International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

Applying Ensemble Machine Learning Techniques for Fake News Identification

Vipul Devendra Punjabi [1, 2 *], Dr. Rajesh Kumar Shukla [1], Dr. B. V. Kiranmayee [3]
Research Scholar, CSE Department,
[1] Oriental University, Indore, MP, India.
[2] R. C. Patel Institute of Technology, Shirpur, MH, India.
0000-0003-4221-7657,
* vipulchaddha@gmail.com
[1] Oriental University, Indore, MP, India.
0000-0002-6602-0229
[3] VNRVJIET, Hydrabad, India

Abstract- The sharing of information has entered an unprecedented era in human history due to emergence of the Internet With the widespread use of social media sites like Facebook and Twitter. As these platforms enjoy extensive use, users are generating and disseminating a wealth of information, some of which is inaccurate and devoid of factual basis. Detecting false or misleading information within textual content poses a significant challenge. Before arriving at a judgment regarding the accuracy of an article, it is imperative to consider various factors within a specific domain. This paper proposes an Ensemble method for the identification of fraudulent news stories. We leverage different textual features found in both authentic and fake news articles. Our dataset comprises 72,134 news articles, with 35,028 being genuine and 37,106 being false, categorized as binary 0s and 1s. To evaluate our approach, we employed well-known machine learning classifiers including Logistic Regression (LR), Decision Tree, AdaBoost, XGBoost, Random Forest, Extra Trees, SGD, SVM, and Naive Bayes.
To enhance the precision of our findings, we devised a multi-model system for identifying fake news the Ensemble approach and the aforementioned classifiers. Experimental analysis conclusively demonstrates that our suggested ensemble learning technique surpasses the performance of individual learners.

Keywords- Fake news, Bogus news, false information, Ensemble, Social media, Web, Machine learning.

1. INTRODUCTION

Fake news, which involves spreading erroneous material that is disguised as actual news is a pervasive issue in contemporary society, despite its historical presence. Various methods, such as machine learning and linguistic analysis, are used. Knowledge-based approaches, topic-agnostic strategies, and hybrid techniques, have been proposed to identify deceptive news sources [1, 10]. Zhou, X. et al. have outlined four criteria for recognizing false news, including the volume of misinformation, dissemination patterns, writing style, and source reliability [11]. Numerous studies by researchers have aimed to identify fake news, misinformation, disinformation, and detection methodologies [3, 4, 5, 16, 17, 18, 19, 20].

In this study, we introduce an ensemble technique for identifying fake news articles by leveraging distinct textual characteristics that separate authentic from false news. We utilized a publicly available dataset comprising 20,800 news articles, with 10,387 classified as true and 10,413 as false, represented by binary labels (0s and 1s). To evaluate our approach, we employed several widely used Decision trees and logistic regression are two examples of machine learning classifiers. XGBoost (XGB), Extra Trees (ET), Naive Bayes (NB), (SGD) Stochastic Gradient Descent, Random Forest (RF), AdaBoost (AB), Support Vector Machine (SVM). We constructed a multi-model system by integrating such classifiers with the ensembles technique, it is possible to detect fake news with greater accuracy. Our experimental results show that our suggested technique was successful. Impressive results, with precision and accuracy reaching 97% and 96.57%, respectively, along with a 97% recall and 97% F1-measure. Notably, the ensemble method outperformed individual learning methods.
The following is a summary of the study's major contributions:
Introducing an ensemble a method-based approach for identifying fake news, demonstrating superior performance compared to individual learning methods.
Employing a diverse range of textual characteristics from genuine and fake news stories to enhance news categorization accuracy.

2. RELATED WORK

Aside from Social media has surpassed conventional media like newspapers and television become a prevalent source of news. However, it also poses a significant risk as a platform for misinformation and false information. Recent reports indicate that Facebook, the most widely used social media network, boasts 1.2 billion users, making it clear that fake news can easily reach a vast audience through such websites. Consequently, identifying false information on social media has become an intricate challenge, prompting researchers to propose numerous techniques for identifying bogus news across different platforms, including Facebook, Twitter, Reddit, Wikipedia, YouTube, and in print media. Here, we provide an overview of some of these methods [29].

Utilising the logistic regression classifier, Aldwairi, M. et al. achieved a remarkable accuracy of 99.4% when identifying bogus news using the heading and post components of social media [23]. In order to vectorize text words, Chauhan, T., et al. used deep learning and an LSTM neural network model with gloVe word embedding, reaching a remarkable 99.88% rate of accuracy [25]. To recognise bogus news, Shu, K. et al. used data mining techniques [26]. FAKEDETECTOR, a cutting-edge automation fake news detection system, was developed by Zhang, J. et al. in the meantime. It represents news stories, writers, and subjects using a deep diffusive network framework and explicit and latent attributes from textual data [19]. Fakeddit, a multimodal datasets with more than one million samples of various kinds of fake news, was first introduced by Nakamura, K. et al. [20]. In order to anticipate the domain & categorisation of fake news, Kumari, S. et al. constructed a model for classification based on BERT, getting a macro F1 grade of 83.76% [27]. (HBSL) Human-in-the-Loop Dependent Swarm Learning was introduced by Dong, X. et al. as a cutting-edge method for incorporating user feedback into fake news identification without jeopardising personal privacy [28]. By classifying news transmission routes, Liu, Y. et al. established a unique methodology for early detection of fraudulent data on social media [29]. An autonomous fake news detection system tailored to Turkish electronics news material was developed by Mertolu, U. et al. [30]. To find fake news in open datasets, Agudelo, G. E. R. et al. used machine learning and the processing of natural language [31].

In order to better understand how time affects the accuracy of contemporary news truthfulness classifiers, Horne, B. D. et al. undertook a study [32]. In order to identify false news, Zhou, X. et al. created a theory-driven approach that looks at the syntax, vocabulary, semantics, & discourse of news material at various levels [33].

Researchers have made tremendous progress in the area of false news identification by utilising a range of datasets and approaches. Aslam and colleagues used the LIAR dataset to create a deep learning technique for classifying news as true or false [35]. By examining linguistic traits that distinguish between authentic and counterfeit information, Ahmad, I., Yousaf, and several others developed an automated learning ensemble method for automatically identifying news items [36].F. Torabi Asr and collaborators introduced the MisInfoText repository to the community and conducted a topic modeling experiment to address gaps and imbalances in existing datasets, guiding future efforts [37]. Mohammed N. offered a revolutionary three-step method for determining the veracity of news that included posture identification, author credibility analysis, and based on machine learning classification. They obtained remarkable F1-score ratings of 93.15%, 92.65%, & 82.60% for precision, accuracy, recall, and 95.71% using the SVM algorithm [38]. Lai, C. M. and colleagues classified fake news on Twitter using a large labeled corpus and combined approaches for NLP, and machine learning (ML). They contrasted their strategy with state-of-the-art ML and models of neural networks based solely on content [40]. S. Stissi. developed a user-friendly and open user interface for automatically detecting fake news [41].

Groh, Epstein, and colleagues compared human observers with a top deepfake detection model for computer vision, finding similar accuracy despite different error types [42]. A. Thota and team provided a using neural network architecture, test data was successfully used to predict the relationship between the article's headline and body with an archiving accuracy of 94.21%. [47].

Mazzeo, V. proposed a method to identify potentially deceptive content by analyzing textual data from search engines [49]. Unsolved fake news detection challenges, especially without historical data, were addressed by another group of researchers [50].

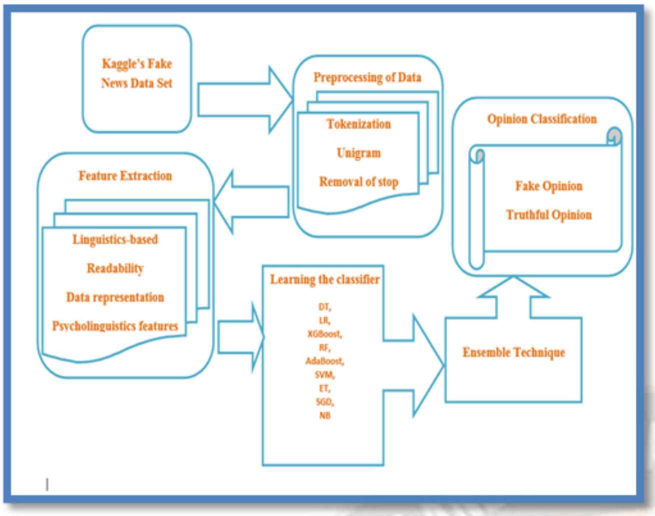
3. MATERIAL AND METHODS

Because of the existence of noise, a lack of data, and unfavourable data formats, developing models for machine learning directly utilising everyday life data is a difficult undertaking. It is essential that you prepare the data by doing data cleaning in order to improve a machine learning (ML) model's accuracy and efficiency [48].

3.1 Proposed Framework

Figure 1 shows our suggested system for identifying bogus news. Pre-processing, extraction of features, and learning are steps in this process. if the information is regarded as real or untrue determines if it is fake news by the reader. The identification of content manipulation in the news material's writing style is achieved through linguistic cue approaches. Prominent techniques within the Linguistic Cue approach include Semantic Analysis, Data Representation, Deep Syntax, and Sentiment Analysis [47].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*



tf(t,d) = count of t in d / number of words in d

Figure 1 illustrates the overarching conceptual approach for detecting counterfeit news.

The initial phase involves gathering data, which includes the compilation and storage of all articles into a consolidated dataset. This dataset then undergoes multiple rounds of processing and analysis to uncover any potential instances of fake news. Prior to utilizing the dataset, preprocessing is essential, eliminating stop words, URLs, unique characters, and and extraneous elements such as advertisements.

The next step involves identifying pertinent features and feeding them into a classifier for training, enabling the determination of the authenticity of the content [47].

Data collection encompasses text mining and scraping. Preprocessing activities consist of tokenization, unigram analysis, and the elimination of stop words, among other tasks. Feature extraction involves various aspects such as linguistic-based features, readability, data representation, semantic analysis, psycholinguistic features, discourse analysis, and deep syntax examination.

3.1.1 Feature Extraction:

In the process of feature extraction, we gather relevant attributes from the text data, which include TF-IDF weights and CountVectorizer representations. Subsequently, these processed textual elements are employed in the feature extraction phase.

3.1.2 TF-IDF Features

Term Frequency is referred to as TF-IDF. A technique for determining the significance of words within a group of documents is called inverse document frequency. Each word receives a score depending on how frequently it appears in a certain document and how important it is overall in the collection of documents. This method is useful for jobs like text mining and information retrieval. Following TF-IDF computation, documents can be transformed into vectors, facilitating various operations like ranking, clustering, and identifying relevant documents.

TF-IDF = Term Frequency (TF) * Inverse Document Frequency (IDF)

A statistical measure called term frequency-inverse documents frequency, or TF-IDF, evaluates a word's significance within a group of texts. The method for calculating this involves multiplying a word's frequency in a file (TF) by the document's inverse frequency (IDF) over the whole collection of documents. We must take the document's length & vocabulary into account while calculating TF. If a word is absent in a document, its TF value will be zero. In cases where a document consists solely of identical words, the TF value can reach a maximum of

The value of the normalised TF lies between [0, 1].

Each document and word has its own TF, and it can be defined as follows:

3.1.3 CountVectorizer Features

Before text can be used for predictive modeling, it undergoes a process called tokenization, which involves analyzing and eliminating specific terms. These words are then encoded as either floating-point values or integers to be used as input for methods based on machine learning. Common names for this encoding technique include feature extraction and vectorization.

3.2 Fake News Detection Using Machine Learning Classifiers

We evaluated the effectiveness of nine various machine learning classifiers as part of our investigation into identifying false information. The techniques used in these classifiers included Decision Tree, (RF) Random Forests, AdaBoost, Logical Regression (LR), XGBoost (XGB), Extra Trees (ET), Stochastic Gradient Descent (SGD), a Support Vector Machine (SVM), and Naive Bayes (NB). We used these classifiers to build a multi-model system for classification with an emphasis on ensemble approaches in order to complete this objective. Python & the scikit-learn library were used in the development of each classifier [51]. Below is a brief description of each classifier.

3.2.1 Detecting Fake News Using Ensemble Methods

3.2.1.1 Ensemble Voting Classifiers

Ensemble voting classifiers are commonly employed in classification tasks due to their ability to amalgamate predictions from multiple learning approaches complete dataset used for training [39]. Every model generates a prediction for a hypothetical data sample, and these predictions are treated as "votes" in favor of a specific class within the framework. The final predictions are then depending on the number of votes received by a given class [32]. Implementing ensemble voting algorithms is generally more straightforward than boosting and bagging methods.

As previously mentioned Bagging methods take data from the source dataset and replace it with samples to produce several random datasets. Each set of data will be utilised to train a different model, and the results from all the models are combined to produce the final product. By modifying weights for incorrectly identified points, each new model in the boosting process builds on the one that came before it, learning from it as it goes.

Contrasted with voting group, integrates diverse models to produce classification results aligned with the majority's collective prediction. In this approach, the conceptual framework is broken down into two or more sub-models, totaling five in this instance. The ensemble technique is utilized to combine predictions from each sub-model. It serves as a meta-classifier that determines the conceptual equivalence or dissimilarity of two machine learning classifiers through a majority vote.

We employ the ensemble technique to predict the final class label, which is the categorical variable typically predicted by classification techniques. The class label "y" is predicted using formula (4), as well as the majority vote from each classification model Cj [33, 37].

3.2.1.2 Decision Tree

Decision tree learning stands as one of the most commonly employed methods for classification tasks. It demonstrates remarkable effectiveness and delivers equivalent to other learning strategies in terms of classification accuracy. The learned categorization process is structuredly represented by the decision tree model. A straightforward methodology is used in this procedure, which evaluates each realistic data splitting test and chooses the one that produces the most useful data [52, 53].

3.2.1.3 Bagging Ensemble Classifiers

A technique known as bootstrap aggregating, or commonly referred to as bagging, is employed early in the realm of ensemble methods to reduce the variance (overfitting) in datasets used for training. One prominent model within this category is the random forest. In the context of bagging, this model is utilized as a classifier. The bagging model selects the challenge based on majority votes predicted by multiple trees, with each tree operating on anonymized data to reduce overall variance. This selection is done through replacement and random sampling from the complete dataset. In regression scenarios, an averaging model is utilized to combine results from multiple estimations.

3.2.1.4 Boosting Ensemble classification algorithms

Boosting is yet another extensively used ensemble technique for enhancing the performance of weak models, operates by making predictions based on the majority votes of a forest of randomized trees specially trained for this purpose. This strategy incrementally improves the classification of data points that are often misclassified by initially assigning equal weights to all data points. Subsequently, these weights are adjusted in successive rounds, increasing for correctly categorized data points and decreasing for those that were misclassified [35]. Each successive tree in a round gains knowledge by accurately classifying previously misclassified data points, reducing errors from prior rounds, and enhancing overall accuracy. Noting, however, that excessive adherence to the training data can be a concern, potentially leading to inaccurate predictions for unseen cases. We used the AdaBoost [38] & XGBoost [37] algorithms for classification.

3.2.1.5 Logistic Regression (LR)

Given the need to classify We use a logistic regression (LR) is framework to analyse text using a huge feature set with either true or false, or true content/fake article outputs. For categorising issues into binary or many classes, this paradigm provides a simple equation [27]. We conducted hyperparameter tuning to optimize model performance for each specific dataset, experimenting with various parameters to achieve the highest accuracy levels.

3.2.1.6 Stochastic Gradient Descent

Especially when dealing with convex loss functions, stochastic descent of gradients (SGD) provides a dependable and straightforward technique for modifying linear classifiers and regressors. These loss functions are widely found in linear models such as (SVM) and Logistic Regression. Although SGD has long been a recognised technique within the field of machine learning ML, its acceptance has grown dramatically in recent years, especially in the area of large-scale learning. With large and sparse datasets being the norm in tasks like text classification & natural language processing, this increase in interest is especially noticeable. The SGD classifier can handle a range of classifications loss functions as well as penalties and contains a key method of learning through SGD. The Scikit-learn framework was developed to make SGD-based categorization easier to apply.

3.2.1.7 Support Vector Machine (SVM)

The SVM, that provides a number of kernel functions, is another approach for solving binary classification issues [28]. The primary goal of an SVM model is to determine a hyperplane or decision-making border for categorising data points using a feature set [29]. The amount of features determines how dimensional the hyperplane is. The goal is to locate the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

hyperplane in a space with N dimensions that maximises the difference between the datasets of the two classes.

3.2.1.8 Voting Ensemble Classifiers

Voting ensemble classifiers are often chosen for classification tasks due to their capability to blend two or more distinct learning approaches that have been based on training data from Complex 5 [39]. With this method, each model makes predictions for a set of data points, and these predictions are treated as individual "votes" in favor of the class the model predicts. After all models have made their predictions, the final classification is determined by a majority vote in favor of a particular class [32]. Unlike bagging and boosting algorithms, voting ensemble is known for its relative simplicity in implementation. Bagging algorithms generate multiple random datasets by selecting and updating the data within them. The final forecast is a compilation of the findings from all models, which are then applied to each dataset to train a model. Contrarily, boosting includes successively training a variety of models, with every model learns from the errors of the previous one by assigning higher weights to incorrectly classified data, resulting in a more accurate general model for classification.
In contrast to both bagging and boosting, a voting ensemble combines multiple distinct models to produce classification results that are collectively determined through a majority vote.

3.3 Performance Metrics

We used a range of evaluation criteria to rate the algorithms' performance, the majority of which were generated using the confusion matrix. Four variables are used to assess the performance of the model for classification on a test sample: the number of false positives, real positives, real negatives, and false negatives. Which are organized in a table known as the confusion matrix.

3.3.1 System Evaluation and Experimental Results

We used a publicly accessible dataset of 72,134 news stories, 35,028 of which were classified as authentic (binary 0), and 37,106 of which were classified as fake (binary 1). Eighty percent of the data were used for training and twenty percent for testing.

Evaluation Metrics:
Although a high level of accuracy normally implies that the model is effective in the case of a system for classification like ours, classifying an item mistakenly as being accurate when it is fake (false positive) might have serious repercussions. Similar to this, labelling an article containing accurate data as fake while it is actually true might damage trust. To account for incorrectly classified observations, we used three additional metrics: precision, recall, and F1-score.

Accuracy = $(T_vP+T_vN)/(T_vP+FP+FN+T_vN)$
Precision = $T_vP/(T_vP+FP)$
Recall = $T_vP/(T_vP+FN)$

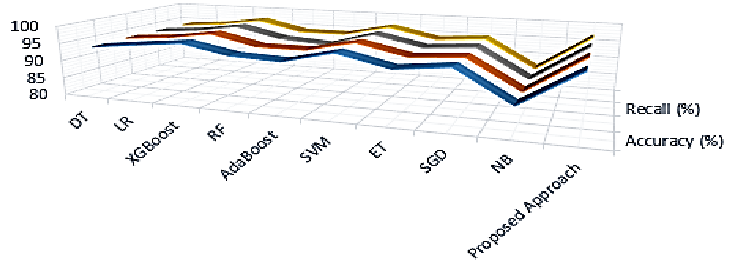
$$\boxed{\textbf{F1 Score} = 2*(\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})}$$

3.3.2 Evaluation of Classifier Performance Using the Fake News Dataset:

The precision, recall, F1-score, and accuracy performance metrics of various classifiers used on the the Kaggle database Fake News dataset are shown in Table 1 and Fig. 3. The accuracy, precision, and recall of the decision-tree classifier were 93.74%, 94%, and 94%, respectively. Similar results were obtained with the XGBoost classifier, which had an accuracy of 96.6% and 97% precision, recall, and F1-score. A 93.74% accuracy rate, 94% preciseness, 94.41% recall, and 94% F1-score were also displayed by the Extra Trees classifier. Along with 94% recall, precision, and F1-score, the random-forest (RF) classifier achieved an accuracy score of 94.06 percent. The AdaBoost classifier produced results with an F1-score of 94%, accuracy of 93.54%, precision of 94%, and recall of 93%. For the logistic regression method (LR), a F1 score of 95% was followed by accuracy, precision, and recall values of 95, 25, and 95, respectively. The SGD classifier achieved a 95% F1-score, 95% recall, 95% precision, and 95.33% accuracy. Additionally, the precision, recall, & F1-score of the support vector machines (SVM) classifier all reached 97%, resulting in a remarkable accuracy of 96.87%. The Naive Bayes classifier, on the other hand, showed statistics of 87.14% accuracy, 87% preciseness, 87% recall, and an 87% F1-score. The accuracy rate, precision, recall, and F1-score of our suggested multi-model classifier, which uses ensemble approaches, are 96.57%, 97%, 97%, and 97%, respectively. In comparison to the aforementioned classifiers, our technique showed greater accuracy and stability.

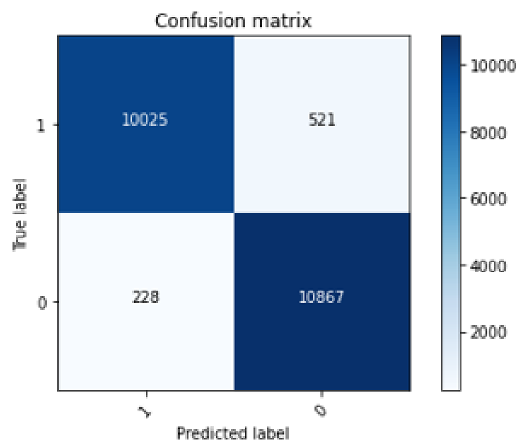
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

Figure 3: Analysis of Classifier Performance on the Fake News Dataset

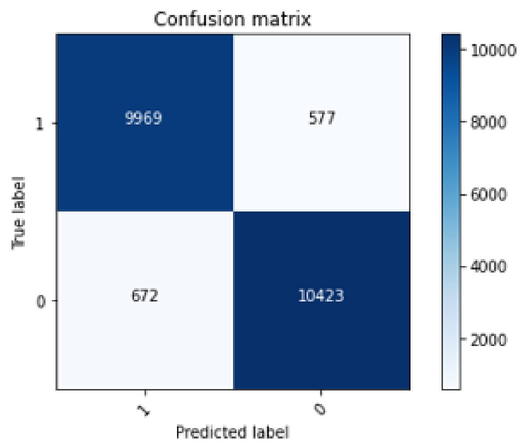


	DT	LR	XGBoost	RF	AdaBoost	SVM	ET	SGD	NB	Proposed Approach
Accuracy (%)	93.74	95.25	96.6	94.06	93.54	96.87	93.74	95.33	87.14	96.57
Precision (%)	94	95	97	94	94	97	94	95	87	97
Recall (%)	94	95	97	94	93	97	94	95	87	97
F1-score (%)	94	95	97	94	94	97	94	95	87	97

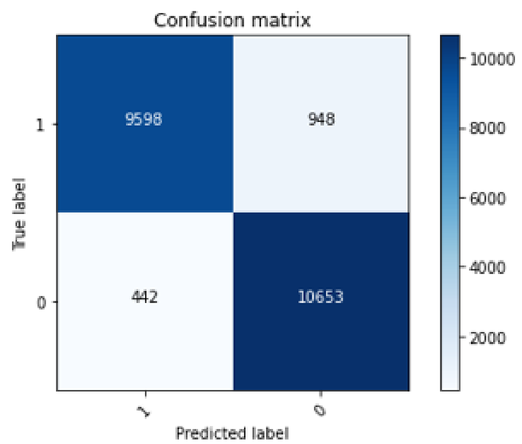
Table 1: Classifier Performance Evaluation Using the Fake News Dataset



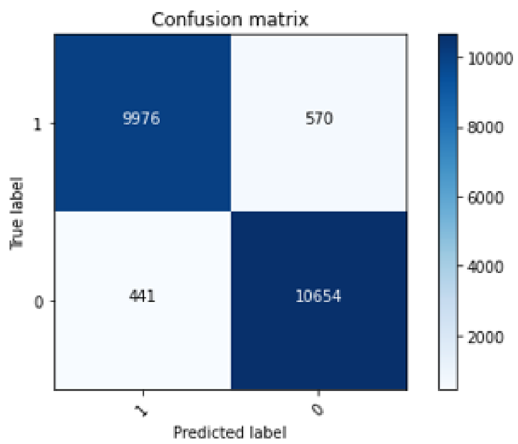
(i)



(ii)

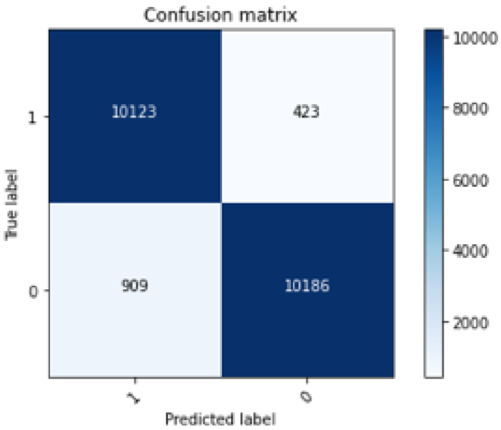


(iii)

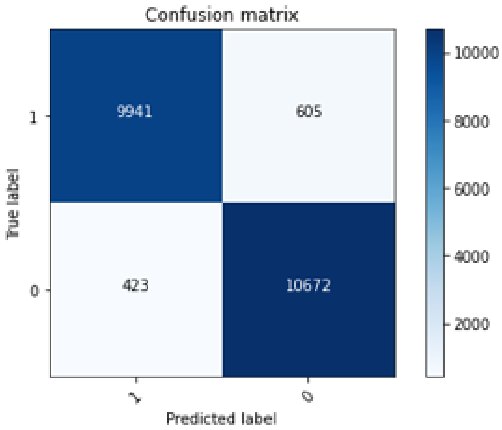


(iv)

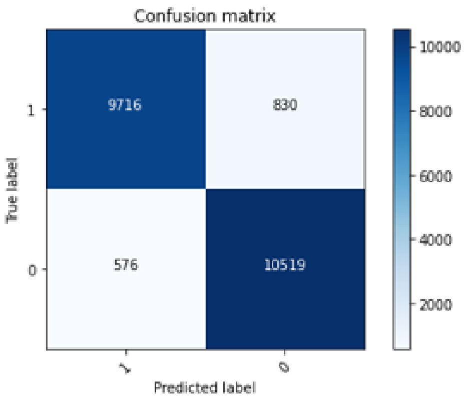
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*



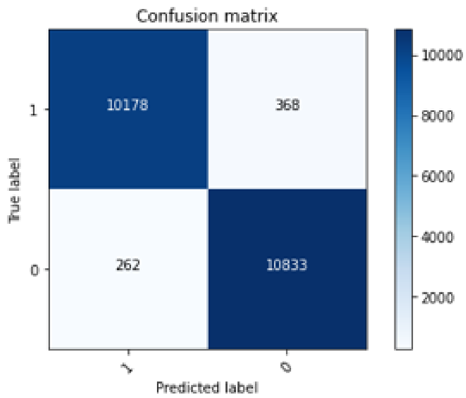
(v)



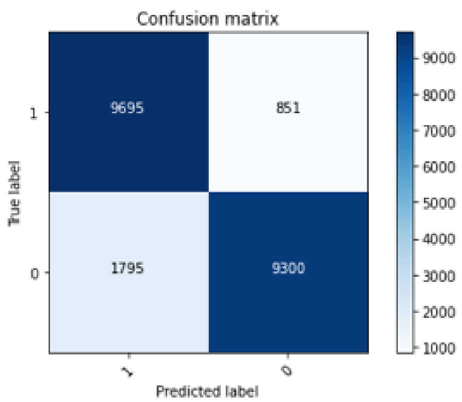
(vi)



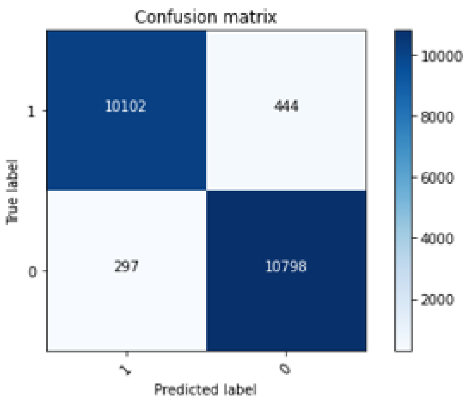
(vii)



(viii)



(ix)



(x)

Figure 3: Confusion Matrix for (i) Decision Tree (DT), (ii) Logistic Regression (LR), (iii) XGBoost (XGB), (iv) Random Forest (RF), (v) AdaBoost (AB), (vi) Support Vector Machine (SVM), (vii) Extra Trees (ET), (viii) Stochastic Gradient Descent (SGD), (ix) Naive Bayes (NB) and (x) recommended method using ensemble method.

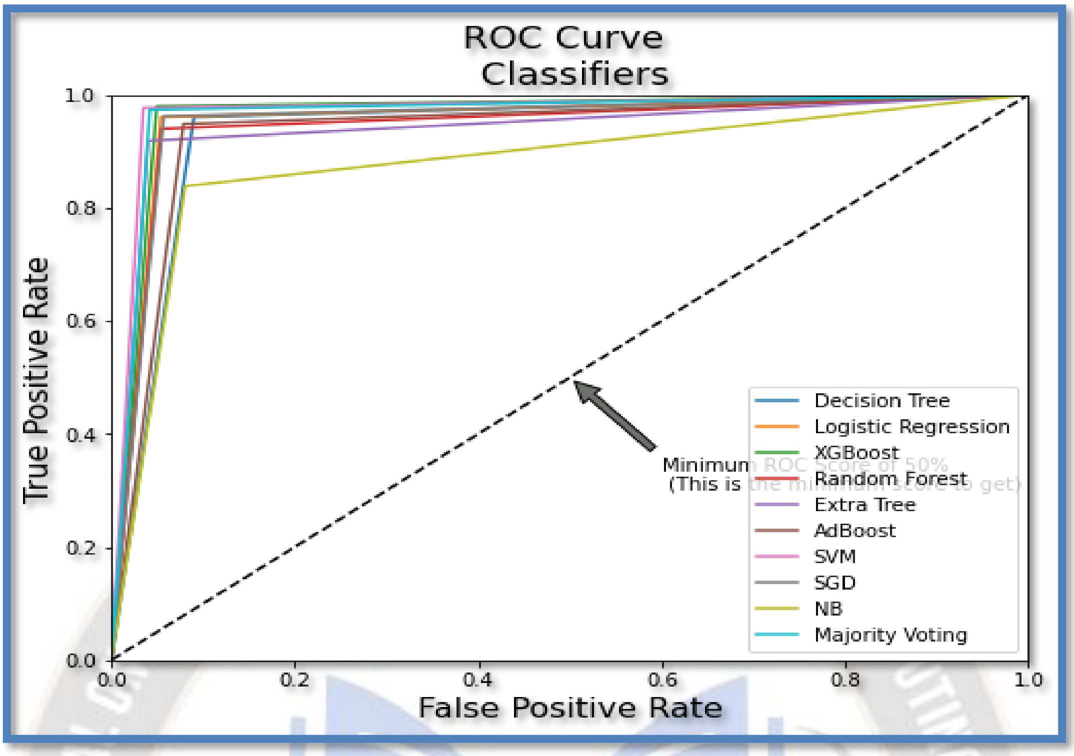


Figure 4: The fake news dataset's ROC curve for classifiers

4. CONCLUSIONS

In this paper, we provide an ensemble-based method for identifying fake news. Our investigation made use of a publicly accessible dataset made up of 72,134 news stories, 35,028 of which were determined to be true and 37,106 to be false using boolean labelling (0s and 1s). Eighty percent of the data were used for training and twenty percent for testing. We used the two CountVectorizer along with TF-IDF features along with well-known machine learning algorithms like Naive Bayes algorithm (NB), the XGBoost AdaBoost, Logical Regression (LR), a Random Forest (RF), the Extra Trees, a Support Vector Machine, as well as Stochastic Gradient Descent to analyse the dataset. We built a false news detection system using numerous models using an ensemble technique, the aforementioned classifiers, and features, with the goal of improving the accuracy of our findings. According to the experimental data, our suggested methodology attained an impressive accuracy and precision rate of 96.57%.

REFERENCES:

[1] Ahmad, M. Yousaf, S. Yousaf and M. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods", *Complexity*, vol. 2020, pp. 1-11, 2020. Available: 10.1155/2020/8885861[Accessed 17 October 2022].

[2] K. Xu, F. Wang, H. Wang and B. Yang, "Detecting fake news over online social media via domain reputations and content understanding", Tsinghua Science and Technology, vol. 25, no. 1, pp. 20-27, 2020. Available: 10.26599/tst.2018.9010139.

[3] "Hyperpartisan News and Articles Detection Using BERT and EL Mo", Ieeexplore.ieee.org, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9034917.

[4] "News labeling as early as possible | Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining", Dl.acm.org, 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3341161.3342957. 1456 Webology (ISSN: 1735-188X) Volume 18, Number 5, 2021 753 http://www.webology.org

[5] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. Choi and B. On, "Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)", IEEE Access, vol. 8, pp. 156695-156706, 2020. Available: 10.1109/access.2020.3019735.

[6] T. Jiang, J. Li, A. Haq, A. Saboor and A. Ali, "A Novel Stacking Approach for Accurate Detection of Fake News", IEEE Access, vol. 9, pp. 22626-22639, 2021. Available: 10.1109/access.2021.3056079.

[7] G. Sansonetti, F. Gasparetti, G. D'aniello and A. Micarelli, "Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection", IEEE Access, vol. 8, pp. 213154-213167, 2020. Available: 10.1109/access.2020.3040604.

[8] S. Ghafari et al., "A Survey on Trust Prediction in Online Social Networks", IEEE Access, vol. 8, pp. 144292-144309, 2020. Available: 10.1109/access.2020.3009445.

[9] Douglas, "News consumption and the new electronic media," 7e International Journal of Press/Politics, vol. 11, no. 1, pp. 29–52, 2006.
[10] Beer, D. D., & Matthee, M. (2020, May). Approaches to identify fake news: A systematic literature review. In International Conference on Integrated Science (pp. 13-22). Springer, Cham.
[11] Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5), 1-40.
[12] Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. Machine Learning with Applications, 4, 100032.
[13] Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770.
[14] Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake news" is not simply false information: A concept explication and taxonomy of online content. American behavioral scientist, 65(2), 180-212.
[15] Mridha, M. F., Keya, A. J., Hamid, M. A., Monowar, M. M., & Rahman, M. S. (2021). A Comprehensive Review on Fake News Detection with Deep Learning. IEEE Access.
[16] Seyam, A., Bou Nassif, A., Abu Talib, M., Nasir, Q., & Al Blooshi, B. (2021, August). Deep Learning Models to Detect Online False Information: a Systematic Literature Review. In The 7th Annual International Conference on Arab Women in Computing in Conjunction with the 2nd Forum of Women in Research (pp. 1-5).
[17] Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., & Vilares, J. (2021). Sentiment analysis for fake news detection. Electronics, 10(11), 1348.
[18] Parikh, S. B., & Atrey, P. K. (2018, April). Media-rich fake news detection: A survey. In 2018 IEEE conference on multimedia information processing and retrieval (MIPR) (pp. 436-441). IEEE.
[19] 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.
[20] D'Ulizia, A., Caschera, M. C., Ferri, F., & Grifoni, P. (2021). Fake news detection: a survey of evaluation datasets. PeerJ Computer Science, 7, e518.
[21] Brașoveanu, A. M., & Andonie, R. (2019, June). Semantic fake news detection: a machine learning perspective. In International Work-Conference on Artificial Neural Networks (pp. 656-667). Springer, Cham.
[22] Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2021). Trends in combating fake news on social media–a survey. Journal of Information and Telecommunication, 5(2), 247-266.
[23] Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. Procedia Computer Science, 141, 215-222.
[24] J. Wong, "Almost all the traffic to fake news sites is from facebook, new data show," 2016.
[25] D. M. J. Lazer, M. A. Baum, Y. Benkler et al., ") e science of fake news," Science, vol. 359, no. 6380, pp. 1094–1096, 2018.
[26] S. A. Garc´ıa, G. G. Garc´ıa, M. S. Prieto, A. J. M. Guerrero, and C. R. Jim´enez, ")e impact of term fake news on the scientific community scientific performance and mapping in web of science," Social Sciences, vol. 9, no. 5, 2020.
[27] V. P´erez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," 2017, https://arxiv.org/ abs/1708.07104.
[28] scikit-learn Machine Learning in Python, available online, https://scikit-learn.org/stable/
[29] Kumari, S. (2021). NoFake at CheckThat! 2021: fake news detection using BERT. arXiv preprint arXiv:2108.05419.
[30] Dong, X., & Qian, L. (2022). Integrating Human-in-the-loop into Swarm Learning for Decentralized Fake News Detection. arXiv preprint arXiv:2201.02048.
[31] Liu, Y., & Wu, Y. F. (2018, April). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).
[32] Mertoğlu, U., & Genç, B. (2020). Automated Fake News Detection in the Age of Digital Libraries. Information Technology and Libraries, 39(4).
[33] Agudelo, G. E. R., Parra, O. J. S., & Velandia, J. B. (2018, October). Raising a model for fake news detection using machine learning in Python. In Conference on e-Business, e-Services and e-Society (pp. 596-604). Springer, Cham.
[34] Horne, B. D., Nørregaard, J., & Adali, S. (2019). Robust fake news detection over time and attack. ACM Transactions on Intelligent Systems and Technology (TIST), 11(1), 1-23.
[35] Zhou, X., Jain, A., Phoha, V. V., & Zafarani, R. (2020). Fake news early detection: A theory-driven model. Digital Threats: Research and Practice, 1(2), 1-25.
[36] Hansrajh, A., Adeliyi, T. T., & Wing, J. (2021). Detection of Online Fake News Using Blending Ensemble Learning. Scientific Programming, 2021.
[37] Aslam, N., Ullah Khan, I., Alotaibi, F. S., Aldaej, L. A., & Aldubaikil, A. K. (2021). Fake detect: A deep learning ensemble model for fake news detection. complexity, 2021.
[38] Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake news detection using machine learning ensemble methods. Complexity, 2020.
[39] Torabi Asr, F., & Taboada, M. (2019). Big Data and quality data for fake news and misinformation detection. Big Data & Society, 6(1), 2053951719843310.
[40] Islam, N., Shaikh, A., Qaiser, A., Asiri, Y., Almakdi, S., Sulaiman, A., ... & Babar, S. A. (2021). Ternion: An Autonomous Model for Fake News Detection. Applied Sciences, 11(19), 9292.Mazzeo, V.
[41] Lai, C. M., Chen, M. H., Kristiani, E., Verma, V. K., & Yang, C. T. (2022). Fake News Classification Based on Content Level Features. Applied Sciences, 12(3), 1116.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

[42] Stissi, S. A Functional and Scale-able User Platform for Automatic Fake News Detection.

[43] Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2022). Deepfake detection by human crowds, machines, and machine-informed crowds. Proceedings of the National Academy of Sciences, 119(1).

[44] Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). Fake news detection: a deep learning approach. SMU Data Science Review, 1(3), 10.

[45] D. Holan, 2016 Lie of the Year: Fake News, Politifact, Washington, DC, USA, 2016. S. Kogan, T. J. Moskowitz, and M. Niessner, "Fake News:

[46] Evidence from Financial Markets," 2019, https://ssrn.com/ abstract=3237763.

[47] Punjabi, V.D. and Shukla, D.R.K. Study on fake news recognition and detection methods using machine learning techniques, Webology. Available at: ttps://www.webology.org/abstract.php?id=1970

[48] Robb, "Anatomy of a fake news scandal," Rolling Stone, vol. 1301, pp. 28–33, 2017.

[49] J. Soll, "e long and brutal history of fake news," Politico Magazine, vol. 18, no. 12, 2016.

[50] J. Hua and R. Shaw, "Corona virus (covid-19) "infodemic" and emerging issues through a data lens: the case of China," International Journal of Environmental Research and Public Health, vol. 17, no. 7, p. 2309, 2020.

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

[52] F. T. Asr and M. Taboada, "Misinfotext: a collection of news articles, with false and true labels," 2019.

[53] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[54] Pardeshi, S.M. and Dr. Dinesh, "A study of sentiment analysis from text through social networking sites", Webology. Available at: http://webology.org/abstract.php?id=1948 (Accessed: November 9, 2022).

[55] S. Vosoughi, D. Roy, and S. Aral, ") e spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[56] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, 2017.

[57] V. L. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in Proceedings of the Second Workshop on Computational Approaches to Deception Detection, pp. 7–17, San Diego, CA, USA, 2016.

[58] H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim, "exBAKE: automatic fake news detection model based on bidirectional encoder representations from transformers (bert)," Applied Sciences, vol. 9, no. 19, 2019.

[59] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pp. 127–138, Springer, Vancouver, Canada, 2017.

[60] W. Y. Wang, Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection, Association for Computational Linguistics, Stroudsburg, PA, USA, 2017.

[61] B. Riedel, I. Augenstein, G. P. Spithourakis, and S. Riedel, "A simple but tough-to-beat baseline for the fake news challenge stance detection task," 2017, https://arxiv.org/abs/1707. 03264.

[62] N. Ruchansky, S. Seo, and Y. Liu, "Csi: a hybrid deep model for fake news detection," in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 797–806, Singapore, 2017.