Detecting Fake News and Fake Reviews through Linguistic Styles

A Dissertation (Major Project-II) submitted for the partial fulfilment of requirement for the award of degree of

Master of Technology

in

Information Systems

Submitted by

CHHAVI JAIN

2K17/ISY/05

Under the Supervision of:

Dr. Dinesh K. Vishwakarma

Associate Professor, Department of Information Technology

Department of Information Technology

Delhi Technological University

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Bawana Road, Delhi – 110042 (India)

June-2019

DECLARATION

I, Chhavi Jain, hereby declare that the work which is being presented in the dissertation (Major Project-II) entitled "**DETECTING FAKE NEWS AND FAKE REVIEWS THROUGH LINGUISTIC STYLES**" by me in partial fulfillment of requirements for the award of degree of Master of Technology (Information System) from Delhi Technological University, is an authentic record of my own work carried out under the supervision of **Dr. Dinesh Kr. Vishwakarma**, Associate Professor, Information Technology Department.

The material contained in the report has not been submitted to any university or institution for the award of any degree.

Place: Delhi **CHHAVI JAIN**

Date: **2K17/ISY/05**

CERTIFICATE

This is to certify that Major Project report-2 entitled "DETECTING FAKE NEWS AND FAKE REVIEWS THROUGH LINGUISTIC STYLES" submitted by **CHHAVI JAIN (Roll No. 2K17/ISY/05)** for partial fulfillment of the requirement for the award of degree Master of Technology (Information System) is a record of the candidate work carried out by him under my supervision.

Place: Delhi
Date:

Dr. Dinesh K. Vishwakarma
Supervisor,
Associate Professor, Department of Information Technology
Delhi Technological University, Delhi

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to **Dr. Dinesh Kr. Vishwakarma,** Associate Professor for his continuous support during this thesis. He has not only been my supervisor but also a very inspirational figure during my master's studies. This is my heartfelt thanks for his motivational lectures which have helped me to improve as a computer science student and pursue the field of machine learning.

Secondly, I am grateful to **Prof. Kapil Sharma**, HOD, Information Technology Department, DTU for his immense support. I would also like to acknowledge Delhi Technological University faculty for providing the right academic resources and environment for this work to be carried out. Last but not the least I would like to express sincere gratitude to my parents and friends for constantly encouraging me during the completion of work.

CHHAVI JAIN

2K17/ISY/05

ABSTRACT

Deceptive content has become challenging to deal with in recent years. Fake reviews continue to misguide customers on the credibility of the product. Since such data can be easily generated and is usually in abundance, fake reviews or the opinion spam problem has now become a growing research area. Also, 2016 US presidential elections proved that fake news can have a huge impact and drew attention of people to this problem. There is a pressing need for fake news detection but it is a challenging problem as well. In this paper, machine learning based classifiers have been used to automatically detect fake content (mainly fake news and fake reviews). 55 features have been extracted from data and 6 classifiers have been used for three datasets. Datasets used are publicly available and they are for fake reviews as well as fake news.

Keywords: *Fake news, Fake Reviews, Text classification, Machine Learning, Opinion spams*

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

S No	Abbreviated Name	Full Name
1	LR	Logistic Regression
2	SGD	Stochastic Gradient Descent
3	KNN	K Nearest Neighbor
4	DT	Decision Tree
6	SVM	Support Vector Machine
7	LSVM	Linear Support Vector Machine
8	ROC	Receiver Operating Characteristic

LIST OF EQUATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Online reviews play an important role in helping customers regarding buying a product online. Opinion spamming is a way to positively manipulate someone's decision towards a product on e-commerce. This is done by adding fake reviews regarding that product. Businesses hire a group of spammers to post fake reviews and hence try to impact the reputation of a product. Generally, there are three types of fake reviews [1]. First one is to impact the reputation of a product in a positive or negative manner. This is done by a group adding reviews in a specific direction i.e. negative or positive manner. Second type is towards targeting a brand. These are used for brand promotion. Third type of false reviews are towards targeting a product. These are generally present in the form of an advertisement. As per the observations, there are simpler methodologies available to identify second and third category of false reviews when compared to methodologies available to identify the first category false reviews.

Apart from fake reviews, fake news is also an alternate opinion spamming way to affect the market and production sale of a product [2]. Social media giants such as Facebook, Twitter etc. can collectively affect mindset of a large chunk of generation within no time. Since, this situation might lead to tremendous change in the market, an active monitoring system on the content is needed. Fake news can be categorized into three major categories broadly as discussed in [2]. Detection of fake news is generally assumed to be tougher than detecting fake reviews. Fake reviews are specifically written to directly impact a target product, brand etc. but sometimes it's a complex task to identify agenda of fake news because the entities being impacted by fake news is generally unclear. Fake reviews and news are the two main classes of opinion spamming which have been explored. Text analysis, n-gram methodology and some other features have been used by the detection model. Six different Machine learning based classification models which are Stochastic gradient descent, logistic regression, Support Vector Machine, Linear Support Vector Machine, KNN and decision tree have been used. Models have been tested on 3 different datasets. These models have been

tested for both fake news and reviews. In Section 2 related work in fake reviews and fake news has been discussed. Detection model along with the details of the techniques and datasets used have been discussed in the section 3. Section 4 presents all the experiments, results and classification algorithm used etc. At last, section 5 concludes the paper and includes the future work.

1.2 Approach Overview

Refining of the datasets is an important task which makes processing of the dataset easier. Datasets were refined by removal of stop words, lowering the case, removal of punctuation, tokenization, segmentation of sentence. Removal of stop words removes insignificant words which might lead to generation of noise in the classification. Stop words include pronouns, conjunctions, articles, prepositions such as a, an, as, are, the, these and so on. Tokens are obtained and changed into a standard form. Lemmatization is understanding meaning of a word in a sentence, according to the context. Classification models which have been used in the systems, were trained on the fake news and fake reviews datasets. Models used are SGD, LR, KNN, DT, SVM, LSVM. Fundamental fake news/reviews model has been shown in Figure 1.

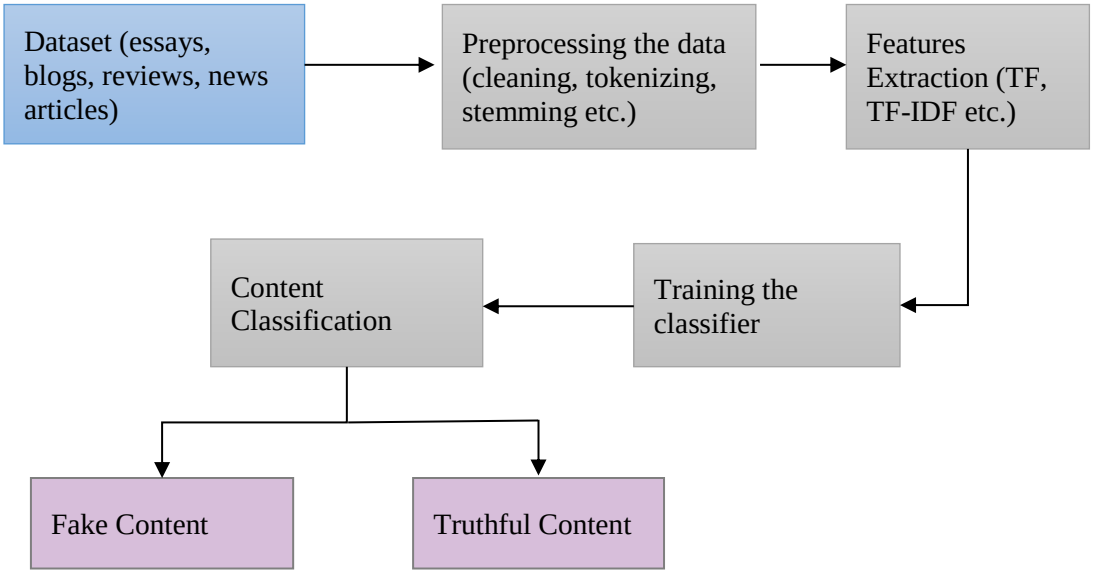

Figure 1: Fundamental model of fake news/reviews detection model [2]

CHAPTER 2
LITERATURE REVIEW

2.1 Fake News Detection

Machine learning based classifiers are generally used in the systems and are trained on some standard fake news datasets such as BuzzFeed and CREDBANK. This type of news can spread in various forms. The fake news can be textual or image based. Textual news and image-based news verification are the two streams in which review of related work has been presented.

2.1.1 Textual News Verification

Various features are there to help in classification of news. There are 3 broad categories of features such as text content features, user features and propagation features.

First category of text content features are the ones obtained from news body either lexically, semantically or statistically. Lexical features consist of features like Ngrams, Punctuation, Psycholinguistic features, readability and syntax [3]. Semantic features consist of opinion words and semantic scores. Second category of semantic mining performance is required to be considered for semantic analysis. Third category of statistical features include stats about news articles such as punctuation, word count, emoticons and hashtags etc. [4]. In [3], classification models were proposed with the help of linguistic features including lexical, semantic and statistical features.

Social media accounts which have posted the suspected news article are used to extract user-based features. Verification type of account, the home page of the user, location and time of the account registration, previous messages posted by the account and number of followers are examples of user features [4]. Reliability of user features can sometimes be very low. In [6], compromised accounts on social media have been analyzed.

Third category of propagational features include stats of the propagation process of the news articles. These also include the degree of root node, number of nodes in propagation graph and related features. Fake news has a few different structural features of propagation network from those of real news. as per the observation in [5]. At early stages of spreading of news,

such differences can be analyzed.

2.1.2 Image-based Verification

Less work has been done in image-based verification as compared to the textual verification. This area has not been sufficiently explored. However, it has been observed that systems that combine both the streams for analysis yield faster and improved results. For instance, it has been noted that user display picture influences the authenticity of the news article [10]. Also, in [10] it was concluded that images associated with the fake news articles are not very diverse as compared to images in real news articles and are limited in amount. However, less analysis has been done on the image features like clarity score or coherence score.

2.1.3 Relevant work along with contribution

Table 1: Relevant and latest work in fake News detection

Ref.	Year	Methodology	Key Contributions/ Performance	Dataset
[7]	2017	What affects spreading of fake news and suggesting solutions.	Highlighting different aspects of fake news detection.	Not used
[8]	2018	Machine-Human (MH) model combining machine linguistic and approaches based on network and a detection tool for human literacy.	Machine's and human's combined efforts.	Not used
[2]	2017	2 extraction techniques of features and 6 ML techniques are compared based on a newly proposed n-gram model.	There was a decrease in accuracy with an increase in n gram size. By mainly using 50 000 and 10 000 features, high accuracy was achieved. Unigram and bigram performed in both datasets in all cases.	Dataset 1: Dataset was built in [39] containing 1600 reviews. Of all these, half reviews are truthful, and rest are fake. Dataset 2: 12,600 truthful articles were from Reuters.com and same number of fake news articles from kaggle.com

[3]	2017	On the basis of linguistic differences, classification models were developed. Features representing properties of text readability were proposed.	An accuracy comparable to human ability to detect fake news was achieved by the best performing models.	A crowdsourced dataset and a web dataset is also created.
[4]	2017	For images, statistical and visual features were proposed.	Other relevant work for better results were combined with work done on images.	From Sina Weibo
[5]	2018	Differences in the propagation network of real and fake news were shown with the help of proposed features.	To spot fake content, collective structural signals can be used	In both, Twitter in Japan and Weibo in China, large databases of fake news and real news
[6]	2017	Compromises of high-profile accounts being identified by techniques.	False positives include 3.6% Twitter accounts and 3.6% Facebook accounts	Crawling of dataset using twitter and Facebook API.
[9]	2016	A four components system: reputation-based, user experience component, credibility classifier engine, and a feature-ranking algorithm	96.0439% accuracy for database Aden and 91.4187% for database Taiz observed	Taiz and Aden
[10]	2017	1. Extension based on web 2. Algorithm to check the fact.	Other detection systems can be combined with querying on search engines and fact checking for better results, considering the reputation of websites.	Not used
[11]	2017	Presentation of two classification techniques has been done out of which, one is logistic regression based and the other is Boolean crowdsourcing algorithm based.	FNC-1 score of 81.72% and accuracy 88.46%	Public posts from selected Facebook pages
[12]	2018	System that consists of similarity and lexical features which are passed to a perceptron with a hidden	Understanding of working and performance was developed because of	FNC Data

		layer that estimates the stance of news article body	it being a simple and straight forward setup	
[13]	2016	A propagation network was built for tweets which is credible and evaluates the news	Accuracy is better than other baseline methods and varies between 0.82 and 0.84.	Authoritative sources made Sina Weibo dataset and fake news obtaining from fake news rank lists. These sources include Xinhua New Agency
[14]	2016	Extraction of writing style, evaluation of the post and identification of user and finally updating of the baseline makes the approach.	Achieved results over 93 %.	Dataset from Twitter composed of tweets of 1000 users.
[15]	2013	Semantic and non-semantic used by detection mechanism analysis for identifying the hidden paid posters	Good results were yielded by the classifier in both supervised and unsupervised learning techniques	Sina dataset and Sohu dataset
[16]	2015	Machine learning and linguistic cue methods were combined by the system with the network-analysis approaches.	Two approaches have been combined.	Not used
[17]	2017	Analysis of fake news language and automatic political fact checking case study.	Performance of all the models was improved by LIWC features. Except for the performance of neural models	PolitiFact.com
[18]	2015	The model that is applied to classify news through discourse feature similarity is vector space.	54% accuracy obtained	Weekly radio show's "Wait, Wait, Don't Tell Me" with its "Bluff the Listener", transcripts were used.

[19]	2018	To decrease the spread of fake news, algorithm Curb is developed. This is done by solving stochastic optimal control problem	Stochastic online optimal control of SDEs and its connection with survival analysis, Bayesian inference and jumps	Twitter and Weibo
[20]	2018	Development of fake news game which helped to reduce the persuasiveness of articles	A multi-player fake news game is to be developed initially that can manage the impact that fake news has on society	Not used
[21]	2015	Serious fabrications, large-scale hoaxes, humorous flakes are the three discussed categories of fake news.	By working category-wise, the task of detecting fake news was tackled	Not used
[22]	2017	A model named CSI comprising of three main modules: Capture, Score, and Integrate, is proposed.	By utilizing the neural networks, different sources of information were used. User's and article's latent representations of are also produced	Twitter and Weibo
[23]	2017	For detection of fake news, review of work was done.	Discussion of State-of-the-art techniques has been done.	Not used
[24]	2017	With respect to the article bodies, stance detection of headlines was done with system based on lemmatization-based n-gram matching.	Accuracy score of 89.59 was achieved and can also be used in a fact-checking too.	Fake News Challenge (FNC1) on stance detection
[25]	2017	To detect and filter fake news on microblogging sites, algorithm was developed.	Presentation of a step-to-step algorithm has been done. Many factors have been taken into account like, a combined steps starting from the sources of news, the administrator of portal etc.	Not used
[26]	2015	Approaches are surveyed for recognition of textual and non-textual click-baiting	Reviewing the current approaches	Not used

		cues.		
[27]	2017	Incorporation of speaker profiles into an attention-based LSTM model	Has better performance than other models by 14.5%	LIAR data set
[28]	2017	Correlations are explored between publisher bias, relevant user engagements and news stance. On this basis, a framework is proposed.	An important feature for the problem was Tri-relationship. Good detection performance was achieved by the framework in early stage of news dissemination.	Buzzfeed and PolitiFact
[29]	2013	To analyze differences between the forced fake reviewers and natural fake reviewers, information theoretic measure and KL-divergence was used. To improve the classification, additional behavioral features are proposed.	For real-life data, behavioral features were proposed. This improved the accuracy	reviews from Yelp.com
[30]	2017	130 thousand news posts were classified as verified or suspicious by the predictive neural network models. Also, four predicted categories of suspicious news are– satire, hoaxes, clickbait and propaganda	Social media interactions and tweet content are considered for classification.	Around 400 twitter accounts
[31]	2018	Label propagation doing causality-based unsupervised framework introduced.	Higher precision (0.75) compared to with random (0.11) and bot detection (0.16)	ISIS related tweets/retweets
[32]	2018	Implicit and explicit profile features were analyzed.	Highlighting of correlation between fake/real news and user profiles has been done.	Buzzfeed and PolitiFact
[33]	2017	Features have been used by classification model to identify fake twitter threads.	Crowdsourced, non-expert workers were leveraged rather than journalists.	CREDBANK and PHEME
[34]	2017	Results of tested dataset on different machine learning algorithms were compared	It was found that Stochastic Gradient Descent model using	Dataset containing news articles from

			TF-IDF feature set was the best performing one.	Signal Media and from OpenSources.co.
[35]	2017	To detect fake news, Naive Bayes classifier was used	74% accuracy was achieved and ways to improve the classifier were discussed.	Facebook news posts
[36]	2017	Comparison was drawn between methods proposed by Facebook and the 'Right-click Authenticate' approach.	Checking would be accelerated by the 'Right-click Authenticate' approach.	Not used
[37]	2014	Hierarchical propagation model	6% of improvement and better results have been seen with multilayered approach.	Microblog datasets: SW-2013 and SW-MH370
[38]	2016	To collect, detect, and analyze the misinformation, a platform named Hoaxy was developed.	The kind of users that spread news has been worked upon. Also, the time when it is spread more is also worked upon.	From sources of misinformation and fact-checking websites like Snopes.com and TruthOrFiction.com, tweets were collected.
[39]	2018	Techniques were reviewed	Techniques were reviewed	Not used
[40]	2017	Along with a classifier, three approaches were developed.	Highlighting of Relationship between deceptive opinion spam and imaginative writing has been done. On insights from psychology and computational linguistics, the approaches were based.	With gold-standard deceptive opinions, an opinion spam dataset has been developed.
[41]	2017	CNN was developed and LIAR dataset was presented.	LIAR, a new larger magnitude dataset was presented.	LIAR

2.2 Content based detection model

In [41], one content-based detection model was built which to classify fake and honest opinions used n-gram term frequency. It also built a “gold-standard” dataset, using truthful opinions from TripAdvisor and fake opinions from Amazon Mechanical Turk and using SVM classifier 86% accuracy was achieved, whereas human judges could achieve only 65% accuracy. Humans are unable to detect fake reviews efficiently was also established from this paper.

In [42], another important content-based model was developed. They argued that it is not as difficult to detect pseudo fake reviews written just for the sake of model testing and are not real-world fake news. Hence, to check Ott et all’s model, they used filtered reviews collected from Yelp. Using the fake reviews generated from AMT it achieved 86% accuracy. However, only 67.8% accuracy was obtained in this model. Though, it was acknowledged that this accuracy is still good, and to detect deceptive content n-grams is still a useful technique.

We can have some content-based detection model stylometric based and some semantic similarity metrics.

2.3 Opinion Spam Detection

Traditionally, web and emails have been used to study spam. Recently, researchers have started studying opinion spam as well. In [1] the opinion spam detection problem was first discussed. About 10 million reviews were investigated on Amazon.com for fake review detection. Lack of labeled data made it difficult to detect fake data. Fake opinion label was given to all duplicate and near duplicate reviews, rest of the reviews were labelled as truthful opinions. For detecting fake reviews Logistics regression (LR), SVM, NB and Decision tree were tried. When using all the features 78% accuracy was achieved and 63% when only text features were used. There is an added benefit in using LR as it also produces the probability that tells the probability of a review to be fake. Also, certain relationships were also established between products, ratings, reviewers and reviews in the study. Psychologically relevant linguistic features were compared manually in [43]. 42 fake hotel reviews and 40 truthful reviews were collected for this purpose. Today however, to build automatic classifiers much larger datasets are generated.

Field of psycholinguistic deception detection has also seen some progress. Two experiments were carried out involving a deceiver and an honest participant in in [44]. The first

experiment was face-to-face discussions and other was computer-based. For classification 16 linguistic features were tested. The discussions were recorded and to form linguistic cues classes later studied. C4.5 DT algorithm was used along with 15-fold cross-validation. With dataset consisting of 72 instances the accuracy obtained was 60.72%. Linguistic features proposed in [44] have been shown in Table 2.

Similarly, in [45] five experimental case studies were conducted with different context, number of participants and different percentage of males and females. The choice if they want to be sincere or deceptive was for the participants to make. Five linguistic cues were proposed after a systematic analysis of these five experimental studies. An accuracy of 67% was given by Logistic regression. Human judges' accuracy was noted to be 52% which is lower than given by model. Studies featured in the paper are present in Table 3.

Features study has been presented again in [46]. Taking Desert Survival Problem as base an experiment was performed. In the experiment, a web-messaging system was used for information exchange. Again, it was the participant's call to be either a deceiver or sincere. Different features were considered for classification and using statistical analysis they were evaluated.

CHAPTER 3

THE PROPOSED WORK

3.1 Problem Denotation

The fake news/reviews detection framework created has a fundamental model for detection. Machine learning based classification models are used in the systems which train on the fake news and fake reviews datasets. Algorithms used are SGD, KNN, LR, DT, SVM and LSVM.

3.2 Flowchart

Dataset which consists of news articles or reviews in textual form are taken and cleaned by various data preprocessing techniques like stop word removal, lower casing etc. Dataset are taken for both fake news and fake reviews. It is a labelled dataset. After preprocessing the dataset, features have been extracted. TF-IDF cosine similarity and 53 other linguistic features have been extracted. Then six classifiers are trained using those features. Once the model is trained it can be used to classify the content as real or fake. Also, various quality metrics have been calculated to test the accuracy of the mode. Precision, recall and ROC curves have been calculated and analyzed to check which model performs better. Flowchart has been shown in figure 2.

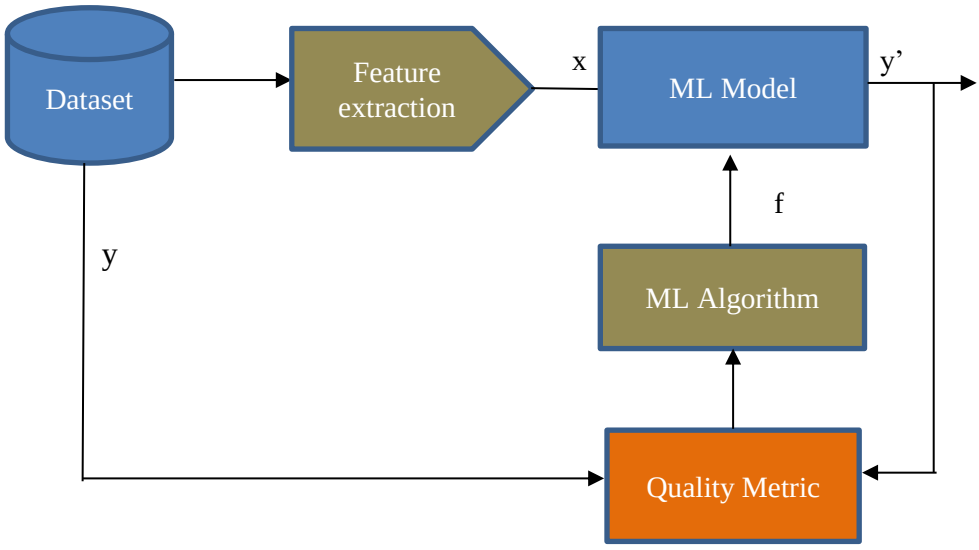

Figure 2: Flowchart of proposed model

3.3 Pseudocode

I. Procedure Posting_list

This is the pseudocode for obtaining posting list.

For all d ϵ docs

text = read(d)

tokens = tokenizer.tokenize(text)

lemmatize = [lemmatize(t) for t in tokens]

words = [l/lower() for l in lemma]

vocab = set(words)

for all v ϵ vocab

post_list[v].append(d).

II. Procedure term_frequency

Following is the pseudocode for obtaining term_frequency

For all d ϵ docs

text = read(d)

tokens = tokenizer.tokenize(text)

lemmatize = [lemmatize(t) for t in tokens]

words = [l/lower() for l in lemma]

vocab = set(words)

for all v ϵ vocab

term_freq[v][doc]+=1

3.3 Data Cleaning and Preprocessing

Before working on data, data needs to be refined so that it is easier to work upon it. Datasets were refined by stop word removal, conversion to lower case, punctuation removal, tokenization, and sentence segmentation. All the steps have been discussed in next sections:

3.3.1 Removal of stop words

Stop words are such words that are not significant and can add error when used as a feature in classification. They are mainly articles, prepositions, conjunctions and pronouns such as a, an, that, what and so on. These words were omitted from the documents and documents are then passed to the next step.

3.3.2 Tokenize

Tokens are usually individual words and tokenization is a task in NLP in which a set or set of text is taken and broken into individual words. These tokens are then used as input for lemmatization.

3.3.3 Lemmatize

Lemmatization involves reducing a word to its base form by usually chopping the ends of the words. In lemmatization, this is done by morphological analysis of words and use of a vocabulary. For example, the word 'saw' is reduced to either 'see' or 'saw' depending on the usage of word. After lemmatization, all the letters of words are converted to lower form.

3.4 Features Extraction

55 features have been extracted for classification. The extraction methods have been discussed briefly in next sections.

3.4.1 TF-IDF Cosine similarity

It is vectorized unigram Term Frequency-Inverse Document Frequency. It is a weighted measure which tells the significance of a term. Importance rises with increasing count of term in that document. However, this is also counteracted by frequency of term in database.

The TF for word *w* in a document *d* is computed by equation (3.1):

$$\mathrm{TF(w)_d} = \frac{n_w(d)}{|d|} \tag{3.1}$$

The inverse document frequency (IDF), denoted by $\mathrm{IDF(w)_D}$, is logarithm of the total count of documents in corpus divided by frequency of documents in which this term is found. It is calculated using (3.2):

$$IDF(w)_D = 1 + \log\left(\frac{|D|}{|\{d:D|w\in d|\}|}\right) \tag{3.2}$$

Hence, it weighs down the TF value of a term, while increasing it for the rare ones.

TF-IDF for a word w in document d and corpus D using (3.3):

$$TF - IDF(w)_{d,D} = TF(w)_d \times IDF(w)_D \tag{3.3}$$

Then cosine similarity is computed for each news article against 60 news articles of each classes i.e. fake & real and then average has been taken along both the classes to get two scores. These two scores are used as features namely real_similarity and fake_similarity.

3.4.2 Linguistic Features

Descriptions of linguistic features have been mentioned in the Table 2,3 and 4. The features mentioned in these three papers have been together used to train the model. They form the feature set for all models in each database along with TF-IDF as this feature set has been mentioned as best feature set in [56] after statistical analysis and tests. Snippets of feature extraction process have been mentioned in Table 5.

Table 2: Linguistic features proposed in [44].

Feature	**Description**
Number of Syllables	A unit of pronunciation having sound of one vowel
Average number of words per sentence	Number of single characters or combination of characters that represent a spoken word in each sentence
Rate of adjectives and adverbs	Adjectives are words that describe a noun, such as sweet or ambitious. Adverbs describe a verb.
Number of words	A single character or group of characters that represent a spoken word
Number of sentences	A word, phrase, clause or group of these which forms a syntactic unit
Count of big words	Words with more than 6 letters
Syllables in each word	Number of units of pronunciation having sound of one vowel in each word
Count of short sentences	Count of sentences in which number of words are less than the average count of words in sentences in whole document
Count of long sentences	Count of sentences in which number of words are more than the average count of words in sentences in whole document
Number of	Number of such words that connect clauses or sentences such

Conjunctions	as and, if, but and so on
Flesh Kincaid grade level	It tells how tough a passage in English is to comprehend
Emotiveness index	This is a metric which is a single unidimensional measure of sentiment of a sentence
Number of affective terms	These are those words which depict positive emotion, negative emotion, anxiety, anger and sadness

Table 3: Total features studied in [45]

Feature	**Description**
Word Count	Number of single characters or combination of characters that represent a spoken word
Count of Words captures, dictionary words	Count of words found in dictionary
Count of Words longer than six letters	Count of comparatively longer words
Total number of Pronouns	These words are used in place of noun phrases and refer to the participants(s) in discourse
Number of First Person Singular	Words like I, me, mine which refer self
Total number of First Person	Words like I, we, us, our in which one includes self
Total number of Third Person	Words like he, she, they
Negations	Words like none, neither, nobody
Number of Articles	Words like the, an, and a, which are used to modify nouns and pronouns
Number of	Words to link nouns and pronouns like of, at, from, among

Prepositions	
Motion Verbs	These words represent movement or transition from one place to another. Examples are come, go, move, arrive
Affective or emotional processes	These words implicate emotional experiences like abandon, sad, happy, terrified
Positive emotions	Words which generate pleasant thoughts like joy, gratitude, hope, love
Negative emotions	Words which generate bitter feelings in a person like hate, anger, disgust
Time	Words which indicate passage of time like session, before, after, end, start
Discrepancy	Words that depict lack of clarity like would, should, vary, could
Cognitive Processes	These words represent processing of information by the human mind like insight, appreciation, intuition, knowing
Space	Words related to physical space occupied like up, down, inside, outside
Tentative	Words that indicate doubt like perhaps, might, maybe
Certainty	Words that imply surety like must, never, forever, always
Social Processes	Words which depict social behavior of humans like meet, talk, mate, them
Inclusive	Words that are inclusive with respect to an object
Exclusive	Words that are exclusive with respect to an object
Insight	These words imply obtaining knowledge regarding something particular like know, realize, think, perceive
Causation	Words which refer to a reason or a consequence like therefore, because, hence, thus, since, due to
Sensory and	These words represent perceiving information from

Perceptual Processes	environment obtained through sensory organs. Words like hear, feel, see are suitable examples.
Past tense verb	Words which showcase any action done in past like did, sat, ate, ran
Present tense verb	Words which showcase any action being done currently like running, doing, walking, dancing
Future tense verb	Words which showcase any action which will occur in the future like will, shall, soon, may

Table 4: Features proposed in [46]

Feature	**Description**
Number of Words	A single character or group of characters that represent a spoken word
Number of Verbs	Word which is grammatical center of the subject and predicate in sentence
Number of Noun Phrases	A phrase consisting of noun, its determiners and modifiers
Number of Sentences	A word, phrase, clause or group of these which forms a syntactic unit
Average noun phrase length	# of words in noun phrases / # of noun phrases words in noun phrases
Average number of clauses	# clauses / # sentences
Uncertainty	A word that indicates lack of sureness [46].
Average sentence length	# words / # sentences
Average word length	# characters / # words
Modifiers	describes a word and can be adverbs or adjectives

Emotiveness	# of adjectives + # of adverbs / # of nouns + # of verbs
Number of Modal Verbs	an auxiliary verb usually used with a verb of predication and expresses a modal modification [46]
Content Word Diversity	# of different content words or terms / # of content words or terms [46]
Passive Voice	a form of the verb used when the subject is being acted upon [46].
Perceptual Information	indicates sensorial experiences [46].
Spatio-temporal information	information of locations or spatial arrangement of people and/or objects [46].
Objectification	an expression given in a form that can be experienced by others and externalizes one's attitude [46]
Generalizing Terms	refers to a person (or object) as a class of persons or objects that includes the person (or object)
Self Reference	first person singular pronoun
Pausality	# punctuation marks / # sentences
Group Reference	first person plural pronoun
Lexical Diversity	# of different words / total # of words
Redundancy	# of function words / # of sentences
Typographical error ratio	# of misspelled words / # of words
Other reference	third person pronoun
Positive Affect	conscious subjective aspect of a positive emotion apart from bodily changes [46]
Negative Affect	conscious subjective aspect of a negative emotion apart from bodily changes. [46]

Table 5: Linguistic features extraction

Feature	**Description**
Number of Syllables	import textstat as ts ts.syllable_count(self.sentence, lang='en_US')
Number of words	len(re.sub('['+string.punctuation+']', '', self.sentence).split())
Number of big words	From LIWC word category (sixltr)
Number of sentences	import textstat as ts ts.sentence_count(self.sentence)
Number of syllables per word	import textstat as ts return ts.syllable_count(self.sentence, lang='en_US')/len(re.sub('['+string.punctuation+']', '', self.sentence).split())
Number of short sentences	import textstat as ts checking this condition for each sentence len(s.split(" ")) <= ts.avg_sentence_length(self.sentence)
Number of long sentences	import textstat as ts checking this condition for each sentence len(s.split(" ")) <= ts.avg_sentence_length(self.sentence)
Number of Words longer than six letters	From LIWC word category (Sixltr)
Number of Words captures, dictionary words	From LIWC word category (Dic)
Flesh Kincaid grade level	import textstat as ts return ts.flesch_kincaid_grade(self.sentence)
Total number of Pronouns	From LIWC word category (pronoun)

Avg number of words per sentence	From LIWC category (WPS)
Number of Conjunctions	From LIWC category (conj)
Rate of adjectives and adverbs	From LIWC word categories
Total number of First Person	From LIWC word category (I, we)
Number of First Person Singular	From LIWC word category (I)
Total number of Third Person	From LIWC word category (he, she, they)
% Negations	From LIWC word category (negate)
Number of affective terms	From LIWC word category (affect)
Emotiveness index	analyzer = SentimentIntensityAnalyzer() vs = analyzer.polarity_scores(self.sentence)
% Articles	From LIWC word category (article)
Positive emotions	From LIWC word category (posemo)
Negative emotions	From LIWC word category (negemo)
Cognitive Processes	From LIWC word category (cogmech)
Insight	From LIWC word category (insight)
Discrepancy	From LIWC word category (discrep)
Inclusive	From LIWC word category (incl)
Exclusive	From LIWC word category (excl)
Time	From LIWC word category (time)

Past tense verb	From LIWC word category (past)
Present tense verb	From LIWC word category (present)
Future tense verb	From LIWC word category (future)
Sensory and Perceptual Processes	From LIWC word category (percept)
Number of Noun Phrases	From pos tags in nltk Search for phrases with noun ("NN", "NNS", "NNP", "NNPS":) its modifiers ("RB", "RBR", "RBS") and determiners ("JJ", "JJR", "JJS")
Average number of clauses	obj.noun_phrases()[0]/(obj.noun_phrases()[1]
Certainty	From LIWC word category (certain)
Average word length	import textstat as ts ts.avg_letter_per_word(self.sentence)
Average noun phrase length	noun_count/noun_phrase [noun phrase are explained in "Number of Noun Phrases"]
Social Processes	From LIWC word category (social)
Pausality	count_punch / ts.sentence_count(self.sentence)
Modifiers	modifiers = adj + adv
Causation	From LIWC word category (cause)
Number of Modal Verbs	Using pos tags "MD"from nltk library
Motion Verbs	From LIWC word category (motion)
Generalizing Terms	From LIWC word category informal languages - "Swear words", "Assent", "NonFluencies" and "Fillers"
Group Reference	From LIWC word category "we"
Lexical Diversity	re.sub('['+string.punctuation+']', '', self.sentence).split()
Content Word	len(list(set(content_word)))/len(content_word)

Diversity	content words basewd on pos tags for adj ("JJ", "JJR" and "JJS")
Tentative	From LIWC word category (tentat)
Redundancy	From LIWC word category “func” divided by ts.sentence_count(self.sentence)
Space	From LIWC word category (space)
Typographical error ratio	from nltk.corpus import words word_list = words.words()
Spatio-temporal information	From geotext library places = GeoText(check_that).country_mentions

3.5 Classification Models

Following algorithms have been used to classify fake and real content. All these models have been applied on the three datasets taken.

3.5.1 Stochastic Gradient Descent

Gradient descent is slope of a function or rate of change of a parameter w.r.t. rate of change of another parameter. This is an iterative method for optimizing an objective function. Stochastic means a system linked with random probability. Method uses randomly selected samples to evaluate the gradients.

3.5.2 Logistic Regression

It is a type of supervised classification algorithm. Target variable can take only discrete values of given set of features. This model builds a regression model to predict if a given data entry belongs to some particular category.

3.5.3 Decision Tree

It is a tree like structure. Each internal node in tree denotes a test on one of the attributes and the branch is the outcome of the test. The leaf nodes hold the class label.

3.5.4 K-Nearest Neighbor

Most basic supervised classification algorithm. Every time we try to find the cluster of a new point. We find k nearest neighbors to this point the class to which maximum of these k neighbors belong to becomes the class of this new one.

3.5.5 Support Vector Machine

Supervised classification algorithm used for analysis of data. Given the labelled training data, SVM algorithm outputs an optimal hyperplane that can classify new examples. The SVM model is a representation of the examples as points in space.

3.5.6 Linear Support Vector Machine

LSVM is fast machine learning algorithm for solving multiclass classification problem from large datasets.

3.6 Performance Measures

Consider the result as positive, when the classifier classifies the news article as fake. Then,

- Number of true positive examples are the articles that are correctly classified as fake.
- Number of false positive examples are the articles that are incorrectly classified as fake.
- Number of true negative examples are the articles that are correctly classified as true.
- Number of false negative examples are the articles that are incorrectly classified as true.

The precision of a classifier is calculated using (3.4):

$$Precision = \frac{t_p}{t_p + f_p} \tag{3.4}$$

where:

t_p and f_p are number of true and false positive examples respectively.

The recall of a classifier is calculated using (3.5):

$$Recall = \frac{t_p}{t_p + f_n} \tag{3.5}$$

where f_n tells the number of false negative examples.

Also, Receiver Operating characteristic curves (ROC curves) have been plotted for each dataset and model used. ROC curves help in performance measurement of binary classifier

system at different threshold settings. In this curve, false positive rate is plotted on X axis and true positive rate is plotted on Y axis. These two are plotted with 100 specificity and various cut off points.

Also, the area under the curve measures discrimination, that is, classifying. More the area, better the classification.

3.7 Software Requirements

3.7.1 Language and Software used

Python has been used for development. Development is done using Spyder IDE. Spyder is a scientific environment which is written in Python. Applications are written in python and advanced functionalities related to debugging, editing, analysis and profiling along with deep inspection, data exploration, beautiful visualization and interactive execution capabilities are provided in Spyder which are part of scientific packages. It is designed by and for engineers, data analysts and engineers.

Apart from these built-in features, many plugins and APIs are also available for Spyder which further extend its abilities. It can be used as a PyQt5 extension library as well, which allows to build upon Spyder's functionality and embed the components.

Among the features of the IDE are:

- Editor
- IPython console
- Variable explorer
- Profiler
- Debugger
- Help

3.7.2 Tools used

3.7.2.1 Linguistic Inquiry and Word Count (LIWC)

LIWC is a learning and research tool that helps in automated text analysis. It learns how words used in everyday language reflect personality, motivations, feelings and thoughts. When given a text, it reads it and counts the number of words that reveal different thinking styles, social

concerns, emotions and parts of speech. It has built-in dictionaries that are mainly used to identify which words reflect which psychologically relevant categories. Text analysis module compares the words in text file against these dictionaries.

For some features LIWC tool has been used because it provides those dictionaries which make it easier to get count for different linguistic features.

3.7.2.2 Weka

Waikato Environment for Knowledge Analysis (Weka) is a collection of machine learning algorithms written in Java. Development started in 1993. It can run on almost any platform. It is a free software. Weka contains several visualization tools and algorithms as well as graphical user interfaces for user friendliness.

Advantages of Weka are:

1. Portability

2. User friendliness due to Graphical user interface

3. Free availability

4. Collection of data analysis and visualization techniques

CHAPTER 4

EXPERIMENTAL WORK AND RESULT

In this section the result of experiments carried out which includes the performance of the models along with accuracy are presented.

This section is organized as follows: Section 4.1 contains description of dataset 1 along with results. Section 4.2 and 4.3 discuss the second and third datasets respectively along with accuracy measures and ROC curves. 10-fold cross validation has been done on each test.

4.1 OpSpam

OpSpam was collected in [41]. It contains 800 truthful and 800 fake reviews. The truthful reviews are collected from TripAdvisor for 20 most popular hotels in Chicago and 400 fake reviews from Amazon Mechanical Turks (AMT). The reviews with less than 150 characters and less than 5 stars were ignored. The information for each review includes:

- Name of hotel
- Review text
- Sentiment of review
- Review label
- Review text length

Table 6. Result on OpSpam with different models

Model	**SGD**	**LR**	**KNN**	**DT**	**SVM**	**LSVM**
Accuracy	85.37	84.6	72.74	75.30	83.77	71.13
Precision	0.854	0.846	0.728	0.753	0.838	0.711
Recall	0.854	0.846	0.727	0.753	0.838	0.711
F1 score	0.854	0.846	0.727	0.753	0.838	0.711

Hence, SGD has 85.37% accuracy which is the highest amongst all the models used. This is followed by LR which gave the accuracy of 84.6%.

Table 7. Comparison of previous work and our work with OpSpam dataset

Model	SGD	LR	KNN	DT	SVM	LSVM	Accuracy
Accuracy in [2]	86	87	78	**73**	**83**	**90**	90
Accuracy in [57]	NA	NA	NA	NA	84.5	NA	84.5
Accuracy in [58]	NA	NA	NA	NA	NA	NA	90.9%
Our accuracy	85.37	84.6	72.74	**75.30**	**83.77**	71.13	85.37

As it can be seen from Table 7, the proposed model beats the accuracy of Decision Tree by 2.30%. Also, the accuracy obtained by SVM is 83.77% which is again slightly higher than accuracy obtained in [2]. One of the main reasons can be the use of only TF-IDF as feature for classification. However, we were unable to get better accuracy than 90% which is their best accuracy. In [57], best accuracy is achieved by SVM for bigrams. However, our best accuracy is more than 84.5%. In [58], a combination of latent Dirichlet allocation (LDA) and word-space model (WSM) gave highest accuracy which is 90.9% and it is higher than our best accuracy. ROC plot of each classifier for OpSpam is as given in Figure 2-8.

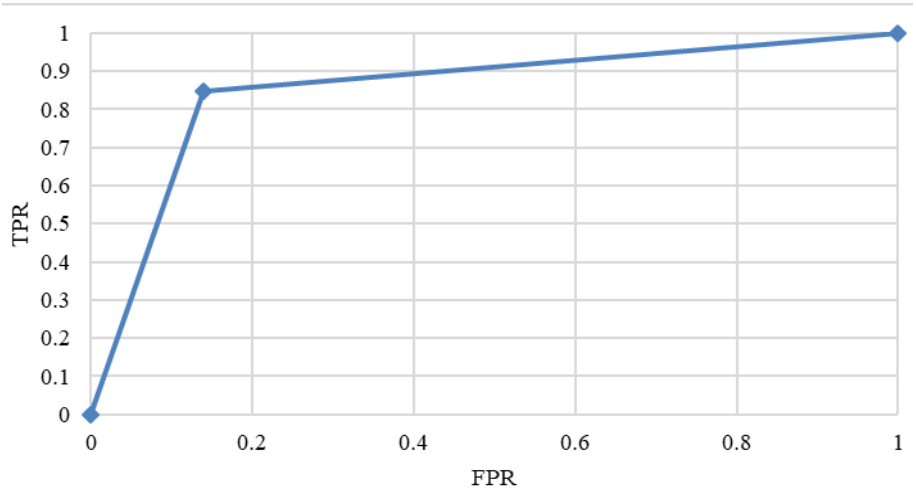

Figure 2. Curve of SGD

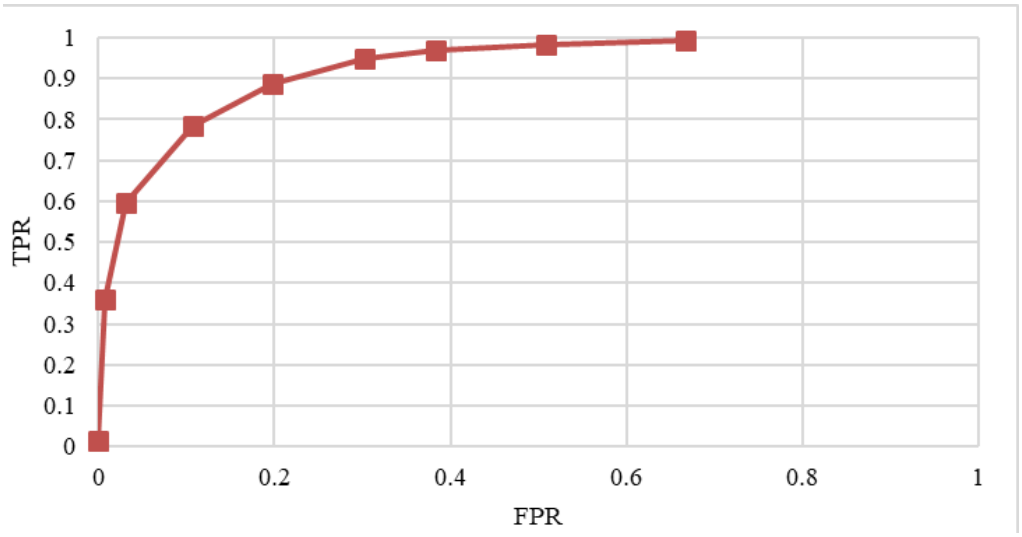

Figure 3. Curve of LR

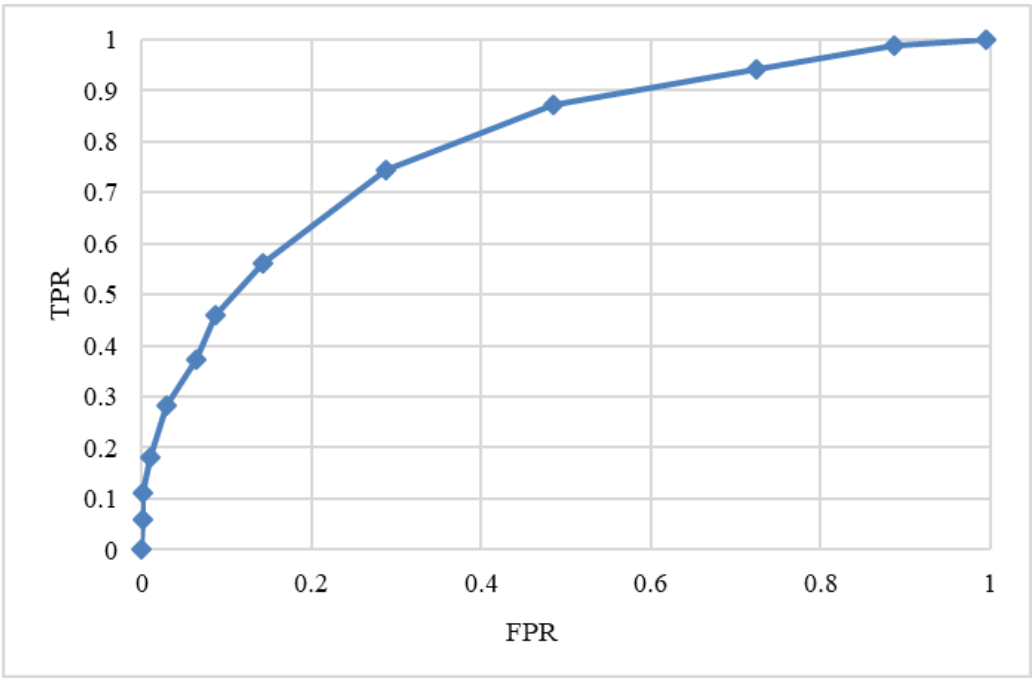

Figure 4. Curve of KNN

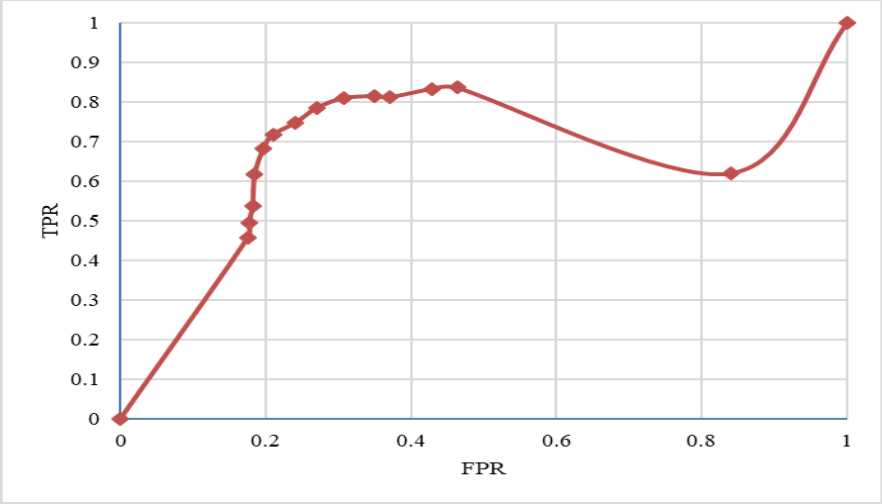

Figure 5. DT ROC Curve for OpSpam

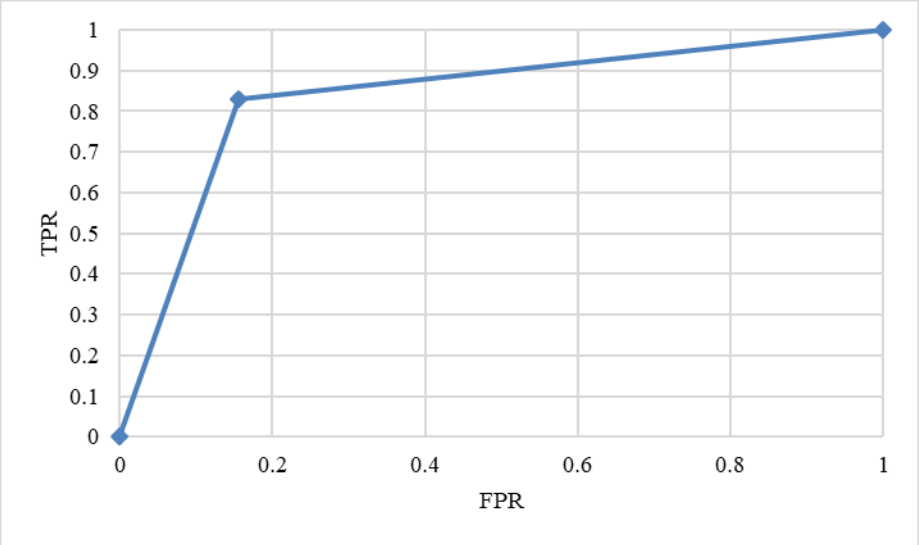

Figure 6. Curve of SVM

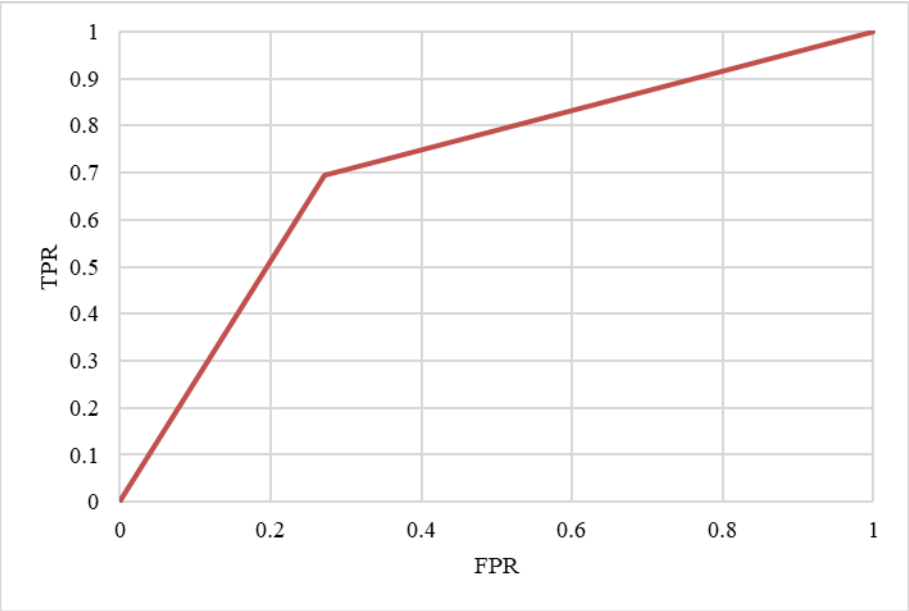

Figure 7. Curve of LSVM

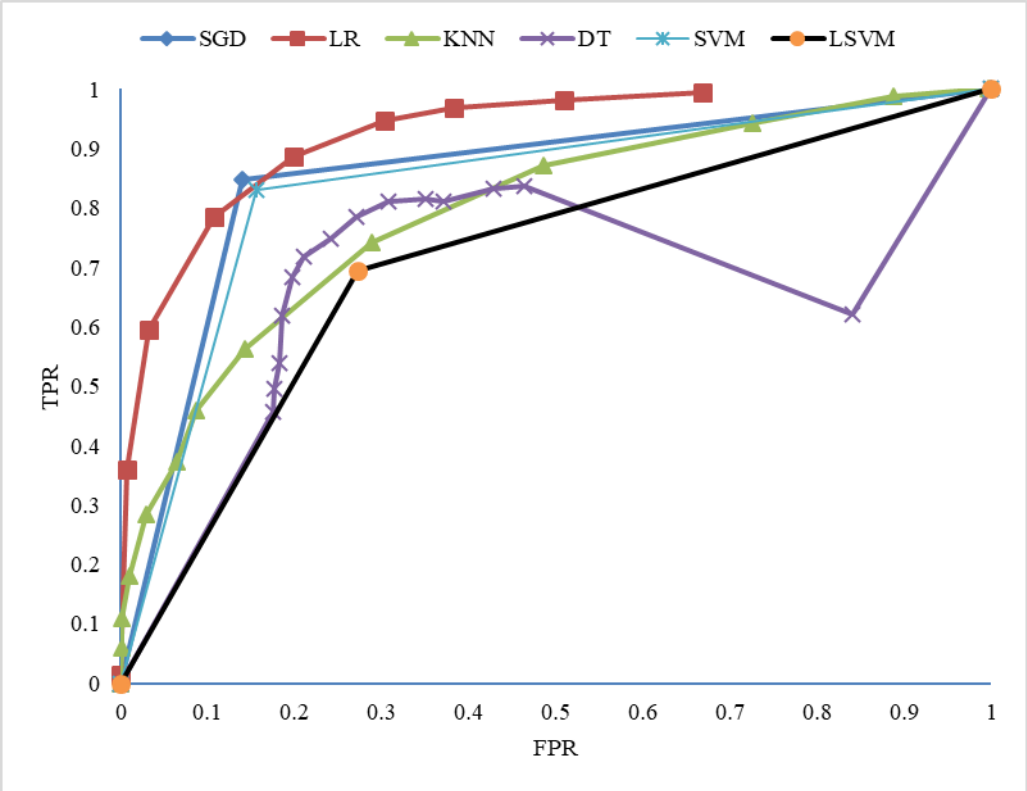

Figure 8. Combined Curve of all Models

The area under the curve of SGD is maximum followed by area under the curve of LR. The

combined ROC has been plotted in Figure 8.

4.2 Horne

In [47], a news dataset was created which consisted of real news articles from Buzzfeed and other news websites. It included satires from satire dataset in [48]. It included text files with titles and content of news articles. In this paper, some observations were made like fake news articles have more nouns and verbs and less stop words and nouns. Also, different features were extracted and grouped in three categories, namely, Complexity, Psychology and Stylistic.

Table 8. Result on Horne with different models

Model	**SGD**	**LR**	**KNN**	**DT**	**SVM**	**LSVM**
Accuracy	**90.87**	85.06	67.63	79.25	**93.36**	73.86
Precision	0.910	0.851	0.757	0.793	0.934	0.739
Recall	0.909	0.851	0.676	0.793	0.934	0.739
F1 Score	0.909	0.851	0.647	0.793	0.934	0.739

From table 8, it can be noted that best accuracy is achieved by using SVM model. 93.36% is the highest accuracy achieved followed by 90.87% accuracy given by SGD.

Table 9. Comparison of previous work and our work with Horne dataset

Model	**SGD**	**LR**	**KNN**	**DT**	**SVM**	**LSVM**	**Best Accuracy**
Accuracy in [2]	NA	NA	NA	NA	NA	**87**	87
Our accuracy	**90.87**	85.06	67.63	79.25	**93.36**	73.86	93.36

From Table 9, it can be noted that the best accuracy obtained for this dataset is through LSVM which is 87%. However, through our proposed model, we could obtain an accuracy of 90.87% by SGD and 93.36% through SVM. The main reason for these results can be the use of only TF-IDF feature for classification in [2]. However, we have used linguistic features also along with TF-IDF. ROC curve for each model has been plotted between True Positive Rate (TPR) and False Positive Rate (FPR) in following figures.

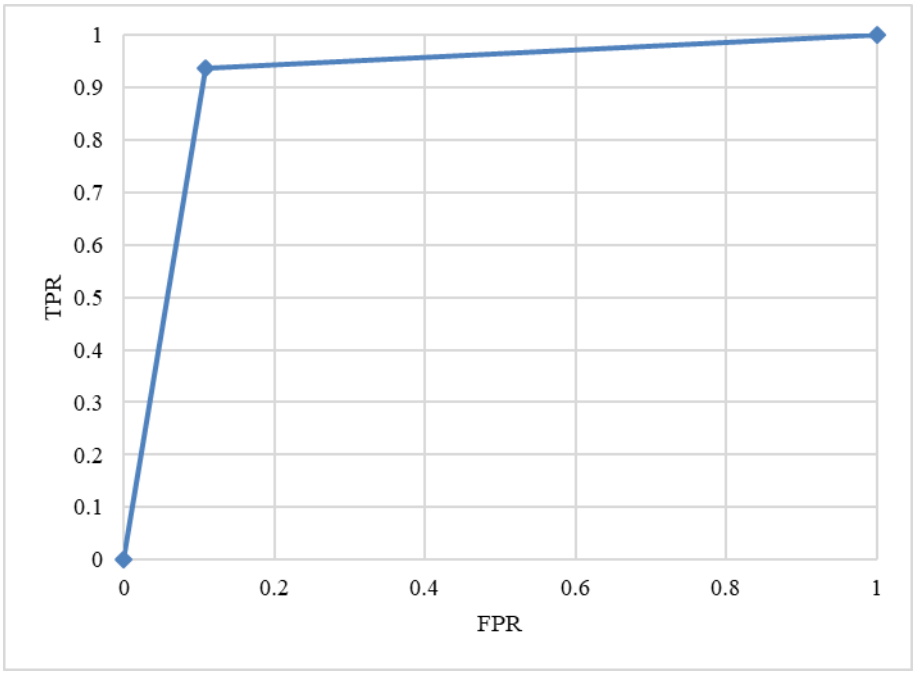

Figure 9. Curve of SGD

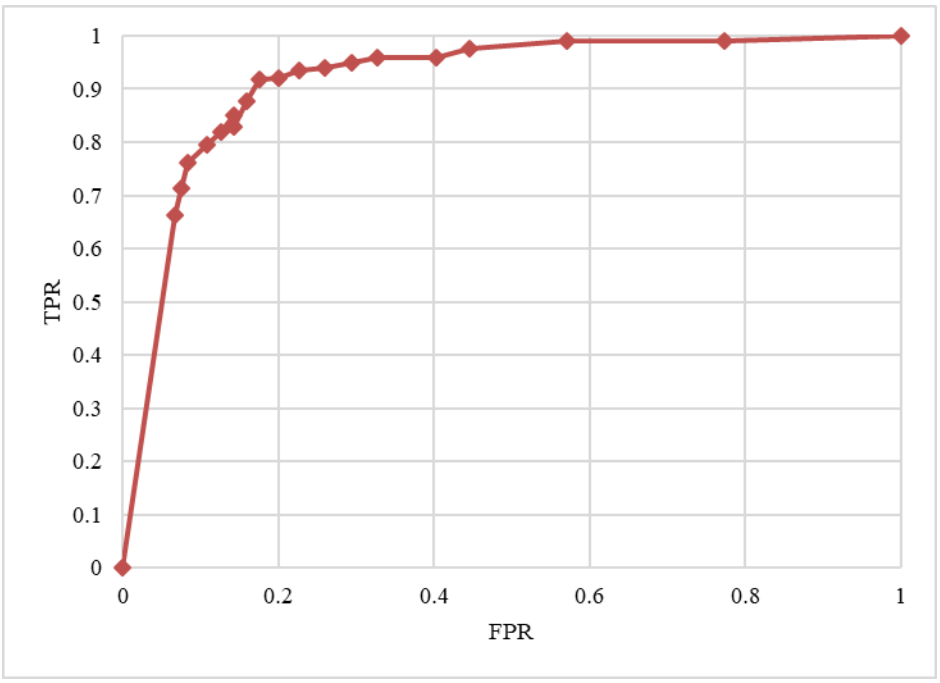

Figure 10. Curve of LR

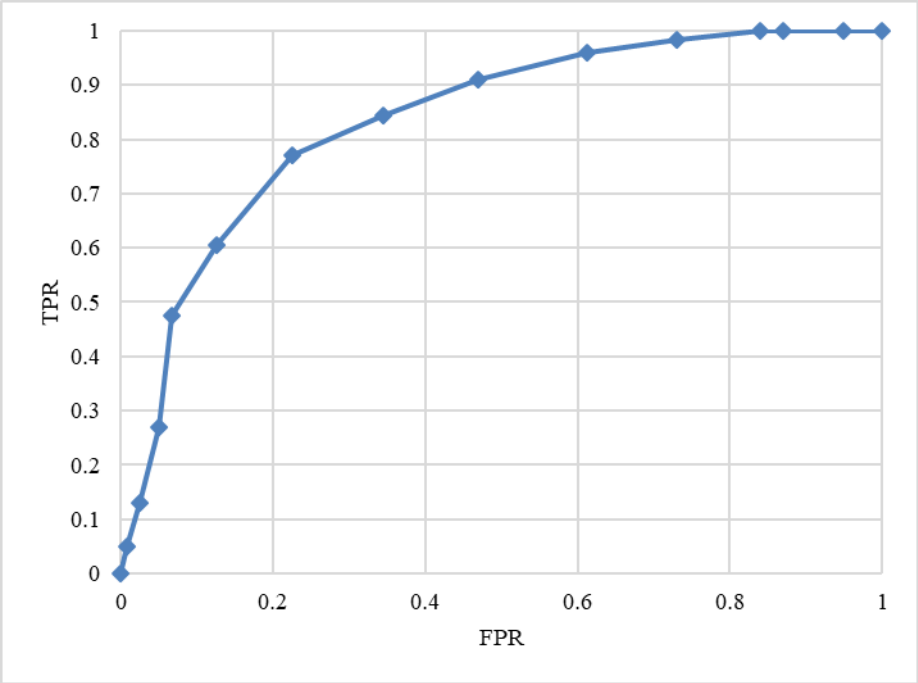

Figure 11. Curve of KNN

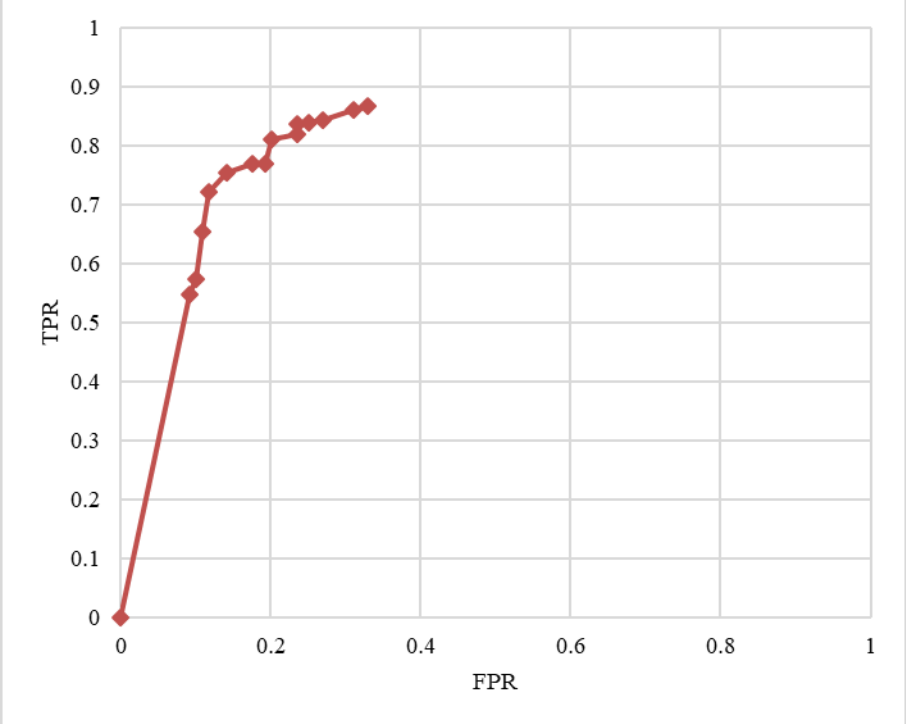

Figure 12. Curve of DT

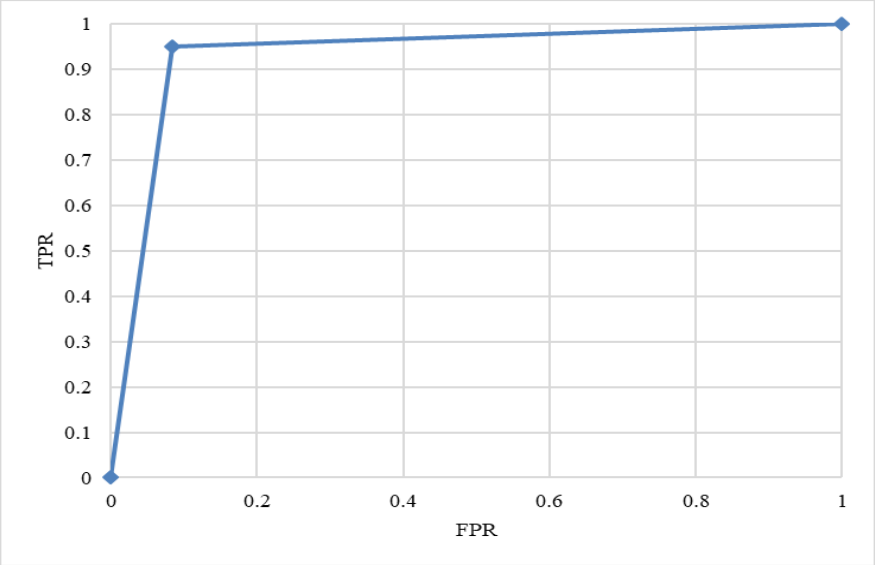

Figure 13. Curve of SVM

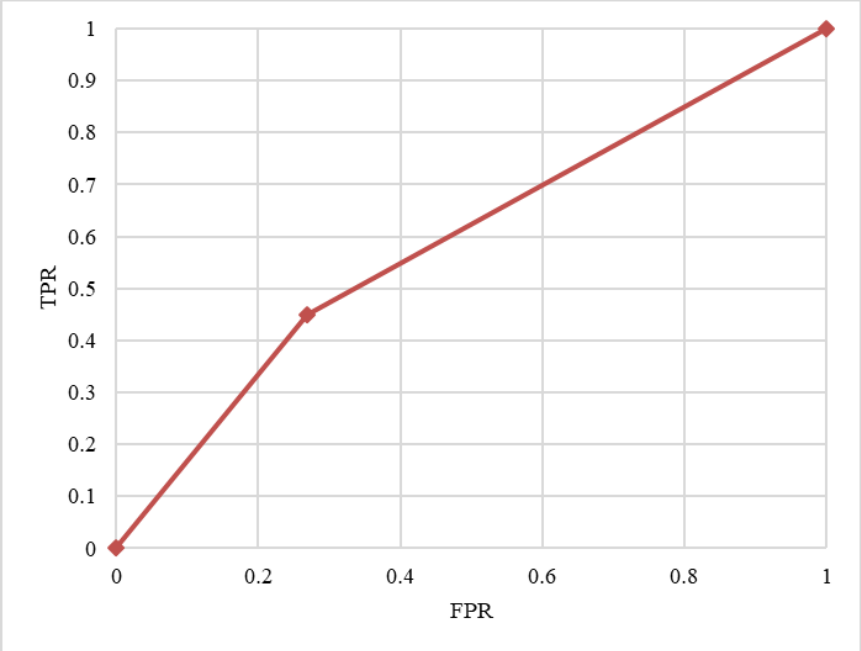

Figure 14. Curve of LSVM

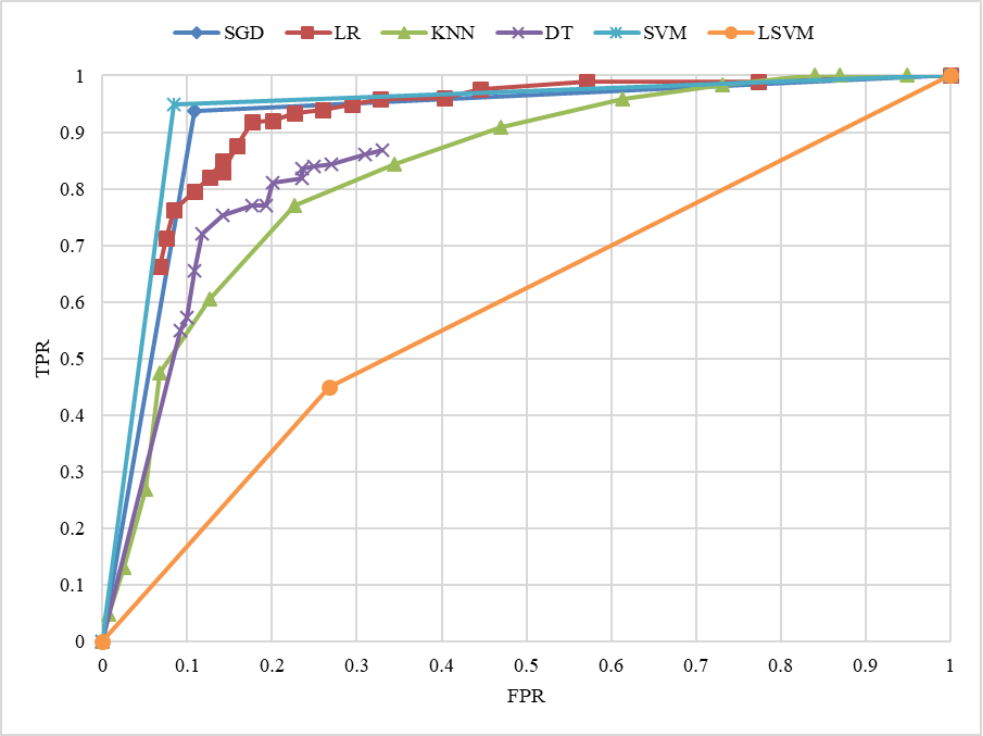

Figure 15. Combined ROC Curve of all models for Horne

The area under the curve is maximum for SVM followed by are under curve of SGD model. The combined ROC curves have been plotted in Figure 15.

4.3 MCIntire

MCIntire dataset is built from Kaggle's fake news dataset and from authentic journalistic organizations. This dataset is available online and the size of dataset satisfies the requirements for extensive text classification.

Table 10. Result on MCIntire dataset with different models

Model	**SGD**	**LR**	**KNN**	**DT**	**SVM**	**LSVM**
Accuracy	**91.94**	**91.94**	82.76	90.03	91.55	82.26
Precision	0.919	0.919	0.836	0.900	0.916	0.822
Recall	0.919	0.919	0.828	0.900	0.916	0.823
F1 Score	0.919	0.919	0.821	0.900	0.916	0.819

Best accuracy is obtained by SGD and LR which give 91.94%, followed by SVM which is

91.55%.

Table 11. Comparison of previous work and our work with MCIntire dataset

Model	**SGD**	**LR**	**KNN**	**DT**	**SVM**	**LSVM**	**Best Accuracy**
Accuracy in [56]	NA	NA	NA	NA	NA	NA	<85%
Our accuracy	91.94	91.94	82.76	90.03	91.55	82.26	91.94

Accuracies obtained by SGD, LR and SVM are more than the highest accuracy obtained in [56] on MCIntire Database. Maximum accuracy obtained in [56] for MCIntire is less than 85%. This suggests that use of TF-IDF feature along with linguistic features gives better results than the results obtained by using only linguistic features or TF-IDF to classify truthful and deceptive content. The ROC curves for each algorithm is given in following figures.

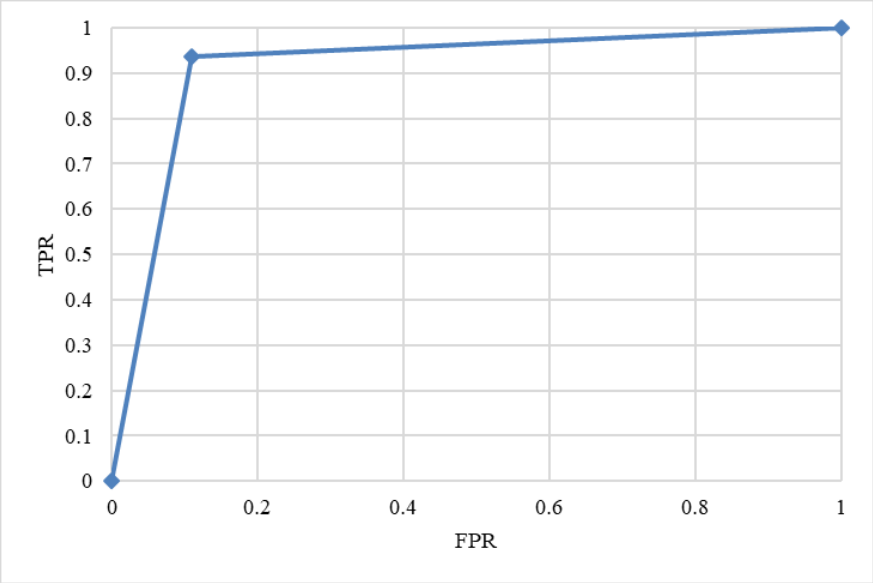

Figure 16: Curve of SGD

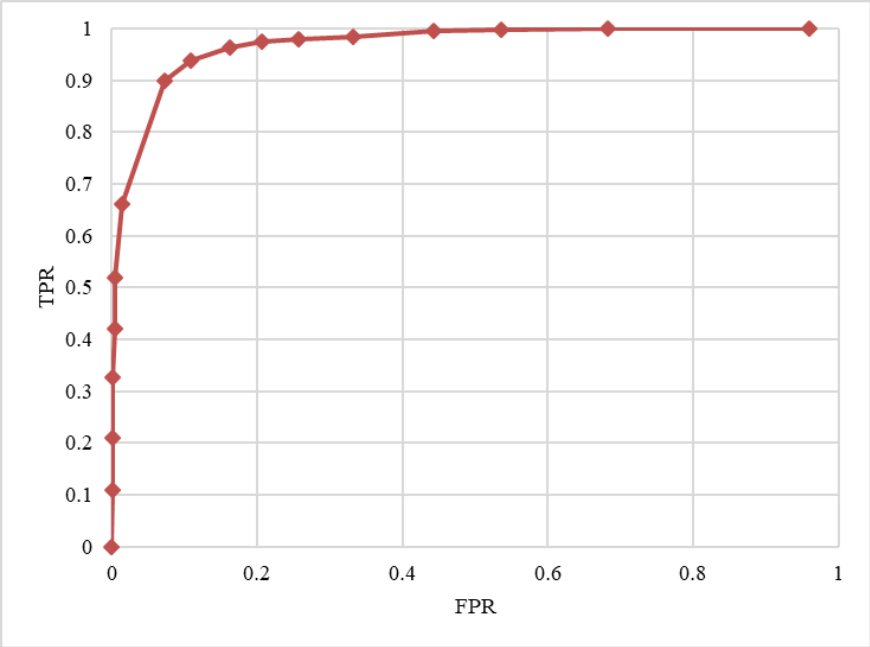

Figure 17. Curve of LR

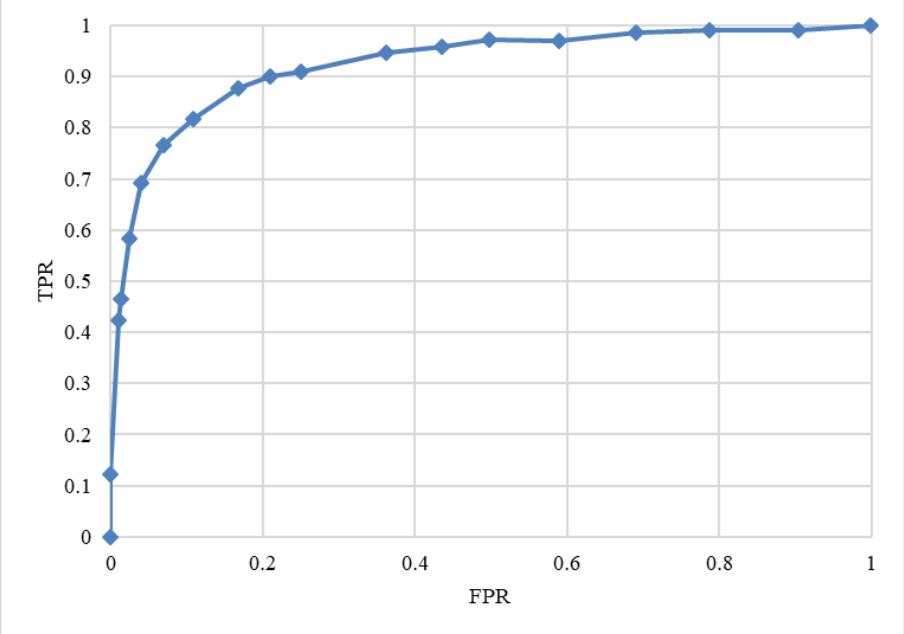

Figure 18. Curve of KNN

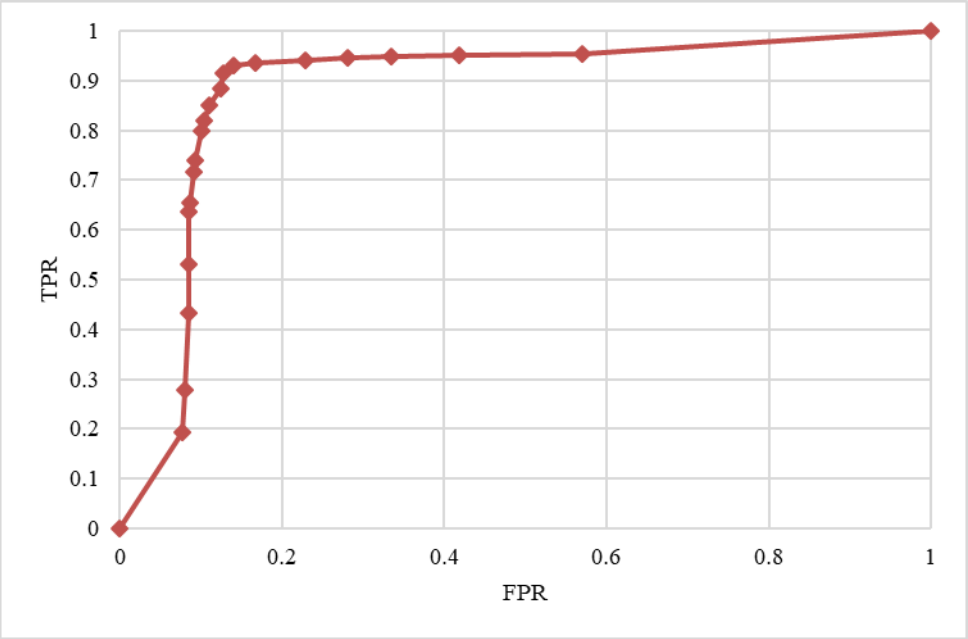

Figure 19. Curve of DT

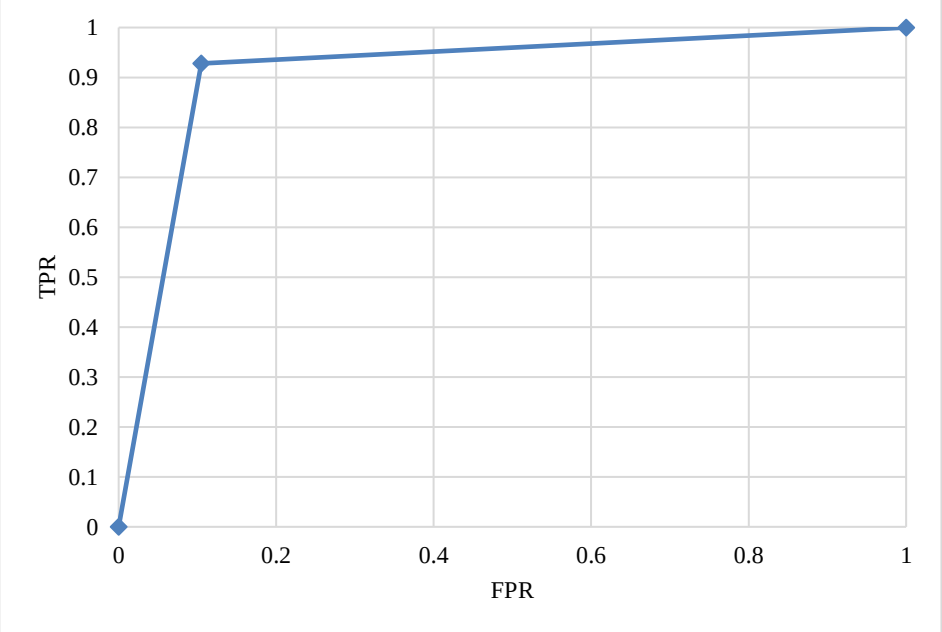

Figure 20. Curve of SVM

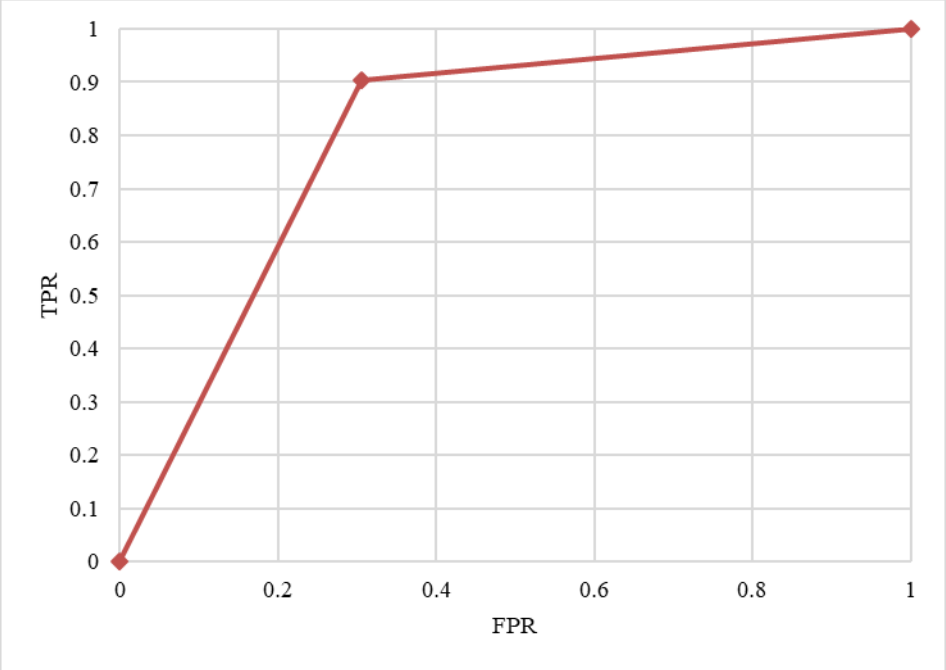

Figure 21. Curve of LSVM

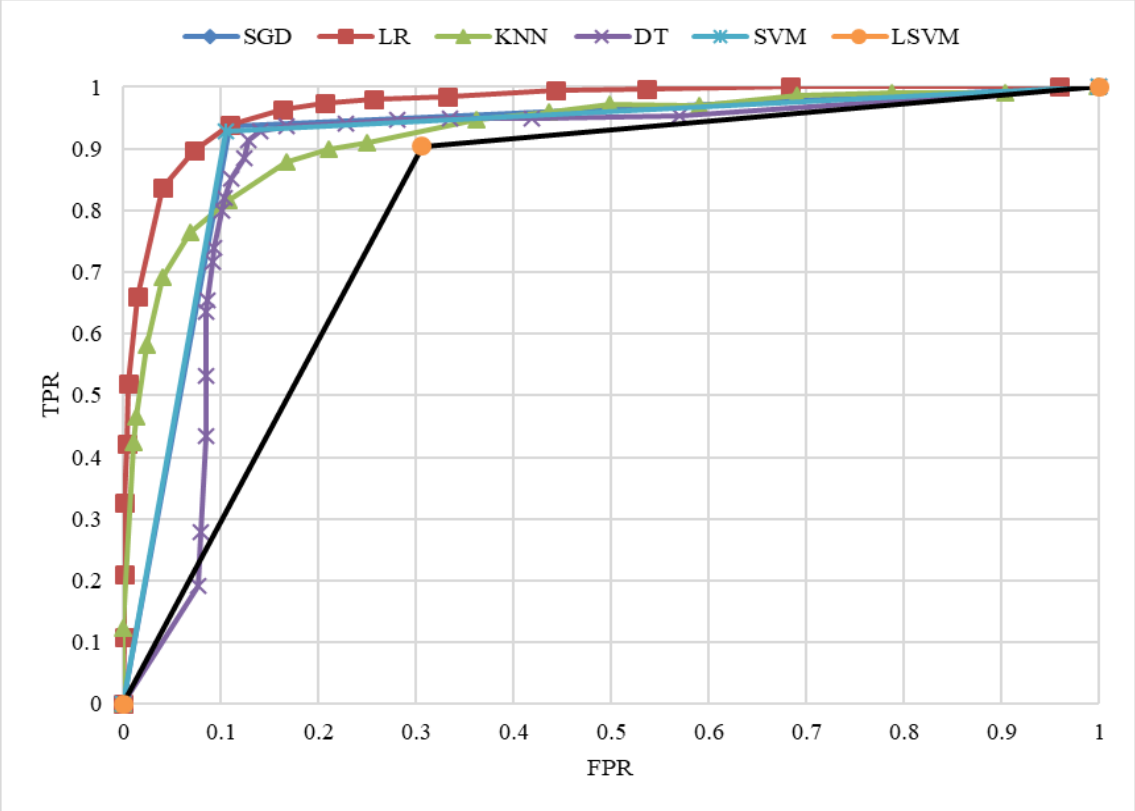

Figure 22. Combined Curve of all models

Area under curve is maximum for SGD and LR, followed by SVM. The combined ROC curve is shown in Figure 22.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

This work used TF-IDF along with many linguistic features which include statistical features and the features that reflect the writer's styles to classify fake and real news or reviews. The experimental result show that the proposed approach helps to classify real and fake content with more accuracy as compared to an approach in which only TF-IDF features are used. Higher accuracy was achieved for OpSpam for some models than accuracy achieved in [2]. SVM gave highest accuracy for Horne and higher than best accuracy achieved in [2] for Horne. Also, accuracy achieved for MCIntire was higher than best accuracy achieved in [56]. Feature extraction process required study of libraries and tools since 55 features were extracted. Extracting linguistic features is sometimes challenging since it is more dependent on definitions and there may be ambiguities which are language specific.

6.2 Future Work

This thesis proposed the use of linguistic features along with TF-IDF to distinguish between fake and real content. However, additional features related to the source of article, author of articles and propagation pattern of such deceptive content can also be added as they usually give insight to the authenticity of text. Also, semi-supervised and unsupervised techniques can be explored to detect fake content since in real world it is tough to collect accurately labeled real datasets. Labelling by experts or journalists is required to ensure that content is authentic which is tough to do.

References

[1] N. Jindal and B. Liu, "Opinion Spam and Analysis," in *Proceedings of the 2008 international conference on web search and data mining*, New York, NY: ACM, 2008.

[2] H. Ahmed, I. Traore and S. Saad, "Detecting opinion spams and fake news using text classification," *Security and Privacy,* vol. 1, no. 1, 2018.

[3] V. P´erez-Rosas, B. Kleinberg, A. Lefevre and R. Mihalcea, "Automatic Detection of Fake News," 2017.

[4] Z. Jin, J. Cao, Y. Zhang, J. Zhou and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE Transactions on Multimedia,* vol. 19, no. 3, pp. 598-608, 2017.

[5] Z. Zhao, J. Zhao, Y. Sano, O. levy, H. Takayasu, M. Takayasu, D. Li, J. Wu and S. Havlin, "Fake news propagate differently from real news even at early stages of spreading.," 2018.

[6] M. Egele, G. Stringhini, C. Kruegel and G. Vigna, "Towards detecting compromised accounts on social networks," *IEEE Transactions on Dependable and Secure Computing,* vol. 14, no. 4, pp. 447-460, 2017.

[7] A. Campan, A. Cuzzocrea and T. M. Truta, "Fighting fake news spread in online social networks: Actual trends and future research directions," in *IEEE International Conference on Big Data (Big Data)*, Boston, MA, 2017.

[8] E. Okoro, B. Abara, U. Alex and Z. Isa, "A hybrid approach to fake news detection on social media," in *Nigerian Journal of Technology (NIJOTECH*, Nsukka, 2018.

[9] M. Alrubaian, M. Al-Qurishi, M. M. Hassan and A. Alamri, "A Credibility Analysis System for Assessing Information on Twitter," *IEEE Transactions on Dependable and Secure Computing,* vol. 15, no. 4, pp. 661-674, 2018.

[10] A. Figueira and L. Oliveira, "The current state of fake news: challenges and opportunities," *Procedia Computer Science,* vol. 121, pp. 817-825, 2017.

[11] E. Tacchini, G. Ballarin, M. L. D. Vedova, S. Moret and L. d. Alfaro, "Some Like it Hoax: Automated Fake News Detection in Social Networks," School of Engineering,

University of California, Santa Cruz, California, 2017.

[12] B. Riedel, I. Augenstein, G. P. Spithourakis and S. Riedel, "A simple but tough-to-beat baseline for the Fake News Challenge stance detection task," 2018.

[13] Z. Jin, J. Cao, Y. Zhang and J. Luo, "News Verification by Exploiting Conflicting Social Viewpoints in Microblogs," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Arizona, 2016.

[14] S. B. Jr, R. A. Igawa and B. B. Zarpelão, "Authorship verification applied to detection of compromised accounts on online social networks," *Multimedia Tools and Applications*, vol. 76, no. 3, p. 3213–3233, 2016.

[15] C. Chen, K. Wu, V. Srinivasan and X. Zhang, "Battling the Internet Water Army: Detection of Hidden Paid Posters," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13)*, NY, USA, 2013.

[16] N. J. Roy, V. L. Rubin and Y. Chen, "Automatic Deception Detection: Methods for Finding Fake News," in *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, St. Louis, Missouri, 2015.

[17] H. Rashkin, E. Choi, J. Yea Jang and S. C. Y. Volkova, "Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017.

[18] V. Rubin, N. Conroy and Y. Chen, "Towards News Verification: Deception Detection Methods for News Discourse," in *ASIST '15 Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Communit*, Missouri, 2015.

[19] J. Kim, B. Tabibian, A. Oh, B. Schölkopf and M. G. Rodriguez, "Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation Reduce," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18)*, NY, USA, 2018.

[20] J. Roozenbeek and S. van der Linden, "The fake news game: actively inoculating against the risk of misinformation," *Journal of Risk*, 2018.

[21] V. L. Rubin, Y. Chen and N. J. Conroy, "Deception Detection for News: Three Types of

Fakes," in *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (ASIST '15)*, MD, USA, 2015.

[22] N. Ruchansky, S. Seo and Y. Liu, "CSI: A Hybrid Deep Model for Fake News Detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*, NY, USA, 2017.

[23] K. Shu, A. Sliva, S. Wang, J. Tang and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," *SIGKDD Explor. Newsl.,* vol. 19, no. 1, pp. 22-36, 2017.

[24] P. Bourgonje, J. Moreno Schneider and G. Rehm, "From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles," in *Association for Computational Linguistics*, Copenhagen, Denmark, 2017.

[25] S. Sirajudeen, N. Azmi and A. Abubakar, "Online Fake News Detection Algorithm," *Journal of Theoretical and Applied Information Technology,* vol. 95, pp. 4114-4122, 2017.

[26] Y. Chen, N. J. Conroy and V. L. Rubin, "Misleading Online Content: Recognizing Clickbait as “False News”," in *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection (WMDD '15)*, NY, USA, 2015.

[27] Y. Long, Q. Lu, R. Xiang, M. Li and C.-R. Huang, "Fake News Detection Through Multi-Perspective Speaker Profiles," in *Asian Federation of Natural Language Processing*, Taipei, Taiwan, 2017.

[28] K. Shu, S. Wang and H. Liu, "Exploiting Tri-Relationship for Fake News Detection," *CoRR*, vol. abs/1712.07709, 2017.

[29] A. Mukherjee, V. Vivek, L. Bing and G. Natalie, "Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews," 2013.

[30] S. Volkova, K. Shaffer, J. Yea Jang and N. Hodas, "Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter," *Association for Computational Linguistics*, no. 10.18653/v1/P17-2102 , p. 647–653, 2017.

[31] S. Elham, R. Guo and P. Shakarian, "Detecting Pathogenic Social Media Accounts without Content or Network Structure," in *1st International Conference on Data Intelligence and Security (ICDIS)*, TX, USA, 2018.

[32] S. Kai, S. Wang and H. Liu, "Understanding User Profiles on Social Media for Fake News Detection," in *IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Miami, FL, 2018.

[33] C. Buntain and J. Golbeck, "Automatically Identifying Fake News in Popular Twitter Threads," in *IEEE International Conference on Smart Cloud (SmartCloud)*, New York, NY, USA, 2017.

[34] S. Gilda, "Evaluating machine learning algorithms for fake news detection," in *IEEE 15th Student Conference on Research and Development (SCOReD)*, Putrajaya, 2017.

[35] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in *IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, Kiev, 2017.

[36] P. Pourghomi, F. Safieddine, W. Masri and M. Dordevic, "How to Stop Spread of Misinformation on Social Media: Facebook Plans vs. Right-click Authenticate Approach," in *2017 International Conference on Engineering & MIS (ICEMIS)*, Monastir, Tunisia, 2017.

[37] Z. Jin, J. Cao, Y.-G. Jiang and Y. Zhang, "News Credibility Evaluation on Microblog with a Hierarchical Propagation Model," in *2014 IEEE International Conference on Data Mining*, Shenzhen, China, 2014.

[38] C. Shao, G. L. Ciampaglia, A. Flammini and F. Menczer, "Hoaxy: A Platform for Tracking Online Misinformation," in *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)*, Republic and Canton of Geneva, Switzerland, 2016.

[39] S. . B. Parikh and P. K. Atrey, "Media-Rich Fake News Detection: A Survey," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Miami, FL, 2018.

[40] M. Ott, Y. Choi, C. Cardie and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1. Association for Computational Linguistics*, Portland, Oregon, 2011.

[41] W. Y. Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News

Detection," in *ACL*, 2017.

[42] A. Mukherjee, V. Venkataraman, B. Liu and N. Glance, "Fake review detection: Classification and analysis of real and pseudo reviews.," UIC-CS-03-2013, 2013.

[43] K.-H. Yoo and U. Gretzel, "Comparison of Deceptive and Truthful Travel Reviews," in *Information and Communication Technologies in Tourism*, Vienna, Springer, 2009, pp. 37-47.

[44] J. K. Burgoon, J. P. Blair, T. Qin and J. F. N. Jr, "Detecting deception through linguistics analysis," in *International Conference on Intelligence and Security Informatics*, Berlin, Heidelberg, 2003.

[45] M. L. Newman, J. W. Pennebaker, D. S. Berry and J. M. Richards, "Lying words: Predicting deception from linguistic styles.," *Personality and social psychology bulletin 29*, vol. 29, no. 5, pp. 665-675, 2003.

[46] L. Zhou, J. K. Burgoon, J. F. Nunamaker and D. Titchell, "Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications.," *Group decision and negotiation*, vol. 13, no. 1, pp. 81-106, 2004 .

[47] B. D. Horne and S. Adali, "This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news.," in *Eleventh International AAAI Conference on Web and Social Media*, 2017.

[48] C. Burfoot and T. Baldwin, "Automatic Satire Detection: Are You Having a Laugh?," in *Proceedings of the ACL-IJCNLP 2009 conference short papers*, Suntec, Singapore, 2009.

[49] A. Hadeer, I. Traore and S. Saad, "Detecting opinion spams and fake news using text classification," in *Security and Privacy*, 2018, p. 1:e9.

[50] V. Pérez-Rosas, B. Kleinberg, A. Lefevre and R. Mihalcea, "Automatic detection of fake news.," 2017.

[51] Z. Jin, J. Cao, Y. Zhang, J. Zhou and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE transactions on multimedia*, vol. 19, no. 3, pp. 598-608, 2016 .

[52] M. Egele, G. Stringhini, C. Kruegel and G. Vigna, "Towards detecting compromised accounts on social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 4, pp. 447-460, 2015.

[53] Z. Zhao, J. Zhao, Y. Sano, O. Levy, H. Takayasu, M. Takayasu, D. Li and S. Havlin, "Fake news propagate differently from real news even at early stages of spreading," 2018.

[54] K. Shuy, A. Slivaz, S. Wangy, J. Tang and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22-36, 2017.

[55] M. Ott, Y. Choi, C. Cardie and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in {Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, Portland, Oregon, Association for Computational Linguistics, 2011, p. 309–319.

[56] Gravanis, Georgios, Athena Vakali, Konstantinos Diamantaras, and Panagiotis Karadais. "Behind the Cues: A benchmarking study for Fake News Detection." Expert Systems with Applications (2019).

[57] Feasley, Eliana, and Wesley Tansey. "Detecting Deception in On and Off-line Communications."

[58] Hernández-Castañeda, Á., Calvo, H., Gelbukh, A. et al. Soft Comput (2017) 21: 585. https://doi.org/10.1007/s00500-016-2409-2

LIST OF PUBLICATIONS BY CANDIDATE

[1] Katarya, Rahul, and Chhavi Jain. "Multilayered Risk analysis of Mobile systems and Apps." In *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 64-67. IEEE, 2018.
[2] Submitted: Fake News Detection: A Review
[3] Submitted: Detecting Fake News and Fake Reviews through Linguistic Styles