Research Article

Detection of Online Fake News Using Blending Ensemble Learning

Arvin Hansrajh ⓘ**, Timothy T. Adeliyi** ⓘ**, and Jeanette Wing** ⓘ

ICT and Society Research Group, Durban University of Technology, Durban 4001, South Africa

Correspondence should be addressed to Timothy T. Adeliyi; timothya2@dut.ac.za

The exponential growth in fake news and its inherent threat to democracy, public trust, and justice has escalated the necessity for fake news detection and mitigation. Detecting fake news is a complex challenge as it is intentionally written to mislead and hoodwink. Humans are not good at identifying fake news. The detection of fake news by humans is reported to be at a rate of 54% and an additional 4% is reported in the literature as being speculative. The significance of fighting fake news is exemplified during the present pandemic. Consequently, social networks are ramping up the usage of detection tools and educating the public in recognising fake news. In the literature, it was observed that several machine learning algorithms have been applied to the detection of fake news with limited and mixed success. However, several advanced machine learning models are not being applied, although recent studies are demonstrating the efficacy of the ensemble machine learning approach; hence, the purpose of this study is to assist in the automated detection of fake news. An ensemble approach is adopted to help resolve the identified gap. This study proposed a blended machine learning ensemble model developed from logistic regression, support vector machine, linear discriminant analysis, stochastic gradient descent, and ridge regression, which is then used on a publicly available dataset to predict if a news report is true or not. The proposed model will be appraised with the popular classical machine learning models, while performance metrics such as AUC, ROC, recall, accuracy, precision, and $f1$-score will be used to measure the performance of the proposed model. Results presented showed that the proposed model outperformed other popular classical machine learning models.

1. Introduction

The increasing use of the Internet coupled with social media platforms has enabled even more people to obtain news from a wide variety of sources instead of old-style news outlets. People who spend a lot of time online are more likely to acquire news and updates through social media with an increased risk of exposure to wide-scale misinformation [1]. This provides fertile ground for fake news as news articles, hoaxes, reviews, rumours, satires, advertisements, reviews, and exaggerated claims proliferate. The widespread distribution of bogus news is capable of producing extremely adverse effects on individuals and humanity [2]. It has now become a part of daily life to hear of the worsening weather crises, political violence, intolerance amongst people of different ethnicity and cultural backgrounds, and even influencing issues of public health. This is often done to advance or foist certain ideas into circulation and is often realised with political agendas. Therefore, all the governments around the world are trying to track and tackle this problem [3]. Bogus news is not a new concept [4] or a product of the digital communication age [5]. It has most recently come to light during the 2016 US presidential election. There have been numerous hoax stories where citizens and governments as well all other social elements are all impacted and influenced by these stories.

Facebook has been at the epicentre of the controversy by the media houses for targeting the population and showing them posts to their support [6]. It has been alleged that bogus news could have been decisive in the 2016 US presidential

election [7]. Nevertheless, we can contend that false news and generally misinformation have become a big problem, which may have a significant social cost in the future. The ubiquitous nature of the Internet enables anybody to spread false and biased information easily. It is virtually impossible to prevent or control fake news from being created or disseminated. Consequently, both online platforms and researchers are very proactive in detecting potential false news. It is a complex problem since false news can present itself in multiple ways making it both physically and automatically challenging to efficiently identify [8].

Headlines in the form of clickbait are used to entice users to view probably subjective articles to make a profit. According to Wang [9], "The problem of fake news detection is more challenging than detecting deceptive reviews, since the political language on TV interviews and posts on Facebook and Twitter are mostly short statements." Therefore, it is very evident that the development of automated solutions for false news detection is imperative and exigent [3]. Prior works have used many classical models. However, several unconventional learning models are not applied although they have proven best in numerous text classification problems [10]. An ensemble approach is proposed to help resolve the identified gap.

Recent studies are demonstrating the effectiveness of ensemble learning approaches with promising results [11]. This study will investigate how natural language processing techniques and machine learning can be combined in a blending ensemble approach to create a model that will use the data of previous news reports and predict a news report as being true or not. The proposed model will be compared with the classical machine learning models using performance metrics, for example, AUC, ROC, recall, accuracy, precision, and f1-score. These measurements will be used to gauge the performance of the model.

The remainder of this paper is outlined as follows. The related literature is discussed in Section 2. Section 3 presents the study materials and methods, while result analysis is presented in Section 4 and the paper is succinctly concluded in Section 5.

2. Related Works

Humans are fairly unimpressive at recognising deception. Most people believe that the information they obtain is factual and trustworthy. They tend to be unaccountably perceptive to knowledge that is not fully understood [12]. Confirmation bias influences people to grasp only what they want to perceive [13]. Therefore, the proliferation and propagation of fake news is a major concern because of its capacity to generate devastating consequences. Diverse machine learning approaches are utilised to combat it. However, the majority focused on a specific category of news without utilising several advanced methods [13, 14].

Numerous neural networks and models based on machine learning have been applied to detect fake news. Models were developed with features designed for specific datasets. Yun and Ahn [15] detected fake news in Korea with machine learning and text mining using a two-step approach.

Initially, the news contents are converted to values by applying text mining, and then classifiers are trained on these values. Aphiwongsophon and Chongstitvatana [16] based their models on identifying fake news using selected data sourced from Twitter. It is likely that these approaches will fall victim to dataset bias and possibly perform poorly on a different category of news [10]. Gilda [17] explored some traditional machine learning approaches. Ahmed et al. [18] investigated and compared six different classification techniques using n-gram analysis on a single dataset using feature extraction. The models were evaluated independently and the linear support vector machine classifier achieved the best score. However, several advanced learning models are not applied although they have excelled in text classification [10].

Research using deep learning to identify fake news works has accomplished encouraging results [10]. Rashkin et al. [19] used linguistic feature analysis and achieved the remarkable outcome of Long Short Term Memory. Wang [9] constructed a hybrid model using a convolutional neural network that outclassed other traditional learning models. Singhania et al. [20] applied a three-level attention network incorporating sentences, words, and headlines. Thota et al. [21] presented a neural network to forecast the stance using the headline and the body of the article. Wang [9] presented a benchmark dataset named Liar and investigated using current models. The evaluation hints at how different types of models perform on data that is structured. Also, some models were prone to being overfit.

Ruchansky et al. [22] built a CSI (capture, score, and integrate) model that used text, article response, and characteristics of the users' behaviour. Ajao et al. [23] developed a framework for classifying and identifying fake news in Twitter posts using a hybrid of neural networks. The tactic intuitively identified pertinent features without considering prior knowledge. Lu and Li [24] developed (GCAN) Graph-aware Co-Attention Networks to determine if a tweet is fake by using the associated sequence of retweet users. Khan et al. [10] analysed the performance of dissimilar approaches on three datasets and showed that Naive Bayes can achieve a similar result as neural network models when working with a dataset containing under 100 thousand articles. Vijayaraghavan et al. [25] applied different models to detect fake news and state that neural networks generally perform consistently and serve as a powerful universal approximator. However, the loss and accuracy come after using too many epochs and thus the issue of overfitting comes into play. In addition, a simpler model using logistic regression also delivered good performance results. Consequently, it does not necessarily follow that the more complicated the model, the better the performance. Furthermore, deep learning is "time-consuming and resource-consuming" [26].

Researchers have studied numerous algorithms for text classification that give good performance. However, some algorithms perform better on some datasets but may give even an average performance on other datasets. Therefore, instead of using a single classifier, it is better to use a group of classifiers and take a collective or team decision, rather than

basing a decision on an individual classifier [27, 28]. This approach called an ensemble approach overcomes the weakness of one classifier by the strength of other classifiers and gives better performance than an individual classifier. The diverse nature of the approach and keeping the variance under control contribute greatly to its success. Furthermore, ensemble learning can result in more robust schemes of classification.

Roy et al. [3] developed models built on a Bidirectional Long Short Term Memory and Convolutional Neural Network. The output from both of these models was input into a Multilayer Perceptron Model to obtain the final result. Al-Ash et al. [29] used a random forest classifier which consists of a decision tree classifier as an ensemble classifier to detect Indonesian fake news. Reddy et al. [26] presented a hybrid approach for fake data detection using an ensemble model. Ahmad et al. [30] explored different textual properties in an ensemble approach to detect fake news. Gutierrez-Espinoza et al. [11] evaluated the performance of ensemble learning using different machine learning techniques for classification in order to identify bogus online information.

Mahabub [31] used a distinct method for detecting fake news in developing an ensemble voting classifier that incorporates many familiar machine learning algorithms. Kaur et al. [32] designed a voting model with multiple levels in automating the detection of fake news by experimenting with several models. Saeed et al. [1] incorporated an ensemble approach to detect spam from Arabic texts. Li et al. [33] applied a pipeline to identify fake news by taking into consideration the headline and article text in a stacked ensemble. In all these studies, the ensemble approach yielded better performance when compared to the individual model in the detection of deceptive information. Therefore, an advanced ensemble approach is adopted to detect fake news. The strategy will integrate blending and machine learning with natural language processing to extend and improve the current approaches.

3. Materials and Methods

In this section, we present the datasets, proposed framework, explanation of the algorithms, and the metrics that are used for performance evaluation. Two datasets have been selected for our experiments which include news from a range of different categories and a combination of fake and truthful articles. Both datasets are publicly available and easily accessible on the web. Categorization of news as "fake news" can be "a very challenging and time-consuming task" [34]. Hence, existing datasets are used in this study. The major challenge to identify false news is the accessibility and calibre of the datasets [35]. Also, finding a corpus of articles related to news is particularly problematic owing to copyright concerns [17]. The Liar [9] and ISOT [36] datasets are used.

The Liar dataset is publicly accessible and has been successfully used [37]. It comprises 12836 short labelled statements from politifact.com. There are six labels for rating the truthfulness of a statement: "pants-fire," "false," "barely-

true," "half-true," "mostly-true," and "true." We focus on classifying news as true or fake. For binary classification of the news, we transform these labels into two labels. "Pants-fire," "false," and "barely-true" are considered as fake and "half-true," "mostly-true," and "true" are considered as true. This dataset largely focuses on politics that contain statements of republicans and democrats, in addition to a substantial quantity of posts from social media [10].

The ISOT Fake News Dataset comprises both truthful and fake news articles sourced from several domains [30]. The true articles were sourced mainly from reuters.com, a well-known news site on the web. The fake news articles were obtained from numerous sources, primarily from websites that have been flagged by politifact.com. The dataset consists of 44,898 articles, 23,481 being fake articles and 21,417 true articles. Each data point consists of a title, text, subject, and date. The text is the actual news article, and the subject or category is any one of Middle East, government news, US news, world news, politics news, left-news, politics, and news.

In the proposed framework, as shown in Figure 1, we are extending the current literature by introducing ensemble learning techniques incorporating blending. News articles from several domains are classified as true or fake by working with different feature sets. Blending ensemble techniques with Term Frequency, Term Frequency Inverted Document Frequency, and n-grams are used in our approach.

Raw texts of the news need to be preprocessed before being fed into the models. Natural language processing techniques will be applied to help improve accuracy. The following operations will be carried out during the preprocessing of the dataset:

(i) Data cleansing: remove irrelevant data that is not required for the analysis.

(ii) Check for missing values that can have an adverse effect on the final result.

(iii) Convert the text to lowercase so that there is consistency.

(iv) Remove all punctuation marks.

(v) Remove stopwords from the textual dataset. These are words that provide no added semantical meaning and are of no significance during natural language processing.

(vi) Stemming (or lemmatization) involves converting words back to their original structure and thus reducing the classes or word types present in the dataset. For example, "Dancing", "Dance," and "Dancer" will be shortened to "dance." Stemming makes classification more efficient and quicker [18]. The Porter Stemmer algorithm will be used due to its accuracy.

Features' design plays a key role in the machine learning models' performance. The extraction of the most relevant or important words and using them as features can be extremely useful. Term Frequency, Term Frequency Inverted
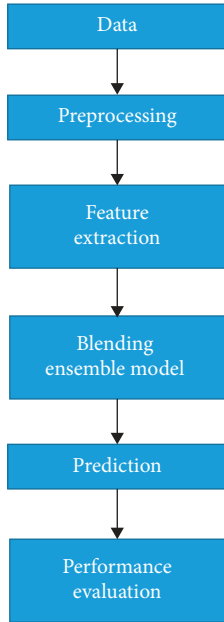
Figure 1: Proposed model.

Document Frequency, and n-grams will be used in the extraction of features from the dataset. This approach has been chosen over word embedding based on the experimental results realised in previous studies. Thota et al. [21] achieved better results using n-grams over word embedding. Vijayaraghavan et al. [25] used Word2Vect embedding and showed that it performed the worst when compared to TF-IDF models. Similar results were also confirmed by Smitha and Bharath [38]. Term Frequency utilises the tallies of words present within the documents to determine the resemblance between documents. A vector with an equal dimension that holds the counts of words is associated with each document. Term Frequency Inverted Document Frequency is a metric frequently used in the processing of natural language and information retrieval. It measures the significance of a term in a document included in the dataset. The n-grams based on words are used in representing the document's context and for features' generation that can be useful in the classification of a document as real or fake. This approach has been used successfully with unigrams and bigrams in fake news detection [10].

3.1. Blending Ensemble Model.

Blending is very closely allied to stacking. Stacking (stacked generalization) involves a learning algorithm being trained to pool the predictions of several other learning algorithms. All the algorithms are trained on the available data. A combiner algorithm is eventually used for the final prediction by taking into account the predictions of the other algorithms [39]. The blending ensemble is a variation of stacking. The prediction blending ensemble variation is based on a holdout dataset validation which was used in this study to fit the meta-model rather than out-of-fold predictions. The model learns to combine the predictions of several contributing ensemble

base models. Models implementing logistic regression, support vector machine, linear discriminant analysis, stochastic gradient descent, and ridge regression are used in the formation of the ensemble. The blending ensemble algorithm is given as follows (Algorithm 1).

3.1.1. Logistic Regression.

A logistic regression model is used since the text is being classified resulting in binary output (0/1 or true/false or true/fake). The hypothesis function can be defined mathematically as follows:

$$h_\theta(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}. \tag{1}$$

A sigmoid function transforms the output into a probability. The goal is to achieve optimal probability by minimizing the cost function as shown as follows [30]:

$$\mathrm{Cost}\left(h_\theta(x), y\right) = \begin{cases} \log\left(h_\theta(x)\right), & y = 1, \\ -\log\left(1 - h_\theta(x)\right), & y = 0. \end{cases} \tag{2}$$

Hence, logistic regression produces a logistic curve that is restricted to values that are between 0 and 1 by utilising a sigmoid function.

3.1.2. Support Vector Machine.

Support vector machine creates a hyperplane to isolate and group features. Support vectors are created on either side of the hyperplane in order to calculate the optimal hyperplane with each vector maximising the distance between them. The greater the vector distance around the hyperplane results in a more accurate decision boundary between the category features [25]. The data points are classified into distinct classes dependent on their position on the hyperplane. The key motive is to maximise the gaps that exist between the hyperplane and data points. The margin is maximised by the loss function. The hyperplane is defined by

$$w^x + b = 0, \tag{3}$$

where w is the weight vector and b is the bias.

$$L(w) = \sum_1 \max\left(0, 1 - y_i\left[w^t x_i + |b|\right]\right) + \lambda \|w\|^2. \tag{4}$$

The errors are calculated by the first term in the loss function. The regularization function is represented by the second term and is used to circumvent overfitting [27].

3.1.3. Linear Discriminant Analysis.

Linear discriminant analysis (LDA) is a machine learning algorithm for classification. It computes the summary statistics, for example, standard deviation and mean, associated with the input features by the class label. The statistics indicate what has been learnt by the model from consuming the training data. Predictions are based on probability estimates of a new case fitting in a class label established for every input feature value. The class that has the greatest probability is allocated to the new case. LDA can be viewed as a straightforward

(1) Split the dataset.
   The dataset is split into test and train sets.
(2) Construct the base models.
(3) Train the blending ensemble.
 Repeat
    Fit on the training set.
    Make prediction on holdout set.
    Store the predictions as input for blending until the end of base models.
 Build a 2D array using the stored predictions.
 Create the blending model.
 Fit the blending model on the predictions from base models.
(4) Make predictions with the blending ensemble.
 Repeat
    Make prediction with base model on test set.
    Store the prediction until the end of base models. Build a 2D array using the stored predictions.
(5) Evaluate the predictions.

ALGORITHM 1: Blending ensemble algorithm.

application of Bayer's theorem aimed at classification. The process can be summarised using three significant steps [37]:

(i) Compute the between-class variance (separability) between the different classes. This is expressed by the following formula:

$$S_b = \sum_{i=1}^{g} N_i (\overline{x_i} - \overline{x})(\overline{x_i} - \overline{x})^T. \tag{5}$$

(ii) Compute the within-class variance using the following formula:

$$S_w = \sum_{i=1}^{g} (N_i - 1)S_i = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (x_{i,j} - \overline{x}_i)(x_{i,j} - \overline{x}_i)^T. \tag{6}$$

(iii) Create the lower-dimensional space to maximise the between-class variance and to minimize the within-class variance. The lower-dimensional space projection (Fisher's criterion) is given by the following equation:

$$P_{\text{lda}} = \arg\max_{P} \frac{\left|P^T S_b P\right|}{\left|P^T S_{wP}\right|}. \tag{7}$$

LDA assumes that we have numeric input variables distributed normally and have the same spread (variance). Otherwise, it may be necessary to transform or normalize the data before modelling. The model is multiclass. It supports double-class classification problems and multiclass classification with no modification.

3.1.4. Stochastic Gradient Descent. Stochastic gradient descent uses an iterative method to optimize an objective function through appropriate smoothness properties such as differentiable or subdifferentiable. The method consumes randomly shuffled or selected samples to gauge the gradients. Therefore, stochastic gradient descent "can be regarded as a stochastic approximation of gradient descent optimization" [6]. The gradient is principally the slope or slant of a function. It is the amount "of change of a parameter with the quantity of change in another parameter" [40]. The greater the gradient, the sharper the slope. Gradient descent is applied iteratively to find the parameter values of a function that will minimize the function value with the maximum quantity. Therefore, the objective is to determine optimal parameter values in order to obtain the minimum value of the cost function.

Mathematically, the details can be expressed (for classification) as follows: given a set of training examples $(x_1, y_1) \ldots (x_n, y_n)$, where $x_i \in \mathcal{R}^m$ and $(y_i \in -1, 1)$, the objective is to learn a linear scoring function $f(x) = w^t + b$ with model parameters $w \in \mathcal{R}^m$ and intercept $b \in \mathcal{R}$. Predictions for binary classification are made by looking at the sign of $f(x)$. To determine the parameters of the model, the regularized training error is minimized and is shown as follows:

$$E(w, b) = \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i)) + \alpha R(w), \tag{8}$$

where L is a loss function and R is a regularization (penalty) term used to penalize the model complexity; $\alpha > 0$ is a hyperparameter that controls the strength of the regularization [37].

3.1.5. Ridge Regression. The regression method serves as a basis for the ridge classifier. For binary classification, the target variable is converted into +1 or −1 dependent on the class to which it belongs, and for multiclass data that uses multioutput regression, the largest value for prediction is acknowledged as the target class. Ridge regression is virtually the same as linear regression except that a small bias is introduced. Consequently, the variance is reduced significantly. So, by beginning with a somewhat worse fit, better predictions in the long term are possible. The added

bias is called the ridge regression penalty. It is computed by finding the product of lambda and the squared weight associated with each feature. The imposition of a penalty based on the size of the coefficients deals with some of the issues of Ordinary Least Squares. Ridge regression addresses some of the problems of Ordinary Least Squares by imposing a penalty on the size of the coefficients. A penalized residual sum of squares is minimized by the ridge coefficients:

$$\min_{w} \|Xw - y\|_2^2 + \alpha\|w\|_2^2. \tag{9}$$

$\alpha \geq 0$ is the complexity parameter that regulates the shrinkage. The greater the value of α, the larger the shrinkage, and therefore, the coefficients turn out to be increasingly robust to collinearity.

4. Results and Discussion

In this study, we present the performance analysis of the traditional machine learning models and the blending ensemble. This is done for both the Liar and ISOT datasets. Six performance measurements have been used in the comparison of the six models. These include ROC AUC, $f1$-score, AUC, precision, recall, and accuracy. The metrics are calculated for both the real and fake classes. Table 1 summarises the experimental results for the Liar dataset gauged in the detection of fake news.

The best performing base model on the Liar dataset is the logistic regression classifier which achieved the best scores for four out of six comparison metrics. These include ROC AUC, AUC, precision, and accuracy. However, overall, the blending ensemble has delivered the top performance. The four best scores obtained out of the six include ROC AUC, AUC, recall, and accuracy. Table 2 summarises the experimental results for the ISOT dataset measured in the detection of fake news.

The linear support vector machine classifier is the best performing base model on the ISOT dataset with the best scores in five out of six comparison metrics. These include ROC AUC, $f1$-score, AUC, recall, and accuracy. However, overall, the blending ensemble is the top-performing model. The four best scores attained include ROC AUC, $f1$-score, recall, and accuracy out of the possible six performance metrics.

4.1. ROC Curve.
The performance of a classification problem can be visualized or verified by using the ROC curve. The true positive rate (on the y-axis) is plotted against the false positive rate (on the x-axis). It is considered a probability curve. The area under the curve is regarded as a key metric for evaluating the model's classification performance. It measures the performance of the classification problem at different threshold settings and indicates the "degree or measure of separability" [6]. Thus, it represents the capacity

of the model in distinguishing between different classes. The higher the measurement of the area under the curve is, the better the model will be able to distinguish true news article from fake news article. The ROC curves for the Liar and ISOT datasets are shown in Figures 2 and 3, respectively. In both instances, we observe that the blending ensemble is the superior model since the area enclosed underneath the curve is the largest. On the other hand, the area under ROC curves for the LDA is the smallest. Therefore, we can conclude that the LDA model is the worst performer on both datasets. The deductions can easily be validated by verifying the respective AUC ROC scores in performance metrics tables.

4.2. Precision-Recall Curve.
The precision-recall curve for the Liar and ISOT datasets are shown in Figures 4 and 5, respectively. It is constructed by computing and then plotting the precision (on the y-axis) against the recall (on the x-axis) for each classifier at various thresholds. The curve summarises the trade-off concerning the true positive rate and the positive predictive label (value) for a classification (predictive) model by consuming varied probability thresholds.

A good classifier maintains both a high recall and high precision throughout the graph and will "hug" the right upper corner in the plots below [41]. This is evident for the ISOT dataset which indicates substantially better performance by all the classifiers when compared to the Liar dataset. This observation is bolstered by the AUC scores in performance metrics tables. Once again, the blending ensemble features very strongly in the comparison plots. It has the top AUC score on the Liar dataset and the second-best score, just 0.001 behind, on the ISOT dataset.

4.3. Confusion Matrix.
A confusion matrix is used for the analysis of a machine learning model. It reflects the data in connection with the true positives, false negatives, false positives, and true negatives [42]. Figures 6 and 7 represent the confusion matrix of blending ensemble on predictions made on the test sets for the Liar and ISOT datasets, respectively.

We can make the following deductions based on the confusion matrix of the Liar dataset:

  (i) 1201 fake news articles have been correctly predicted as fake

 (ii) 2702 news articles that are true (real) have been correctly predicted as true

(iii) 1687 fake news articles have been incorrectly predicted as true

 (iv) 828 true news articles have been incorrectly predicted as fake

Similarly, we can make the following inferences based on the confusion matrix of the ISOT dataset:

  (i) 11646 fake news articles have been correctly predicted as fake

TABLE 1: Performance metrics for the Liar dataset.

	Liar dataset using n-gram					
	ROC AUC	*f*1-score	AUC	Precision	Recall	Accuracy %
Logistic regression (LR) classifier	0.634	0.673	0.666	0.616	0.742	60.346
Linear discriminant analysis (LDA) classifier	0.553	0.573	0.596	0.589	0.558	54.269
Stochastic gradient descent classifier (SGDC)	0.616	0.682	0.648	0.592	0.805	58.725
Ridge classifier (RC)	0.598	0.685	0.626	0.580	0.836	57.682
Linear support vector machine (SVM) classifier	0.609	0.682	0.642	0.589	0.810	58.523
Blending (BLD) ensemble	**0.634**	**0.682**	**0.668**	**0.616**	**0.765**	**60.813**

TABLE 2: Performance metrics for the ISOT dataset.

|	ISOT dataset using n-gram					
	ROC AUC	*f*1-score	AUC	Precision	Recall	Accuracy %
Logistic regression (LR) classifier	0.998	0.979	0.997	0.981	0.978	98.036
Linear discriminant analysis (LDA) classifier	0.997	0.977	0.997	0.986	0.969	97.875
Stochastic gradient descent classifier (SGDC)	0.998	0.982	0.997	0.980	0.983	98.254
Ridge classifier (RC)	0.998	0.980	0.997	0.979	0.981	98.071
Linear support vector machine (SVM) classifier	0.998	0.983	0.998	0.983	0.983	98.387
Blending (BLD) ensemble	**0.998**	**0.984**	**0.997**	**0.984**	**0.984**	**98.481**
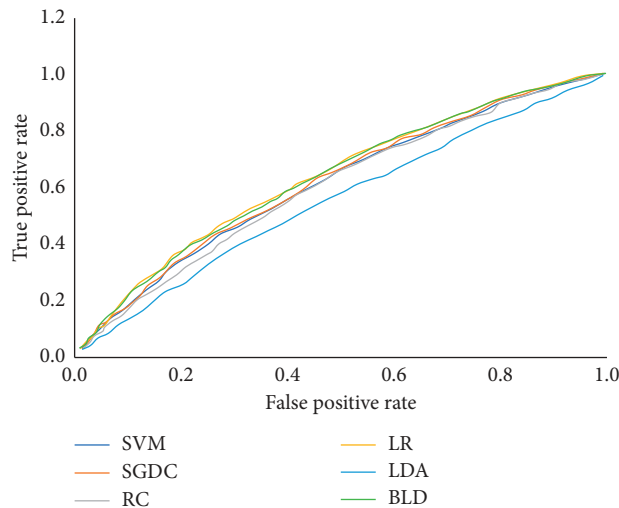
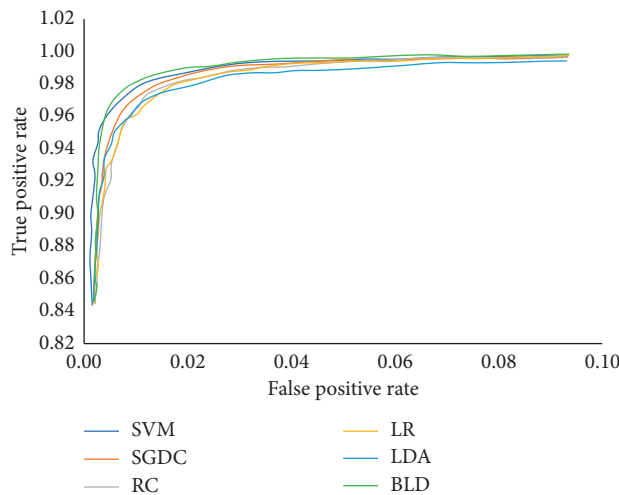FIGURE 2: Receiver operating curve (ROC) curve for the Liar dataset.

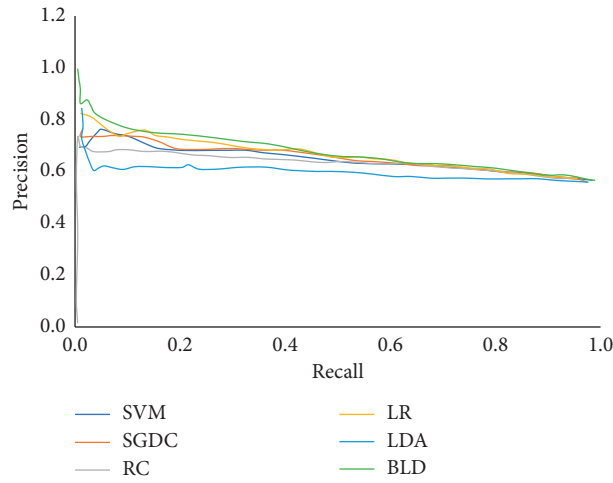FIGURE 3: Receiver operating curve (ROC) curve for the ISOT dataset.

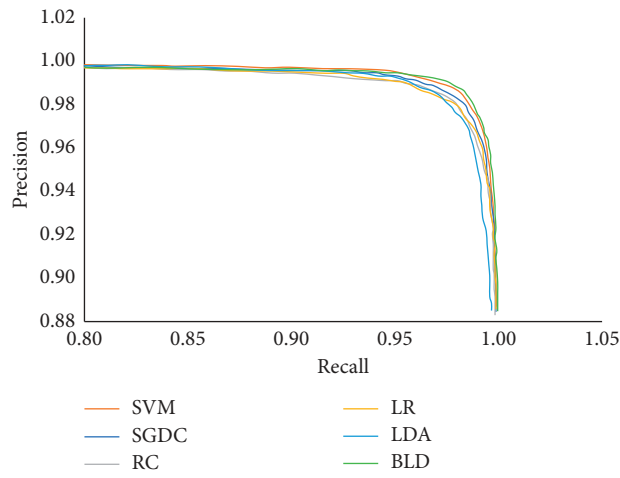Figure 4: Precision-recall curve for Liar dataset.

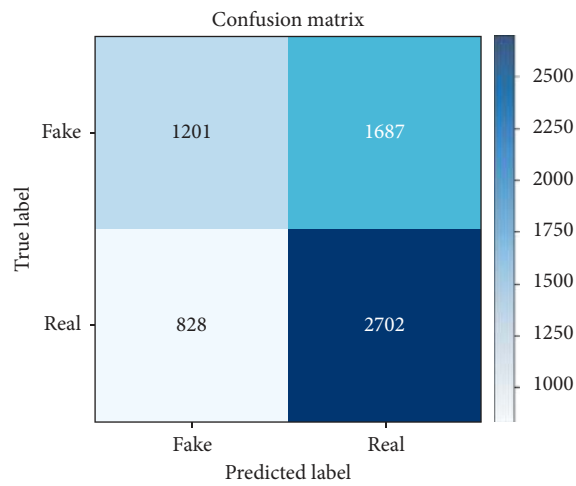Figure 5: Precision-recall curve for ISOT dataset.

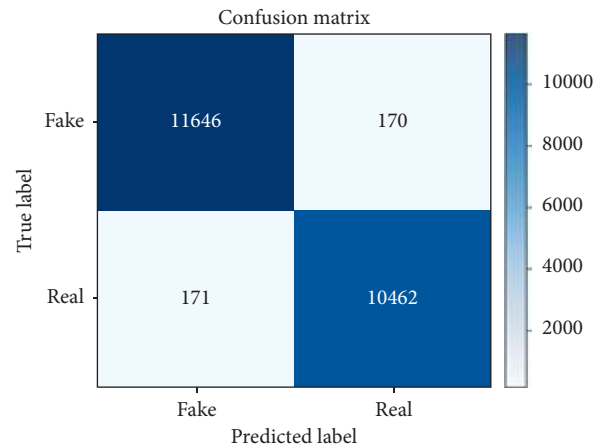Figure 6: Confusion matrix of the Liar dataset.

Figure 7: Confusion matrix of the ISOT dataset.

(ii) 10462 news articles that are true (real) have been correctly predicted as true

(iii) 170 fake news articles have been incorrectly predicted as true

(iv) 17 true news articles have been incorrectly predicted as fake

5. Conclusions

The paper has presented the application of six machine learning models according to TF-IDF vectors as features (n-gram level TF-IDF) for the goal of discovering fake news. By building the classifiers and conducting the experiments, we can conclude that the blending ensemble is the best performing model on the Liar and ISOT datasets. This has been validated by employing a variety of metrics to measure performance. This model is saved and will be used for prediction. The results of the blending ensemble are compared favourably with other studies. It is a performance improvement compared to many of the ensemble models used by Ahmad et al. [30] on the ISOT dataset including the two benchmark models Wang-CNN and Wang-Bi-LSTM. We presented a blending ensemble model for detecting fake news as a solution to improve the current approaches. Our plans include experimenting with other and larger datasets and varying the type, combination, and number of base models for the ensemble. We will also consider examining current trends on social media connected to fake news to incorporate them in our model for detection. However, the associated limitations are that the data are often inconsistent, thus adding to the mistakes or anomalies of the prediction model.

Data Availability

All the data are available from the authors.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] U. Saeed, H. Fahim, and F. Shirazi, *Profiling Fake News Spreaders on Twitter*, Notebook for PAN at CLEF, Clerkenwell, UK, 2020.

[2] X. Zhou, R. Zafarani, K. Shu, and H. Liu, "Fake news: fundamental theories, detection strategies and challenges," in *Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, Melbourne, Australia, January 2019.

[3] A. Roy, K. Basak, A. Ekbal, and P. Bhattacharyya, "A deep ensemble framework for fake news detection and classification," 2018, https://arxiv.org/abs/1811.04670.

[4] D. de Beer and M. Matthee, "Approaches to identify fake news: a systematic literature review," in *Proceedings of the International Conference on Integrated Science*, Kep, Cambodia, May 2020.

[5] V. L. Rubin, Y. Chen, and N. K. Conroy, "Deception detection for news: three types of fakes," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.

[6] Y. Shukla, N. Yadav, and A. Hari, "An unique approach for detection of fake news using machine learning," *Proceedings of the International Journal for Research in Applied Science and Engineering Technology*, vol. 7, no. 6, pp. 491–496, 2019.

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

[8] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Information Sciences*, vol. 497, pp. 38–55, 2019.

[9] W. Y. Wang, ""Liar, liar pants on fire": a new benchmark dataset for fake news detection," 2017, https://arxiv.org/abs/1705.00648.

[10] J. Y. Khan, M. Khondaker, T. Islam, A. Iqbal, and S. Afroz, "A benchmark study on machine learning methods for fake news detection," 2019, https://arxiv.org/abs/1905.04749.

[11] L. Gutierrez-Espinoza, F. Abri, A. S. Namin, K. S. Jones, and D. R. Sears, "Fake reviews detection through ensemble learning," 2020, https://arxiv.org/abs/2006.07912.

[12] G. Pennycook, J. A. Cheyne, N. Barr, D. J. Koehler, and J. A. Fugelsang, "On the reception and detection of pseudo-profound bullshit," *Judgment and Decision Making*, vol. 10, pp. 549–563, 2015.

[13] 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 2nd Workshop on Computational Approaches to Deception Detection*, San Diego, CA, USA, June 2016.

[14] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, "Some like it hoax: automated fake news detection in social networks," 2017, https://arxiv.org/abs/1704.07506.

[15] T. U. Yun and H. Ahn, "Fake news detection for Korean news using text mining and machine learning techniques," *Journal of Information Technology Applications and Management*, vol. 25, pp. 19–32, 2018.

[16] S. Aphiwongsophon and P. Chongstitvatana, "Detecting fake news with machine learning method," in *Proceedings of the 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 18–21, Chiang Rai, Thailand, July 2018.

[17] S. Gilda, "Evaluating machine learning algorithms for fake news detection," in *Proceedings of the 2017 IEEE 15th Student Conference on Research and Development (SCOReD)*, pp. 13-14, Putrajaya, Malaysia, December 2017.

[18] 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. 25–27, Vancouver, Canada, October 2017.

[19] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: analyzing language in fake news and political fact-checking," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 7–11, Copenhagen, Denmark, September 2017.

[20] S. Singhania, N. Fernandez, and S. Rao, "3HAN—a deep neural network for fake news detection," in *Proceedings of the International Conference on Neural Information Processing*, Guangzhou, China, November 2017.

[21] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake news detection: a deep learning approach," *SMU Data Science Review*, vol. 1, p. 10, 2018.

[22] N. Ruchansky, S. Seo, and Y. C. Liu, "A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 6–10, Singapore, November 2017.

[23] O. Ajao, D. Bhowmik, and S. Zargari, "Fake news identification on twitter with hybrid CNN and RNN models," in *Proceedings of the 9th International Conference on Social Media and Society*, pp. 18–20, Copenhagen, Denmark, July 2018.

[24] Y. J. Lu and C. T. Li, "GCAN: graph-aware co-attention networks for explainable fake news detection on social media," 2020, https://arxiv.org/abs/2004.11648.

[25] S. Vijayaraghavan, Y. Wang, Z. Guo et al., "Fake news detection with different models," 2020, https://arxiv.org/abs/2003.04978.

[26] P. B. P. Reddy, P. K. Reddy, G. V. M. Reddy, and K. Mehata, "Fake data analysis and detection using ensembled hybrid algorithm," in *Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication*, pp. 27–29, Erode, India, March 2019.

[27] A. Agarwal and A. Dixit, "Fake news detection: an ensemble learning approach," in *Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems*, pp. 13–15, Madurai, India, May 2020.

[28] M. Granik, V. Mesyura, and A. Yarovyi, "Determining fake statements made by public figures by means of artificial intelligence," in *Proceedings of the 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies*, Lviv, Ukraine, September 2018.

[29] H. S. Al-Ash, M. F. Putri, P. Mursanto, and A. Bustamam, "Ensemble learning approach on Indonesian fake news classification," in *Proceedings of the 3rd International Conference on Informatics and Computational Sciences*, pp. 29-30, Semarang, Indonesia, October 2019.

[30] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," *Complexity*, vol. 2020, Article ID 8885861, 11 pages, 2020.

[31] A. Mahabub, "A robust technique of fake news detection using ensemble voting classifier and comparison with other classifiers," *SN Applied Sciences*, vol. 2, pp. 1–9, 2020.

[32] S. Kaur, P. Kumar, and P. Kumaraguru, "Automating fake news detection system using multi-level voting model," *Soft Computing*, vol. 24, no. 12, pp. 9049–9069, 2020.

[33] S. Li, K. Ma, X. Niu et al., "Stacking-based ensemble learning on low dimensional features for fake news detection," in *Proceedings of the 2019 IEEE 21st International Conference on High Performance Computing and Communications*, pp. 10–12, Zhangjiajie, China, August 2019.

[34] V. Agarwal, H. P. Sultana, S. Malhotra, and A. Sarkar, "Analysis of classifiers for fake news detection," *Procedia Computer Science*, vol. 165, pp. 377–383, 2019.

[35] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," 2018, https://arxiv.org/abs/1811.00770.

[36] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Security and Privacy*, vol. 1, no. 1, p. e9, 2018.

[37] Digital Vidya. https://www.digitalvidya.com 2021.

[38] N. Smitha and R. Bharath, "Performance comparison of machine learning classifiers for fake news detection," in *Proceedings of the 2020 2nd International Conference on Inventive Research in Computing Applications*, pp. 15–17, Coimbatore, India, July 2020.

[39] K. Kowsari, M. Heidarysafa, D. E. Brown, K. J. Meimandi, and L. E. R. Barnes, "Random multimodel deep learning for classification," in *Proceedings of the 2nd International Conference on Information System and Data Mining*, Lakeland, FL, USA, April 2018.

[40] H. Garg, A. Goyal, and M. A. Joshi, "Techniques of fake news detection," *Journal of Advanced Research in Instrumentation and Control Engineering*, vol. 7, pp. 8–11, 2020.

[41] N. Mqadi, N. Naicker, and T. Adeliyi, "A SMOTE based oversampling data-point approach to solving the credit card data imbalance problem in financial fraud detection," *International Journal of Computing and Digital Systems*, vol. 10, no. 1, pp. 277–286, 2021.

[42] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 6–13, 2020.

[43] P. Mohale and W. S. Leung, "Fake news detection using ensemble machine learning," in *Proceedings of the 18th European Conference on Cyber Warfare and Security*, Coimbra, Portugal, July 2019.