2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON) | 978-1-6654-9602-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/COM-IT-CON54601.2022.9850560

Fake News Detection using Machine Learning

[1]Varun Gupta, [2]Rohan Sahai Mathur, [3]Tushar Bansal, [4]Anjali Goyal

[1,2,3] *Department of Computer and Science and Engineering,*
Amity School of Engineering and Technology
Amity University, Noida, Uttar Pradesh, India
[1]varun.gupta18@student.amity.edu
[2]rohan.mathur2@student.amity.edu
[3]tushar.bansal3@student.amity.edu

[4]*Department of Computer and Science and Engineering*
School of Engineering and Technology
Sharda University, Uttar Pradesh, India
[4]anjaligoyal19@yahoo.in

***Abstract*— With the advent of the World Wide Web and the swift advocacy of online platforms paved the way for news propagation that has never been seen in the past. With the present situation of social media platforms, users are developing and sharing more information when compared to the last five years, some of them are not even related to real life. Classifying the text automatically is a tedious and tough job to do. To give a verdict on the truthfulness of an article, a professional too needs to explore multiple aspects of the domain first. Machine learning algorithms are popularly being used to detect the truthfulness of a piece of text. In present scenario, different performance metrics are used to compare and evaluate the effectiveness of various machine learning algorithms. The study examines various textual properties that can be utilized to differentiate the fake and real news. Natural Language Processing techniques are used for data pre-processing which increases the accuracy of the machine learning models. Further, the extracted and preprocessed properties are used to train various ML classifiers with all possible combinations and the built models are then evaluated using various performance metrics.**

Keywords*— *Natural Language Processing, Random Forest, Fake News Detection, Logistic Regression, k-Nearest Neighbor, Machine Learning, SVC.

I. INTRODUCTION

The exponentially rising number of web users has led to exponential growth of unreliable news digitally. Misleading news also gains a lot of attention. It proliferates faster when compared to the reliable news. The main problem in today's era is social platforms have become the main place for campaigns of misleading information that questions the credibility of the whole news system [1][2][3]. As the social platforms are now becoming one of the enormous news sources most of the people tend to believe these sources. For example, during United States election of 2016 there were many news sources that were circulating unreliable news on social platforms which may have altered the decision of the voter making it biased towards a particular candidate [4]. It is a difficult job for a human to detect the fake news from the real ones so the only way to forecast the news is real or not is by having an in-depth knowledge on the news issue. The social platform has given an opportunity to spread any information which leads to even wide spread of fake information.

The main reason for creating fake news by the fake websites is to affect the public opinion towards something. Fake news is something closer to fake product reviews. In fake product reviews, the reviews influence the mindset of the people which inclines their opinion towards buying a particular product. Similarly in the case of fake news, it influences people towards a particular thing which could benefit the person spreading the fake news. Many reports suggest it is simpler and cheaper for political parties and politicians to buy the services of fake websites to influence people and change the election outcome in their favour [5][6][7]. It is correctly being said in the 21st century that elections are now fought over social networking platforms. Various parties deploy a large amount of money and manpower to spread real and fake information on social networking platforms which basically helps to make their visibility, engagement and influence over voters.

Many scientists believe that the problem of fake news spread can be controlled through the means of artificial intelligence (AI) and machine learning (ML) technologies. Many technical developments in the recent years suggest AI and ML algorithms achieve acceptable performance when used in classification problems similar to fake news detection like text and voice detection [8][9], automated essay grading [10][11], conflicting statement detection [12], healthcare [13], mail phishing detection [14], etc. Machine learning as a broad term basically means supervised learning in which a classifier is fed with the dataset which helps in learning a model and gain experience. This helps the classifier to predict the output for new instance [15][16].

In the present paper, execution of various machine learning algorithms like naïve bayes (NB), logistic regression (LR), Support Vector Classifier (SVC), etc. has been compared on a fake news dataset. The classifiers are judged based on various performance metrics like accuracy score, confusion matrix, classification report etc.

Authorized licensed use limited to: MANIPAL INSTITUTE OF TECHNOLOGY. Downloaded on October 18,2025 at 07:07:37 UTC from IEEE Xplore. Restrictions apply.

The dataset has been taken from the Kaggle website which consist of different sets of reliable and unreliable news. The dataset is then pre-processed, and the textual data is also converted to numerical data using TF-IDF Vectorizer.

In data processing, many attributes of natural language processing (NLP) are also being used. The word processing file was broken into a training and testing dataset. Then, the machine learning algorithms were applied to them. The experimental judgment was done on the models which yielded very good results. In the present paper, the working of various classifiers has also been discussed.

The rest of the paper is organized as follows: Section II presents the review of literature in the broad domain of fake news detection. Section III explains the methodology used in present work. Section IV deals with the model implementation. Section V presents the results of classification. Section VI provides the conclusions drawn from research work and Section VII enlists some interesting future research directions for the readers.

II. LITERATURE SURVEY

Fake news detection has evolved as a hot research topic among various scientists across the globe. Over the past decade, there has been a lot of research on the topic of unreliable news observation using ML. Various authors proposed their work in the field of fake news detection using ML and devised numerous methods for using ML for detecting the fake news. Below are some of the proposed ideas by different authors on the topic of using ML for detecting fake news.

Bali et al. [17] compared the performance of 7 different ML algorithms i.e., Random Forest (RF), Gaussian Naïve Bayes (GNB), Multi-Layer Perceptron (MLP), Gradient Boosting (XGB), K- Nearest Neighbour (KNN), AdaBoost (AB) and Support Vector Classifier (SVC) on 3 different datasets using different features of NLP and validated the results using F1-score and accuracy values. From all the above ML algorithms, the accuracy score of Gradient Boosting (XGB) came out to be the highest.

Shabani et al. [18] proposed the idea of using a hybrid machine-crowd technique to tackle the problem of fake news detection. Their approach is better than machine learning algorithms as it provides better accuracy with low cost and latency as compared to machine learning algorithms. The approach also works in places where the machine learning algorithm fails. The paper mainly focuses on distinguishing Fabricated and Satire content by applying the approach mentioned above.

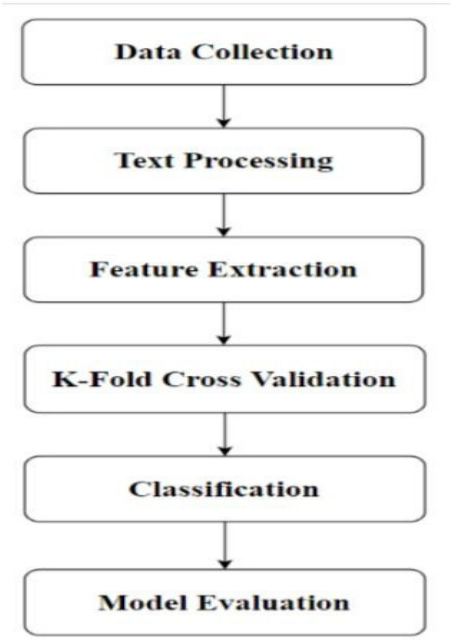
Shu et al. [19] initiated a framework named dEFEND, Explainable Fake News Detection. It has four parts, first is the news piece encoder, second is the user comment encoder, third is a sentence-comment co-attention component and fourth is fake news predictor. In the report the author studies the expandability of fake news detection.

Saikh et al. [20] used Textual Entailment (TE) and ML. The project was broken into two parts, first it uses the features of TE with various ML classifiers and the second part is to integrate features of ML with Deep Learning (DL) network and use it as a load for feed-forward neural networks. The author used supervised learning classifiers like SVC and Multi-Layer Perceptron (MLP) for the experiment. The TE features used in the model are Modal verbs, Longest Common Overlap, Cosine Similarity, Numerals, Hypernyms etc [21].

The present works on the similar lines and presents an empirical study on evaluating the performance of six machine learning classifiers for fake news detection.

III. METHODOLOGY

The model mentioned in current work is used to determine the fake news by using various supervised learning algorithms. In present project, various ML classifiers are used to distinguish between reliable and



unreliable news. The first step is to collect the data that may be used to train and test the classifiers. After the collection of data, it is followed by processing of data, then converting the textual data to numerical data using Term Frequency – Inverse Document Frequency (TF-IDF). The k-fold cross validation is then used to divide the dataset into k number of equal parts in which the k-1 part is used as the training dataset and the residual part as the test dataset. Afterwards the training dataset and test dataset are fed to the machine learning algorithms. The ML classifiers used in the project are LR, SVM, NB, KNN, RF and DT. Then the models are judged using a confusion matrix, accuracy score and the classification report.

In data processing, things like stop words, stemming etc. which are the part of the Natural language processing toolkit (NLTK) are used to clean the dataset for further processing. The objective of the project is to find the finest matched algorithm from the above figure which could be used to determine the reliable and unreliable news efficiently.

Authorized licensed use limited to: MANIPAL INSTITUTE OF TECHNOLOGY. Downloaded on October 18,2025 at 07:07:37 UTC from IEEE Xplore. Restrictions apply.

IV. MODEL IMPLEMENTATION

The model used in the present work has been depicted in fig. 1. The model implementation could be divided into following points:

- The initial step is to import all the necessary libraries in model. Afterwards, it is important to load the dataset to the panda's data frame as it represents the data better and it can handle large data efficiently.
- The second step is to count the no. of missing values in the dataset as it is possible that there may be missing data in the dataset, and it is also important to fill these missing values with empty string.
- The third step is to operate the dataset before feeding it to the ML algorithms. It is important as the quality of data which we are feeding to the model affects the results which makes it a key step.
- The fourth step is the conversion of textual data to numerical data using TF-IDF. This step is important because computers can only understand numerical data.
- The fifth step is to use k-fold cross validation to divide the dataset into 10 folds.
- The sixth step is to feed the dataset to the 6 ML classifiers i.e. Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbor, Random Forest and Decision Tree.
- The last step is evaluating the model using a confusion matrix, classification report and cross_val_score.

A. Dataset

The dataset used in present work has been taken from the Kaggle website which consists of five columns i.e., id, title, author, text, and label. The dataset contains 20,800 news articles out of which 10,387 are reliable/real news and 10,413 are unreliable/fake news. There are also 558 news articles that do not contain a title, 1,957 news articles do not have author names and 39 news articles do not have the text in them.

B. Text Processing

Text processing is done using the following steps:

a.) *Removing all irrelevant characters*: The first step for text processing is to remove all the irrelevant characters like numerical numbers, punctuations, and set of symbols that are not needed at the time of analysing.

b.) *Converting uppercase characters to lowercase:* The next step is to convert all the characters from uppercase to lowercase as it is important because two words like "Shop" and "shop" will be viewed differently by the computer. So, to avoid the duplication all the text is converted to lowercase.

c.) *Tokenization:* Tokenization is the process which is used to divide the raw text into some meaningful data called tokens. It helps in understanding the given data as it divides it into smaller parts. For example, it divides "I am playing" to "I", "am", "playing".

d.) *Stop words:* are those words which do not carry much meaning with them, so they are removed from the raw text. Removing these unnecessary text makes the text more efficient for processing.

e.) *Stemming:* Stemming is the way or process to convert the similar words to its root word. For example, it converts similar words like "waited", "waiting", "waits" to "wait" which is the root word for all the given words.

f.) *Converting tokens to string:* At the last step, the tokenized text is converted back to string.

C. Feature Extraction

In the present model, feature extraction is done using TF- IDF vectorizer. It is a kind of method used to convert textual data to some meaningful numerical data [22-24]. Once the text-based data is changed to numeral data they are used as input for ML classifier for further prediction. TF-IDF Vectorizer has two parts:

a.) *Term Frequency (TF):* It refers to the number of times a word occurs in a particular report or line of the corpus.

b.) *Inverse Document Frequency (IDF):* It mentioned the no. of times a word occurs in different reports or lines all over the corpus.

D. Machine Learning Algorithms

There exist various algorithms for classification in ML. Present work uses following six algorithms:

a.) *Logistic Regression*- is a supervised learning technique which is used to predict dependent variables using a set of independent variables. As it predicts the output of the dependent variable the outcome should come in discrete values like 0 or 1, true or false but it can also give values between 0 and 1.

b.) *Support Vector Machine*- In the SVM algorithm, it makes the boundary line which divides the n-dimensional space into parts so that the new data can be classified easily. The line dividing the space is known to be a hyperplane. The algorithm chooses the most extreme points in the space that helps in creating the hyperplane. These points are also known as support vectors. SVM are of two types- Linear and Nonlinear. In linear, the dataset can be classified using a straight line while in non-linear it cannot be classified using a straight line.

c.) *Naïve Bayes*- The algorithm predicts based on the probability of happening of particular event. The working of the can be explained in three steps, first is to convert the dataset into a frequency table. Then, the second is to create a likelihood table which can help in predicting the probability of the features. The last step is to apply the Bayes theorem to calculate the probability.

d.) *K-Nearest Neighbor*- In KNN the new data is classified based on the existing data which means it first checks the similarities between the new data and existing data; afterwards it categorizes the new data. The advantage of using KNN is that it is easy to use, and it is very

Authorized licensed use limited to: MANIPAL INSTITUTE OF TECHNOLOGY. Downloaded on October 18,2025 at 07:07:37 UTC from IEEE Xplore. Restrictions apply.

effective against large datasets. The algorithm firstly stores the dataset because it cannot learn immediately during the training set and afterwards it performs the necessary actions while at the time of classification.

e.) Random Forest- Random Forest is based on the ideology of ensemble learning which means that it is a combination of various classifiers which is used to solve problems that are very complex. The classifier contains various subsets of the dataset which are used to take the average to enhance the accuracy of the model. The more subsets there are, the better will be the accuracy of the model and also it will help to prevent the overfitting problems.

f.) Decision Tree- Decision Tree is a type of Supervised Learning technique mainly used for solving problems related to classification in machine learning, but it can also be used for solving Regression problems also. Decision Tree has three parts-

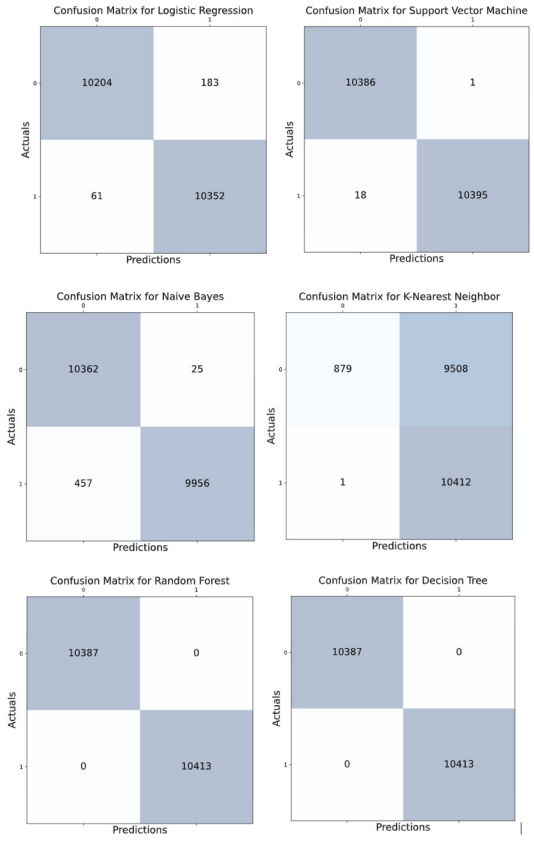


Fig. 2. Confusion Matrix of different classifiers

Decision/Internal node: Decision or the Internal node are the nodes used to make decisions. They act as the attribute of the dataset and can have multiple branches.

Branches: Branches act as the decision rules for the classifier.

Leaf node: Leaf nodes act as the outcome or output of the internal node so they cannot have multiple branches. The structure makes it a tree-like classifier as it starts from the root node, and it extends by adding branches like a tree.

E. Model Evaluation

a.) K-Fold Cross Validation: It is used to evaluate the skills and performance of the models or classifiers. In it the dataset is divided into an equal number of samples. These samples are also known as folds. The dataset is divided into k- numbers of parts in which k-1 parts are for the training dataset and one part for the test set is kept aside. The steps to implement k-fold cross validation are as follows:

i. Divide the given dataset into k numbers of equal folds.

ii. Keep one-fold for the test set.

iii. Use the remaining k-1 folds as the training set.

iv. Train the model using the training dataset and judge the model using the test dataset.

The cross_val_score is used to learn the accuracy of the model using cross validation. Cross validation is better than train/test split because every fold is used for training the classifier and testing it in case of cross validation.

b.) Confusion Matrix: is used to judge the performance of the Model or the classifier. Confusion Matrix has 4 components:

i. True Positive (TP) – It means Prediction is yes, and the actual answer is also yes.

ii. True Negative (TN) – It means Prediction is no, but the actual answer is yes.

iii. False Positive (FP) - It means Prediction is yes, but the actual answer is no.

iv. False Negative (FN) - It means Prediction is no, and the actual answer is also no.

c.) Classification Report: It is used to show the chief classification metrics which are used to evaluate the model or the classifier. The main classification metrics are as follow-

i. Precision: Precision is calculated by the ratio of the true positive to the 47 additions of true and false negatives.

$$\text{Precision} = TP / (TP + FP)$$

ii. Recall: Recall is calculated by the ratio of the true positive to the addition of true positive and false negative.

$$\text{Recall} = TP / (TP + FN)$$

iii. F1-score: F1-score is the total accuracy of the model or the classifier.

$$F1\_score = 2 \times [(P \times R) / (P + R)]$$

where, P refers to Precision

Authorized licensed use limited to: MANIPAL INSTITUTE OF TECHNOLOGY. Downloaded on October 18,2025 at 07:07:37 UTC from IEEE Xplore. Restrictions apply.

And R refers to Recall

iv. Support: Support is the number of data items which have been evaluated. Imbalanced support suggests that the model is weak.

V. Result and Analysis

After learning about different metrics used in the project, the confusion matrix, and the accuracy score of the six classifiers was taken out. The result of the confusion tree is depicted in Figure 2. From figure 2, it can be said that the Random Forest classifier and the Decision Tree classifier gave the best performance while K-Nearest Neighbour failed to meet up the expectations with the worst performance. The Precision, Recall and F1-Scores of all the six classifiers are presented in Table I. In terms of F1-Score, Naïve Bayes algorithm achieved the best performance by scoring 99.72%. The worst F1-Score was achieved by k-Nearest Neighbor with 15.6%.

The accuracy score of all the six classifiers is represented in Figure 3. From the figure 3, we can conclude that Decision Tree algorithm performed the best by achieving 99.36% accuracy.

TABLE I. Performance Comparison of Various Classifiers

Classifiers	Precision	Recall	F1- Score
Logistic Regression	0.9940	0.4964	0.6620
Support Vector Machine	0.9982	0.4997	0.6659
Naive Bayes	0.9976	0.9576	0.9972
K – Nearest Neighbor	0.0846	0.9988	0.1560
Random Forest	1	0.4997	0.6668
Decision Tree	1	0.4997	0.6668

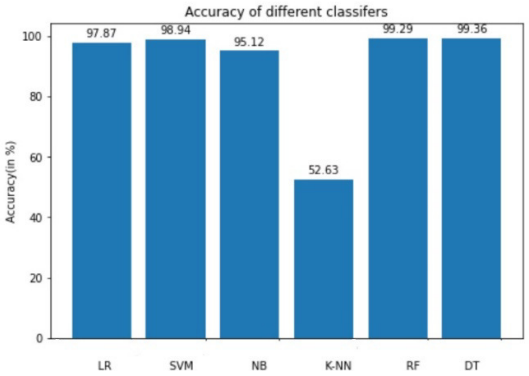


Fig.3. Accuracy Score of different classifiers (in percentage)

VI. Conclusion

In present project, we successfully implemented supervised machine learning algorithms to check if the news is unreliable or not. The paper also talked about different research papers using the theory of deep learning for detection of fake news. Different evaluation methods were used to determine the best classifier for fake news detection. Natural Language Processing (NLP) was used in the project for text processing like stemming and stop words. Many libraries of python were also used for best results like NumPy for numerical calculations and pandas for loading the dataset into a data frame. The ML classifiers also performed well as the Logistic Regression achieved an accuracy which is equal to 97.87%, Support Vector Machine achieved an accuracy of 98.97%, Naïve Bayes achieved an accuracy of 95.12%, K-Nearest Neighbour achieved an accuracy of 52.63%, Random Forest achieved 99.29%, and Decision Tree achieved 99.36%. In comparison, tree-based algorithms performed better in the present case.

VII. Future Work

Though the classifiers performed well, there is still room for improvement. In future, more classifiers can be used for unreliable news detection and feature selection can also be added to the model for better results. Furthermore, more feature extraction methods can also be implemented like bag-of- words, Word2Vec etc. for better performance of the classifiers. More datasets or data can also be taken for evaluation of the model. Text Processing can also be improved using more tools such as the Natural Language Toolkit (NLTK). Different methods can also be adopted like Linguistic feature- based methods, deception modelling-based methods, clustering based methods, content cue- based methods etc. for fake news detection [16].

References

[1] J. C. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Supervised learning for fake news detection," IEEE Intelligent Systems, vol. 34, no. 2, pp. 76–81, 2019.

[2] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," arXiv preprint arXiv:1902.06673, 2019.

[3] R. K. Kaliyar, A. Goswami, and P. Narang, "Fakebert: Fake news detection in social media with a bert-based deep learning approach," Multimedia tools and applications, vol. 80, no. 8, pp. 11 765–11 788, 2021.

[4] S. Gilda, "Notice of violation of ieee publication principles: Evaluating
machine learning algorithms for fake news detection," in 2017 IEEE 15th student conference on research and development (SCOReD). IEEE, 2017 , pp. 110–115.

[5] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in International conference on intelligent, secure, and dependable systems in distributed and cloud environments. Springer, 2017, pp. 127–138.

[6] S. Paul, J. I. Joy, S. Sarker, S. Ahmed, A. K. Das et al., "Fake news detection in social media using blockchain," in 2019 7th International Conference on Smart Computing & Communications (ICSCC). IEEE, 2019, pp. 1–5.

[7] M. K. Elhadad, K. F. Li, and F. Gebali, "Fake news detection on social
media: a systematic survey," in 2019 IEEE Pacific Rim conference on communications, computers and signal processing (PACRIM). IEEE, 2019, pp. 1–8.

[8] M. Granik and V. Mesyura, "Fake news detection using naive bayes classifier," in 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON). IEEE, 2017, pp. 900–903.

[9] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in 2018 IEEE conference on multimedia information processing and retrieval (MIPR). IEEE, 2018, pp. 430–435.

Authorized licensed use limited to: MANIPAL INSTITUTE OF TECHNOLOGY. Downloaded on October 18,2025 at 07:07:37 UTC from IEEE Xplore. Restrictions apply.

[10] D. S. V. Madala, A. Gangal, S. Krishna, A. Goyal, and A. Sureka, "An empirical analysis of machine learning models for automated essay grading," PeerJ Preprints, Tech. Rep., 2018.

[11] S. Sharma and A. Goyal, "Automated essay grading: An empirical analysis of ensemble learning techniques," in Computational Methods and Data Engineering. Springer, 2021, pp. 343–362.

[12] V. Lingam, S. Bhuria, M. Nair, D. Gurpreetsingh, A. Goyal, and A. Sureka, "Deep learning for conflicting statements detection in text," PeerJ Preprints, Tech. Rep., 2018.

[13] P. Khurana, S. Sharma, and A. Goyal, "Heart disease diagnosis: Perfor-
mance evaluation of supervised machine learning and feature selection
techniques," in 2021 8th International Conference on Signal Processing
and Integrated Networks (SPIN). IEEE, 2021, pp. 510–515.

[14] P. Verma, A. Goyal, and Y. Gigras, "Email phishing: Text classification using natural language processing," Computer Science and Information Technologies, vol. 1, no. 1, pp. 1–12, 2020.

[15] N. Garg and K. Sharma, "Text pre-processing of multilingual for sentiment analysis based on social network data." International Journal
of Electrical & Computer Engineering (2088-8708), vol. 12, no. 1, 2022.

[16] N. Garg and K. Sharma,, "Annotated corpus creation for sentiment analysis in code-mixed hindi-english (hinglish) social network data," Indian Journal of Science and Technology, vol. 13, no. 40, pp. 4216–4224, 2020.

[17] A. P. S. Bali, M. Fernandes, S. Choubey, and M. Goel, "Comparative performance of machine learning algorithms for fake news detection," in International conference on advances in computing and data sciences. Springer, 2019, pp. 420–430.

[18] S. Shabani and M. Sokhn, "Hybrid machine-crowd approach for fake news detection," in 2018 IEEE 4th International Conference on Collab-
oration and Internet Computing (CIC). IEEE, 2018, pp. 299–306.

[19] L. Cui, K. Shu, S. Wang, D. Lee, and H. Liu, "defend: A system for explainable fake news detection," in Proceedings of the 28th ACM international conference on information and knowledge management, 2019, pp. 2961–2964.

[20] T. Saikh, A. Anand, A. Ekbal, and P. Bhattacharyya, "A novel approach towards fake news detection: deep learning augmented with textual entailment features," in International Conference on Applications of Natural Language to Information Systems. Springer, 2019, pp. 345–358.

[21] S. B. Parikh and P. K. Atrey, "Media-rich fake news detection: A survey," in 2018 IEEE conference on multimedia information processing and retrieval (MIPR). IEEE, 2018, pp. 436–441.

[22] N. Garg and K. Sharma, " A review of classification and clustering techniques for sentiment analysis based on social network data." International Journal of u- and e-service, Science and Technology, vol. 13, no. 2, 2022.

[23] N. Garg and K. Sharma, "Sentiment Analysis of Events on Social Web." International Journal of Innovative Technology and Exploring Engineering (IJITEE, vol. 9, no. 6, 2020.

[24] S. Jindal, A.K.Tyagi and K. Sharma, "Twitter user's behavior and events affecting their mood." International Journal of Engineering & Advanced Technology, vol. 9, no. 3, 2020.

Authorized licensed use limited to: MANIPAL INSTITUTE OF TECHNOLOGY. Downloaded on October 18,2025 at 07:07:37 UTC from IEEE Xplore. Restrictions apply.