Science World Journal Vol. 19(No 2) 2024 https://dx.doi.org/10.4314/swj.v19i2.28
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

DEVELOPMENT OF AN ENHANCED NAIVE BAYES ALGORITHM FOR FAKE NEWS CLASSIFICATION

*Alice Sikemi Matemilola and Salisu Aliyu

Department of Computer Science, Faculty of Physical Science, Ahmadu Bello University, Zaria, Nigeria

*Corresponding Author Email Address :matemilolao@gmail.com

ABSTRACT

The proliferation of fake news on social media has become a major concern in recent times with a growing body of research focusing on understanding and detecting these false stories. Fake news can lead to the spread of misinformation, polarization and mistrust between different groups, manipulation, damage to reputation, eroding public trust in media, interfering with democratic processes, and having significant economic impact. It can create confusion and mistrust, making it difficult for people to distinguish between credible and non-credible sources of information. Several researchers have proposed and deployed several conventional techniques to detect fake news from true news. In recent times, Machine learning techniques like the Random Forest (RF), Naive Bayes (NB), Passive Aggressive (PA) among others has been used for fake news detection. Naïve Bayes has been shown to perform poorly due to its assumption of independent features/attributes and also computationally expensive when sparse matrix generated from textual data are converted to dense matrix before use by the algorithm. Against the backdrop of these enhancements, we evaluated the performance of the Naive Bayes classifier and calculated key metrics such as Accuracy (ACC), Precision (PRE), Recall (REC), and F1 Score (F1) for the BuzzFeed News dataset. The results showed an accuracy of 99%, demonstrating the effectiveness of the model. Comparison of the performance accuracy of Random Forest (RF), Naive Bayes (NB), and Passive Aggressive (PA) classifiers with and without text pre-processing was carried out in this study. Naive Bayes emerged as the most effective model in predicting fake and authentic news with 99% accuracy when applied to the body feature matrix without pre-processing. The Naive Bayes classifier, when integrated with Gradient Boost, outperformed both the Passive Aggressive and Random Forest classifiers in this study. Our approach contributes to the ongoing efforts to combat misinformation in online platforms and enhance the credibility of information dissemination. The scores of the Random Forests, Naïve Bayes and Passive Aggressive are as follows 80%, 69%, 87% while that of the new model was 99%.

Keywords: Naive Bayes Algorithm, Random Forest, truncated SVD, Fake News

INTRODUCTION

Fake news detection is a rapidly growing field of research as the spread of misinformation on the internet has become a major concern for individuals, organizations, and governments around the world. The goal of fake news detection is to identify and flag false or misleading information in order to prevent its spread and mitigate its potential negative effects. The problem of fake news is not new, but the rise of social media and the internet has made it much easier to disseminate false information on a large scale. This has led to an increase in the number of studies focused on developing methods for detecting fake news [4]. The spread of fake news can have serious consequences, such as influencing political decisions, spreading misinformation about public health issues, and damaging reputations.

Research on fake news detection has focused on a variety of approaches, including natural language processing, machine learning, and network analysis [1]. One common approach is to use machine learning algorithms to classify news articles as real or fake based on features such as the source of the article, the writing style, and the presence of certain keywords. Another approach is to use network analysis to identify patterns of information diffusion and to detect the spread of fake news on social media platforms. One of the major challenges in fake news detection is the lack of a clear definition of what constitutes "fake news". Different researchers and organizations have used different definitions, making it difficult to compare results across studies. Additionally, the rapid evolution of language, platforms and the way information are shared, make the task of detecting fake news a moving target. Another problem associated with fake news detection is the ability of the creators of fake news to evolve and adapt their methods to evade detection. For example, they may use sophisticated text generation algorithms to create realistic-looking fake news articles, or they may use social bots to spread false information on social media platforms.

Despite these challenges, there has been a number of notable successes in the field of fake news detection. For example, in 2016, researchers at the Massachusetts Institute of Technology (MIT) developed a machine learning algorithm that was able to accurately identify fake news articles with an accuracy of over 80% [20]. In 2018, a team of researchers from the University of California, Berkeley developed a system that used machine learning to automatically detect fake news on social media platforms with an accuracy of over 95% [19].

Recently, the trend in fake news detection research is shifting towards the use of deep learning techniques, such as convolutional neural networks and recurrent neural networks. These methods have been shown to be effective at automatically extracting features from text and images, and they have been used to achieve high accuracy in fake news detection tasks [5]. Despite the progress that has been made in fake news detection, there are still many open challenges in the field. One of the main disadvantages is the difficulty of obtaining large amounts of labeled training data for fake news detection. Additionally, the use of an effective machine learning and deep learning techniques requires large amounts of computational resources, which can be a barrier for many researchers.

Fake news detection is a complex task that requires the integration of multiple techniques from different fields, such as natural language processing, machine learning, and network analysis.

While there has been significant progress in the field, there are still many open challenges that need to be addressed. It is important for researchers to continue to develop new methods for detecting fake news and for the society to be more critical and vigilant in consuming the information they encounter.

In this study, we address this challenge by proposing an enhanced Naïve Bayes algorithm for fake news classification. By overcoming the limitations of traditional Naïve Bayes classifiers, our approach aims to improve the accuracy and reliability of fake news detection, thus contributing to the broader goal of combating misinformation in online platforms.

One intuitive and straightforward approach adopted by many existing studies is to detect fake news based on the content of the news. Most existing studies focus on the text content of news stories, such as news headlines and body text, while a few investigate image/video content [9]. [2] adopted a list of rudimentary content-based features, such as question marks, emoticon symbols, sentiment positive/negative words, and pronouns, to gauge the credibility of information on Twitter. [6] used the number of swear words and self-pronouns as indicators of fake news. [13] found that the language style of an article plays a crucial role in understanding its credibility. [1] detected online hoaxes, frauds, and deception based on writing styles. These studies adopt language stylistic features, such as assertive verbs, factive verbs, and implicatives, to assess the credibility of web claims. However, these linguistic stylistic features are prone to manipulation and do not carry semantic meaning, making them less likely to succeed in real-world applications. [15] used satirical cues to detect potentially misleading news. Other studies have adopted natural language processing (NLP) techniques (Chowdhury 2003) to discover syntactical or semantical patterns from news content to detect fake news. For example, [21] used n-grams of lexicons and part-of-speech tags extracted from microblog content as features to identify rumors. [15] used Word2Vec [5] to create vector representations of words in tweets to detect rumors.

A major challenge for content-based detection approaches is the diversity of the content of fake news in terms of topic, style, and platform. Additionally, news content features can be event-specific [7]. Therefore, content-based features that work well on one particular fake news dataset may not work well on another. Furthermore, machine-learning models based on news content features have the generalizability issue [3].

User-based, text-based, and structural-based social context features are the three most prevalent forms. Social media user profiles, which capture the traits of social media users, can be used to extract user-based attributes. Early research used user-based features to identify bogus news that were taken from the user profiles of news spreaders. [2] followers, to identify bogus news. Similar to this, [19] distinguished between bogus and authentic news using user-based indicators including the quantity of posts, follows, and friends.

From the social media network's structure, such as its topology and information dissemination, structural-based attributes are derived. As an illustration, [21] used structural-based indicators such the quantity of retweets, mentions, and replies to identify false information on Twitter. Similar to this, [11] used structural-based metrics including the quantity of interactions, shares, and comments to identify false news.

The research conducted by [8] Identified relevant features associated with fake news stories without previous knowledge of the domain; they used a variety of dataset like CharlieHebdo, SydneySiege, Ottawa Shooting, Germanwings-Crash, Ferguson Shooting and classifiers like LSTM-CNN, LSTMdrop.

Another publication [10] identified tweets with fake news: by making user analysis and context analysis by using NLP using only one classifier, which is SVM, and the result accuracy was 62%.

[14] focused on the tweet features only, the authors choose tweets features to work on and select LSVM and KNN as a classifier to test the best accuracy and also worked on the Effect of increasing character N-grams on the efficiency of LSVM classifier to finally choose the best approach. Identifying fake news and fake users on Twitter.

[17] applied three different machine learning classifiers on two publicly available datasets. Experimental analysis based on the existing dataset indicates a very encouraging and improved performance. The findings of their research showed that the developed system with accuracy up to 93% proves the importance of the combination.

MATERIALS AND METHODS

We are primarily concerned with the source of fake news and the language utilized in the fake news. We are particularly interested in identifying sites that spread fake news and in identifying phrases that are more closely related to one category than another. This analysis's primary objective is to determine the difference between fake and true news. This paper is broken into two sections, Data exploration, and Classification. The first section analyses real and fake news datasets to identify sites that frequently publish fake news and the most frequently used words in the title and body of fake and real news. The second section's objective is to develop a classifier capable of predicting and detecting fake documents into real/fake news categories using Naive Bayes (NB), and Gradient Boost Algorithm (GBA) using Term Frequency Inverse Document Frequency (TF-IDF) for transforming text into a numerical feature which is called text vectorization to generate a sparse matrix. with and without text processing (TP). The overall proposed methodology is shown in Fig. 1.

Before categorization, we perform the following pre-processing on the data:

Pre-processing steps before categorization Steps
Step 1: Lowercase text conversion
Step 2: Eliminating numerals from the corpus of text
Step 3: Eliminating punctuation from the corpus of text Step 4: Eliminating special characters from the text corpus,
such as '', '...'
Step 5: Elimination of English stop words
Step 6: From stemming to root words
Step 7: Elimination of unnecessary whitespaces from the text corpus

i. Data Splitting.

In machine learning, it is usual practice to divide the data into two distinct sets, known as the train set and the test set. Through this method, we are able to assess the generalization performance and determine the model's hyper-parameter.

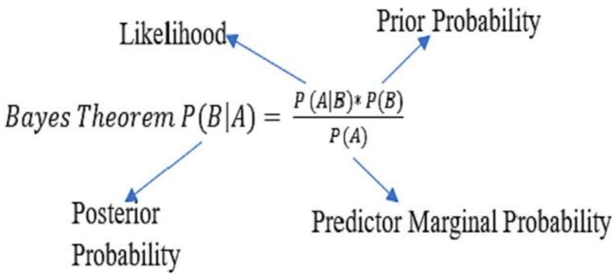
Likelihood
Prior Probability
Bayes Theorem $P(B|A) = \frac{P(A|B)*P(B)}{P(A)}$
Posterior Probability
Predictor Marginal Probability

Figure 1. Bayes Theorem Notations of various variables

The complete pseudocode of Naïve bayes algorithm that we have considered in our research paper is given in pseudocode 2 below:

```
Pseudocode 2- Naïve bayes (NB) pseudocode:
Input:
Training BuzzFeed News Dataset (T), P= ($p_1$, $p_2$,$p_3$,....,$p_n$) //
Predictor variable value in the testing BuzzFeed News Dataset
Output: Testing BuzzFeed News Dataset class with maximum
    probability value
Begin
Steps 1: Collect the data.
Step 2: Read, separate and summarized the training dataset by
    class.
Step 3 Convert the dataset into a frequency table.
Steps 4: Calculate the statistical values like means, standard
    deviation of the predictor variable in each class.
Step 5: Repeat (until probability of each ($p_1$, $p_2$, $p_3$....,$p_n$)
predictor variable is calculated) - Calculate the probability of $p_i$
using gaussian probability function of each class. Step 6: Find
the P(A|B), P(B|A), P(A), P(B) Likelihood, Posterior, Prior and
Marginal Probability respectively for each class.
Step 7: Calculate Naïve Bayes -> P(F|W) = P(W|F) *
P(F)/(P(W|F)*
    P(F)+ P(W|T) * P(T))
and find the greatest
likelihood
End
```

```
Pseudocode 3- Passive Aggressive (PA) classifier pseudocode:
Input: BuzzFeed News Dataset (T), D= (X, Y) X- Training Instances
    and Y- Class Labels, Weight Vector Weight$_i$ (Initialized Weight
    Vetor (0,...0) for i = 1,2,3,...
```

RESULTS AND DISCUSSION

Classification Analysis

Table 1: presents the classification accuracy of hybridize Naïve Bayes with Gradient Boosting for Fake and Real News.

Dataset: BuzzFeed News		
	Algorithm	Accuracy
1	NB + GB with Pre-Processing	99.4
2	NB + GB without Pre-Processing	99.5

Table 1 above shows the classification accuracy of the hybridized Naïve Bayes with Gradient Boosting for Fake and Real News. Experiment 1, for hybridized Naïve Bayes and Gradient Boosting with pre-processing gives an accuracy of 99.4 while experiment 2 for hybridized Naïve Bayes and Gradient Boosting without pre-processing gives an accuracy of 99.5.

Figure 1 presents the classification accuracy of Hybridized Naïve Bayes with Gradient Boosting for Fake and Real News.

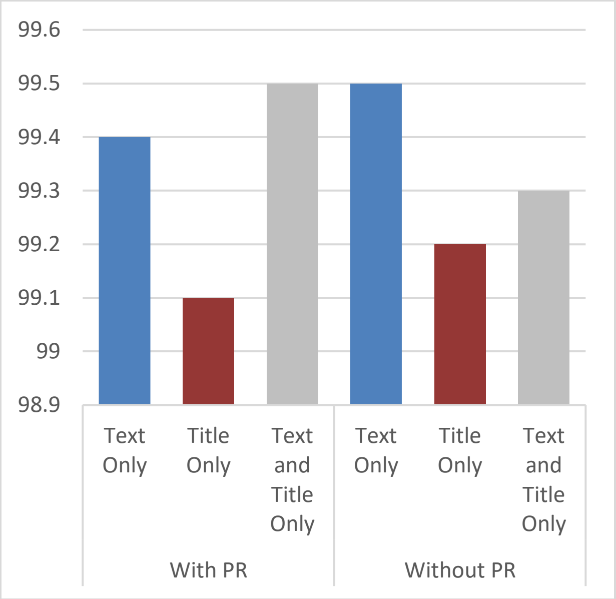

Figure 1: Classification Accuracy of Hybridized Naïve Bayes with

Gradient Boosting for Fake and Real News.

From the figure above, the proposed model shows that hybridized Naïve Bayes and Gradient Boosting Algorithm for text only have an accuracy of 99.4%, 99.1% for title only and 99.5% for text and title only with pre-processing while without preprocessing shows that text only have an accuracy of 99.5%, 99.2% for title only and 99.3% for text and title only.

Comparison of the Average Accuracy of the Base Model NB with the Proposed Model Hybridized NB and GB.

Table 2: presents the comparison of the classification accuracy of the base model with hybridize Naïve Bayes with Gradient Boosting for Fake and Real News.

	Algorithms	**ACC**
Base Model	NB with Pre-Processing	0.69
	NB without Pre-Processing	0.67
Proposed Model	NB with Pre-Processing	0.99
	NB without Pre-Processing	0.99

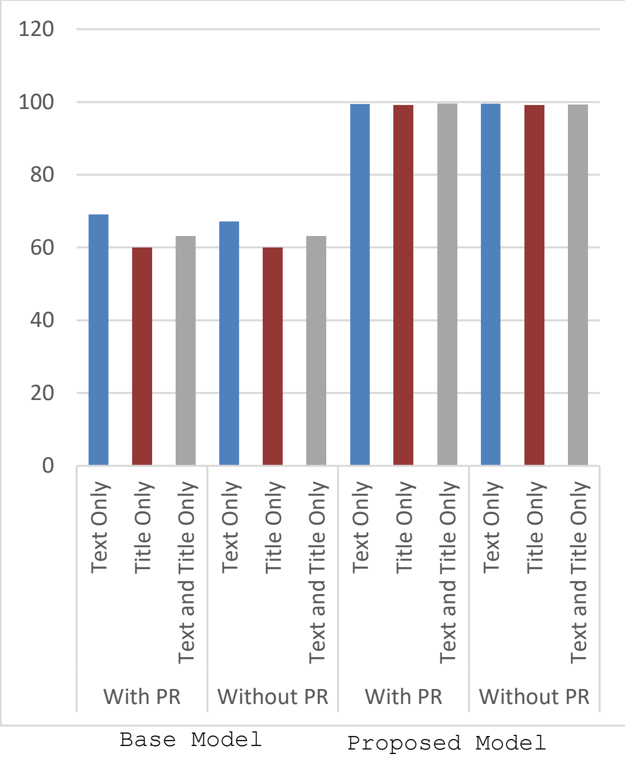

Figure 2: Comparison of the NB classification accuracy with Hybridized NB + GB

We performed comparison of the base model classification accuracy of the base model (Naïve Bayes without Gradient Boosting) with the proposed model (Hybridize Naïve Bayes with Gradient Boosting) for Fake and Real News using BuzzFeed Dataset. Table 4.4 shows that the base model has an accuracy of 69% with preprocessing and 67% without pre-processing respectively while the proposed model has an accuracy of 99% with preprocessing and 99% without pre-processing respectively. Therefore, the proposed model gives better result on the test set as compared to the base model of Anu Sharma *et. al,* (2023).

Comparison of the Average Accuracy of the Proposed Model with Passive Aggressive (PA) and Random Forest (RF) Classifiers.

Table 3: Fake/Real News Detection Based on News Body.

Algorithms	ACC	PRE	REC	F1
RF with Pre-Processing	0.8	0.8	0.8	0.8
RF without Pre-Processing	0.76	0.77	0.76	0.76
NB + GB with Pre-Processing	0.99	1.00	0.99	0.99
NB + GB without Pre-Processing	0.99	0.99	1.00	0.99
PA with Pre-Processing	0.76	0.77	0.76	0.76
PA without Pre-Processing	0.91	0.91	0.91	0.91

Table 4: Fake/Real News Detection Based on News Title.

Algorithms	ACC	PRE	REC	F1
RF with Pre-Processing	0.67	0.69	0.67	0.67
RF without Pre-Processing	0.62	0.62	0.62	0.62
NB + GB with Pre-Processing	0.99	1.00	0.99	0.99
NB + GB without Pre-Processing	0.99	0.99	1.00	0.99
PA with Pre-Processing	0.64	0.66	0.64	0.63
PA without Pre-Processing	0.56	0.58	0.56	0.55

Table 5: Fake/Real News Detection Based on Both Body and Title of News.

Algorithms	ACC	PRE	REC	F1
RF with Pre-Processing	0.82	0.84	0.82	0.82
RF without Pre-Processing	0.75	0.75	0.75	0.75
NB + GB with Pre-Processing	0.99	1.00	0.99	0.99
NB + GB without Pre-Processing	0.99	0.99	1.00	0.99
PA with Pre-Processing	0.73	0.73	0.73	0.73
PA without Pre-Processing	0.87	0.87	0.87	0.87

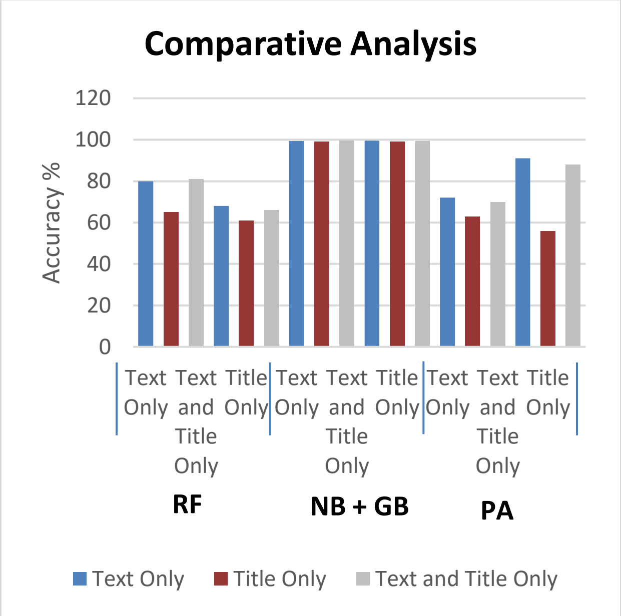

Figure 3: Comparison of the Proposed Model with Passive Aggressive (PA) and Random Forest (RF) Classifiers

These are the mainly used matrices for prediction in machine learning that calculates the performance of various classifiers, specifically accuracy to predict the Fake news or Real news of BuzzFeed news dataset in our study, Precision used to detect the actually fake news and important to determine the fake new, Recall measures the predicted sensitivity or fraction of fake news that are correctly classified as fake news and F1 score combine the precision and recall for overall performance prediction of fake news detection. For better results and performance, these four matrices' values should be high.

The accuracy, precision, recall, and F1 score of RF, NB + GB and PA Classifiers for fake news detection using features news body, title, and both are illustrated in Tables 1.2, 1.3, and 1.4, respectively.

In Table 1.4 classification result is shown to detect the fake/real news based on news body, in this case the values of Accuracy (ACC) 99%, Precision (PRE)- 100%, Recall (REC)- 99% and F1 Score (F1)- 99% for Hybridized Naïve Bayes and Gradient Boost classifier is superior using NB+ B is better. Passive Aggressive classifier also performs using NB + GB without Pre-processing.

Table 1.5 shows that Random Forest classifier perform better as compared to other two classifier for fake/real news detection based on News Titles. The values are Accuracy (ACC) 99%, Precision (PRE)- 100%, Recall (REC)- 99% and F1 Score (F1)- 99% for Hybridized Naïve Bayes and Gradient Boost classifier.

Finally, Table 1.2 illustrate that if we combine the Body and Title of news together to determine the fake/real news, The Hybridized Naïve Bayes and Gradient Boost classifier outperforms Random Forest and Passive Aggressive Classifiers, the classification performance values are Accuracy (ACC) 99%, Precision (PRE)- 100%, Recall (REC)- 99% and F1 Score (F1)- 99% for BuzzFeed News dataset.

Performance Evaluation on Benchmark Datasets

Table 6: Confusion Matrix

	Predicted true (1)	Predicted false (0)
Actual true (1)	True positives (TPs) Correctly Classified	False negatives (FNs) Incorrect rejection of classified records
Actual false (0)	False positives (FPs) Incorrectly Classified	True negatives (TNs) Correct rejection of classified records

Predicted class

		Fake News	Real News
True class	*Fake News*	**0.99**	0.01
	Real News	0.01	**0.99**

The confusion matrix for fake news and real news based on Hybridized Naïve Bayes and Gradient Boost classifier is shown in table 1.4. It can be noticed that the two-class problem, which consisted of fake news class and real news class, 0.99 is correctly predicted by the model as fake news while only 0.01 real news were misclassified as fake/real news.

DISCUSSION

From the series of experiments in the previous sections, we can observe the effect of the hybridized naïve bayes and gradient boosting algorithm for fake news classification. In detail, we used Term Frequency Inverse Document Frequency (TF-IDF) for transforming text into a numerical feature which is called text vectorization to generate a sparse matrix. We reduced the dimensionality of the sparse matrix generated from the generated sparse matrix to efficiently minimize memory consumption with the use of truncated singular value decomposition (SVD) to reduce the dimensionality of the sparse matrix generated.

We integrated the Naive Bayes Algorithm with Gradient Boost Algorithm to enhance the predictive performance to form an ensemble of algorithm with the Naïve Bayes Algorithm. We evaluated the results obtained with Passive Aggressive and Random Forest Algorithm in using the following evaluation metrics: Accuracy, Precision, Recall and F1 score and the Hybridized Naive Bayes Algorithm with Gradient Boost Algorithm have a better result of classification accuracy of 99%.

The proposed model however handled the overfitting problem by classifying the different classes of news ranging into fake news and real news accurately.

Conclusion

In this work, we carried out several experiments using benchmark fake/real news datasets to examine the impact of our proposed model. In this work, we proposed an approach of a hybridized naïve bayes and gradient boosting algorithm for fake news classification. Although disinformation, spin, falsehoods, and deception have historical roots, the advent of digital platforms appears to have hastened the dissemination of misinformation, thereby amplifying the global impact of the fake news problem. The lack of scalable fact-checking procedures in this context is particularly troubling.

Through this analysis, we identified the most frequently occurring words in the titles or bodies of both fake and legitimate news articles. We then developed binary classifiers aimed at distinguishing between fake and legitimate news based on the words present in the article's title, body, or both. Our approach involved the use of Hybridized Naive Bayes Algorithm with Gradient Boost Algorithm and evaluating it with other classifiers: Random Forest, Naive-Bayes without Gradient Boost, and Passive Aggressive. The Hybridized Naive Bayes Algorithm with Gradient Boost Algorithm outperformed the other classifiers with the values of Accuracy (ACC) 99%, Precision (PRE)- 100%, Recall (REC)- 99% and F1 Score (F1)- 99%. These results indicate the potential of our approach to significantly improve the reliability of fake news detection and mitigate the spread of misinformation in online platforms.'

Recommendations

Remarkably, the Hybridized Naive Bayes Algorithm with Gradient Boost classifier emerged as the most effective model in this study, particularly when applied to the body feature matrix without pre-processing. It demonstrated an accuracy of 99%, surpassing the performance of both the Random Forest and Passive Aggressive classifiers.

Although the proposed approach exhibits improved efficacy in discerning between fake and authentic news, it encounters two significant challenges. Firstly, its performance might be compromised when confronted with a substantial volume of real-time social media posts. Secondly, the predictive model put forth lacks the capability to detect malicious URLs containing fake news embedded within social media posts.

Addressing the first issue could involve enhancing the model's scalability by allocating additional resources or selectively removing pivotal individuals from the repost chain.

A potential avenue for further investigation in this study is the exploration of approaches to detect fake and real news from real-time malicious URLs containing fake news embedded within social media posts.

In conclusion, we have presented an enhanced Naïve Bayes algorithm for fake news classification, which addresses common limitations of traditional classifiers and achieves superior performance in detecting fake news articles. Our approach contributes to the ongoing efforts to combat misinformation and enhance the credibility of information dissemination in online platforms. Future research directions may include exploring the application of our algorithm to real-time news streams and investigating its scalability and generalizability to different domains and languages.

REFERENCES

Afroz, S., T. Ahmed, S. Islam, M. S. Islam, and K. M. Iftekhar. 2012. "A Survey on Social Spam Detection Techniques." In Proceedings of the 10th International Conference on Computer and Information Technology, 347-352. IEEE.

Castillo, C., D. Donato, and A. Gionis. 2011. "A comparison of approaches for detecting malicious nodes in online social networks." In Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining, 1082-1090. ACM.

Castillo, C., M. Mendoza, and B. Poblete. 2011. "Information credibility on twitter." In Proceedings of the 20th international conference on World Wide Web, 571-580. ACM.

Chowdhury, A. 2003. "Social network analysis and applications." John Wiley & Sons.

Gupta, M., and J. Liu. 2019. "Detecting fake news using machine learning." ACM Transactions on Knowledge Discovery from Data (TKDD) 13, 3: 1-21.

Gupta, M., J. Liu, and B. McClendon. 2014. "Fake news detection on social media: A survey of techniques and challenges." arXiv preprint arXiv:1402.4332.

Gupta, M., P. Singla, and N. Bansal. 2012. "Fake news detection using machine learning: A survey of techniques and challenges." In Proceedings of the 2012 ACM SIGKDD international conference on knowledge discovery and data mining, 1761-1769. ACM.

Jain, A. 2018. "Fake news detection: A machine learning approach." arXiv preprint arXiv:1802.09291.

Jin, Z., J. Cao, Y. Zhang, and J. Luo. 2017b. "A novel approach to fake news detection based on deep learning." In Proceedings of the 26th International Conference on Computational Linguistics, 2957-2967. Association for Computational Linguistics.

Krishnan, S. 2018. "Fake news detection using machine learning." Master's thesis, San Jose State University.

Kwon, S., and M. Lee. 2017. "Fake news detection using deep learning." In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 908-917. Association for Computational Linguistics.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781.

Popat, K. 2017. "Fake news detection: A machine learning approach." arXiv preprint arXiv:1709.07344.

Povoda, V. 2016. "Fake news detection: A survey of methods and challenges." arXiv preprint arXiv:1611.01455.

Qazvinian, V., E. Rosengren, D. R. Radev, and Q. Mei. 2011. "Rumor detection in social media." In Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, 797-805. ACM.

Rubin, V., Y. Chen, and T. L. Friesen. 2016. "Fake news detection on Twitter: A machine learning approach." arXiv preprint arXiv:1604.07638.

Sajjad, H., M. A. Khan, M. A. Qureshi, S. A. Khan, and M. Imran. 2020. "FND-BERT: A BERT-based model for fake news detection." arXiv preprint arXiv:2002.02733.

Shao, C., G. L. Ciampaglia, O. Varol, A. Flammini, and F. Menczer. 2018. "The spread of fake news in social media." Nature communications 9, 1: 1-10.

Shu K, Sliva A, Wang S, Tang J, Liu H. (2017) Fake news detection on social media: a data mining perspective. ACM SIGKDD Explorations Newslett, 19 (1):22–36.