ANALYSIS AND DETECTION OF FAKE NEWS USING MACHINE LEARNING

Sathish Kumar P
Computer and Communication Engineering
Rajalakshmi institute of technology
Chennai-124.
sathishkumar.p@ritchennai.edu.in

P. Suthanthiradevi
Data Science and Business Systems,
School of Computing, SRM institute of Science and Technology,
Kattankulathur, Chennai-603203
suthantp@srmist.edu.in

Arul Stephen C
Electronics and Communication Engineering
Vels Institute of science Technology and Advanced studies
Chennai- 117
arulstephenc@gmail.com

Ebenezer Abishek. B
Electronics and Communication Engineering
Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College,
Avadi, Chennai-600062.
ebenezerabishek@gmail.com

S. Sivakumar
Electronics and Communication Engineering
BIHER
Chennai-73
sivacud@gmail.com

Mathiyarasu. M
Computer and Communication Engineering
Rajalakshmi institute of technology
Chennai.
mathiyarasu.m.2021.cce@ritchennai.edu.in

2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT) | 979-8-3503-7212-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIIoT58432.2024.10574761

***Abstract*—The Fake news becomes a more critical issue nowadays, it may reduce the trust of the public about the particular information. This project focuses on the advancement of fake news detection model using machine learning. The main goal of this project is to develop, analyze a model that can effectively classify the news articles based on their Subject or contents. Natural preprocessing Algorithm and features such as Stemming, Tokenization, Stop word and TF-IDF Vectorization are used to train the machine learning model. Then the Machine learning Algorithms likeLogistic Regression and Random Forest are used to classifying the dataset which gives 98% accuracy and 99% of accuracy score. For this Project we have taken a dataset of 22413 Real news and 23481 Fake News from the online resources. This model has a potential to reduce the impact of fake news by grant the promotion of accurate and reliable information in a society.**

Keywords—fake news, model, news articles, trust classify information.

I. INTRODUCTION

Fake news refers to intentionally false or misleading information presented as factual news and manipulate public opinion. The consequences of fake news can serve, leading to erosion of public trust in media sources and even influencing political decisions. To reduce the spread fake news, automated detection model is crucial. Machine learning with natural language processing (NLP) techniques, has a promising approach to identify and classify fake news articles effectively. Python, with its libraries and tools, provides a suitable platform for developing such model. The main goal of this project is to develop a machine learning model to classify fake news using various algorithms. NLP techniques will be employed to extract meaningful information from textual data, including lexical analysis, sentiment analysis, and syntactic parsing. The development of this model has various steps, including data collection, preprocessing, feature extraction, model training and testing. Python libraries such as NLTK (Natural language toolkit), scikit-learn, and machine learning algorithms. The model is trained on a large dataset of labelled news articles.

II. LITERATURE SURVEY

The study introduced a fake news detection on social media through Data mining technique. The authors uses a machine learning techniques, including decision trees and Support vector Machine algorithms, to classify news articles. In this Study, they use data extraction technique like user based and content-based features to classify the fake news [1].

Data mining approach for detecting fake news on social media platforms. The members in this study, use machine learning algorithms, such as Logistic regression, decision tree to determine the fake news from the given data. They persistent the importance of features selection and introduce a group of features based on content, and propagation patterns [2].

Rohit et al., shows the modern techniques used in fake news detection on social media platforms. It Deliberate the importance of text-based analysis, modeling, network analysis to recognize misleading data. The authors focus on the role of machine learning algorithms, such as deep learning models and ensemble methods like Bagging, Boosting to improve the detection accuracy [3].

The systematic literature review highlights the application of machine learning for fake news detection. It provides an outline of the datasets, data features and algorithms applied in above mentioned studies. In this review an author analyze the performance of different machine learning techniques (Naïve Bayes, Support Vector machines, and Neural networks), to classify the fake news and real news [4].

Uma et al., explores identification and mitigation techniques for detecting the news which is fake. It deliberate the use of machine learning algorithms, which includes supervised, unsupervised and semi-supervised techniques, in identifying and classifying fake news. The authors also underline the dispute of large-scale data processing and the integration of textual document of information [5].

Rada et al., had implemented the detection of fake news using certain algorithms like conventional neural networks,

979-8-3503-7212-0/24/$31.00 ©2024 IEEE

long short-term memory, K nearest neighbor, Decision Tree, Naive Bayes, Random Forest techniques. The authors have found that using the k nearest neighbor algorithm reduces the accuracy value and also the f1 score of the model. Whereas by using the convolutional neural network algorithm with greater depths of the layer, the accuracy value can be improved and the authors had received a final accuracy of 91.3% using the convolutional neural network algorithm [6]. The authors have referred a recent research paper related to this work in order to develop a more effective model than the previous work. The survey provides the knowledge about the fake news system.

III. METHODOLOGY

The python language being one of the most commonly used languages for machine learning, possesses many libraries and packages in it which is easy to learn and implement. Natural Language Processing is a branch of AI which enables the computer to process and understand the native human languages. This has been implemented with the help of NLTK (Natural language toolkit), which is a collection of certain libraries and programs that are related to the natural language processing and also includes Scikit learn- a ML library for regression, classification and clustering. In addition to the other libraries, few other libraries such as numerical python and Seaborn are included in the program of the study.

First, we have to collect the dataset, then we performing the following processes. (1). Data preprocessing that involved transforming file dataset into the python object in which the dataset is used efficiently to build a model. In this stage, the noise in a dataset is removed, stop words are removed, stemming techniques are applied to the dataset, vectorization techniques are implemented and other process are performed to make a dataset more convenient to use in the further stages. (2) Data Split: The dataset is split into train and test categories from which the model will learn from the training set and prediction or classification are done through the test set. (3) Build a Model: Then, the suitable ML algorithm (that is, Logistic Regression, Random Forest, Decision Tree, Support Vector Machines and so on) is selected to build a model. Scikit learn package contains above mentioned ML algorithms. Then, calculate the matrices such as accuracy scores, f1 score, confusion matrices and so on. If the accuracy score is more (nearer to 100%), then the model is considered to be a best model.

IV. SYSTEM ARCHITECTURE

The proposed system architecture is shown in figure 1. The System Architecture consists of following stages

- Corpus Design
- Preprocessing
- Individual Classifier
- Ensemble Classifier

Corpus Design:

In this stage, the collected dataset(article) is translated into required language if it is necessary. It is done by translation API. Therefore, all text is in the similar language, so the inappropriate consequences due to dissimilar text languages in an article is avoided

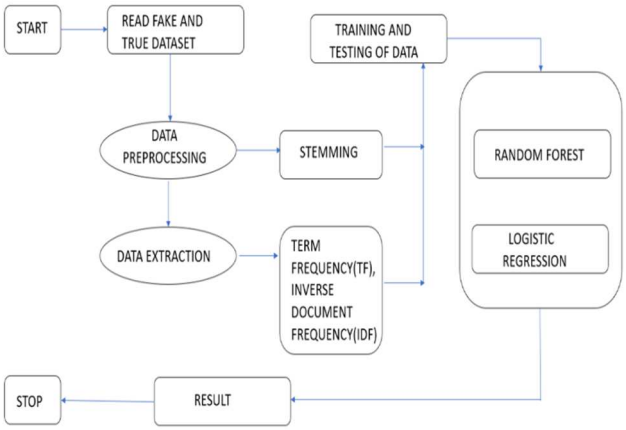

Fig. 1. System Architecture

Preprocessing:

In Preprocessing Stage, the text from the dataset is tokenized (replacing sensitive data), then the special characters are removed from the dataset. Next, the stop words such as 'is', 'through', 'hence' are eliminated and the processed dataset is extracted by vectorization. Here, we use TF-IDF vectorization [7].

Classification Model:

Then, the dataset is split into train and test set. The training set is trained by using the Logistic Regression and Random Forest Algorithms. These algorithms are used to detect the fake news from the dataset. Finally, calculate the accuracy scores, f1 score and confusion matrices. The more accuracy score gives better model.

V. IMPLEMENTATION

A. Data Collection:

We can get a various dataset from different sources like social media, websites, news agency websites and so on. There is a publicly available dataset for fake news detection in the internet like Kaggle, Google's Datasets Search engine, Earth Dataset. In our project, we have used the ISOT dataset which contains a large amount of real news and fake news which were collected from various news sites and also from sites that Politifact.com marked as unreliable. The dataset mainly focuses on the two major categories, that is political news and World news topics [12].

The dataset consists of two csv files; "True.csv" and "Fake.csv". Th

e "True.csv" file consists of 22413 articles and the second file "Fake.csv" contains more than 23487 articles collected from various resources. Table 2 represents the dataset and

their descriptions. Each article in the dataset contains the following information: article title, text, type and the date the article was published [8]. The output of the article is depicted in figure 2 as the head of the dataset. There is large number of articles in the dataset which is mentioned in table1. There is news from various sources are considered to check veracity of the news.

TABLE1: CONTENT OF A DATASET

News	Size	Contents	
Real news	22413	Type	Article Size
		Political News	11272
		World News	10145
Fake news	23481	Type	Article size
		Politics	1570
		Government News	774
		Middle East	645
		News	9040
		Left News	4256
		US news	783

title	text	subject	date
As U.S. budget fight looms, Republicans flip the	WASHINGTON (Reuters) - Th	politicsNews	December 31, 2017
U.S. military to accept transgender recruits on	WASHINGTON (Reuters) - Tr	politicsNews	December 29, 2017
Senior U.S. Republican senator: 'Let Mr. Muell	WASHINGTON (Reuters) - Th	politicsNews	December 31, 2017
FBI Russia probe helped by Australian diploma	WASHINGTON (Reuters) - Tr	politicsNews	December 30, 2017
Trump wants Postal Service to charge 'much m	SEATTLE/WASHINGTON (Re	politicsNews	December 29, 2017
White House, Congress prepare for talks on sp	WEST PALM BEACH, Fla./WA	politicsNews	December 29, 2017
Trump says Russia probe will be fair, but timeli	WEST PALM BEACH, Fla (Reu	politicsNews	December 29, 2017
Factbox: Trump on Twitter (Dec 29) - Approval	The following statementsÂ v	politicsNews	December 29, 2017
Trump on Twitter (Dec 28) - Global Warming	The following statementsÂ v	politicsNews	December 29, 2017
Alabama official to certify Senator-elect Jones	WASHINGTON (Reuters) - Al	politicsNews	December 28, 2017
Jones certified U.S. Senate winner despite Mo	(Reuters) - Alabama officials	politicsNews	December 28, 2017

Fig. 2. Head of the Dataset

B. Data Preprocessing

In Preprocessing, the dataset is transformed or data is changed through the sequence of procedures. Figure 3 shows the visualization of the dataset. The dataset is cleaned from the null values which may reduce the overall accuracy of a model. The dataset is visualized in the form of histogram. The matplot library is imported to work with a data visualization technique [13]. If a dataset size is too large, then it will take more time to execute. Hence, next step is to drop the unnecessary column from a dataset. In our dataset, the date when the article is published does not give any useful information to our model. So, the date column is eliminated or dropped from the dataset. Figure 3.1 and 3.2 shows the sample data for the labels '0' and '1'. The labels which hold '1' for True news and '0' for Fake news [9].

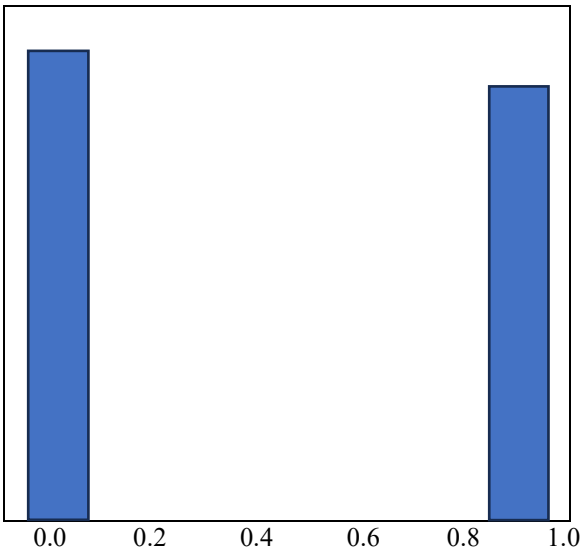

Fig..3. Data Visualization

- **Label 1 – Real News**
- **Label 0- Fake News**

title	text	subject	Label
As U.S. budget fight looms, Republicans flip the	WASHINGTON (Reuters) - Th	politicsNews	1
U.S. military to accept transgender recruits on	WASHINGTON (Reuters) - Tr	politicsNews	1
Senior U.S. Republican senator: 'Let Mr. Muell	WASHINGTON (Reuters) - Th	politicsNews	1
FBI Russia probe helped by Australian diploma	WASHINGTON (Reuters) - Tr	politicsNews	1
Trump wants Postal Service to charge 'much m	SEATTLE/WASHINGTON (Re	politicsNews	1
White House, Congress prepare for talks on sp	WEST PALM BEACH, Fla./WA	politicsNews	1

Fig 3.1. Drop Date Column from Real news Dataset

title	text	subject	Label
Donald Trump Sends Out Embarrassing N	Donald Trump just couldn t w	News	0
Drunk Bragging Trump Staffer Started Rus	House Intelligence Committe	News	0
Sheriff David Clarke Becomes An Internet	On Friday, it was revealed tha	News	0
Trump Is So Obsessed He Even Has Obam	On Christmas day, Donald Tru	News	0
Pope Francis Just Called Out Donald Trum	Pope Francis used his annual	News	0
Racist Alabama Cops Brutalize Black Boy	The number of cases of cops	News	0

Fig. 3.2. Drop Date Column from Fake news Dataset

STOP WORDS:

There are numerous words and sentences are in every category in each article in the dataset. This may increase the complexity of a model and the execution time. Hence, we have to decrease the unnecessary words or useless words such as ‘is’, ‘the’, ‘through’, ’those’. Such words are called Stop words. These Stop word function are in the natural language toolkit (NLTK) packages. So, we import Stop watch from the NLTK. The output of the NLTK is shown in Table 2. Now, the dataset size is reduced when compared to the previous one. The work is not till completed. Another preprocessing technique is used to make the information more precise, that is Stemming [14].

TABLE 2. OUTPUT OF NLTK

Stop words	‘as’, ‘through’, ‘by’, ‘these’, ’are’, ’at’, ‘then’, ‘for’, ‘is’, ‘from’, ‘what’, ‘will’, ‘when’, ‘where’, ‘of’, ‘in’, ‘was’, ‘how’, ‘that’, ‘an’, etc.

STEMMING:

We don’t need any gerund words, degree of comparison word and so on. Here, we have to convert those words into form of its verb. Figure 4 depicts the data before stemming. Example, if there is a word ’Running’, ‘Saying’ then it will convert into its verb form ‘Run’, ‘Say’ and at the same time there is a word ‘Runner’ it also converted into ‘Run’. The results of the stemming is shown in figure 4.1. The Natural Language Toolkit contain the Stemming method. Hence, we import Porter Stemmer function from NLTK library. This technique allows us to make convenient dataset which will be loaded into the ML algorithms for classification. In this project, the stemming is applied to each column in the dataset that is ‘Title’, ‘Text’ [15].

title	text	subject	Label
As U.S. budget fight looms, Republicans flip the	WASHINGTON (Reuters) - Th	politicsNews	1
U.S. military to accept transgender recruits on	WASHINGTON (Reuters) - Tr	politicsNews	1
Senior U.S. Republican senator: 'Let Mr. Muell	WASHINGTON (Reuters) - Th	politicsNews	1
FBI Russia probe helped by Australian diploma	WASHINGTON (Reuters) - Tr	politicsNews	1
Trump wants Postal Service to charge 'much m	SEATTLE/WASHINGTON (Re	politicsNews	1
White House, Congress prepare for talks on sp	WEST PALM BEACH, Fla./WA	politicsNews	1

Fig. 4. Before Stemming

title	text	subject	Label
us budget fight looms, republicans flip their fis	washington reuters head c	politicsNews	1
us military accept transgender recruits monda	washington reuters transge	politicsNews	1
senior u.s republican senator mr mueller do hi	washinton reuters the speci	politicsNews	1
fbi russia probe help australian diplomat tip-o	washington reuters trump c	politicsNews	1
trump want postal service charge much more t	seatle washington reuters pr	politicsNews	1
white house congress prepare talks on spendi	west palm beach fla washing	politicsNews	1

Fig. 4.1. After Stemming

C. Data Extraction:

The above dataset is in the textual form which will not understand by the machine. The Machine understand only the numerical data. Hence, the textual Data are converted into numerical data which the Machine can understand easily by applying various technique like Word Count, Bag of Words and so on. The main aim of this technique is to count the frequency of words in the particular Dataset and convert it into numerical data. In this Project, we use TF-IDF vectorization to convert textual document into numerical data based on the frequency of the wors in the dataset [16].

TF-IDF VECTORIZATION:

TF-IDF stands for Term Frequency and Inverse Document Frequency. In this Vectorization technique, the textual document is converted into numerical data by using the frequency of words. That is, it checks the number of occurrences of particular words in the dataset and then assign the corresponding numerical data which is further fed into the classifier. Now, the Machine understand the numerical data and then Classifier is fit to the training dataset [17]. Figure 5 represents the dataset before vectorization.

$$\text{Time Frequency (TF)} = \frac{\text{Number of times this word occurred}}{\text{Number of words in a document}}$$

The TF is calculated by Dividing the Number of words occurred in a particular article to the Total number of words in the Dataset we used.

$$\text{Inverse Document Frequency (IDF)} = \log \frac{\text{Number of total Document}}{\text{Number of Documents were this word occurred}}$$

The Inverse Document Frequency is calculated by the Taking Logarithm of Dividing the Number of Document with the frequency of words occurred in the number of Documents [21]. The below Figure 6 shows the Textual Dataset which is Before the TF-IDF vectorization.

title	text	subject	Label
us budget fight looms, republicans flip their fis	washington reuters head c	politicsNews	1
us military accept transgender recruits monda	washington reuters transge	politicsNews	1
senior u.s republican senator mr mueller do hi	washinton reuters the speci	politicsNews	1
fbi russia probe help australian diplomat tip-o	washington reuters trump c	politicsNews	1
trump want postal service charge much more t	seatle washington reuters pr	politicsNews	1
white house congress prepare talks on spendi	west palm beach fla washing	politicsNews	1

Fig. 5. Before TF-IDF Vectorization

```
(0, 88056)    0.04290871328649399
(0, 86770)    0.05621765821806731
(0, 86633)    0.02252673077878987
(0, 86440)    0.18195126526861122
(0, 86180)    0.04043486614937066
(0, 85926)    0.08335411656082627
(0, 85690)    0.09196526133959089
(0, 85198)    0.03450175924513437
(0, 85065)    0.031369045131426035
(0, 84841)    0.03293714738513028
(0, 84561)    0.061286351070381405
(0, 84467)    0.05046194770560223
(0, 84007)    0.18004503046358855
(0, 83998)    0.03526683604683736
(0, 83348)    0.03153784460074233
(0, 82185)    0.0990497064631502
(0, 81772)    0.06741867440112577
(0, 81042)    0.07422082640085517
(0, 79543)    0.1131795616828624
(0, 79263)    0.034406269443037804
(0, 78757)    0.048607181186643894
(0, 77960)    0.052772415926680034
(0, 77683)    0.038233924708487736
(0, 77657)    0.04851778582002006
(0, 76141)    0.1423555138848486
```

Fig. 6. After TF-IDF Vectorization

Based on the vector formula which was mentioned above is used to translate the textual document into the machine or computer understandable language (Numerical data). The above figure 6 shows the conversion of textual Dataset into Numerical Dataset by using TF-IDF vectorization. These forms of Datasets are easily understood by the Machine and it is fed into the classifier [19].

D. Data Spliting:

The dataset is split into two parts; Train and Test Set. The Trained dataset is used to train the Machine Learning Model like Logistic Regression, Random Forest, Decision Tree and so on. These algorithms learned a dataset by using a training dataset. Hence, it gains a knowledge about our Dataset. Then, our model is tested with a Test dataset which predict the result that is classify the data into Real and Fake [19]. Generally, maximum portion are split for Training dataset and the Test dataset holds a minimum size. Hence, the model learns more data and accuracy score will be better. In this project, we use 80% of a dataset for training the model and 20% for testing. The Model Selection from Scikit Learn library contains Train and Test Split so we import this library. Then the dataset is fed into the ML Algorithm (Classification) [20].

E. Model Selection:

In this Stage, we choose the best Machine Learning Algorithm for text Classification. It is important to select the most efficient Machine Learning algorithm. The various Machine learning Classification Algorithms are Random Forest, Decision Tree, Logistic Regression, Naïve Bayes and so on. Figure 8 shows the RF classifier model is fit with hyperparameter tuning. We select Logistic Regression and Random Forest Algorithms for this project. Both algorithms give the better accuracy scores (98% and 99%). A Scikit-Learn library contains all Machine Learning Algorithms hence, from this library we import the Logistic Regression and Random Forest [10]. The results of the Logistic Regression and Random Forest is shown in figure 7 and 9. In Logistic Regression, the actual output and predicted output will be almost similar, hence it shows the accuracy level of 0.98 (98%) and f1 score will be 0.98752. On the other hand, the Random Forest Algorithms which uses Ensemble technique like Bagging (Bootstrap Aggregation) to predict the output. Random Forest is a collection of Decision Tree. Each decision tree gives its own output and those output are aggregated by using voting or averaging the output to predict the result [11]. In Random Forest, the accuracy score will be 0.99 (99%) and f1 score will be 0.99147. Both the algorithms give the better accuracy score and hence the model is considered to be a best model.

OUTPUT

```
model=LogisticRegression()
model.fit(x_train,y_train)

model.score(x_test,y_test)

0.9857461024498887
```

Fig. 7. Output of Logistic regression

```
model1=RandomForestClassifier()
model1.fit(x_train,y_train)
```

Fig. 8. Fitting RF classifier to model

```
0.9913140311804008

array([[4669,   27],
       [  51, 4233]])
```

Fig. 9. Output of Random Forest

The following table 3 shows the evaluation metrics of the machine learning algorithm results. The performance of the proposed is compared with the existing models [21]. This work uses same dataset and obtained accuracy of 93.8 using machine learning model.

TABLE 3: MACHINE LEARNING RESULTS

Evaluation Metric	Logistic Regression	Random Forest
Accuracy Score	0.9857 (98%)	0.9913 (99%)
Confusion Matrix	[4637, 59 48, 4236]	[4663, 33 40, 4244]
F1 Score	0.98752	0.99147

Our model outperforms compared to the previous models. This model is trained on data several times randomly to achieve good accuracy. Provides an efficient method of dealing with missing data. The proposed system is compared with the existing work is decipted in table 4.

TABLE 4. COMPARISON OF EXISTING MODELS

Author	Proposed Approach	Accuracy
Sawinder et al. [21]	Classify news article as fake or real using various machine learning classifiers.	93.8
Proposed model	Rumor message detection system is proposed to achieve high accuracy.	99.14

CONCLUSION

From the early days, the Fake news are circulated in the internet in any forms either misunderstanding or intentionally. This Project considered the news articles having the title, text to achieve the better accuracy score of a model. From the above two algorithms, Random Forest gives the accuracy score of 99% and Logistic Regression gives the accuracy score of 98%. Both the machine learning algorithms performed well in which we obtained a best model. This project shows the small web application for detecting the Fake news from the Dataset. However, this project only focuses on the 22413 Real news and 23481 Fake news due to limited knowledge. For this project, we collect information about Political and some World news from the various online resources. Later, we will segregate more information in the wide ranges like Sports news, Celebrity news and so on. With this information we will design a special system which will has the potential to classify the fake news from the articles when user can give any data as an input. At the same time, in future we will work on other Machine Learning algorithms such as Naïve Bayes, Support Vector Machine, Decision Tree for detecting the Fake news. Not all news in the internet or from some other sources are accurate or Real. Hence, it is more important to note that “Be aware of Fake News”.

REFERENCES

[1] Detecting Fake News on Social Media: A Data Mining Perspective by Castillo et al. published in the journal ACM Transaction on Intelligence System and Technology, October (2011).

[2] Kushal Agarwal, Shubhan Nandan, Varun Anil Nair, D. Deva Hema, “Fake News Detection using Machine Learning and Natural Language Processing”, International Journal of Recent Technology and Engineering (2019)

[3] Rohit Kumar Kaliyar developed A Deep Neural Network techniques for Fake news detection: Published in IEEE(Institute of Electrical and Electronic Engineering).(2020)

[4] Raising a model for fake news detection using machine learning in Python Gerardo Rolang Agudelo, Octavia Jose Salcedo and Valendia.(2018)

[5] Uma, Sidarth Saran, Shankar M. Phil;” Fake News Detection using machine learning algorithms” Bharati Vidyapeeth college of engineering, Navi Mumbai, Mumbai University, Indira Gandhi College of Engineering, Navi Mumbai. (2020)

[6] Rada Mihalcea, Carlo Strapparava, The lie Detector: Exploration in the automated recognition of Deceptive language: International Journal of Recent Technology and Engineering (2020)

[7] Research | ISOT research lab (uvic.ca)

[8] M. Ghirwal: “Fake News Detection,” Internal Journal Of Advance Research, Innovation in Technology, volume. 4 (2018)

[9] Understanding the Random Forest algorithm. | by Anirudh Palaparthi | Medium

[10] Understanding Logistic regression. | by Anirudh Palaparthi | Analytics Vidhya | Medium

[11] Anjali Jain, Avinash Shakya, Harsh Khatter, Amit Kumar Gupta; A smart System for Fake News Detection, Conference: International Conference on Issues and Challenges Computing Techniques (ICICT). (2019)

[12] Dey, R. Z. Rafi, S. Hasan Parash, S. K. Arko, and A. Chakarabarthy, “Fake news pattern recognition using linguistic analysis,”; 7th International Conference of Information, Electron (2019).

[13] Conroy, N. Rubin, V. Mesuyra, “Automated deception detection: Method for finding news”, at proceeding of the Association for Information Science and Technology.

[14] Combining Fake News:A Survey on Identification and Mitigation Techniques by Zhang et al; International Journal of Recent Technology (2020)

[15] Fake News Detection Using Machine Learning:A Systematic Literature Review by Silva et al.(2020).

[16] [Aaysh Ranjan, “Fake News Detection using Machine learning”, Department of Computer Science and Technology.

[17] Kudarvalli, H. Fiaidhi: Experiment on Detecting Fake News using Machine learning Algorithm. International Journal of Reliable information and Assurence.(2020)

[18] Ni, B. Guo, Z., Li, J., and Jiang: Improving Generalizability of Fake News Detection method using Propensity Score matching, Social and information Networks.

[19] Kesarwani, A., Chauhan, and Nair: Fake news Detection on Social media using K-Nearest Neighbor Classifier.2020 International Conference on Advance in Computer and Communication Engineering.

[20] Kurasinki. L (2020). Machine Learning explainability in text classification for fake news detection.

[21] Kaur, Sawinder, Parteek Kumar, and Ponnurangam Kumaraguru. "Automating fake news detection system using multi-level voting model." Soft Computing (2019): 1-21