An Ensemble Machine Learning Approach for Fake News Detection and Classification Using a Soft Voting Classifier

Y. B. Lasotte, E. J. Garba, Y. M. Malgwi, and M. A. Buhari

***Abstract* — Fake news has grown in popularity and spread as a result of increased insecurity, political events, and pandemics, among other things. This study used an ensemble machine learning technique to better predict fake news on social media based on the content of news articles. The proposed model used a soft voting classifier to aggregate four machine learning algorithms, namely, Naïve Bayes, Support Vector Machine (SVM), and Logistic Regression, for the classification of news articles as fake or real. GridSearchCV was used to fine-tune the algorithms to get the optimal results during the training process. A Kaggle dataset was used for the experiment; it was comprised of both false and true news. Performance evaluation metrics were used to measure the performance of the base learners and our proposed ensemble technique on the dataset. The results of our experiment show that the proposed ensemble approach produced the highest accuracy, precision, recall, and F1_score values of 93%, 94%, 92%, and 93%, respectively, on the dataset as compared to the individual learners. This approach may also be used in other classification techniques for spam detection, sentiment analysis, and prediction of loan eligibility, among other things.**

***Keywords* — classification, ensemble machine learning, fake news detection, social media, soft voting.**

I. INTRODUCTION

Due to the accessibility, affordability, and convenience of use of social media platforms, a large number of individuals utilize social media to get their daily news**.** However, news on social media is not always reliable when compared to print news sources. Fake news has grown in popularity and spread as a result of increased insecurity, political events, and pandemics, among other things. The term fake news is often used by many researchers as well as laypeople to denote disinformation or misinformation. Fake news is a piece of news that is deliberately and verifiably false and could mislead consumers [1]. Authenticity and intent are two vital characteristics in this definition. According to this definition, a news article must be verified, and the intent must be known before it can be flagged as fake. They chose this definition primarily to eliminate ambiguities between false news and similar notions. This definition was embraced by many researchers, including [2]. Despite the fact that satire is frequently created for entertainment and recognizes its deception to consumers, some papers regard satire news as fake news because the contents are false [3]. Other authors classify deceptive news as fake news, including satires, hoaxes, and serious fabrications [4]. According to Shu *et al.* [5], the following concepts are not fake news:

i. Satire news that is properly contextualized, which has no intention of misleading consumers, and is unlikely to be misinterpreted as truthful;
ii. rumors that did not originate from news events;
iii. Conspiracy theories, which are hard to verify as true or false;
iv. misinformation that is not intended to mislead consumers; and
v. Hoaxes created only for the purpose of amusement or to defraud certain individuals.

Ahmed *et al.* [6] grouped fake news into three categories. The first is fake news, which is information that is completely incorrect and created by the authors of the articles. The second category is fake satire news, which is fake news with the primary objective of providing humor to consumers. The third category is poorly written articles, which have some degree of real news, but they are not completely true.

With the introduction of the Internet and the rapid adoption of social media platforms, anyone can easily create and publish news online. Social media platforms have brought together families, associates, like-minded people, and alumni associations, among other things. It allows users to speedily and easily access current news. Also, users can use social media to express their dissatisfaction or satisfaction with their leaders. Users are creating and sharing more information than ever before, some of which is deceptive [7]. The spread of information on social media is increasing at an exponential rate, and most individuals nowadays read their news on social media rather than in print media. There is a good chance that traditional news media may go out of business as a result of people's growing acceptance and use of social media.

Every phenomenon has advantages and disadvantages. Social media has its own challenges, despite its numerous benefits. The spread of fake news on social media has generated a lot of mistrust among its consumers. Some people are suspicious of practically every piece of news they see on social media as a result of fake news, while others believe virtually everything they see on social media. The spread of fake news has negative impacts on both target individuals and society as a whole [8]. People fabricate news for a variety of

Submitted on January 24, 2022.
Published on March 03, 2022.
Y. B. Lasotte, Federal University of Technology, Minna, Niger-Nigeria.
(corresponding e-mail: y.lasotte@futminna.edu.ng)
E. J. Garba, Modibbo Adama University, Yola, Adamawa-Nigeria.
(e-mail: e.j.garba@mautech.edu.ng)
Y. M. Malgwi, Modibbo Adama University, Yola, Adamawa-Nigeria.
(e-mail: yumalgwi@mautech.edu.ng).
M. A. Buhari, Airforce Institute of Technology, Nigerian Airforce Base, Nigeria.
(e-mail: muri.ameen@gmail.com)

reasons, including political and economic benefits. Fake news is spread to promote substandard products or to undermine political opponents. Elections are negatively impacted by fake news. For example, during the 2016 US presidential election, fake news was intentionally circulated on social media rather than accurate information [9]. In addition, there are also many occasions where cleverly fabricated fake news has wreaked havoc by inciting communal, religious, or ethnic crises [10]. One piece of fake news can spark a crisis, especially when it comes to sensitive topics like religion and ethnicity. False news has caused numerous ethnic and religious crises in Nigeria, disrupting the peaceful coexistence of a lot of communities. A religious crisis has been an obstacle to growth, economic prosperity, peaceful coexistence, and national integration. Religious, ethnic, and community crises have had a devastating effect on Nigerian educational standards, ranging from incessant school closures, destruction of school infrastructure, and abduction, maiming, killing, and rapping of schoolchildren, which have a serious effect on their psychology [11].

The surge in the popularity and spread of fake news has become a global issue that even major tech companies like Facebook, Instagram, and Twitter are striving to address. Fake news on social media is often made to look like real news, making it difficult for humans to distinguish. Therefore, the negative impacts posed by fake news can be mitigated to the barest minimum by several state-of-the-art machine learning algorithms. Machine learning algorithms have become so popular that they are now used in practically every scientific field [12]. In this paper, we offer a better method for classifying news articles as fake or real based on their content using the ensemble machine learning technique. We first, identified and used four popularly used machine learning algorithms: Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, and Random Forest. The algorithms were parameter-tuned using GridSearchCV to obtain optimal results during the training process. The data we used in our work was collected from Kaggle, a publicly available dataset. The best parameters were identified and used in the machine learning algorithms for the classification of news. The proposed system develops a better model by aggregating the base learners into a model that predicts by majority vote using a soft voting classifier. The performance of the algorithms was measured using popular evaluation performance matrices: accuracy, precision, recall, and f1-score.

II. Literature Review

In recent years, several studies on fake news detection have been conducted using machine learning algorithms, deep learning approaches, and natural language techniques. Scholars have applied machine learning ensemble approaches to solve many related problems. Ali *et al.* [13] suggest that when compared to single models, ensemble models have a higher acceptance in terms of accuracy. Ensemble techniques have proven to be very reliable in many domains due to their ability to cancel weaknesses in some machine learning algorithms, hence increasing the predictive power of a model. Kumari *et al.* [14] in their paper "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier" proposed an ensemble soft voting classifier for predicting diabetes mellitus using three machine learning algorithms: Logistic Regression, Random Forest, and Naive Bayes for the classification. Elgeldawi *et al.* [12] in their paper "Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis" suggest that the choice and settings of a machine learning model's hyperparameters can have a significant impact on the model's performance. Ngoc *et al.* [15] in their paper "Hyper-parameter Optimization in Classification: To-do or Not-to-do" propose a framework for deciding whether to use hyper-parameter optimization or the default hyper-parameter settings when dealing with the problem of whether or not to apply hyper-parameter optimization. Their empirical evaluation outcomes indicate that the framework technique can be used to systematically and incrementally determine the problem "to-tune-or-not-to-tune."

Probierz *et al.* [16] proposed an approach to classify news as fake or real based on the title without analyzing the whole content of the article. They compared their findings to a classification based on the entirety of the news article. Their work aims to propose a method that balances data analysis time and classification quality in fake news detection. The approach was based on natural language processing methodologies and machine learning techniques. The authors in [17] adopted a new automated algorithm. In their work, the authors have built a classification model based on the combination of lexical, syntactic, and semantic information. In [18], the authors proposed an emotion-based method for fake news detection. Their method combines the publisher's emotion and social emotion. Authors in [19] discovered specific traits that distinguish real news from false news. Their findings reveal that false news article titles are longer, contain more capitalized words, and use fewer stop words; while the body content of fake news articles is shorter, repetitious, has fewer nouns, and contains analytical and technical words.

The authors in [20] created a model that can predict whether a piece of news is false or not based on its content. The problem was handled entirely from the standpoint of deep learning using recurrent neural networks (vanilla, GRU) and long short-term memory (LSTM). After applying the results to the LIAR dataset, they reported their findings. GRU, LSTM, and vanilla achieved results of (0.217), (0.2166), and (0.215) respectively. GRU is the best among them. Gurav *et al.* [21] created a classifier using Naïve Bayes algorithm that can predict whether a piece of news is false based on its data source. This model takes news events as input and predicts whether the news is false or real based on Twitter reviews and classification algorithms. In [22], the authors proposed a system that uses Naïve Bayes algorithm with TF-IDF for fake news detection; the model gives a promising result. Gadekar [23] proposed two different classifiers which are Support Vector Machine (SVM) and Naïve Bayes. The SVM and Naïve Bayes achieved accuracies of 60.97% and 59.76% respectively. Ahmed *et al.* [6] built a model that uses n-gram analysis and machine learning techniques. Their experimental evaluation produces the best performance using Term Frequency Inverted Document Frequency as a feature extraction technique, and Linear

Support Vector Machine (LSVM) as a classifier, with an accuracy of 92%.

Gaydhani *et al.* [24] proposed a method for automatically categorizing tweets on Twitter into three categories: hateful, offensive, and clean. The Twitter dataset were used for the experiments. They used n-grams as features and fed their term frequency-inverse document frequency (TFIDF) values to several machine learning models. They performed a comparative analysis of the models using several values of n in n-grams and TFIDF normalization methods.

III. Methodology

This research work is based on improving the performance and accuracy of fake news detection. We have proposed an ensemble machine learning technique for the binary classification of news articles as true or false. The proposed ensemble technique with a soft voting classifier is depicted in Fig. 1 as a flow diagram.

A. Data Collection

The most significant thing in this work is the collection of input data, which is news in this case. Though, the domain of fake news detection is a relatively new area of study; Kaggle has some publicly available datasets that can be used to study fake news detection. Kaggle was used to obtain the data for this study. Kaggle is a Google LLC subsidiary that allows people to search and exchange datasets online. The data was split into two parts: training and test datasets. The training and test datasets contained 70% and 30% of the samples, respectively. To ensure an equal distribution of fake and real news articles, the news articles were shuffled.

B. Data Visualization

To understand the structure of the dataset, the dataset's visualization was carried out in the forms of pie chart and word cloud. The dataset contains 6335 news articles in a comma-separated values (CSV) format and four columns. The dataset includes both true and false information. 3171 news articles, which represent 50.1% of the dataset, were labeled as real news, while the other 3164, which represent 49.9% of the dataset, were labeled as fake news. Fig. 2 shows a pie chart to visualize our dataset. Fig. 3 and fig.4 represent the word clouds for fake news and real news for the Kaggle dataset. The word clouds displayed the top 500 most frequently used words that appear in fake and real news. Word cloud visualizations display the most commonly used words in a text, from small to large, based on how often they appear.

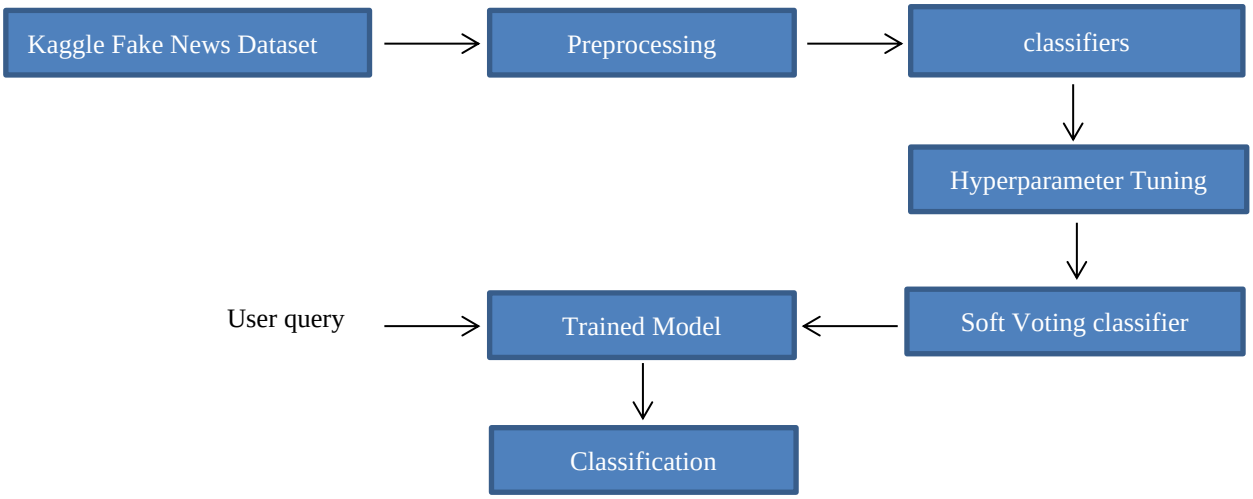

Fig.1. The flow diagram of the proposed ensemble technique using soft voting classifier.

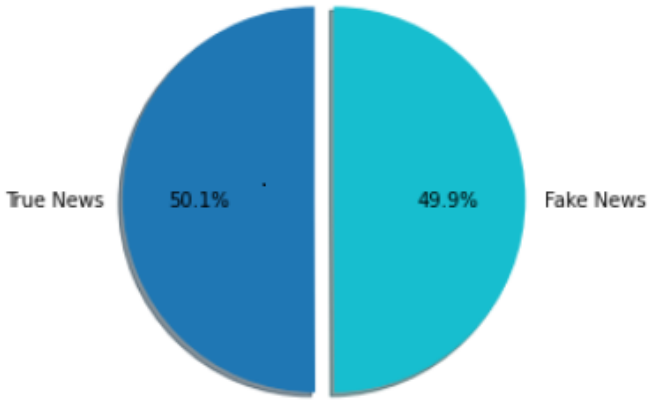

Fig. 2. Pie-Chart shows the percentage of fake and real news.

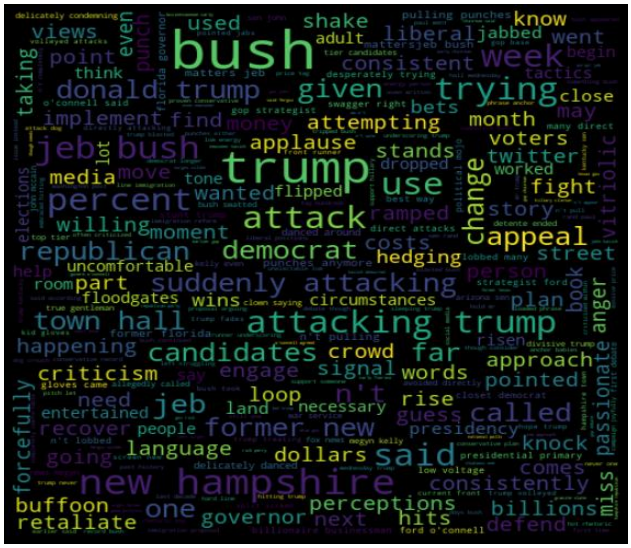

Fig. 3. Kaggle dataset real news word-cloud.

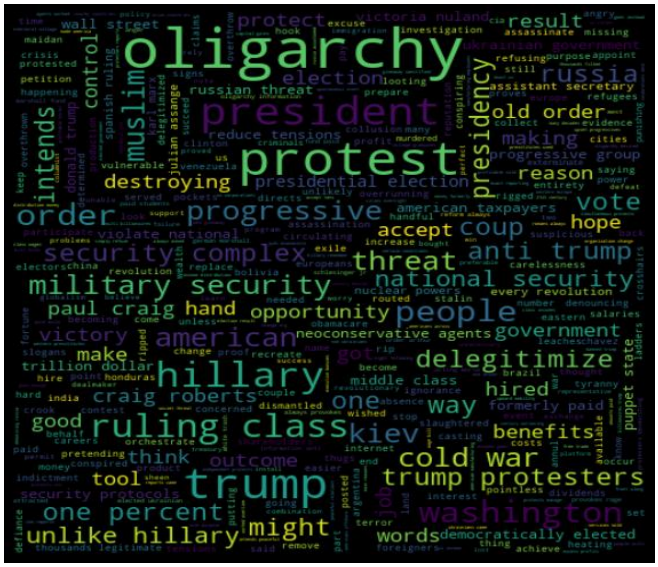

Fig. 4. Kaggle dataset fake news word-cloud.

C. Data pre-processing

Data Preprocessing is a crucial phase in the machine learning process that changes the data into a usable and efficient format so that it can be inputted into the algorithms. The first technique used for data preprocessing is label encoding. This technique is applied to the dependent variable, which is the news article, which is either real or fake. The labels were converted to binary digits with REAL as 1 and FAKE as 0. The next preprocessing of datasets is carried out as follows:

1. *Tokenization*: This process splits the given text into tokens (smaller parts) and removes all the punctuation from the textual data. nltk.tokenize method (an inbuilt function in the nltk library) was used in this research work for tokenization.

2. *Stop Words Removal:* Stop words are unimportant words in a language that affect the accuracy, efficiency, and performance of machine learning algorithms. These are words that are often employed in sentences to complete sentences or connect expressions. Conjunctions, articles, prepositions, and some pronouns are stop-words. In English, there are about 400-500 stop-words [25]. Some of the stop words are a, where, above, an, until does, will, who, when, that, what, but, by, on, about, once, after, too, again, all, am, and any, against, and so on. These terms will be eliminated from each document, and the processed document will be sent to the next phase.

3. *Stemming:* Stemming is the process of transforming a word's grammatical forms, such as its noun, adjective, verb, adverb, and so on, into its root form (also known as lemma). The main purpose of stemming is to get the basic forms of terms whose meanings are the same. For example, words such as select, selection, selections, selective, selecting, and selected can be stemmed to their lemma, which is the word "select". This can be done using the porter stemmer algorithm from the NLTK library, which is the most frequently used stemming algorithm.

D. Feature Extraction

Feature extraction is used to improve the accuracy of a model. Irrelevant features in a dataset can reduce model accuracy and performance while increasing training costs. Selecting a large number of features can increase the training time of the models [26]. Feature extraction should be utilized to reduce the quantity of text features in order to ensure efficient processing. In this thesis, the feature extraction technique adopted is the countvectorizer. The scikit-learn toolkit in Python has a great utility called CountVectorizer. It is used to convert a text into a vector based on the frequency (count) of each word in the text. It also allows you to preprocess a given text data before creating the vector representation, making it a highly flexible text feature representation module.

E. Hyper-parameter Tuning

Hyper-parameter tuning is the process of selecting a suitable set of parameters for a learning model. Parameters are the default values of the machine learning models that control the learning process. Every model comes with the default values of parameters, but this does not guarantee optimal performance. Determining the optimal values of hyper-parameters before the learning process is practically infeasible. Thus, different combinations were tried to determine the best hyper-parameters. GridSearchCV hyper-parameter tuning method was used.

F. Model Architecture

The proposed system uses four machine learning algorithms: Naïve Bayes, Support Vector Machine, and Logistic Regression for the classification of news. Hyper-parameter tuning was performed to get the optimal parameter values for the algorithms based on the dataset. Different numbers of estimators were used in a grid search to produce the best models that could predict the outcomes with high accuracy. The base learners were ensembled with a soft voting classifier to improve accuracy. This section briefly discussed these algorithms as well as the ensemble technique.

1. Naïve Bayes algorithm

Naïve Bayes is used to compute the conditional probability, which is defined as the probability that something will happen, given that something else has already occurred**.** It is a classification method that uses Bayes' Theorem and assumes predictor independence. The presence of one feature in a class is independent of the presence of any other feature, according to the Naive Bayes classifier. Naïve Bayes is easy to build, relatively fast, and useful for large datasets. It can be used for binary or multiclass classifications, making it reliable for text classification problems.

2. Support Vector Machine (SVM) Algorithm

Support Vector Machine (SVM) is a supervised machine learning algorithm that can solve both classification and regression problems. It is, however, commonly used in classification problems. An SVM classifier is a high-performing machine learning technique that works by dividing data into distinct regions. The SVM's goal is to discover a maximum margin that splits the dataset into two groups and to determine which category any new data falls under. Many people highly prefer the support vector machine because it produces significant accuracy while using less computing power. It performs exceptionally well on datasets that are smaller and more concise. Support vector machines are also capable of handling high-dimensional spaces and are memory-efficient [27].

3. Logistic regression

Logistic regression is a supervised machine learning algorithm that is used to solve classification problems. It is used for predicting the categorical dependent variable using a given set of independent variables. The predictions of logistic regression are done in terms of the probabilities of an event occurring. Logistic regression uses a sigmoid function to transform the output to a probability value; the aim is to minimize the cost function to achieve an optimal probability. The sigmoid function is as follows.

$$sigmoidequation = \frac{1}{1 + e\hat{}(-x)}$$

4. Random Forest Classifier

Random forest is a supervised machine learning technique that is versatile, simple, and diversified. It can solve classification and regression problems. The forest it builds is an ensemble of decision tree models to achieve better prediction results. In classification, the decision trees work individually to predict the outcome of a class, where the final prediction is the class that has the highest majority votes [28].

5. Ensemble Learners

Ensemble learners are primarily used to improve the performance of a model. The ensemble technique combines the predictions of two or more classifiers to create a model that can provide a more accurate prediction. The logic behind ensemble modeling is similar to that which we are already used to in our daily lives, such as obtaining the opinions of many experts before taking a final decision. As a result, ensemble-based machine learning is a method for reducing risk in decision-making. An example of such an approach is using voting classifiers, in which the final classification is based on the major votes provided by all algorithms [29]. Ensemble learning has been used in diverse applications such as spam detection, text categorization, optical character recognition, and face recognition, etc. Anywhere machine learning techniques can be used, ensemble learning can be used. The voting ensemble classifier will be used in this study as an ensemble technique.

6. Voting ensemble classifier

The voting ensemble is often employed for classification problems as it allows the aggregation of two or more learning models trained on the whole dataset [30]. It is s a machine learning model that learns from an ensemble of multiple independent models and predicts an output class based on the highest probability. The voting classifier uses two types of voting techniques. They are explained as follows:

i. Hard voting

In hard voting, the predicted output class is the one that receives the highest number of votes and has the highest probability of being predicted by each of the classifiers. Assume that three classifiers predicted the output class (A, A, B) and that the majority of them predicted A as the result. Hence, A will be the outcome.

ii. Soft voting

The output class in soft voting is the prediction based on the average probability assigned to that class. Suppose given some input to three models, the prediction probability for class A = (0.30, 0.47, 0.53) and B = (0.20, 0.32, 0.40). The averages for classes A and B are 0.4333 and 0.3067, respectively. The selected class is A since it had the highest probability averaged by each classifier.

IV. Performance Metrics

A variety of evaluation metrics were used to assess the model's performance in detecting fake news. Evaluation metrics are often employed in the machine learning community to measure the effectiveness of a classifier. A confusion matrix was used to evaluate the performance of our model, as shown in Table I. The confusion matrix is a tabular representation of a classification model's performance on the test set, which consists of four parameters: true positive, false positive, true negative, and false negative.

1. **True Positive (TP):** Denoting the number of positive instances that a classifier correctly predicted as positive.
2. **False Positive (FP):** Indicating the number of negative instances that a classifier erroneously predicted as positive;
3. **True Negative (TN):** Denoting the number of negative instances that a classifier correctly predicted as negative and
4. **False Negative (FN):** indicating the number of positive instances that a classifier mistakenly predicted as negative.

TABLE I: Confusion Matrix

	Predicted true	Predicted false
Actual true	True positive (TP)	False negative (FN)
Actual false	False positive (FP)	True negative (TN)

The performance of our model was evaluated based on four criteria, namely accuracy, precision, sensitivity, and F1_score:

i. **Accuracy:** In the case of the fake news detection problem, accuracy is the rate of accurately predicted news across all of the samples. It displays the percentage of messages correctly predicted as fake or real. The formula for calculating accuracy is as follows:

$$Accuracy = \frac{|TN|+|TP|}{|TN|+|FN|+|TP|+|FP|}$$

ii. **Recall:** recall represents the total number of positive classifications out of true class. In our case, it denotes the number of articles predicted as true out of the total number of true articles. The formula for the recall is as follows:

$$Recall = \frac{|TP|}{|FN|+|TP|}$$

iii. **Precision**: A precision score, on the other hand, is the ratio of true positives to all events predicted as true. In our case, precision refers to the number of articles that are marked as true out of all the positively predicted (true) articles:

$$Precesion = \frac{|TP|}{|FP|+|TP|}$$

iv. **F-measure:** F-measure is used to combine precision and recall, resulting in a general prediction

performance for fake news detection. The formula for the f-measure is as follows:

$$F - measure = 2 \times \frac{Recall \times Precesion}{Recall + Precesion}$$

In many circumstances, a high accuracy value signifies a good model. However, while training a classification model, an article that was predicted as true while it was actually false (false positive) can have negative consequences; similarly, an article that was predicted as false while it was actually true can create trust concerns. Therefore, we have considered precision, recall, and F1-score that take into account the wrongly classified observation. Note that the higher the values for Precision, Recall, F1, and Accuracy, the better the performance.

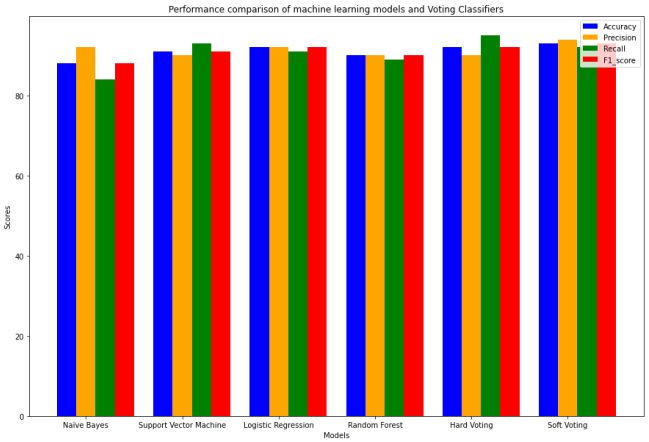

Fig. 5. Comparative graph of Accuracy, Precision, Recall, and F1_score of individual learners and voting classifiers.

V. Results and Discussions

Table II illustrates the comparison between the base learners and ensemble voting classifiers (hard and soft voting), and Fig. 5 shows the bar chart of the individual learners and voting classifiers based on accuracy, precision, recall, and f1_score. We can conclude from Table II, that the ensemble soft voting classifier achieved better accuracy, precision, recall, and f1_score of 93%, 94%, 93%, 94, and 93%, respectively as compared to the individual learners. It can be observed from the table that Logistic Regression is the best among the base learners, while Random Forest is the least. To get a clearer picture of the predictions, a confusion matrix for the soft voting classifier was created. Fig. 6 depicts the confusion matrix for the soft voting classifier. It shows the number of instances that are correctly or incorrectly predicted by the proposed ensemble soft voting classifier. The confusion matrix shows that 854 was true negative, 57 was false-negative, 79 was false positive, and 911 was true positive.

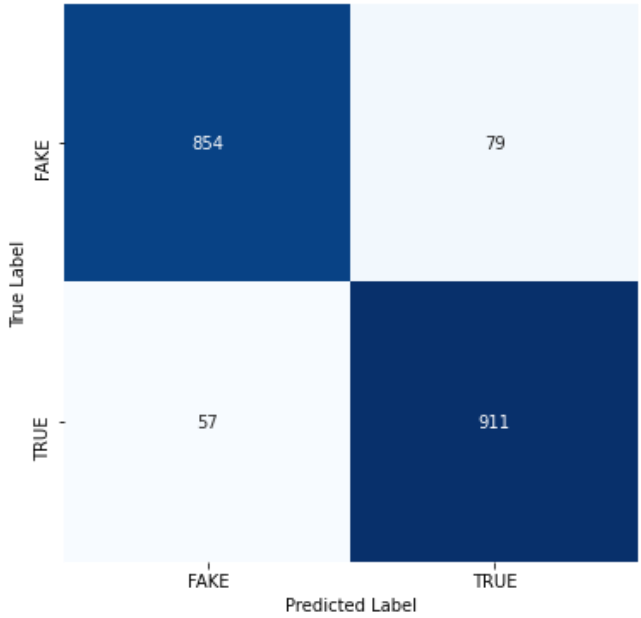

Fig. 6. Confusion matrix for voting classifier.

TABLE II: Comparison between Machine Learning Models and Voting Classifiers (Kaggle Dataset)

Algorithm	Accuracy Rate	Precision Fake news	Precision Real news	Recall Fake news	Recall Real news	F1_score Fake news	F1_score Real news
Naïve Bayes	88	92	86	84	93	88	89
SVM	91	90	93	93	90	91	91
Logistic Regression	92	92	92	91	92	92	92
Random Forest	90	90	90	89	90	90	90
Hard Voting	92	90	94	95	90	92	92
Soft Voting	93	94	92	92	94	93	93

V. Conclusion

Social media contributes to the increase in the popularity and propagation of fake news due to its flexibility, convenience, and affordability. The primary goal of this research work was to improve the accuracy of fake news predictions. We first identified and used four popular supervised machine learning algorithms, namely, Naïve Bayes, SVM, Logistic Regression, and Random Forest. The algorithms were trained and parameter-tuned using GridSearchCV to obtain optimal results. The data we used in our work was collected from Kaggle, a publicly available dataset. The best parameters were identified and used in the machine learning algorithms for the classification of news. The proposed system creates a better model by aggregating the base learners into a model that predicts by majority vote using a soft voting classifier. Performance metrics were used to measure the performance of the base learners and the soft voting technique. The experimentation evaluation showed that the soft voting classifier has better results as compared to the individual learners. The proposed model has a 93% accuracy rate. The voting classifier has been shown to be a more powerful classifier that balances the weaknesses of the individual learners on the Kaggle dataset. This accuracy might be improved in the future by employing various deep learning models. Proper hyperparameter tuning of a machine learning classifier gives better accuracy. We therefore recommend that various hyperparameter tuning techniques such as Random Search, Particle Swarm Optimization (PSO), Bayesian Optimization, and Genetic Algorithm (GA) among

others should be tried to see which one gives a promising result.

CONFLICT OF INTEREST

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Zhang Y, Su Y, Weigang L, Liu H. Rumor and authoritative information propagation model considering super spreading in complex social networks. *Physica A.* 2018; 395-411.
[2] Conroy NJ, Rubin VL, Chen Y. Automatic Deception Detection: Methods for Finding Fake News. *in ASIS&T Annual Meeting: Information Science with Impact.* St. Louis, MO, USA, 2015.
[3] Balmas M. When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. *Communication Research.* 2014; 41(3): 430-454.
[4] Rubin VL, Chen Y, Conroy NJ. Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology, 52.* 2015.
[5] Shu, K, Sliva A, Wang S, Tang J, Liu H. Fake news detection on social media: A data mining perspective. 2017; KDD exploration newsletter.
[6] Ahmed H, Traore I, Saad, S. Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. 2017; Security and Privacy.
[7] Ahmad I, Yousaf M, Yousaf S, Ahmad, MO. Fake news detection using machine learning ensemble methods. *Hindawi. 2020.*
[8] Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PK, Khan WZ. An ensemble machine learning approach through effective feature extraction to classify fake news. *Elsevier.* 2021; 47-58.
[9] Bovet A, Makse HA. Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications.* 2019; *10*(1): 1-14.
[10] Khan YJ, Khondaker ST, Iqbal A, Afroz SA. Benchmark Study on Machine Learning Methods for Fake News Detection. 2019.
[11] Apuwabi OO. The Effects of Religious Crisis on Economic Development in Nigeria. *International Journal of Academic Research in Business and Social Sciences.* 2018; *8*(6): 321–330.
[12] Elgeldawi E, Sayed, A, Galal AR, Zaki AM. Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. Informatics 2021, 8, 79. https://doi.org/10.3390/informatics8040079. 2021.
[13] Ali S, Tirumala SS, Sarrafzadeh A. Ensemble learning methods for decision making: status and future prospects. 2015.
[14] Kumari S, Kumar D, Mittal M. An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering*. 2021; 40-46.
[15] Ngoc T, Jean-Guy S, Ingo W. Hyper-parameter Optimization in Classification: To-do or Not-to-do. Article *in* Pattern Recognition · July 2020. DOI: 10.1016/j.patcog.2020.107245.
[16] Probierz B, Stefa´nski P, Kozaka J. Rapid detection of fake news based on machine learning methods. *25th International Conference on Knowledge-Based and Intelligent Information & Engineering System*, pp. 2893–2902, Poland, 2021.
[17] Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R. Automatic Detection of Fake News. 2017.
[18] Guo C, Cao J, Zhang X, Shu K, Yu M. Exploiting emotions for fake news detection on social media.2019.
[19] Horne BD, Adali S. This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. 2017.
[20] Girgis S, Amer E, Gadallah M. Deep Learning Algorithms for Detecting Fake News in Online Text. 2018.
[21] Gurav S, Sase W, Shinde S, Wabale P, Hirve S. Survey on Automated System for Fake News Detection using NLP & Machine Learning Approach. *International Research Journal of Engineering and Technology (IRJET).* 2019; *6*(1).
[22] Poovaraghan RJ, Priya MV, Vamsi PV, Mewara M, Loganatha S. Fake news accuracy using naive bayes classifier. *International Journal of Recent Technology and Engineering (IJRTE).* 2019; 8(1C2): 2277-3878.
[23] Gadekar PS. Fake News Identification using Machine Learning. *International Journal for Research in Applied Science & Engineering Technology (IJRASET). 2019; 7(V): 2321-9653.*
[24] Gaydhani A, Doma V, Kendre S, Bhagwat L. Detecting Hate Speech and Offensive Language on Twitter using Machine Learning: An N-gram and TFIDF based Approach. 2018.
[25] Dharmendra S, Suresh J. Evaluation of stemming and stop word techniques on text classification problem. *International Journal of Scientific Research in Computer Science and Engineering.* 2015; 3: 1-4.
[26] Reddy GT, Reddy MPK, Lakshmanna KVR. Rajput DS, Srivastava G, Baker T. (2020). Analysis of dimensionality reduction techniques on big data. *IEEE Access.* 2020; 8: 54776–54788.
[27] Ray S, Srivastava T, Dar P, Shaikh F. *Understanding Support Vector Machine algorithm from examples (along with code).* Available from https://www.analyticsvidhya.com/blog/2020/09/understaing-support-vector-machine-example-code/. [Accessed 15th August, 2020].
[28] Pal M. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*. 2005; 26(1): 217–222.
[29] Ruta D, Gabrys B. Classifier selection for majority voting. *Information Fusion. 2005; 6*(1): 63–81.
[30] Lam L, Suen SY. (1997). Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics - Part A.* 1997; 27(5): 553–568.

Lasotte Yakubu Boyi-Musa is a graduate assistant in the Department of Computer Science, Federal University of Technology, Minna Niger State Nigeria. He received his Bachelor's degree in Computer Science from Modibbo Adama University, Yola, Adamawa State Nigeria in the year 2015. He is pursuing his Master's degree in Computer Science from Modibbo Adama University, Yola Adamawa State, Nigeria. His research areas are data science and machine learning.

Yusuf Musa Malgwi is a Lecturer with the Department of Computer Science Modibbo Adama University, Yola, and Adamawa state-Nigeria. He graduated with a Bachelor of Technology in Computer Science with honors from Federal University of Technology, Yola in 2006. M.Sc. Computer Science in 2014 from Adamawa State University, Mubi-Nigeria, and PhD in Computer Science in 2019 from Modibbo Adama University Yola Adamawa State- Nigeria. He specializes in Machine Learning/Medical Informatics. The main courses of teaching and learning are in the area of Computer Science. He has supervised 5 PhD candidates as Co- Supervisor and 9 M.Sc candidates to completion as main supervisor. His experience in postgraduate supervision has been recognized as he was appointed the Postgraduate Co-coordinator for the Department of Computer Science of the Modibbo Adama University, Yola. He has also published more than 20 Journals in indexed publication especially those with high impact factors.