

An Effective Optimized Fake News Detection System based on Machine Learning Techniques

A DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE AWARD OF DEGREE

OF

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

RAVISH

2K20/CSE/19

Under the supervision of

Dr. Rahul Katarya

(Professor)



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

JUNE, 2022

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

CANDIDATE'S DECLARATION

I, Ravish, RollNo. 2K20/CSE/19 student of M.Tech(Computer Science and Engineering), hereby declare that the project Dissertation titled “**An Effective Optimized Fake News Detection System based on Machine Learning Techniques**” 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 and Degree, Diploma Associate ship, Fellowship or other similar title or recognition.

Place: Delhi

Ravish

Date:

2K20/CSE/19

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled “**An Effective Optimized Fake News Detection System based on Machine Learning Techniques**” which is submitted by Ravish, RollNo. 2K20/CSE/19, 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 a record of the project work carried out by the students 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

Date:

Dr. Rahul Katarya

Professor

Department of CSE

ACKNOWLEDGMENT

The success of this project requires the assistance and input of numerous people and the organization. I am grateful to everyone who helped in shaping the result of the project.

I express my sincere thanks to **Dr. Rahul Katarya**, my project guide, for providing me with the opportunity to undertake this project under his guidance. His constant support and encouragement have made me realize that it is the process of learning which weighs more than the end result. I am highly indebted to the panel faculties during all the progress evaluations for their guidance, constant supervision and for motivating me to complete my work. They helped me throughout with new ideas, provided information necessary and pushed me to complete the work.

I also thank all my fellow students and my family for their continued support.

RAVISH

2K20/CSE/19

ABSTRACT

Fake News generates misleading suspense information that may be discovered. This promotes dishonesty about something like a country's position or overstates the price of specific tasks for a government, eroding democracy in particular places, such as with the Arab Spring. Organisations such as the "House of Representatives and the Background check project" try to address problems such as publisher accountability. But, as they depend exclusively on human detection by people, their coverage is small. This is not sustainable nor practicable in a world where billions of things are removed or uploaded every second. So in this publication, we suggested a strategy by employing Multi-SVM (MSVM) to identify bogus news with higher dependable accuracy and For the purpose, we are using multi-layer PCA for selecting features. The principal component analysis decreases the dimension for dataset having a huge number of the connected variables and remembers the largest change in real data. The key characteristics will be picked using firefly optimised algorithm. Several experiments have been undertaken to increase the standardised firefly algorithm's competency and adjust it to the nature of the situation. Good aspect for the suggested strategy is that it will straighten the algorithms to get a fantastic accuracy of 99.64 percent and lowered 20 percent execution time. Therefore, it delivers superior results for fake news detection performance measurements.

CONTENTS

Candidate's Declaration	i
Certificate	ii
Acknowledgement	iii
Abstract	iv
Contents	v
List of Figures	viii
List of Tables	ix
List of Abbreviations	x
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 PRIOR WORK	4
CHAPTER 3 FUNDAMENTALS	10
3.1 Fake News Detection System	10
3.2 Categories of Fake News	10
3.3 Limitation of Fake News	11
3.4 Application of Fake News	11
3.5 SVM(support vector machine)	11

CHAPTER 4 PROPOSED WORK	13
4.1 Problem statement	13
4.2 Motivation	13
4.3 METHODOLOGY	14
4.3.1 Proposed Methodology	14
4.3.2 Data Gathering & Data Pre-Processing	14
4.3.3 Extract the Features, Select the Features, and Classify	14
CHAPTER 5 Data Evaluation	16
5.1 DATASET	16
CHAPTER 6 EXPERIMENTS AND RESULTS	20
6.1 Analysis and Visualization of the Experimental Result	20
6.1.1 Accuracy Test on Poltifact dataset	20
6.1.2 Accuracy Test on FakeNewsNet dataset	21
6.1.3 Accuracy Test on BUZZFEED dataset	21
6.1.4 Accuracy Test on Weibo dataset	22
6.1.5 Accuracy Test on LIAR dataset	22
6.1.6 Accuracy Test on PHEME dataset	23
6.1.7 Performance Test on GossipCop dataset	23

6.1.8 Performance Test on Twitter Election Integrity dataset	24
6.1.9 Performance Test on ISOT dataset	24
6.1.10 Performance Test on COVID19 dataset	25
6.2. Performance Graphs on Fake or Real Dataset	25
6.2.1 Histogram based on News Distribution (Fake and Real)	25
6.2.2 Training Dataset	26
6.2.3 Parameter : Accuracy Rate	27
6.2.4 Precision	27
6.2.5 Recall	28
6.2.6 Specificity	28
CHAPTER 7 CONCLUSION AND FUTURE SCOPE	29
REFERENCES	29
LIST OF PUBLICATIONS	33

LIST OF FIGURES

1.1	Need of Fake News Detection	1
4.1	Flowchart of proposed model	25
5.1	Fake Dataset wordcloud	26
5.2	Real Dataset wordcloud	26
6.1	Proposed Model Comparison on: Poltifact	30
6.2	Proposed Model Comparison on: FakeNewsNet	31
6.3	Proposed Model Comparison on: BUZZFEED	31
6.4	Proposed Model Comparison on: Weibo	32
6.5	Proposed Model Comparison on: LIAR	32
6.6	Proposed Model Comparison on: PHEME	33
6.7	Proposed Model Comparison on: GossipCop	33
6.8	Proposed Model Comparison on: Twitter Election Integrity	34
6.9	Proposed Model Comparison on: ISOT	34
6.10	Proposed Model Comparison on: COVID19	35
6.11	Histogram for news distribution	35
6.12	Training Dataset	36
6.13	Accuracy	37
6.14	Precision	37
6.15	Recall	38
6.16	Specificity	38

LIST OF TABLES

1.1	Summarized review of literature papers	14
5.1	Dataset information	27

LIST OF ABBREVIATIONS

1. FN: Fake News
2. MSVM: Multi-support vector machine
3. PCA: Principal Component Analysis
4. SM: Social Media
5. FND: Fake News Detection
6. FB: Facebook
7. CNN: Convolutional Neural Network
8. OCL: One-Class Learning
9. SVM: Support Vector Machine
10. FE: Feature Extraction
11. FA: Firefly Optimization

CHAPTER 1

INTRODUCTION

Generation invests most of the time conversing on SM, as the universal popularity of clever phones generates their access available practically everywhere and whenever. They ease contact with family members, relatives, friends, and strangers via the review or comment basis, through debate, comments/quickly like and dislike buttons [1].

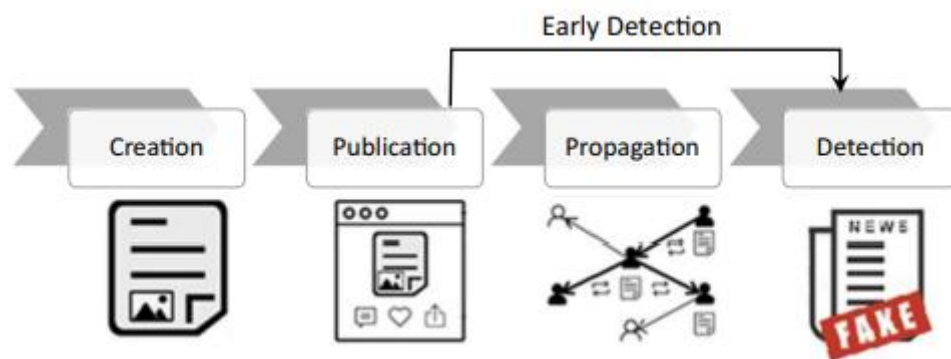


Figure 1.1. Need of Fake News Detection [2]

Figure 1.1 showcases the need for FND. After creation of fake news, it takes very less time for propagation, but if we detect it before propagation than we can save people from getting influenced by the propaganda of the news, e.g. following the 2016 presidential elections, it was found that influence of “fake news” has been a huge problem. Based on a large investigation of ~126,000 verified true and fraudulent news items on Twitter from 2006 to 2017, Vosoughi and colleagues revealed that fake articles spread more frequently and quickly compared to authentic news stories [2]. As stated by the key theories on fake news in psychology and social science [3], the more a false news story spreads, the higher the probability of social media users spreading and believing it due to repeated exposure and/or peer pressure. Such levels of trust and beliefs may quickly be multiplied and perpetuated within social media thanks to its echo chamber effect [2]. Hence, a large inquiry has been undertaken on the effective identification of false news to prevent its transmission on social media. Fake news

detection approaches may be typically divided into (1) content-based and (2) social-context-based tactics. The primary difference between these two types of innovations is whether or not they revolve around social context information: the data on how the news has been shared on social media on social media, where rich secondary research about people on social media engagement and their connections/networks can sometimes be exploited. Many unique and substantial strategies [2], [4] have been presented to use social context information. As we can see in Figure 1 after production and publishing, the most hazardous aspect is about propagation, we need to halt false news at its base before it spreads its branches, thus we utilise FND to block this at its roots.

According to the study, in 2020, fifty-three percent of United States individuals said they got news from SM "occasionally" or "often," with fifty-nine percentage of Twitter consumers and forty percent of FB (FACEBOOK) customers spending news on the site regularly. New and creative ways, fifty-nine percent of people who achieved news on SM indicated they predicted such news to be significantly imprecise[2]. SM provides least cost, quick access, and faster data streaming. The bulk of consumers examine news through SM contrasted with the established institutions these days. Basically, on the one hand, SM has performed and become a big source of data and carried people together. On the other side, it is damaging society.

FN detection on SM confronts several problems. Initially, it was not simple to acquire FN information. Also, It is not possible to mark FN manually. It is closed messaging apps because they are purposefully built to trick readers. The distortion shared by trustworthy news/ their family and friends is not simple to regard as a FAKE. It is tough to recognise the dependability of fresh emerging and restricted news since they are not simple to train the app database[3]. Essential approaches to designate reputable consumers, extract reliable or useful news characteristics, and progress a security-based data system are dependable study fields that need future research.

Various strategies to handle the problem of disinformation on SM. Statistical tools check the link among numerous data aspects; pattern analyses the dispersion. Usually, ML-based algorithms categorise dubious material and examine the accounts that disseminate such stuff. Several approaches are involved with developing security or authentication mechanisms and specialised case analysis [4].

So we have introduced our model in which we are using optimised MSVM algorithms with the help of feature attraction using PCA and the feature selection will

be done by the optimised firefly algorithm we have tested our algorithm on more than 8 different datasets which includes some famous datasets FakeNewsNet, PHEME, LIAR, covid19 2020 twitter dataset, PolitiFact, ISOT, Weibo and many more the performance we are getting on our algorithm is more reliable than any other algorithm which we are comparing in this article. In this research article first, we are discussing the different models which are present in the market with multiple datasets and we discussed where these algorithms are lacking in which our model is outperforming and we will discuss our model and implementation in forth section and fifth section is all about the different datasets we have used and compare the results of performance with other models the FND model's performance is measured by their accuracy, precision, recall, f1-score, etc. In the concluding part, we explained our conclusion and the outcome. Some notable contribution are:

-
- *Machine learning-based system optimized MSVM is proposed for detecting fake news.*
 - *Proposed model is used to detect fake news on the given dataset of news article tweets or any other social media post.*
 - *Firefly and PCA are used for enhancing the feature selection and feature extraction in this system.*
 - *Experimented results show some improved accuracy than many existing models on 10 popular datasets*
-

CHAPTER 2

PRIOR WORK

With the growth and verity of technology, there are plenty of suggested models which are employing diverse technologies. Others of the literature studies are employing machine learning methods, some utilising deep learning models, and many of them are using hybrid methodologies. There are some literature papers which we have studied and tested with our models on different datasets as we have mentioned their accuracy also concerning their dataset we have made this table consist of references which will be helpful for the reader to get a better understanding about technology they are using on which dataset and what accuracy they are getting through these models and after this table we will discuss about limitation of these model. The summary of the complete literature research, combined with their strengths and shortcomings, are presented in Table 2.1.

Table 2.1. Summarized review of literature papers

S.No	Ref.	Methodology	Dataset	Advantages	Disadvantages
1.	[2]	Text-CNN is utilized in Similarity-Aware Fake news detection technique (SAFE) which enables it to analyses multi-modal (textual and visual) important data of media articles.	PolitiFact GossipCop	Getting good performance on visual based dataset.	Not so much accurate on text base dataset.
2.	[5]	WELFake is a two-phase benchmark model for FND that uses machine learning classification and is built on word embedding over linguistic data.	FakeNewsNet BUZZFEED	Give better performance on test base dataset as compare to above.	Used synthetic dataset not good for visual dataset.

3.	[6]	In UFNDA for feature extraction GRU(Bi-GRU) layer to detect fake news	weibo	Used for unsupervised FND for text and media both.	Feature combination is very complex and that make its accuracy less.
4.	[7]	Used two phase learning one is unsupervised learning for that SSL is proposed with CNN for supervised learning.	LIAR PHEME	Good performance on less labeled or semi labeled dataset.	Accuracy is not that much good.
5.	[8]	One-Class Learning (OCL) and Positive and Unlabeled Learning (PUL) seem to be promising approaches for detecting content-based false news using a smaller number of labelled data than typical machine learning algorithms.	FakeNewsNet Fact-checked news	Semi-supervise ml algorithm and strategy used for unlabeled learning.	Accuracy is not that much high as compare to other dataset.
6.	[9]	BERT technique utilised for mutual learning framework that integrates relational characteristics classification (RFC) for named entity recognition (NER).	PolitiFact Pymedia	To simplify the problem BERT is used.	Time complexity is very high than expected.
7.	[10]	Fakenews detection technique based on Natural Language Inference (NLI) methodology with the aid of BERT is employed.	FakeNewsNet LIAR	Good for text base dataset.	Not good for typical dataset like liar.
8.	[11]	Through ensemble approaches, CNN was utilised to combine stylometric characteristics and text-based word vector representations.	Politifact	Used assembled approach of deep learning makes it less complex.	Accuracy is good but precision, recall is not good.
9.	[12]	One-Class Learning (OCL) and Positive and Unlabeled Learning (PUL) seem to be promising approaches	Politifact	Used Bi-LSTM model for high performance.	Higher execution time for every epoch

		for detecting content-based false news using a smaller number of labelled data than typical machine learning algorithms.			
10.	[13]	Through ensemble approaches, CNN was utilised to combine stylometric data with text-based word vector representations.	Poltifact Gossipcop	Deep learning is used make algorithmic problem simpler	Accuracy is not up to the mark
11.	[14]	SVM two-layer consensus utilizing Byzantine fault-tolerant (BFT) and asynchronous tolerance widely accepted coin protocols employed.	Poltifact	Better quality result	Accuracy can be increase ca be use MSVM
12.	[15]	crawler and labeling module's features vector extraction by using conspiracy classifier. BORJIS using NLP	FakeNewsNet	Better accuracy	Old algorithm tested on only one dataset
13.	[16]	In ASSO-OSIW Salp Swarm Optimization (SSO) based on a non-linear significantly reduced rate of change and oscillating inertia weight is employed for implementation.	ISOT LIAR Buzzfeed Poltifact	Accuracy is greater than most of the algorithm.	Accuracy is good but precision and recall is not that good.
14.	[17]	BERT tool designed for Arabic text	FakeNewsNet	Great accuracy	Old tech.
15.	[18]	Multi Neural Network is used in the proposed method	PHEME Weibo	Proposed method make problem simple.	Accuracy is not that good.
16.	[19]	a Bi-LSTM based (Bidirectional long short term memory) deep learning strategy by putting self-attention on top of it.	FakeNewsNet	Accuracy is getting better in every epoch	Loss is also great
17.	[20]	LSTM	FakeNewsNet	Accuracy is good.	Only tested on one dataset.
18.	[21]	Applied multiple machine learning a models for feature	FakeNewsNet	Make problem simple for	Tested on only one dataset so cannot make it

		attraction.		feature attraction	trustful.
19.	[22]	Using the multiplicative fusion approach, CT-BERT and RoBERTa were applied to the COVID-19 false news dataset.	COVID-19 pandemic (2020)	Best accuracy for bert algorithm in this table	Except accuracy precision and recall
20.	[23]	Use of CNN with FT-MELMo embedding system	FakeNewsNet	Accuracy is good	Algorithm have high execution time
21.	[24]	Flair library is used with DNN	ISOT	Time complexity is less	Accuracy is not that much good.
22.	[25]	To predict social networking behavior, a hybrid LSTM-SVM classifier is used.	COVID-19 pandemic (2020)	Accuracy is good	Not good for typical dataset like liar.
23.	[26]	Term Frequency–Inverse Document Frequency multi-level voting ensemble model.	FakeNewsNet	Accuracy is good	Tested on only one dataset so cannot make it trustful.
24.	[27]	several GCN, GAT, and GraphSAGE model are used	COVID-19 pandemic (2020)	Best accuracy on this dataset	Tested on only one dataset so cannot make it trustful.
25.	[28]	Fake BERT is a deep learning approach based on BERT (Bidirectional Encoder Representations from Transformers), which mixes numerous parallel blocks of a single-layer deep (CNN) with varying kernel sizes and filters with the BERT.	FakeNewsNet	Best accuracy for BERT algorithm	Tested on only one dataset so cannot make it trustful.
26.	[29]	A combination of deep neural network with optimal hyper-parameters.	Poltifact FakeNewsNet Buzzfeed	Make problem simple for feature attraction	Not good for typical dataset like liar.
27.	[30]	An enfeeblement machine learning classifier (XGBoost) and a deep neural network model (DeepFakeE)	Buzzfeed FakeNewsNet Poltifact	Time complexity is less	Accuracy is not that much good.
28.	[31]	They've proposed Credit Rank as an	Poltifact Gossipcop	Performance is good for	Time complexity is

		algorithm for assessing publications' social media reputation.	FakeNewsNet LIAR IFND	every dataset	more than expected.
29.	[32]	a basic network design that combines the usage of contextual embedding like word embedding and leverages attention methods with pertinent metadata	LIAR	Time complexity is less	Accuracy is not good on typical dataset.
30.	[33]	Pre-trained encoders are used to create a content graph based on Edge-weighted Graph Attention.	PolitiFact Gossipcop FakeNewsNet LIAR	Accuracy is good on typical data set	Time complexity is more than expected.
31.	[34]	MCNN-TFW, a various convolution neural infrastructure false news detection	ISOT	Accuracy is good	Tested on only one dataset so cannot make it trustful.
32.	[35]	Multi network and CNN is used to increase the accuracy.	PHEME Weibo	Time complexity is less	Accuracy is not good on typical dataset.
33.	[36]	A semantic false news detection system based surround relational characteristics like emotion, entities or facts collected straight from manuscript using NLP.	PolitiFact LIAR	Accuracy is good at typical dataset	This is not good for visual base dataset

Fake news detection is divided into three types: first is textual data second is image related data and third is video related data there are lots of models are used to detect fake news from all of these outcomes like in [12] blockchain and Bi-LSTM is used to achieve the highest accuracy in all of the given research articles for the dataset of PolitiFact, Gossipcop which is a smaller dataset the CNN approach used by [31] is great in all the other research articles, on the twitter election dataset XLM-RoBERTa CNN approach is used in is giving maximum accuracy among all the other models, and on the FakeNewsNet dataset BERT approach is used in [28] is giving maximum accuracy among all the other models, on BUZZFEED dataset ASSO-OSSIW approach is used in [16] is giving maximum accuracy among all the other models, on the Weibo dataset CNN approach is used in [35] is giving consistent accuracy in every epoch among all the other models, on the LIAR dataset NLP approach is used in [36] is giving maximum accuracy among all the other models, on PHEME dataset MN

approach is used in [35] is giving maximum accuracy among all the other models. AS like this some specific model therefore more accuracy in the special dataset however, there is a problem which is not handled like The present FND algorithms suffer from numerous limitations such as limited datasets and high processing costs. A false news classification model job to evaluate whether an item of information is accurate or wrong is the primary case of misinformation identification. However, the binary classifier strategy is ineffective when the data is partly right and partially erroneous. False news detection also may be regarded as a fine-grained multi-classification task by providing many categories to data sets to combat this restriction.

- **Absence of rumors tracking:** The brighter side of the FND is to get track of rumours additionally which are more damaging as opposed to false news, but unhappily, no intervention made a meaningful effort in this area. Although various solutions focused on the user's qualities related to social status, depth of comprehension, half-truths, and a particular sort of false claims, in this research, we applied the MSVM and PCA, and firefly techniques to get this problem out up to certain degree. We also utilised trust, a critical factor, for developing a strong relationship on the social media platform.
- **Small datasets:** The utilisation of correct datasets is significantly useful to the procedure's effectiveness. Most of the other possibilities, as we've seen, selected online blogs, a restricted number of tweets and posts, and a small number of news articles for their trial, evaluation, and findings. For evaluation considerations, only a few approaches utilised the true social network dataset. We utilised 10 datasets for experimental analysis in this study; however, the recommended approach may be applied with any form of dataset for evaluation.
- **Manipulated dataset:** there are a lot of articles that are using manipulated datasets according to their model's comfort to get more accuracy manipulated datasets are not reliable and we cannot trust the synthetic dataset because this dataset is not official and not picked a random ratio of fake and real news so once you can achieve more accuracy on these kinds of model they are not reliable.

CHAPTER 3

FUNDAMENTALS

3.1 Fake News Detection System

"News" indicates to data relating to recent happenings. This may be performed in numerous methods, including oral word, publishing, postal services, transferring, DC (digital communication), as well as the testimony of incident participants and onlookers [2]. Fake news is the deliberate spread of erroneous material through regular news channels or social networking platforms. An erroneous comment spreads fairly rapidly. In our digitally connected world, FNs (false news) has become one of the biggest hurdles. Fake news spreads at extraordinary rates, influencing enormous numbers of individuals every day in the way of enticement and trigrams. As a consequence, finding FN has become a significant topic that has inspired extensive study. Detecting FN on social media is a novel problem every time. It was disseminated on social media to fool people [3]. FNs propagated quicker on Facebook during the 2016 Presidential campaign in the US than actual news. The identification of false news on social media has caught the curiosity of politicians and academics alike. Fake news detection on social networking platforms is crucial. This is due to false news having the power to alter people's thinking, societies, and nations.

3.2 Categories of FNs (Fake News)

The several categories of FNs are discussed below:

- **User-based:** Fake accounts create this form of fake news that is focused on a certain population, which may include people of a given age, religion, language, or political affinity.
- **Visual based:** Graphics, such as manipulated graphics, digitally altered videos, or a fusion of both, is used more for these false media articles as material.
- **Stance-based:** It presents factual statements in such a manner that their context and intention are altered.

Knowledge-based: Such kinds of posts provide comprehensive (so-called) explanations for some unaddressed situations, leading individuals to assume the

information is genuine—for instance, natural therapies for high blood sugar levels in humans [4].

3.3 Limitations of FNs

However, on SM (social media) sites, the range and repercussions of sharing data are significantly influenced and happen at such a quick pace that skewed, erroneous, or fraudulent data has a great possibility to have real-world ramifications for millions of subscribers within seconds. Users' capacity to identify what's phoney or what is genuine while reading and actively engaging in information-overloaded platforms has become a societal concern [5].

3.4 Applications of Fake News

Since the Roman period, deceptive data, twisted details, disinformation details, and fun false facts have been features of interpersonal interaction. Falsified content now has a significantly larger reach owing to the influence and ubiquity of social networking sites. The introduction of clever gadgets and incredibly low-cost broadband has broadened its popularity. Even the most distant communities in India have access to digital internet and phone services. However, these solutions have considerable advantages, and they come at the price of the speedy transmission of false data and correct facts [6]. The relationship with social networking and micro-blogging has expanded within the previous 10 years. The information/text available on such websites in the form of propaganda, articles, comments, views, ideas, conversations, and opinions, among other things, helps to the development of methods and procedures for assuring the authenticity of such assertions. Several research have been done in which machine learning has been applied to spot false news reports automatically.

3.5 SVM (support vector machine) classifier

This classifier relies on the learning of statistical theory. Various scientists created support vector machines [13] for many presentations categorising data records or PR (pattern recognition) (pattern recognition). This model includes theoretical concepts are as follow:

- Structural Risk Minimization is the notion, which indicates the dispersion of risk/ the possibility of learning exceptions. This learning approach regulates the decision-making method to maximise the means square error rate. Kernel function plays a critical part in supporting a VM (vector machine) technique. This function translates

information from input space to feature space to produce non-linear decision-making approaches to information in the learning space [18].

- The central notion is specified as the best margin HYPERPLANE of the SVM support approach. SVM technique is a learning approach to find the plane with the maximum margin. It may split the information into two groups and alleviate the issue of over-fitting.

CHAPTER 4

PROPOSED WORK

4.1 PROBLEM STATEMENT

The existing fake news detection methods suffer from different issues such as limited dataset high computational cost. A fake news classification model task to determine whether a bit of information is correct or incorrect is the fundamental instance of misinformation identification. However, the binary classifier approach is inadequate when the data is partially correct and partially false. False news detection also can be treated as a fine-grained multi-classification challenge by introducing multiple categories to data collections to tackle this limitation. The datasets supplied feature distinct Ground Truth Labels, and the regression construction gets difficult because translating the distinct labels to numerical scores appears to be a complex operation [1]. Therefore, there is a need for a methodology to overcome the existing issues.

4.2 MOTIVATION

Fake News generates erroneous suspense information that may be discovered. This promotes a dishonesty about a country's standing or overstates the price of specific tasks for a government, eroding democracy in particular nations, such as the Arab Spring. Associations such as the "House of Commons and the Crosscheck project" try to address problems such as publisher accountability. But, as they depend exclusively on manual detection by people, their coverage is small. This is not sustainable nor practicable in a world where billions of things are removed or uploaded every second. An approach may be the building of a system that offers a trustworthy automated index score or rating for diverse sources' trustworthiness and media settings [1]

4.3 METHODOLOGY

The existing FND methods suffer from different issues such as limited datasets and high computational costs. A fake news classification model task to determine whether a bit of information is correct or incorrect is the fundamental instance of misinformation identification. However, the binary classifier approach is inadequate when the data is partially correct and partially false. False news detection also can be treated as a fine-grained multi-classification challenge by introducing multiple categories to data collections to tackle this limitation. There is a need for a methodology to overcome the existing issues.

4.3.1 Proposed Methodology

The suggested system leverages a unique database to generate a decision model based on the MSVM classification approach. The suggested model will be used to categorise or detect the news as phoney or authentic, as we can see in Figure 4.1.

4.3.2 Data Gathering & Data Pre-Processing

Generally, it will collect the reviews or comments with different sections like fake or real. Most of the researchers can gather the data from online sites. After the gathering phase, it will upload the input comment data and verify the feasibility of the news dataset when the dataset will explore. Before classifying the comments, reviews, the following steps are applied to datasets:

- Data cleaning
- Stemming

4.3.3 Extract the Features, Select the Features, and Classify:

After the data pre-processing stage, the suggested model will go on to the next phase is, feature extraction (FE) (FE). It will extract the features and dependable optimised feature sets with the aid of the FE technique employing optimised PCA. The retrieved feature set data is subsequently processed using the feature optimization approach. Here, the improved PCA and FA (firefly Optimization) algorithm procedure takes place. This step analyses the dataset delivers the optimum feature sets for classification models and optimises exceptions. The optimization process is solely used to increase the categorised accuracy rate, error rate, etc. This FE model isolates each component from a dataset, it should compute just many eigenvalues and vectors. It is more calculation and practical to extract the feature sets. The labels or groups are the titles of reviews or comments where the feature sets take place. The training set will employed to categorise the

outcome for the testing module. Then, it is preserved within the recent execution directory of the research model. Normally, the best result produced during optimization is estimated to be the fitness solution to the issue search by the proposed model is known as an optimised, feature-based FA using the MSVM model. This suggested model has automatically researched the relevant feature vectors without human intervention. This suggested approach attains the advantages of offering optimum performance. At the conclusion of the suggested model, several performance metrics are examined and compared with the current strategies to find the advancements in the provided method.

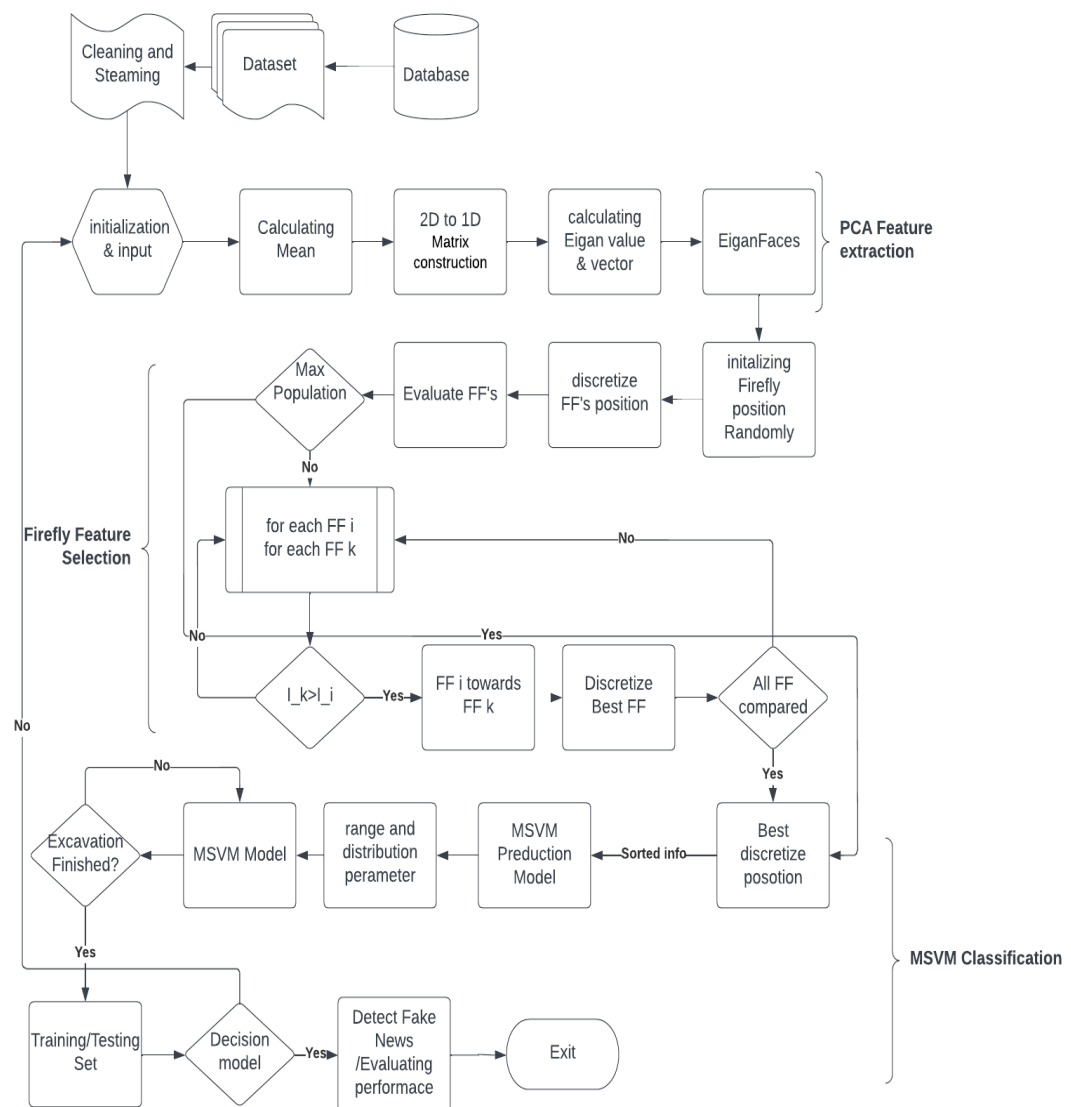


Figure 4.1 Flowchart of proposed model

CHAPTER 5

DATASET EVALUATION

For using the suggested algorithm, we have employed 10 datasets which are called PolitiFact, Gossipcop, Twitter’s Election Integrity, FakeNewsNet, BuzzFeed, Weibo, Liar, PHEME, ISOT, and COVID-19 pandemic (2020). We utilised python 3.0 on an Intel i5 9th-generation CPU with 8 GB memory and NVIDIA GTX 1650 GPU with 4GB memory to acquire the experimental results for our suggested technique. Now, we have reviewed the comprehensive description of datasets that we have compared in the following section.

5.1 DATASET



Figure 5.1 fake Dataset wordcloud



Figure 5.2 Real Dataset wordcloud

In this part, we have done an overview of the details of the datasets Figure 5.1 and Figure 5.2. We have given the dataset word line for false and true news. We utilised the 10 genuine datasets in Table no (2): PolitiFact, Gossipcop, Twitter’s Election Integrity, FakeNewsNet, BuzzFeed, Weibo, Liar, PHEME, ISOT, and COVID-19 pandemic (2020). (2020). Now, we have discussed the structure and configuration of all datasets in Table 5.1.

Table 5.1. Dataset information

Ref.	Dataset	Fake News	Real News	Total News
[46]	Politifact	432	624	1056
[47]	Gossipcop	5323	16817	22140
[48]	Twitter's Election Integrity	6,235,123	7,621,331	13,856,454
[49]	FakeNewsNet	211	211	422
[50]	Buzzfeed	45	56	101
[51]	Weibo	4749	4779	9528
[52]	LIAR	5500	7300	12800
[53]	PHEME	27,992	64,507	92,499
[54]	ISOT	23,481	21,417	44,848
[55]	COVID-19 Pandemic (2020)	3060	3360	6420

POLTIFECT: The news stories in the datasets are all taken from PolitiFact. PolitiFact is a well-known non-profit in the United States that fact-checks political claims and reports. This dataset feature total 1056 posts divided into 432 FN and 624 real news [46].

GOSSICOP: GossipCop is used to gather news stories for the datasets. The website GossipCop (<https://www.gossipcop.com/>) fact-checks celebrity reports and entertainment articles in magazines and newspapers. PolitiFact news items were released between May 2002 and July 2018, whereas GossipCop news pieces were produced between July 2000 and December 2018[47]. This dataset feature total 22140 posts divided into 5323 FN and 16817 real news [47].

Twitter's Election Integrity: The initial misinformation dataset comes from Twitter's Voter Suppression Hub⁴, where three layers of false news were uncovered in late summer 2019. This dataset features a total of 13,856,454 posts on twitter and contains 31 variables that show tweet-related facts about not only the tweet's content and the person [48].

FakeNewsNet: FakeNewsNet consists of only 422 news articles with incomplete classification in real and fake categories [49].

Buzzfeed: BuzzFeed News is a data set with less number of news dataset contain 101

news it is used to check system on shorter dataset[50].

Weibo: WEIBO dataset [51] data was gathered from Xinhua News Agency¹ and Weibo.² The former is a reliable news source, whereas Chinese microblogging platform. The information was gathered between May 2012 and January 2016. There are 9528 posts in all, including 4749 fraudulent posts, 4779 true posts, and 9528 photos relating to postings in the dataset. Each post in the WEIBO dataset includes text and a unique picture. Xinhua News Agency verifies if the posts in the dataset are fraudulent or true news.

LIAR: LIAR comprises different hard-to-classify social media postings and speeches owing to the absence of verification process sources or knowledge bases [52]. This dataset feature total 12,800 posts divided into some of them are full truth some are half-truth are remain are whole FN.

PHEME: The PHEME dataset [53] includes of data based on five breaking stories, encompassing charliehebdo, Robertson, German wings crash, ottawashooting, and sydneyseige. Each news contains a series of postings, containing a substantial quantity of prose and graphics correlating to that same tweets with labels. This dataset feature total 92,499 posts divided into 27,992 FN and 64,507 real news.

ISOT: There are two categories of news in ISOT data: genuine and bogus. There are a total of 44,848 news pieces in the data collection, including 21,417 true news and 23,481 fraudulent news [54].

COVID-19 Pandemic (2020): The COVID-19 false news English dataset [55] was given with the id, tweet, and label ("Fake" and "Real") in the form of TVs during the COVID-19 epidemic (2020). This dataset feature total 6420 posts divided into 3060 FN and 3360 real news.

CHAPTER 6

EXPERIMENTS AND RESULTS

6.1 Analysis and Visualization of the Experimental Result

Now, we deployed the provided method on the supplied ten dataset in which we have compared accuracy with the other models which are utilising the same dataset thus we are comparing. Further, to support the usefulness of the proposed method, we compared the experimental data with the Fake News diagnostic parameters results in terms of accuracy, precision, recall, and F1-score [15].

Thus, initially we are comparing our suggested model with the other research papers on the top 6 datasets which are commonly used and we compare the performance of these dataset on the bases of accuracy for these datasets so the comparison on the bases of datasets are provided below.

6.1.1 Accuracy Test on Politifact dataset: Accuracy comparison of proposed model on this dataset is shown in figure 6.1.

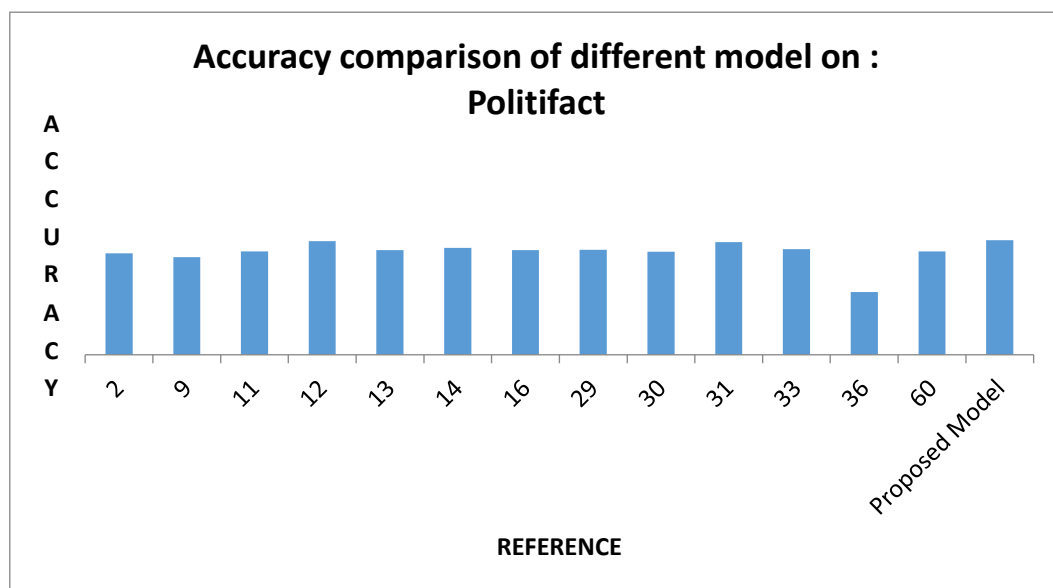


Figure 6.1 Proposed Model Comparison on: Politifact

6.1.2 Accuracy Test on FakeNewsNet dataset: Accuracy comparison of proposed model on this dataset is shown in figure 6.2.

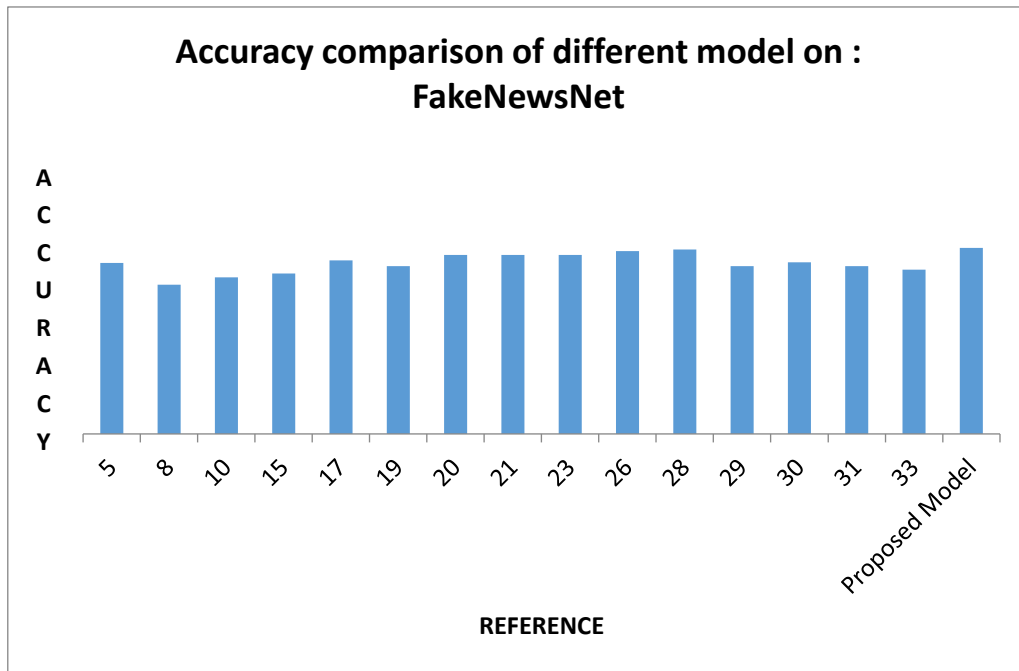


Figure 6.2 Proposed Model Comparison on: FakeNewsNet

6.1.3 Accuracy Test on BUZZFEED dataset: Accuracy comparison of proposed model on this dataset is shown in figure 6.3.

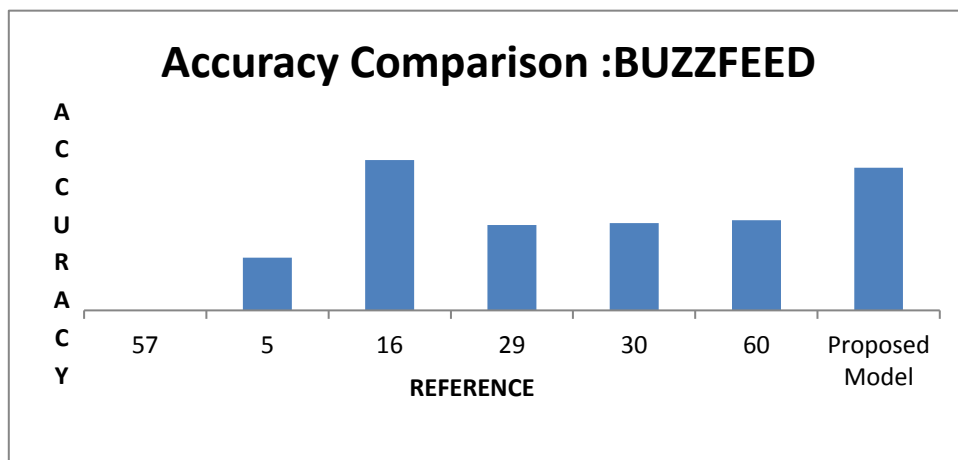


Figure 6.3 Proposed Model Comparison on: BUZZFEED

6.1.4 Accuracy Test on Weibo dataset: Accuracy comparison of proposed model on this dataset is shown in figure 6.4.

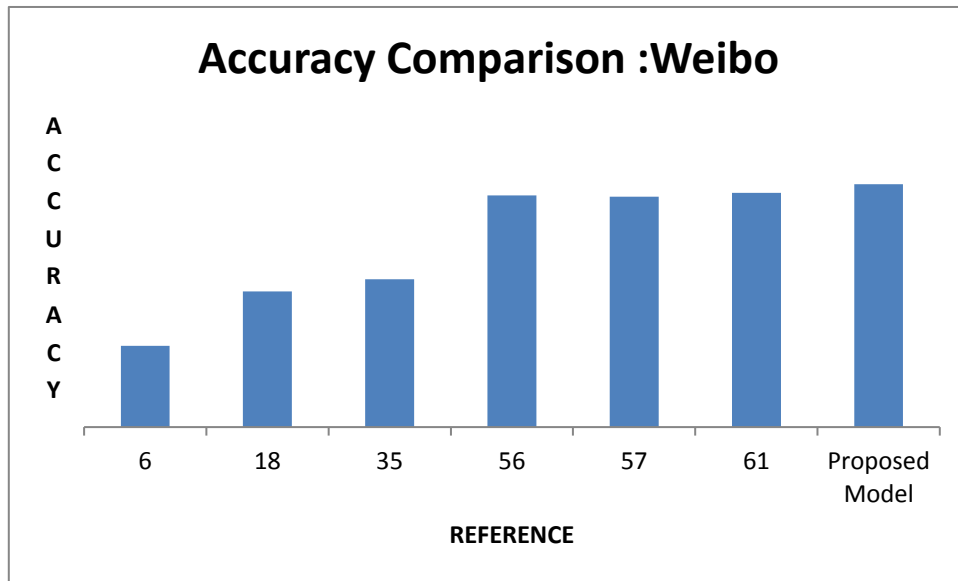


Figure 6.4 Proposed Model Comparison on: Weibo

6.1.5 Accuracy Test on LIAR dataset: Accuracy comparison of proposed model on this dataset is shown in figure 6.5.

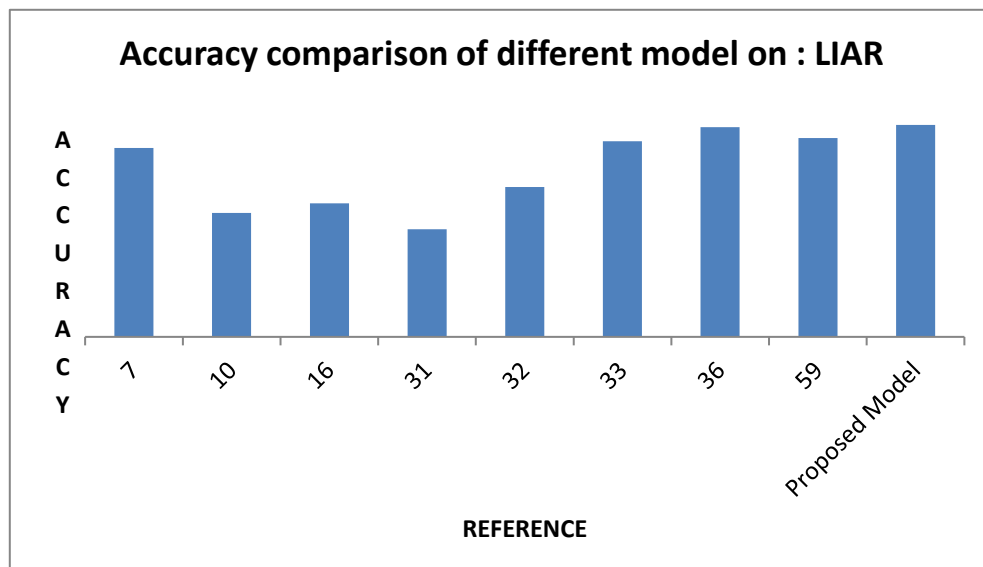


Figure 6.5 Proposed Model Comparison on: LIAR

6.1.6 Accuracy Test on PHEME dataset: Accuracy comparison of proposed model on this dataset is shown in figure 6.6.

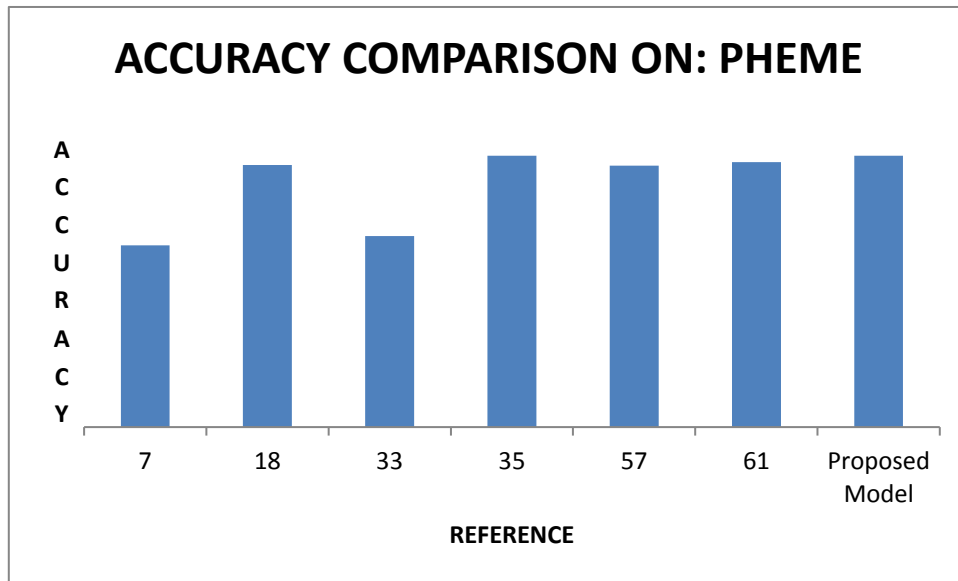


Figure 6.6 Proposed Model Comparison on: PHEME

6.1.7 Performance Test on GossipCop dataset: Performance comparison of proposed model on this dataset is shown in figure 6.7.

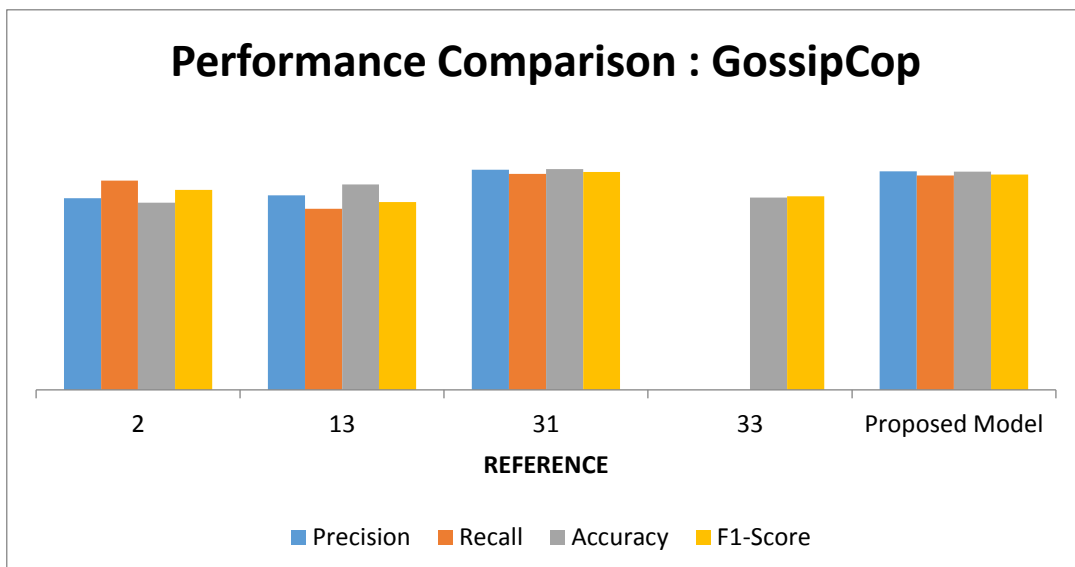


Figure 6.7 Proposed Model Comparison on: GossipCop

6.1.8 Performance Test on Twitter Election Integrity dataset: Performance comparison of proposed model on this dataset is shown in figure 6.8.

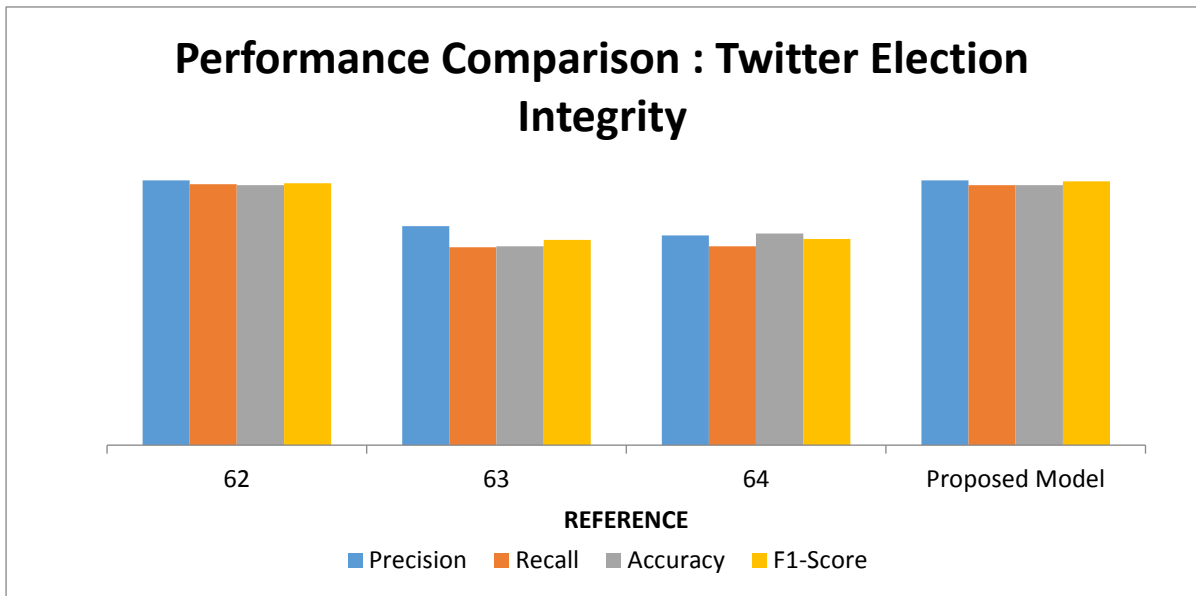


Figure 6.8 Proposed Model Comparison on: Twitter Election Integrity

6.1.9 Performance Test on ISOT dataset: Performance comparison of proposed model on this dataset is shown in figure 6.9.

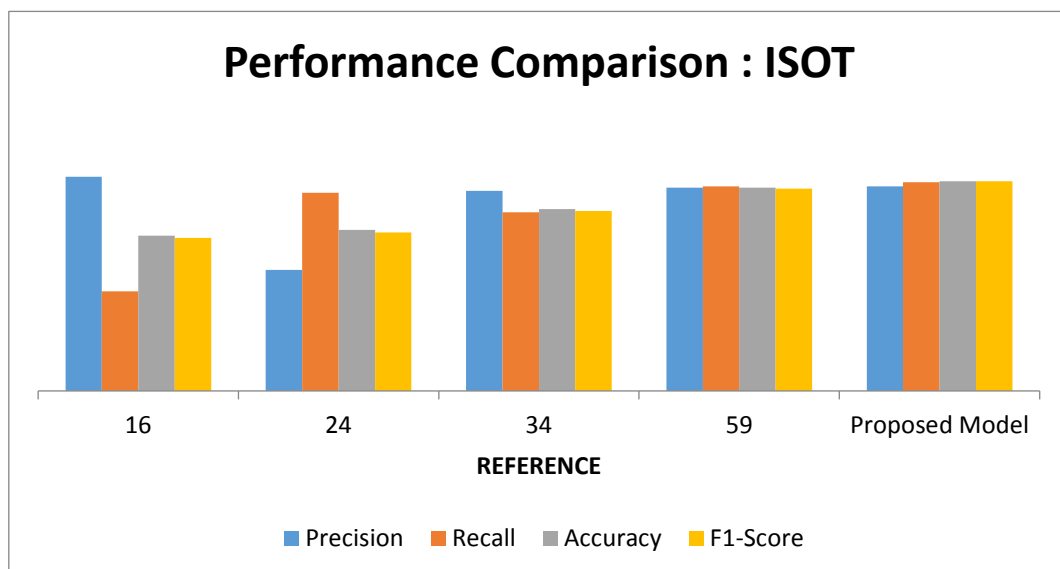


Figure 6.9 Proposed Model Comparison on: ISOT

6.1.10 Performance Test on COVID19 dataset: Performance comparison of proposed model on this dataset is shown in figure 6.10.

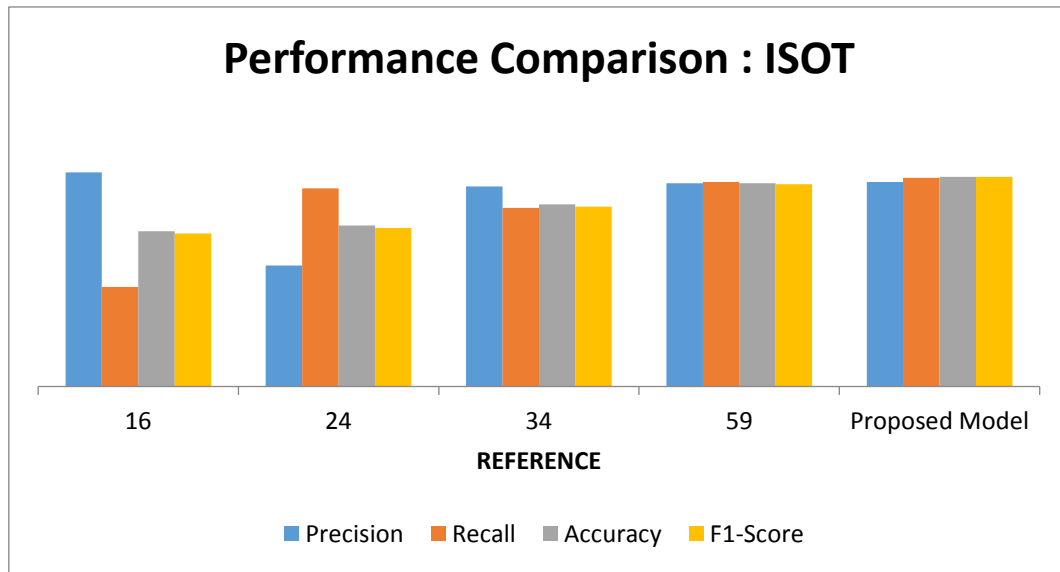


Figure 6.10 Proposed Model Comparison on: COVID19

6.2 Performance Graphs on Fake or Real Dataset

Performance Evaluation on the bases on kaggle dataset (Fake or Real) [65]. We have mention Accuracy, Precision, Recall graph in it.

6.2.1 Histogram based on News Distribution (Fake and Real): Histogram of news distribution is shown in figure 6.11.

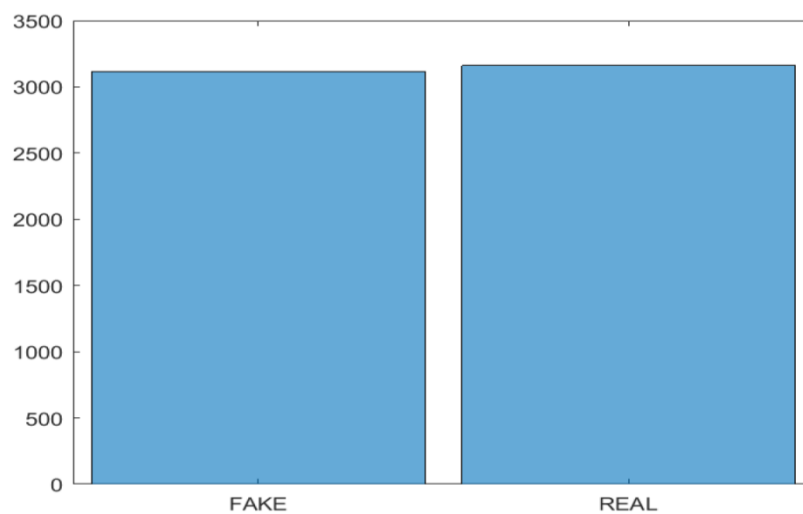


Figure 6.11 Histogram for news distribution

6.2.3 Accuracy Rate: Accuracy vs number of reviews for FakeorReal Dataset is shown in fig. 6.13.

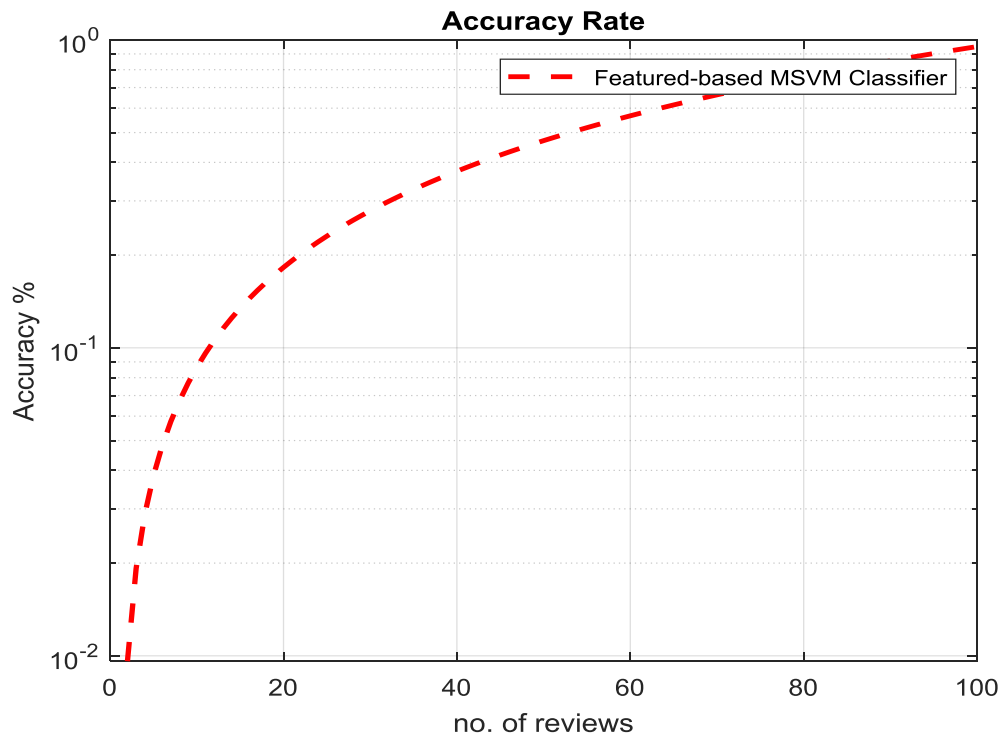


Figure 6.13 Accuracy

6.2.4 Precision: Precision vs number of reviews for FakeorReal Dataset is shown in fig. 6.14.

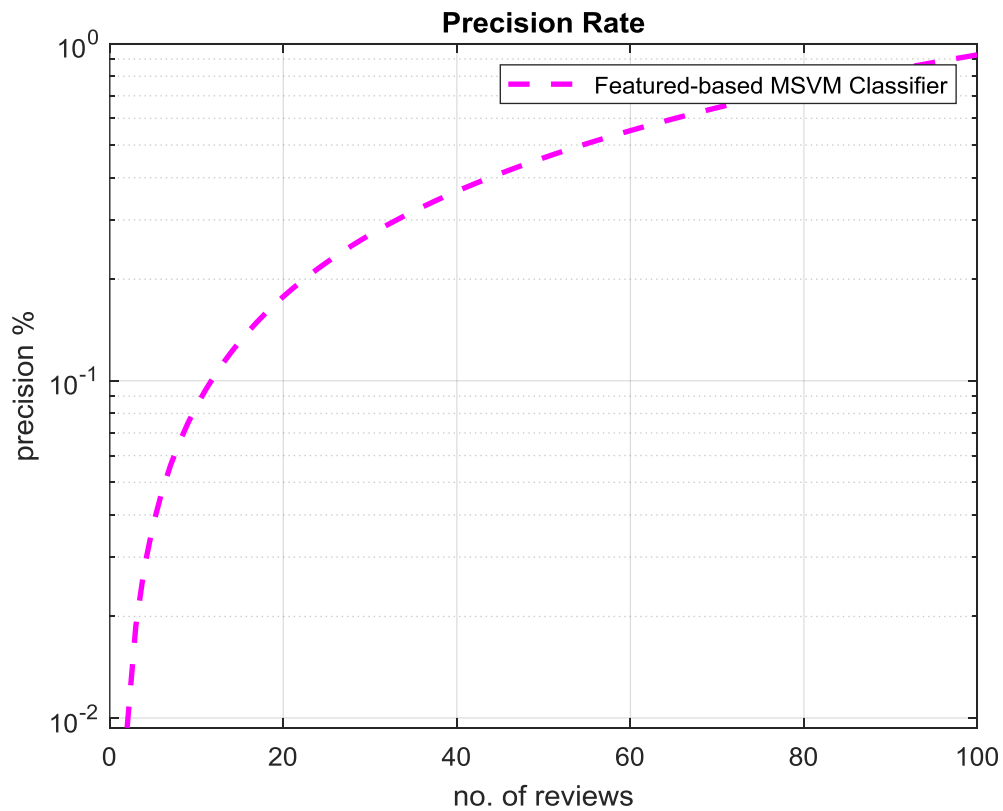


Figure 6.14 Precision

6.2.5 Recall: Recall vs number of reviews for FakeorReal Dataset is shown in fig. 6.15.

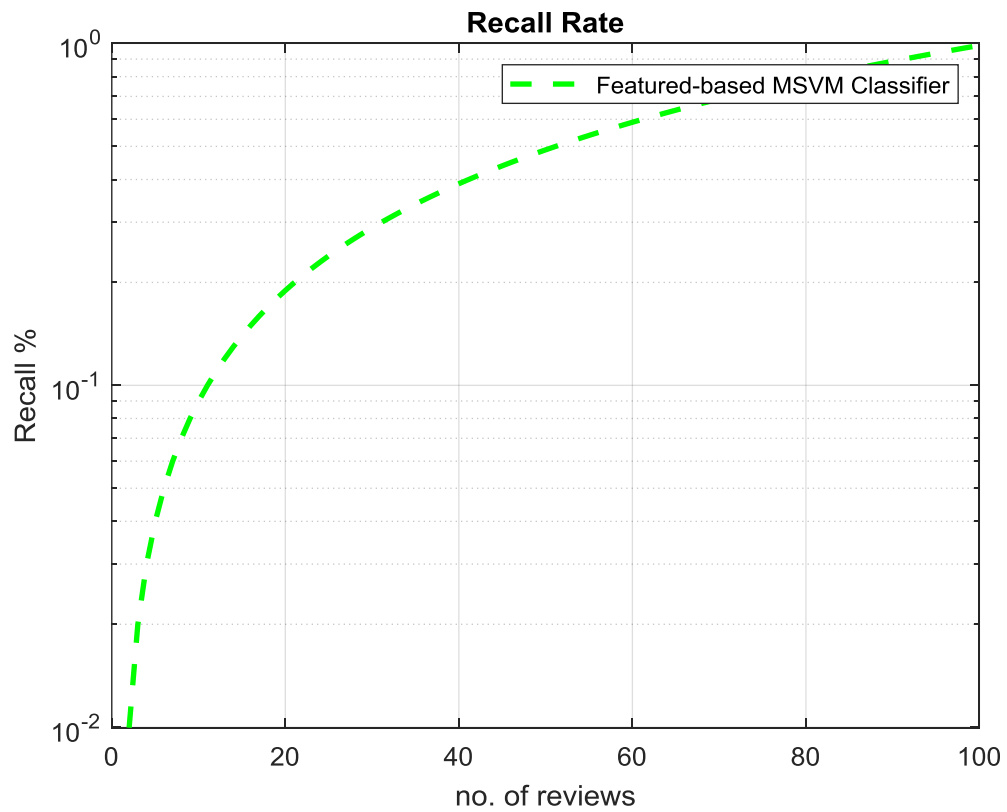


Figure 6.15 Recall

6.2.6 Specificity: Specificity vs number of reviews for FakeorReal Dataset is shown in fig. 6.16.

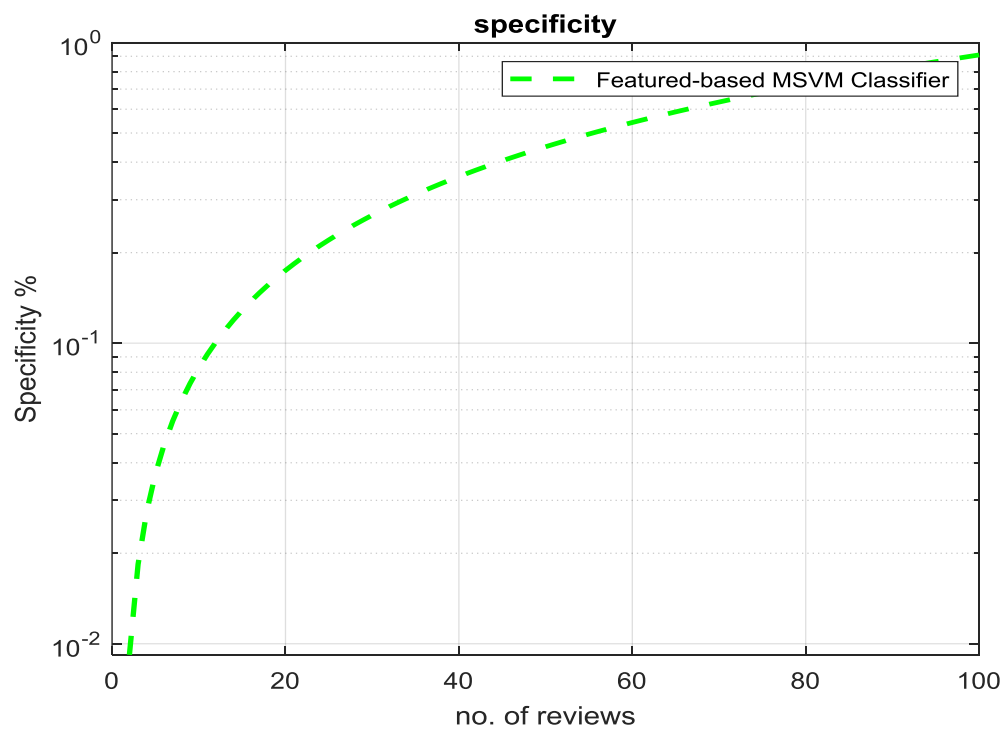


Figure 6.16 Specificity

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

Deceptive material, such as misleading news and fraudulent reviews, has been an increasing concern to internet users in recent years. Fake reviews have affected both consumers and companies. Hired writers are increasingly being utilised to manufacture false reviews to promote sales. Consumers find it difficult to discern between outstanding and terrible things based on evaluations. Furthermore, in 2016, the subject of fake news got significant attention, notably in the aftermath of the 2016 US presidential election. According to current research and surveys, 62 percent of Americans obtain their news from social media. The bulk of popular fake news pieces got more Facebook shares than the most popular mainstream news items. A considerable majority of persons who read fake news items claim they believe them more than news from the mainstream media. Fake news, according to Dewey, had a substantial effect on the 2016 US election and continues to affect people's beliefs and decisions [15].

In this research, we have developed an MSVM based technique for identification of bogus news. We have also utilised a distinct strategy for feature selection and feature extraction for this model for making it more trustworthy than any other model on numerous kinds of datasets. For feature extraction, we have utilised a machine learning approach's optimal model which is an improved two-phase PCA which is also called TP-PCA we used this approach to extract features and send this to the next block of a model whose work is feature selection so for feature selection we have used one more optimised popular algorithm which is used to improve the accuracy and the performance of any model called Firefly model so we have used optimised Firefly model for feature selection and after that multi-class SVM is used for the final work, this approach makes this algorithm more reliable.

In the future, we can use deep learning models tuned with optimised machine learning multi-model to achieve more accuracy.

REFERENCES

- [1] Shearer, E., & Mitchell, A. (2021). News use across social media platforms in 2020.
- [2] Zhou X, Wu J, Zafarani R. Safe: similarity-aware multi-modal FND (2020) Preprint. arXiv. 2020;200304981.
- [3] Zhou, X., Zafarani, R.: Fake news: a survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315 (2018).
- [4] Qian, F., Gong, C., Sharma, K., Liu, Y.: Neural user response generator: FND with collective user intelligence. In: IJCAI, pp. 3834–3840 (2018).
- [5] Verma, Pawan Kumar, Prateek Agrawal, Ivone Amorim, and Radu Prodan. "WELFake: word embedding over linguistic features for FND." *IEEE Transactions on Computational Social Systems* 8, no. 4 (2021): 881-893.
- [6] Li, Dun, Haimei Guo, Zhenfei Wang, and Zhiyun Zheng. "Unsupervised FND based on autoencoder." *IEEE Access* 9 (2021):
- [7] Dong, Xishuang, Uboho Victor, and Lijun Qian. "Two-path deep semisupervised learning for timely FND." *IEEE Transactions on Computational Social Systems* 7, no. 6 (2020): 1386-1398.
- [8] de Souza, Mariana Caravanti, Bruno Magalhães Nogueira, Rafael Geraldeli Ross Ricardo Marcondes Marcacini, Bruce Neves Dos Santos, and Solange Oliveira Rezende. "A network-based positive and unlabeled learning approach for FND." *Machine Learning* (2021): 1-44.
- [9] Shishah, W., 2021. Fake News Detection Using BERT Model with Joint Learning. *Arabian Journal for Science and Engineering*, 46(9), pp.9115-9127.
- [10] Sadeghi, Fariba, Amir Bidgoly, and Hossein Amirkhani. "Fake News Detection on Social Media Using A Natural Language Inference Approach." (2020).
- [11] Reddy, H., Raj, N., Gala, M. and Basava, A., 2020. Text-mining-based FND using ensemble methods. *International Journal of Automation and Computing*, 17(2), pp.210-221.
- [12] ani, P., Jain, V., Shokeen, J. *et al.* Blockchain-based rumor detection approach for COVID-19. *J Ambient Intell Human Comput* (2022). <https://doi.org/10.1007/s12652-022-03900-2>.
- [13] Palani, Balasubramanian, Sivasankar Elango, and Vignesh Viswanathan K. "CB-Fake: A multimodal deep learning framework for automatic FND using capsule neural network and BERT." *Multimedia Tools and Applications* (2021): 1-34.
- [14] Shan, Guohou, Boxin Zhao, James R. Clavin, Haibin Zhang, and Sisi Duan. "Poligraph: Intrusion-tolerant and distributed FND system." *IEEE Transactions on Information Forensics and Security* 17 (2021): 28-41.
- [15] Galende, Borja Arroyo, Gustavo Hernández-Peñaloza, Silvia Uribe, and Federico Álvarez García. "Conspiracy or Not? A Deep Learning Approach to Spot It on Twitter." *IEEE Access* 10 (2022): 38370-38378.
- [16] Ozbay, Feyza Altunbey, and Bilal Alatas. "Adaptive Salp swarm optimization algorithms with inertia weights for novel FND model in online social media." *Multimedia Tools and Applications* 80, no. 26 (2021): 34333-34357.
- [17] Nassif, Ali Bou, Ashraf Elnagar, Omar Elgendy, and Yaman Afadar. "Arabic FND based on deep contextualized embedding models." *Neural Computing and Applications* (2022): 1-14.
- [18] Ying, Long, Hui Yu, Jinguang Wang, Yongze Ji, and Shengsheng Qian. "Multi-

- Level Multi-Modal Cross-Attention Network for Fake News Detection." *IEEE Access* 9 (2021): 132363-132373.
- [19] Mughaid, Ala, Shadi Al-Zu'bi, Ahmed Al Arjan, Rula Al-Amrat, Rathaa Alajmi, Raed Abu Zitar, and Laith Abualigah. "An intelligent cybersecurity system for detecting fake news in social media websites." *Soft Computing* (2022): 1-15.
- [20] Mohapatra, Asutosh, Nithin Thota, and P. Prakasam. "Fake news detection and classification using hybrid BiLSTM and self-attention model." *Multimedia Tools and Applications* 81, no. 13 (2022): 18503-18519.
- [21] Meesad, Phayung. "Thai Fake News Detection Based on Information Retrieval, Natural Language Processing and Machine Learning." *SN Computer Science* 2, no. 6 (2021): 1-17.
- [22] Malla, SreeJagadeesh, and P. J. A. Alphonse. "Fake or real news about COVID-19? Pretrained transformer model to detect potential misleading news." *The European Physical Journal Special Topics* (2022): 1-10.
- [23] Malhotra, Pooja, and Sanjay Kumar Malik. "An Efficient Fake News Identification System Using A-SQUARE CNN Algorithm." *Wireless Personal Communications* (2022): 1-26.
- [24] Kula, Sebastian, Michał Choraś, Rafał Kozik, Paweł Ksieniewicz, and Michał Woźniak. "Sentiment analysis for FND by means of neural networks." In *International Conference on Computational Science*, pp. 653-666. Springer, Cham, 2020.
- [25] Kiruthika NS, Thailambal DG. Dynamic Light Weight Recommendation System for Social Networking Analysis Using a Hybrid LSTM-SVM Classifier Algorithm. *Optical Memory and Neural Networks*. 2022 Mar;31(1):59-75.
- [26] Kaur, Sawinder, Parteek Kumar, and Ponnurangam Kumaraguru. "Automating FND system using multi-level voting model." *Soft Computing* 24, no. 12 (2020): 9049-9069.
- [27] Karnyoto, Andrea Stevens, Chengjie Sun, Bingquan Liu, and Xiaolong Wang. "Augmentation and heterogeneous graph neural network for AAAI2021-COVID-19 FND." *International journal of machine learning and cybernetics* (2022): 1-11.
- [28] Kaliyar RK, Goswami A, Narang P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia tools and applications*. 2021 Mar;80(8):11765-88.
- [29] Kaliyar, Rohit Kumar, Anurag Goswami, and Pratik Narang. "EchoFakeD: improving FND in social media with an efficient deep neural network." *Neural computing and applications* 33, no. 14 (2021): 8597-8613.
- [30] Kaliyar, Rohit Kumar, Anurag Goswami, and Pratik Narang. "DeepFakeE: improving FND using tensor decomposition-based deep neural network." *The Journal of Supercomputing* 77, no. 2 (2021): 1015-1037.
- [31] Jarrahi, A. and Safari, L., 2022. Evaluating the effectiveness of publishers' features in FND on social media. *Multimedia Tools and Applications*, pp.1-27.
- [32] Jain, Vedit, Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, and Yashvardhan Sharma. "AENeT: an attention-enabled neural architecture for FND using contextual features." *Neural Computing and Applications* (2021): 1-12.
- [33] Galli A, Masciari E, Moscato V, Sperlí G. A comprehensive Benchmark for FND. *Journal of Intelligent Information Systems*. 2022 Mar 21:1-25.
- [34] Fayaz, Muhammad, Atif Khan, Muhammad Bilal, and Sanaullah Khan. "Machine Learning Fake News Classification with Optimal Feature Selection." (2021).
- [35] Ying, Long, Hui Yu, Jinguang Wang, Yongze Ji, and Shengsheng Qian. "Fake News Detection via Multi-Modal Topic Memory Network." *IEEE Access* 9 (2021): 132818-132829.
- [36] Braşoveanu, Adrian MP, and Răzvan Andonie. "Integrating machine learning techniques in semantic FND." *Neural Processing Letters* 53, no. 5 (2021): 3055-3072.

- [37] Khanam, Z., Alwasel, B. N., Sirafi, H., & Rashid, M. (2021, March). Fake news detection using machine learning approaches. In IOP Conference Series: Materials Science and Engineering (Vol. 1099, No. 1, p. 012040). IOP Publishing.
- [38] Vishwakarma, D. K., & Jain, C. (2020, June). Recent state-of-the-art of FND: A review. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.
- [39] Choudhary, M., Jha, S., Saxena, D., & Singh, A. K. (2021, May). A Review of Fake News Detection Methods using Machine Learning. In 2021 2nd International Conference for Emerging Technology (INCET) (pp. 1-5). IEEE.
- [40] Fattah, Mohamed Abdel. "The use of MSVM and HMM for sentence alignment." *Journal of Information Processing Systems* 8, no. 2 (2012): 301-314.
- [41] C. Darwin, *The Origin of Species*, P.F. Collier, New York, 1859.
- [42] Tilahun, S. and Ong, H., 2012. Modified Firefly Algorithm. *Journal of Applied Mathematics*, 2012, pp.1-12.
- [43] Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.
- [44] Yang, X. S., & He, X. (2013). Firefly algorithm: recent advances and applications. *International journal of swarm intelligence*, 1(1), 36-50.
- [45] Franc, V., & Hlavác, V. (2002, August). Multi-class support vector machine. In *Object recognition supported by user interaction for service robots* (Vol. 2, pp. 236-239). IEEE.
- [46] Politifact.com. 2022. PolitiFact. [online] Available at: <<https://www.politifact.com/>> [Accessed 22 May 2022].
- [47] Suggest. 2022. *homepage*. [online] Available at: <<https://www.gossipcop.com/>> [Accessed 22 May 2022].
- [48] Transparency.twitter.com. 2022. *Information Operations - Twitter Transparency Center*. [online] Available at: <<https://transparency.twitter.com/en/information-operations.html>> [Accessed 22 May 2022].
- [49] Shu, K., Wang, S. and Liu, H., 2019, January. Beyond news contents: The role of social context for FND. In *Proceedings of the twelfth ACM international conference on web search and data mining* (pp. 312-320).
- [50] Horne, Benjamin, and Sibel 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 *Proceedings of the international AAAI conference on web and social media*, vol. 11, no. 1, pp. 759-766. 2017.
- [51] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," in Proc. 25th ACM Int. Conf. Multimedia, Oct. 2017, pp. 795–816.
- [52] Wang, William Yang. "'liar, liar pants on fire': A new benchmark dataset for FND." *arXiv preprint arXiv:1705.00648* (2017).
- [53] A. Zubiaga, M. Liakata, and R. Procter, "Exploiting context for rumour detection in social media," in Proc. Int. Conf. Social Inform. Cham, Switzerland: Springer, 2017, pp. 109–123.
- [54] Ahmed, Hadeer, Issa Traore, and Sherif Saad. "Detecting opinion spams and fake news using text classification." *Security and Privacy* 1, no. 1 (2018): e9.
- [55] Patwa P, Sharma S, Pykl S, Guptha V, Kumari G, Akhtar MS, Ekbal A, Das A, Chakraborty T. Fighting an infodemic: Covid-19 fake news dataset. In *International Workshop on Combating On line Hostile Posts in Regional Languages during Emergency Situation* 2021 Feb 8 (pp. 21-29). Springer, Cham.
- [56] Liu, Yang, and Yi-Fang Brook Wu. "Fned: a deep network for fake news early detection on social media." *ACM Transactions on Information Systems (TOIS)* 38, no. 3

- (2020): 1-33.
- [57] Li, Ke, Bin Guo, Jiaqi Liu, Jiangtao Wang, Haoyang Ren, Fei Yi, and Zhiwen Yu. "Dynamic Probabilistic Graphical Model for Progressive Fake News Detection on Social Media Platform." *ACM Transactions on Intelligent Systems and Technology (TIST)* (2022).
- [58] Chen, Mu-Yen, Yi-Wei Lai, and Jiunn-Woei Lian. "Using Deep Learning Models to Detect Fake News About COVID-19." *ACM Transactions on Internet Technology* (2022).
- [59] Rajalaxmi, R. R., Narasimha Prasad, L. V., Janakiramaiah, B., Pavankumar, C. S., Neelima, N., & Sathishkumar, V. E. (2022). Optimizing Hyperparameters and Performance Analysis of LSTM Model in Detecting Fake News on Social media. *Transactions on Asian and Low-Resource Language Information Processing*.
- [60] Zhou, Xinyi, Atishay Jain, Vir V. Phoha, and Reza Zafarani. "Fake news early detection: A theory-driven model." *Digital Threats: Research and Practice* 1, no. 2 (2020): 1-25.
- [61] Qian, Shengsheng, Jun Hu, Quan Fang, and Changsheng Xu. "Knowledge-aware multi-modal adaptive graph convolutional networks for FND." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 17, no. 3 (2021): 1-23.
- [62] Zervopoulos, Alexandros, Aikaterini Georgia Alvanou, Konstantinos Bezas, Asterios Papamichail, Manolis Maragoudakis, and Katia Kermanidis. "Deep learning for FND on Twitter regarding the 2019 Hong Kong protests." *Neural Computing and Applications* 34, no. 2 (2022): 969-982.
- [63] Zervopoulos, A., Alvanou, A.G., Bezas, K., Papamichail, A., Maragoudakis, M. and Kermanidis, K., 2020, June. Hong kong protests: Using natural language processing for FND on twitter. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 408-419). Springer, Cham.
- [64] Ni, Shiwen, Jiawen Li, and Hung-Yu Kao. "MVAN: Multi-View Attention Networks for Fake News Detection on Social Media." *IEEE Access* 9 (2021): 106907-106917.
- [65] Kaggle.com. 2022. Fake or Real News. [online] Available at: <<https://www.kaggle.com/datasets/jillanisofttech/fake-or-real-news>> [Accessed 28 May 2022].
- [66] I. Strumberger, E. Tuba, N. Bacanin, M. Zivkovic, M. Beko and M. Tuba, "Designing Convolutional Neural Network Architecture by the Firefly Algorithm," 2019 International Young Engineers Forum (YEF-ECE), 2019, pp. 59-65, doi: 10.1109/YEF-ECE.2019.8740818.

LIST OF PUBLICATIONS

- 1 Ravish, Dr. Rahul Katarya “systematic study of fake news detection systems using machine learning algorithms”. Accepted at the **3rd International Conference on Data Science and Applications (ICDSA 2022)**

Indexed by scoups

Paper id:358



3rd International Conference on Data Science and Applications (ICDSA 2022)

<https://www.icdsa22.scrs.in/>



E-CERTIFICATE OF PARTICIPATION

Ravish Kumar
presented the paper titled
**A systematic study of fake news detection systems using machine learning
algorithms**
authored by
Ravish Kumar and Rahul Katarya
in the International Conference on Data Science and Applications (ICDSA 2022)
held during March 26-27, 2022 in India.



Prof. Sarbani Roy
(General Chair)



Dr. Chandreyee Chowdhury
(General Chair)



Dr. J. C. Bansal
(General Secretary, SCRS)

Organized by
Jadavpur University, Kolkata, India
Technically Sponsored by
Soft Computing Research Society

SCRS/ICDSA2022/358

- 2 Ravish, Rahul Katarya, “Detection of Fake News using Optimized Support Vector Machine”. Accepted for presentation and publication at the **4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) 2022.**

Indexed by Scopus.

Paper Id: 816

Abstract- In today’s ever-growing internet, information and news spreads rapidly like wildfire. A fraction of this news is fake or misleading, mostly for political purposes. This fake news need to be monitored. Since the amount of information generated every day is gigantic, it is only viable to automate the fake news detection procedure. These FN may be publicity against individuals, organizations, society, etc. Generally, a human being is not able to find all these FN. ML (machine learning) has played an essential role in classifying or detecting the data with some disadvantages. We have proposed an optimized machine learning algorithm. We have used optimized multi- support vector machine (MSVM) for getting better performance as compare to some existing models. We are getting 95.8% accuracy, 92.5% Precision and 98.6% recall on a Fake_or_Real dataset available on kaggle.

- 3 Ravish, Rahul Katarya, “**Fake News Detection System Using Featured-Based Optimized MSVM Classification**”. Submitted to **The International Journal of Machine Learning and Cybernetics (IJMLC)**, Electronic ISSN: 1868-808X, Print ISSN: 1868-8071

Indexed by Scopus, EBSCO, INSPEC, Science Citation Index Expanded (SCIE), EBSCO Discovery Service, etc.

Paper Id: “EMID:6a54d3b93c7b2aa1”

Abstract- Fake News creates erroneous suspense information that can be identified. This spreads dishonesty about a country's status or overstates the expense of special functions for a government, destroying democracy in certain countries, such as in the Arab Spring. Associations such as the "House of Commons and the Crosscheck project" attempt to address concerns such as publisher responsibility. But, since they rely entirely on manual detection by humans, their coverage is minimal. This is neither sustainable nor possible in a world where billions of items are withdrawn or posted every second. So in this manuscript, we propose an approach by using Multi-Support Vector Machine (MSVM) to detect fake news with more reliable accuracy, and for implementation purposes, we are using multi-layer principal component analysis PCA for feature selection. PCA reduces the dimension for the dataset comprising a large amount of the related variables and recalls the maximum change in actual data. The essential features will be selected with a firefly optimized algorithm. Several experiments have been conducted to improve the standardized firefly algorithm's competence and adapt it to the nature of the problem. The advantage of the proposed approach is that it streamlines the algorithm. Our proposed model achieves high accuracy of upto 99.64% on the considered datasets.

PAPER NAME

ravish thesis_1.pdf

AUTHOR

RAVISH M.TECH

WORD COUNT

9701 Words

CHARACTER COUNT

52437 Characters

PAGE COUNT

47 Pages

FILE SIZE

1.4MB

SUBMISSION DATE

May 29, 2022 10:26 AM GMT+5:30

REPORT DATE

May 29, 2022 10:29 AM GMT+5:30**● 13% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database.

- 8% Internet database
- 6% Publications database
- Crossref database
- Crossref Posted Content database
- 8% Submitted Works database

● Excluded from Similarity Report

- Bibliographic material
- Small Matches (Less than 10 words)