since, every document size varies, a term may appear more in a long sized document that a short one. Thus, the length of the document often divides Term frequency.

$$\boldsymbol{TF(t,d) = \frac{Number\ of\ times\ t\ occurs\ in\ a\ document\ 'd'}{Total\ word\ count\ of\ document\ 'd'}}$$

IDF stands for Inverse Document Frequency: A word is not of much use if it is present in all the documents. Certain terms like “a”, “an”, “the”, “on”, “of” etc. appear many times in a document but are of little importance. IDF weighs down the importance of these terms and increase the importance of rare ones. The more the value of IDF, the more unique is the word.

$$\boldsymbol{IDF(t) = \log_e(\frac{Total\ number\ of\ documents}{Number\ of\ documents\ with\ term\ t\ in\ it})}$$

TF-IDF – Term Frequency-Inverse Document Frequency: TF-IDF works by penalizing the most commonly occurring words by assigning them less weightage while giving high weightage to terms, which are present in the proper subset of the corpus, and has high occurrence in a particular document. It is the product of Term Frequency and Inverse Document Frequency.

$$\boldsymbol{TFIDF(t,d) = TF(t,d) * IDF(t)}$$

TF-IDF is a widely used feature for text classification. In addition, TF-IDF Vectors can be calculated at different levels i.e. Word level and N-gram level, which I have used in this project.

i) *Word level TF-IDF*: Calculates score for every single term in different documents.

ii) *N-gram level TF-IDF*: Calculates score for the combination of N terms together in different documents.

5.3 Algorithms used for classification

This section deals with training the classifier. Different classifiers were investigated to predict the class of the text. I explored specifically five different machine-learning algorithms – Multinomial Naïve Bayes Passive Aggressive Classifier, Logistic regression, Linear Support Vector machines and Stochastic

Gradient Descent. The implementations of these classifiers were done using Python library Sci-Kit Learn.

Brief introduction to the algorithms

Naïve Bayes: This classification technique is based on Bayes theorem, which assumes that the presence of a particular feature in a class is independent of the presence of any other feature. It provides way for calculating the posterior probability.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

P(c|x)= posterior probability of class given predictor
P(c)= prior probability of class
P(x|c)= likelihood (probability of predictor given class)
P(x) = prior probability of predictor

Passive Aggressive Classifier: The Passive Aggressive Algorithm is an online algorithm; ideal for classifying massive streams of data (e.g. twitter). It is easy to implement and very fast. It works by taking an example, learning from it and then throwing it away.

Logistic Regression: Logistic regression is a classification algorithm, used to predict the probability of occurrence of an event (0/1, True/False, Yes/No). It uses sigmoid function to estimate probabilities.

Support Vector Machine: In this algorithm, each data item is plotted as a point in n-dimensional space (n is the number of features). Values of each feature are the value of each co-ordinate. It specifically extracts a best possible hyper-plane or a set of hyper-planes in a high dimensional space that segregates two classes. Linear kernel was used for SVM in this work.

Stochastic Gradient Descent: A SGD algorithm starts at a random point, updates the cost function with each of the iteration using one data point at a time and builds a classifier with progressively higher accuracy given a large dataset. In SGD, a sample of training set or one training value is used to calculate parameters, which are much faster than other gradient descent.

5.4 Metrics used to access the Performance of Model

In this section, I have explored some of the most significant metrics by which a machine learning model performance is measured. These metrics measures how well our model is able to classify or evaluate predictions. The below metrics introduction were used in this project.

Classification Accuracy: It is the most common evaluation metric for classification problems. It is defined as the number of correct predication as against the number of total predictions. However, this metric alone cannot give enough information to decide whether the model is a good one or not. It is suitable when there are equal numbers of observation in every class.

Area under ROC-curve: Area under ROC curve is a performance metric used for binary classifications. It tells a model's ability to disseminate between the two classes. If the Area under curve or AUC is 1.0 then, it means it has made all predictions correctly whereas the AUC of 0.5 is good as the random predictions. ROC can be further classified into Sensitivity and Specificity. A binary Classification problem is a tradeoff between these two factors.

Sensitivity: It is called as "Recall" and is defined as number of instances from the positive class that are actually predicted correctly. This phenomenon is called as True Positive Rate. In this work, "Fake" was selected as positive class and "Real" as negative.

Specificity: It is the number of instances in the negative class that are actually predicted correctly. It is called as True negative rate.

Confusion Matrix: It is also known as Error matrix, which is a table representation that shows the performance of the model. It is special kind of Contingency table having two dimensions- “actual”, labeled on x-axis and “predicted” on y-axis. The cells of the table are the number of predictions made by the algorithm.

Table 5.2 Confusion Matrix

Total Instances		Predicted	
		Yes	**No**
Actual	**Yes**	*True Positive*	*False Negative*
	No	*True Negative*	*True Negative*

True Positives: It is correctly predicted positive values.

True Negatives: It is correctly predicted negative values.

False Positives: It is incorrectly predicted negative values as positive values.

False Negatives: It is incorrectly predicted negative values as positive values.

Classification Report: Scikit-learn provides a convenience report when working on classification problems which outputs precision, recall, F1 score and support for each class.

Precision: Precision is the ratio of correctly predicted positive instances to the total predicted positive instances. High precision means low False Positive rate.

$$Precision = \frac{TP}{TP + FP}$$

Recall (Sensitivity): Recall is the ratio of correctly predicted positive instances to the all instances in actual class - Yes.

$$Recall\ = \frac{TP}{TP + FN}$$

F1-Score: It is the weighted average of Precision and Recall. Therefore, it takes into consideration both false positives and false negatives. F1 score is usually more useful than accuracy, especially when there is uneven class distribution. Accuracy performs best if false positives and false negatives have similar instances or cost. If the cost of false positives and false negatives differs widely, then it is better to look at both Precision and Recall.

$$F1\,Score\ =\ 2 * \frac{(Recall\ *\ Precision)}{(Recall\ +\ Precision)}$$

CHAPTER 6

EXPERIMENT, RESULTS & ANALYSIS

Experiments were performed using the above algorithms using Vector features-Count Vectors and Tf-Idf vectors at Word level and Ngram-level. Accuracy was noted for all models. I used K-fold cross validation technique to improve the effectiveness of the models. In the First phase of my experiment, I applied text classification on the articles body in two different publicly available datasets [][]. In the second phase, Experiment was performed on the responses collected on a set of Fake news and Real news claims extracted from Twitter and Facebook.

6.1 Dataset split using K-fold cross validation

This cross-validation technique was used for splitting the dataset randomly into k-folds. (k-1) folds was used for building the model while k^{th} fold was used to check the effectiveness of the model. This was repeated until each of the k-folds served as the test set. I used 3-fold cross validation for this experiment where 67% of the data is used for training the model and remaining 33% for testing.

6.2 Set of Experiments Conducted

6.2.1 Experiment (Proposed method)

Responses were classified using Count Vector and Tf-Idf vector at two levels:

Word level – Single word was chosen as token for this experiment.
N-gram level – I kept the range of N-gram from 1 to 3 i.e. from one word to at most 3

Words (bigram, trigram), which was considered as token and experiment was performed.

Maximum document frequency was also used in this experiment as a parameter with Tf-Idf vector. This parameter removed all those tokens that appeared in say X% of the Responses. Initially X was set to 0 i.e. no parameter was set but later X was increased with step “0.1” i.e. 10%, and the results were noted down.

Classification Accuracy at Word Level

Table 6.1 Classification accuracy at Word-level

Accuracy	*Linear SVM*	*Stochastic Gradient Descent*	*Passive Aggressive Classifier*	*Multinomial Naïve Bayes*	*Logistic Regression*
Using Count Vector	72.7	86.4	83.3	85.6	86.4
Using Tf-Idf Vector	92.4	91.7	93.2	78	84.1

Classification Accuracy at N-gram Level:

Table 6.2 Classification accuracy at N-gram level

Accuracy	*Linear SVM*	*Stochastic Gradient Descent*	*Passive Aggressive Classifier*	*Multinomial Naïve Bayes*	*Logistic Regression*
Using Count Vector	69.7	89.4	81.8	86.4	85.6
Using Tf-Idf Vector	91.7	90.9	90.9	77.3	81.1

Classification Accuracy at Word level performed better than N-gram level as we can see from the above tables. The accuracy for Multinomial Naïve Bayes with Tf-Idf at N-gram level was the lowest at 77.3% while Linear SVM, Stochastic Gradient Descent and Passive Aggressive Classifier, using Tf-Idf vectors performed well at both levels and their accuracy was above 90%. Since, Classification accuracy alone is not sufficient to determine the effectiveness of the model; other metrics was also explored especially for these three algorithms at word level, using Tf-Idf Vectors. In another experiment, I included the MDM Parameters described above. With the increase of MDM from 0 to 1 in step of 0.1, classification accuracy of the three models increased significantly as depicted by the table below.

Classification Accuracy using MDM (X = 0 to 1) in step of 0.1

Table 6.3 Classification Accuracy using MDM

Classification Accurarcy		*Linear SVM*	*Stochastic Gradient Descent*	*Passive Aggressive Classifier*
Maximum Document Frequency	***0.1***	76.5	78	77.3
	0.2	84.8	86.4	86.4
	0.3	84.1	87.9	88.6
	0.4	85.6	90.9	88.6
	0.5	88.6	88.6	87.1
	0.6	89.4	90.2	90.9
	0.7	93.2	92.4	92.4
	0.8	92.4	91.7	90.9
	0.9	92.4	91.7	91.7
	1	92.4	91.7	91.7

Best performing model was Linear SVM with 93.2% at MDM(X = 0.7) and close to this model was Stochastic Gradient Descent and Passive Aggressive Classifier with 92.4%. Beyond, 0.7 the algorithms did not show improvement. So, MDM with 0.7 was chosen as optimal value.

Henceforth, I obtained the Classification reports including precision, recall, f-score of all three models at MDM(X=0.7)

Classification Error: It means overall, how often the model is incorrect, also called as Misclassification Rate.

Classification Error for Linear SVM-TFIDF = 100– 93.2 = 6.8%

Classification Reports:

Linear SVM-TFIDF

Table 6.4 Classification Report for LinearSVM-TFIDF

Classification Report	***Precision***	***Recall***	***F1-score***	***Support***
Fake	94	92	93	23
Real	92.6	93.3	92.6	21

SGD-TFIDF

Table 6.5 Classification Report for SGD-TFIDF

Classification Report	***Precision***	***Recall***	***F1-score***	***Support***
Fake	93	92	92.33	23
Real	92.6	91.6	92	21

PAC-TFIDF

Table 6.6 Classification Report for PAC-TFIDF

Classification Report	***Precision***	***Recall***	***F1-score***	***Support***
Fake	87.3	92	89.6	23
Real	92	85.3	88.6	21

Precision value for Linear SVM-TFIDF at 94% is higher than SGD-TFIDF, which is 93% and Recall values (Sensitivity) was calculated as 92% for both models.

Confusion Matrix

Linear SVM-TFIDF

TP = 16, TN =24, FP = 1, FN = 3

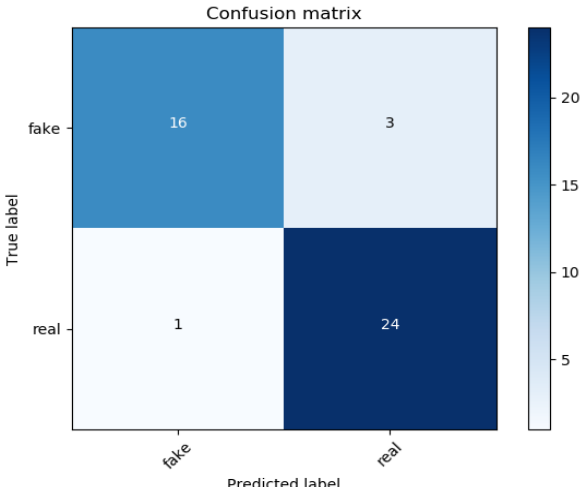


Figure 6.1 Confusion Matrix for Linear SVM-TFIDF, Split 1

TP = 24, TN =18, FP = 1, FN = 1

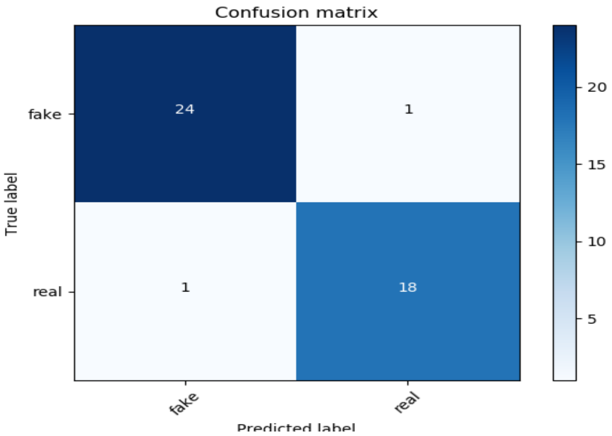


Figure 6.2 Confusion Matrix for Linear SVM-TFIDF, Split 2

TP = 24, TN =17, FP = 2, FN = 1

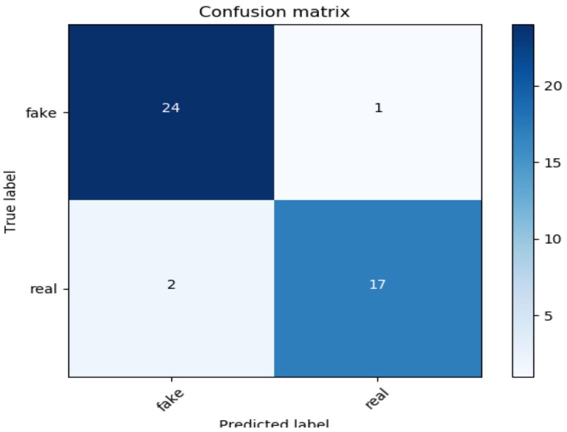


Figure 6.3 Confusion Matrix for Linear SVM-TFIDF, Split 3

SGD-TFIDF

TP = 16, TN =24, FP = 1, FN = 3

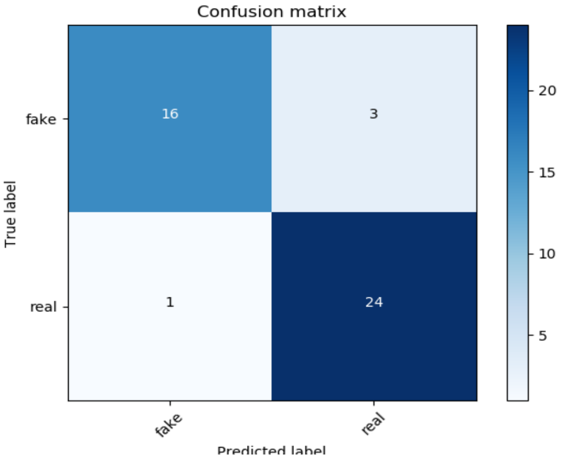


Figure 6.4 Confusion Matrix for SGD-TFIDF, Split 1

TP = 24, TN =18, FP = 1, FN = 1

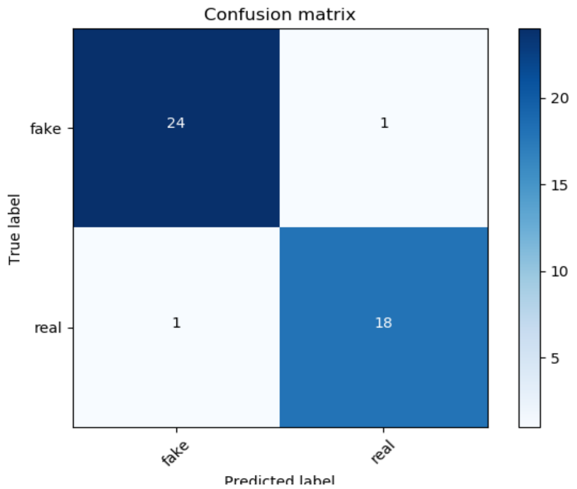


Figure 6.5 Confusion Matrix for SGD-TFIDF, Split 2

TP = 24, TN =16, FP = 3, FN = 1

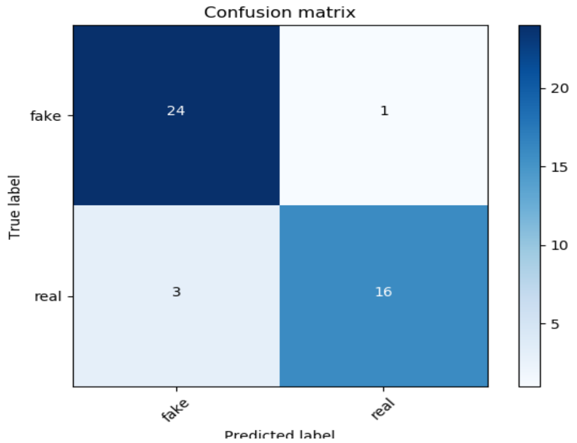


Figure 6.6 Confusion Matrix for SGD-TFIDF, Split 2

Sensitivity tells how sensitive is the classifier to detect fake news, while *Specificity* tells how selective or specific the model in predicting real news is.

Choosing the metric depends on what kind of application is developed. The positive class in this binary classification is class “Fake”. Therefore, Sensitivity should be higher, because False positives are more acceptable than False negatives in classification problems of such applications.

The sensitivity is high for both the models and is having equal value. By optimizing more for Sensitivity, We can get better results.

By decreasing the threshold for predicting fake news, we can increase the Sensitivity of the classifier. This would increase the number of True Positives. In this work, threshold is set to 0.5 by default but we can adjust it to increase sensitivity or specificity depending on what we want.

ROC Curve (Receiver Operating Characteristics Curve)

It is a way to check how various thresholds affect sensitivity and specificity, without actually changing the threshold.

Linear SVM (for three splits):

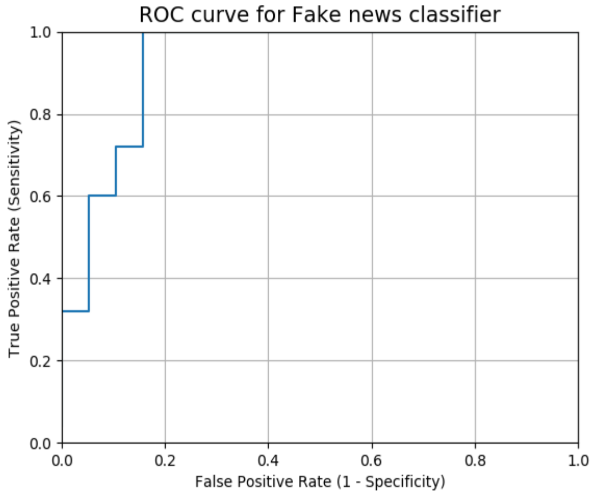


Figure 6.7 ROC Curve for Linear SVM-TFIDF, Split 1

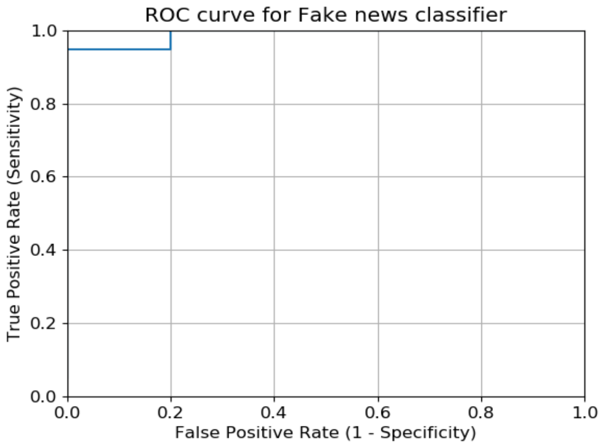


Figure 6.8 ROC Curve for Linear SVM-TFIDF, Split 2

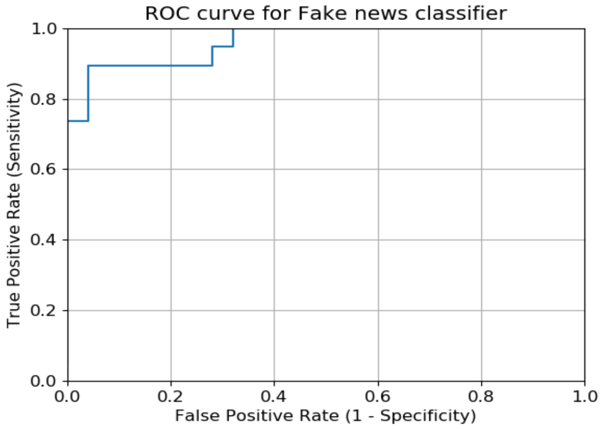


Figure 6.9 ROC Curve for Linear SVM-TFIDF, Split 3

SGD-TFIDF (for three splits)

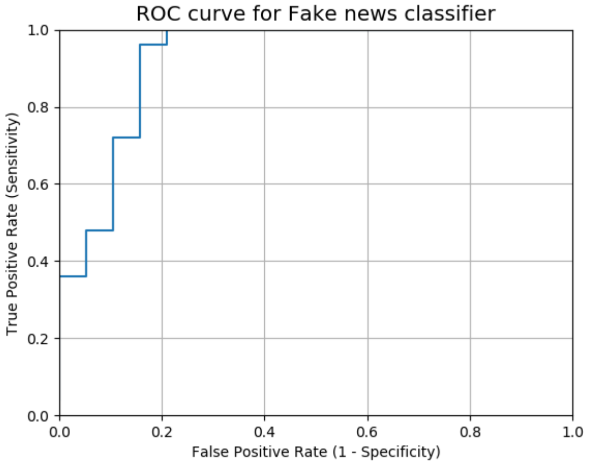


Figure 6.10 ROC Curve for SGD-TFIDF, Split 1

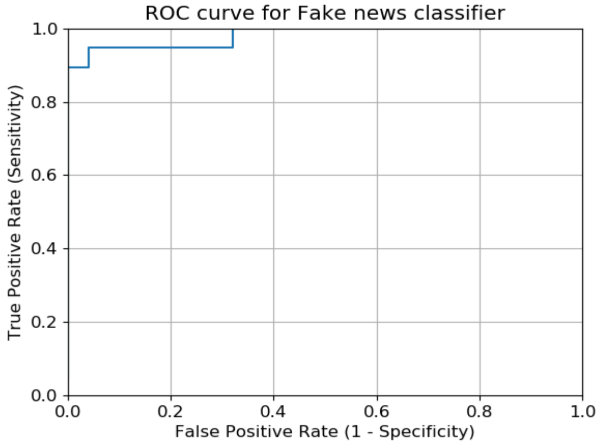


Figure 6.11 ROC Curve for SGD-TFIDF, Split 2

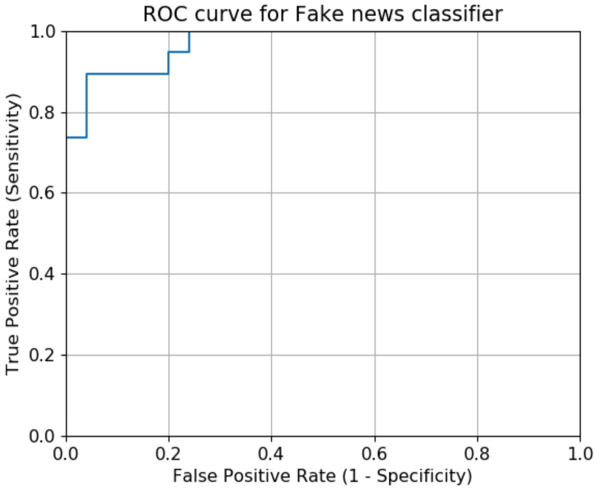


Figure 6.12 ROC Curve for SGD-TFIDF, Split 3

ROC Area Under Curve Score

ROC is the percentage of the ROC plot that is underneath the curve. Higher the value of AUC better is the classifier. AUC is very useful when there is high imbalance of classes. AUC Score for both the models is shown below:

SCHOTASTIC GRADIENT DESCENT-TFIDF_ROC-AUC-SCORE: ***95.7%***

LinearSVM-TFIDF_ROC-AUC-SCORE: ***96.0%***

Cross Validation Score for both models:

LinearSVM-TFIDF_CROSS-VAL-SCORE: ***97 %***

SCHOTASTIC GRADIENT DESCENT-TFIDF_CROSS-VAL-SCORE: ***96.9%***

From the above experiments and results, it was concluded that Linear SVM algorithm, using Term-Frequency Inverse Document Frequency vector (Word level) at maximum document frequency of 0.7, gave the best performance. Finally, it was chosen as the best model to determine the Veracity of the News.

6.2.2 Other Experiments

Content (Body of the article) was classified using Count vector and TF-idf vector on datasets with varying length of text content.

Experiment was performed on the two publicly available datasets. The First dataset, contained more news items but short text length while the second dataset contained less news items and long texts. Accuracy was noted down for both of them. It was found that accuracy given by models on second dataset was higher than the first one.

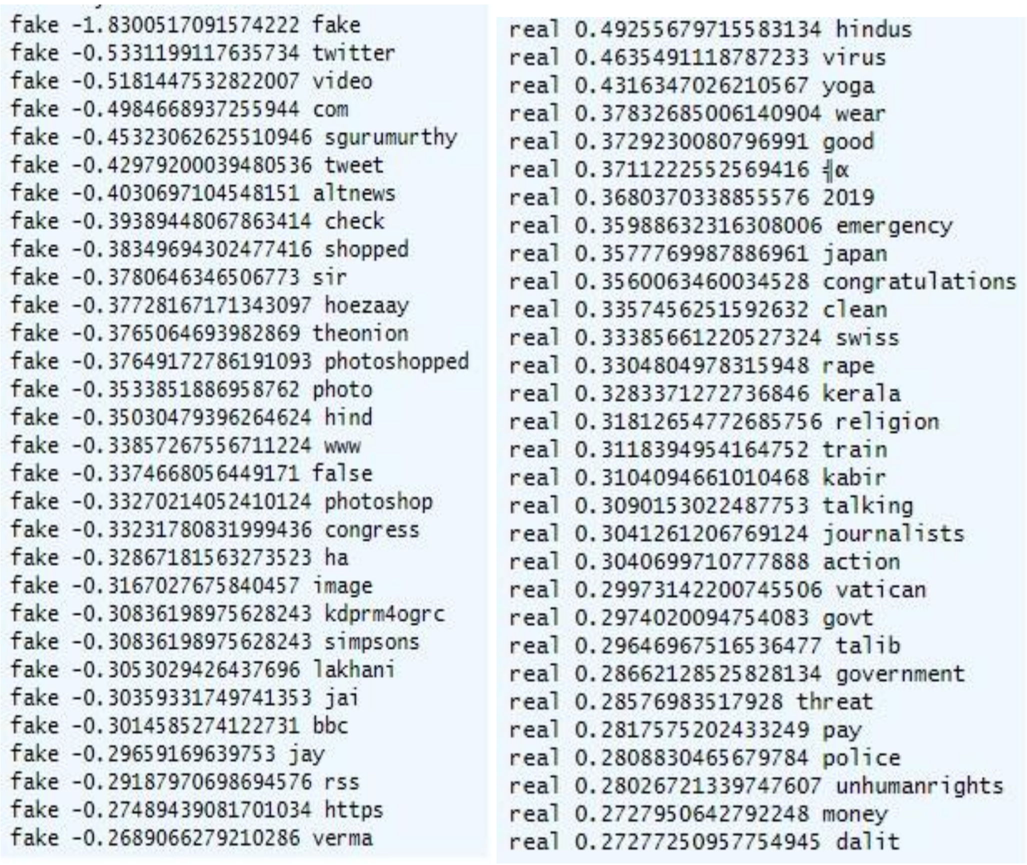
Experiment to count the number of Fake related words or a combination of two or more words in the responses was performed.

This was the most useful experiment which proved that Response based detection has significant advantage over the text based detection on the article's body. In this experiment, I calculated the frequency of words signifying Fakeness for e.g. "Fake news", "Misinformation", "Hoax", "Photo-shopped" etc. in the responses collected. The general idea is that if there is more number of such words used in a response then that news has high probability of being Fake. If no such words are present, then that article is most probably a real article.

Experiment to find Most informative tokens was performed.

It was an incredibly useful experiment, which was performed in the end. It finds out the most informative features / tokens in the collection of responses that affects the news veracity (fake/real).

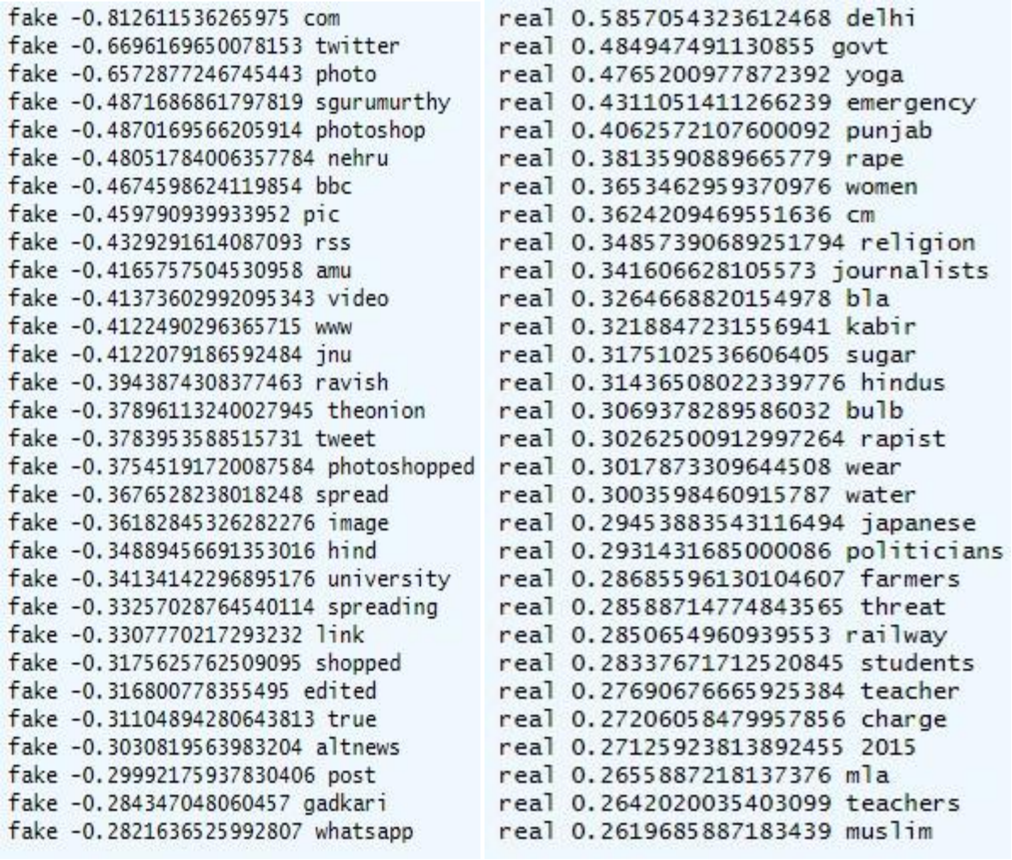
Most informative tokens for SVC-TFIDF: The image below shows the top 30 tokens for three splits of dataset, sorted by TFIDF values.



```
fake -1.8300517091574222 fake
fake -0.5331199117635734 twitter
fake -0.5181447532822007 video
fake -0.4984668937255944 com
fake -0.45323062625510946 sgurumurthy
fake -0.42979200039480536 tweet
fake -0.4030697104548151 altnews
fake -0.39389448067863414 check
fake -0.38349694302477416 shopped
fake -0.3780646346506773 sir
fake -0.3772816717134 3097 hoezaay
fake -0.3765064693982869 theonion
fake -0.37649172786191093 photoshopped
fake -0.3533851886958762 photo
fake -0.35030479396264624 hind
fake -0.33857267556711224 www
fake -0.3374668056449171 false
fake -0.33270214052410124 photoshop
fake -0.33231780831999436 congress
fake -0.32867181563273523 ha
fake -0.316702767584045 7 image
fake -0.30836198975628243 kdprm4ogrc
fake -0.30836198975628243 simpsons
fake -0.3053029426437696 lakhani
fake -0.30359331749741353 jai
fake -0.3014585274122731 bbc
fake -0.29659169639753 jay
fake -0.29187970698694576 rss
fake -0.27489439081701034 https
fake -0.2689066279210286 verma
```

```
real 0.49255679715583134 hindus
real 0.463549111878 7233 virus
real 0.4316347026210567 yoga
real 0.37832685006140904 wear
real 0.3729230080796991 good
real 0.3711222552569416 ╡α
real 0.3680370338855576 2019
real 0.35988632316308006 emergency
real 0.3577769987886961 japan
real 0.3560063460034528 congratulations
real 0.3357456251592632 clean
real 0.33385661220527324 swiss
real 0.3304804978315948 rape
real 0.3283371272736846 kerala
real 0.31812654772685756 religion
real 0.3118394954164752 train
real 0.3104094661010468 kabir
real 0.3090153022487753 talking
real 0.3041261206769124 journalists
real 0.3040699710777888 action
real 0.29973142200745506 vatican
real 0.2974020094754083 govt
real 0.29646967516536477 talib
real 0.28662128525828134 government
real 0.28576983517928 threat
real 0.2817575202433249 pay
real 0.2808830465679784 police
real 0.28026721339747607 unhumanrights
real 0.2727950642792248 money
real 0.27277250957754945 dalit
```

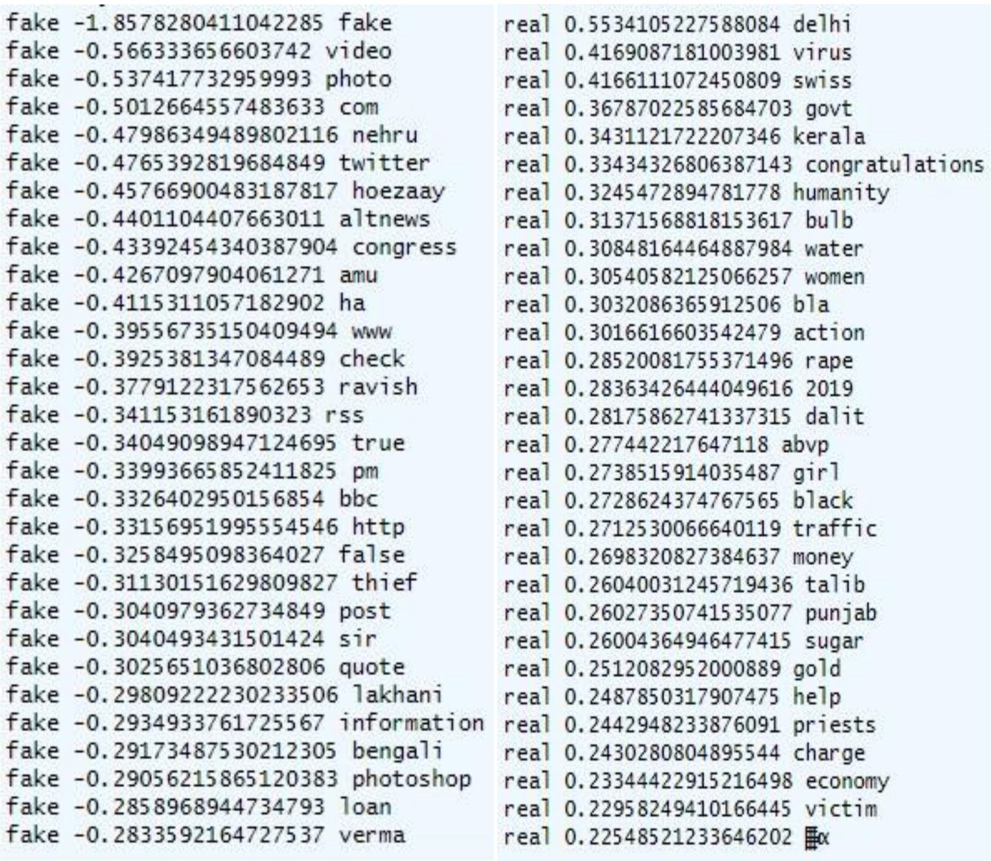
Figure 6.13: Top 30 informative tokens for SVC-TFIDF Split 1



```
fake -0.812611536265975 com
fake -0.6696169650078153 twitter
fake -0.657287724674544 3 photo
fake -0.4871686861797819 sgurumurthy
fake -0.4870169566205914 photoshop
fake -0.48051784006357784 nehru
fake -0.4674598624119854 bbc
fake -0.459790939933952 pic
fake -0.4329291614087093 rss
fake -0.4165757504530958 amu
fake -0.41373602992095343 video
fake -0.4122490296365715 www
fake -0.4122079186592484 jnu
fake -0.3943874308377463 ravish
fake -0.37896113240027945 theonion
fake -0.3783953588515731 tweet
fake -0.37545191720087584 photoshopped
fake -0.367652823801824 8 spread
fake -0.36182845326282276 image
fake -0.34889456691353016 hind
fake -0.34134142296895176 university
fake -0.33257028764540114 spreading
fake -0.3307770217293232 link
fake -0.3175625762509095 shopped
fake -0.316800778355495 edited
fake -0.31104894280643813 true
fake -0.3030819563983204 altnews
fake -0.29992175937830406 post
fake -0.284347048060457 gadkari
fake -0.282163652599280 7 whatsapp
```

```
real 0.5857054323612468 delhi
real 0.484947491130855 govt
real 0.4765200977872392 yoga
real 0.4311051411266239 emergency
real 0.4062572107600092 punjab
real 0.3813590889665779 rape
real 0.3653462959370976 women
real 0.3624209469551636 cm
real 0.34857390689251794 religion
real 0.341606628105573 journalists
real 0.3264668820154978 bla
real 0.3218847231556941 kabir
real 0.3175102536606405 sugar
real 0.31436508022339776 hindus
real 0.3069378289586032 bulb
real 0.30262500912997264 rapist
real 0.3017873309644508 wear
real 0.3003598460915787 water
real 0.29453883543116494 japanese
real 0.2931431685000086 politicians
real 0.28685596130104607 farmers
real 0.2858871477484356 5 threat
real 0.2850654960939553 railway
real 0.28337671712520845 students
real 0.27690676665925384 teacher
real 0.27206058479957856 charge
real 0.27125923813892455 2015
real 0.2655887218137376 mla
real 0.2642020035403099 teachers
real 0.2619685887183439 muslim
```

Figure 6.13: Top 30 informative tokens for SVC-TFIDF Split 2



```
fake -1.8578280411042285 fake         real 0.5534105227588084 delhi
fake -0.566333656603742 video         real 0.4169087181003981 virus
fake -0.537417732959993 photo         real 0.4166111072450809 swiss
fake -0.5012664557483633 com          real 0.36787022585684703 govt
fake -0.47986349489802116 nehru       real 0.343112172207346 kerala
fake -0.4765392819684849 twitter      real 0.33434326806387143 congratulations
fake -0.45766900483187817 hoezaay     real 0.3245472894781778 humanity
fake -0.4401104407663011 altnews      real 0.31371568818153617 bulb
fake -0.43392454340387904 congress    real 0.30848164464887984 water
fake -0.4267097904061271 amu          real 0.30540582125066257 women
fake -0.4115311057182902 ha           real 0.3032086365912506 bla
fake -0.39556735150409494 www         real 0.3016616603542479 action
fake -0.3925381347084489 check        real 0.28520081755371496 rape
fake -0.377912231756265 3 ravish      real 0.28363426444049616 2019
fake -0.341153161890323 rss           real 0.28175862741337315 dalit
fake -0.34049098947124695 true        real 0.27744221764 7118 abvp
fake -0.33993665852411825 pm          real 0.2738515914035487 girl
fake -0.332640295015 6854 bbc         real 0.2728624374767565 black
fake -0.33156951995554546 http        real 0.2712530066640119 traffic
fake -0.3258495098364027 false        real 0.269832082738463 7 money
fake -0.31130151629809827 thief       real 0.26040031245719436 talib
fake -0.304097936273484 9 post        real 0.26027350741535077 punjab
fake -0.3040493431501424 sir          real 0.26004364946477415 sugar
fake -0.3025651036802806 quote        real 0.251208295200088 9 gold
fake -0.29809222230233506 lakhani     real 0.2487850317907475 help
fake -0.2934933761725567 information  real 0.2442948233876091 priests
fake -0.29173487530212305 bengali     real 0.243028080489554 4 charge
fake -0.29056215865120383 photoshop   real 0.23344422915216498 economy
fake -0.2858968944734793 loan         real 0.22958249410166445 victim
fake -0.2833592164727537 verma        real 0.22548521233646202 ▦α
```

Figure 6.13: Top 30 informative tokens for SVC-TFIDF Split 2

As clear from the above images, the “fake” word has the maximum negative tf-idf value of -1.85 in the Fake class. A trivial response by users in case of Fake claims. Some other important words are:

1) Video: Users generally comments “Fake Video” or “Morphed Video” etc.

2) Tweet: The words like False tweet or Misleading tweet are used in response to any Fake tweets.

3) Altnews: This word has tf-idf value as -0.403 and has high influence in determining the fake news. Altnews.com is fact-checking agency, which busts fake news circulating on Social media. Users in their response give reference to articles of such agencies debunking the Fake claims. Therefore, this word appeared in the top 30 important tokens. Other Fact-checking popular agencies are Smhoaxslayer.com, Snopes.com, Boomlive.com, Politifact.com etc.

4) Check: This word is used in response when people ask the tweeter to Fact check before Tweeting Or in the sentences like “Please check the facts before posting it.” Etc.

5) Theonion: It has tf-idf values 0.376. It is a popular satirical news website.

6) Photo shopped, Photoshop: For any morphed/modified images circulating on social media, users terms it as Photo shopped images in the response. Therefore, it has high influence.

7) Images: It is used with words like “Fake Images” or “Photo shopped Images” etc.

8) Spread/Spreading: Sentences like “Please don’t spread misinformation.” or “Why are you spreading this Fake article?” appear mostly in comments.

CHAPTER 7

CONCLUSION

User’s opinion on social media posts can be well applied to determine the veracity of news. Dissemination of Fake news on social media is very fast and therefore this method, can serve as a basic building block for Fake news detection. With highest classification accuracy of 93.2%, sensitivity of 92% and ROC AUC score of 97%, Linear Support Vector machine with Tf-Idf vector served as a better model as compared to others. In this work, the classification was performed on small number of news items. Adding more data to the dataset will test the consistency of the performance thereby increasing trust of users on the system. In addition, gathering real news that almost appears as Fake news will improve the training of the model. More linguistic based features can be applied on responses to determine the news veracity. Social media plays an important role in the news verification process, however if the news is recent and is published in a few news outlets only in the beginning, then social media cannot be used as an additional resource. The shift from traditional media to social media and fast dissemination of news, checks this limitation. Therefore, by exploring more social media features in our experiments, and combining them we can create an effective and reliable system for detecting Fake news.

REFERENCES

[1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Huan Liu. "Fake News Detection on Social Media", ACM SIGKDD Explorations Newsletter, 2017

[2] https://en.wikipedia.org/wiki/Fake_news

[3] https://en.wikipedia.org/wiki/Machine_learning

[4] https://en.wikipedia.org/wiki/Natural_language_processing

[5] FAKE NEWS IDENTIFICATION CS 229: MACHINE LEARNING : GROUP 621 Sohan Mone, Devyani Choudhary, Ayush Singhania

[6] Emilio Ferrara, Onur Varol, Clayton Davis, FilippoMenczer, and Alessandro Flammini. The rise of social bots. Communications of the ACM, 59(7):96{104, 2016.

[7] *Carlos Merlo (2017)*, "Millonario negocio FAKE NEWS", *Univision Noticias*

[8] Chang, Juju; Lefferman, Jake; Pedersen, Claire; Martz, Geoff (November 29, 2016). "When Fake News Stories Make Real News Headlines". Nightline. ABC News.

[9] https://www.cjr.org/analysis/facebook-rohingya-myanmar-fake-news.php

[10] https://blog.paperspace.com/fake-news-detection/

[11]Eni Mustafaraj and Panagiotis Takis Metaxas. The fake news spreading plague: Was it preventable? arXiv preprint arXiv:1703.06988, 2017.

[12]Niall J Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology.

[13] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendor, and Benno Stein. A stylometric in-quiry into hyperpartisan and fake news. arXiv preprint, arXiv:1702.05638, 2017.

[14]David O Klein and Joshua R Wueller. Fake news: A legal perspective. 2017.

[15] Andrew Ward, L Ross, E Reed, E Turiel, and T Brown. Naive realism in everyday life: Implications for social conict and misunderstanding. Values and knowledge, pages 103{135, 1997.

[16]RaymondSNickerson.Conrmation bias: A ubiquitous phenomenon in many guises. Review of general psychology, 2(2):175, 1998.

[17] Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 us presidential election online discussion. First Monday, 21(11), 2016

[18] Michele Banko, Michael J Cafarella, Stephen Soder-land, Matthew Broadhead, and Oren Etzioni. Open information extraction from the web. In IJCAI'07.

[19] Amr Magdy and Nayer Wanas. Web-based statistical fact checking of textual documents. In Proceedings of the 2nd international workshop on Search and mining user-generated contents, pages 103{110. ACM, 2010.

[20] Giovanni Luca Ciampaglia, Prashant Shiralkar,Luis M Rocha, Johan Bollen, Filippo Menczer, andAlessandro Flammini. Computational fact checking from knowledge networks. PloS one, 10(6):e0128193,2015.

[21] You Wu, Pankaj K Agarwal, Chengkai Li, Jun Yang, and Cong Yu. Toward computational fact-checking. Proceedings of the VLDB Endowment, 7(7):589{600, 2014 [22] Baoxu Shi and Tim Weninger. Fact checking in het-erogeneous information networks. In WWW'16

[23] https://www.huffingtonpost.in/2018/04/25/facebook-says-its-fact-checkers-will-stop fake-news-in-the-karnataka-election-well-just-have-to-believe-them_a_23420278/

[24] Christina Boididou, Symeon Papadopoulos, Markos Zampoglou, Lazaros Apostolidis, Olga Papadopoulou, Yiannis Kompatsiaris. "Detection and visualization of misleading content on Twitter", International Journal of Multimedia Information Retrieval, 2017

[25]Cody Buntain, Jennifer Golbeck. "Automatically Identifying Fake News in Popular Twitter Threads", 2017 IEEE International Conference on Smart Cloud (SmartCloud), 2017

[26] Zhiwei Jin, Juan Cao, Yongdong Zhang, Jianshe Zhou, Qi Tian. "Novel Visual and Statistical Image Features for Microblogs News Verification", IEEE Transactions on Multimedia, 2017

[27] Detection and visualization of misleading content on Twitter
Christina Boididou1,2 ·Symeon Papadopoulos2·Markos Zampoglou2·Lazaros Apostolidis2·
Olga Papadopoulou2·Yiannis Kompatsiaris

[28] https://www.analyticsvidya.com

[29] https://www.ritchieng.com/machine-learning-evaluate-classification-model/

[30] https://machinelearningmastery.com/

[31] Automatically Identifying Fake News in Popular Twitter Threads

[32] "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, William Yang Wang

[33] https://github.com/likeaj6/FakeBananas

[34] https://github.com/nishitpatel01/Fake_News_Detection

[35] Song Feng, Ritwik Banerjee, and Yejin Choi. Syntactic stylometry for deception detection. In ACL'12.

[36]"Buzzfeednews:2017-12-fake-news-top-50,"https://github.com/BuzzFeedNews/2017-12-fake-news-top-50.

[37] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, 2015, pp. 1–4.

[38] V. L. Rubin, Y. Chen, and N. J. Conroy, "Deception detection for news: three types of fakes," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, 2015, pp. 1–4.

[39] Dong ping Tian et al. A review on image feature extraction and representation techniques. International Journal of Multimedia and Ubiquitous Engineering, 8(4):385–396, 2013

[40] Local tampering detection in video sequences Paolo Bestagini, Simone Milani, Marco Tagliasacchi, Stefano Tubaro