www.ijarp.com ISSN 2456-9992 Page: 01-12

International Journal Advanced Research Publications

Review Article

Volume: 02
Issue: 04

FAKE NEWS DETECTION USING MACHINE LEARNING

*[1]Mohd Shadab,[2]Mohd Maaz, [3]Saad Jamal, [4]Mr. Aaftab Alam

[1,2,3]Research Scholar, Department of Computer Science Engineering, Integral University, Lucknow.

[4]Assistant Professor, Department of Computer Science Engineering, Integral University, Lucknow.

Article Received: 13 March 2026, Article Revised: 02 April 2026, Published on: 22 April 2026

***Corresponding Author: Mohd Shadab**

Research Scholar, Department of Computer Science Engineering, Integral University, Lucknow.

DOI: https://doi-doi.org/101555/ijarp.5081

ABSTRACT

The rapid growth of online news media, along with the use of social media such as Facebook, Twitter, and WhatsApp, has significantly impacted the global process of information dissemination. Currently, it is possible to reach millions of people in a matter of seconds, bypassing the traditional fact-checking process, which is generally required in such situations. Misinformation is spreading rapidly through the social media network. The new technologies provide opportunities for free expression, but they also provide opportunities for the rapid spread of misinformation, disinformation, and fake news. These trends have the potential to influence a variety of results, such as public opinion, the spread of disease, financial markets, etc. Considering the speed at which information is disseminated in modern society, it is important to use sophisticated methodologies, bypassing the traditional fact-checking process.

Machine Learning (ML) and Natural Language Processing (NLP) provide data-driven, automated, and scalable approaches to address this problem [1], [4], [5]. In the current study, three conventional machine learning algorithms, Support Vector Machines (SVM), Naïve Bayes (NB), and Random Forest (RF), were examined for binary classification of fake news and legitimate news. Machine learning algorithms were found to perform well in conventional fake news detection problems involving text data [4], [5], [6]. The data set contains 44,898 labeled data points, which is sufficient to develop an effective model. A range of natural language processing operations were performed during data preprocessing, which were found to be effective for text data classification, as reported by previous studies [1], [4].

In order to measure the efficiency of the models, various parameters were used, out of which the most important was the accuracy level of the models. Among the models used, the highest accuracy level was obtained by the Random Forest model, for which the accuracy level was 99.77%. The next highest accuracy level was obtained by the Support Vector Machine model, for which the accuracy level was 99.56%. The Naive Bayes model had an accuracy level of 93.16%. The high level of accuracy obtained by the first two models in identifying fake news is due to the efficiency of the SVM model in making use of the TF-IDF feature in a high-dimensional space, in addition to its ability to generate results with minimal overfitting, making it more efficient than the others, as supported by the results for the efficiency of the models used in the study, as provided in references [4], [5].

The appropriateness of the support vector machine model was further demonstrated in relation to its deployment in the problem domain. The results further confirm the viability of traditional machine learning approaches in the classification of fake news, especially in consideration of emerging trends in the deployment of deep learning models in this classification problem [2], [5].

KEYWORDS: Fake News Classification, Machine Learning, Natural Language Processing, TF-IDF, Support Vector Machine, Random Forest, Naïve Bayes, Text Classification.

INTRODUCTION

The world seems to be accelerating, and some people are adamant that there is no way to slow it down. The digital world has turned the way we consume information upside down. In the past, news was the domain of professionals who reviewed and fact- checked what was published. Today, information dissemination through the internet happens in real time, whereby the information has the ability to go viral in a matter of minutes in the absence of editorial filters, which were traditionally associated with conventional news media [1], [2].

This freedom, however, also leads to the creation of a large amount of false information, commonly referred to as fake news. Fake news is described as information intended to be false but appearing to be true, where the intention of the creator of the information is to make it appear as if it is coming from a credible source [2]. The reasons for creating fake news are many, ranging from political ideologies to financial gain through click-based advertisements, as well as the intention to cause social unrest [1], [2].

In situations where there is a high incidence of certain events, such as elections or the spread of a disease, misinformation campaigns can significantly impact public perception and

behavior. Social media bots and groups help to quickly spread misinformation, enabling it to reach a large audience in a short period of time.

Machine learning is an scalable technique to fight misinformation by analyzing the linguistic characteristics of articles, such as the choice of words, sentence formation, and the use of certain words, to identify legitimate news and fake news.

The objective of the present study is to create an optimal framework for the classification of fake news through the application of natural language processing techniques for data cleansing and preprocessing of extensive datasets for the application of machine learning techniques. The methodology used in the present study is the application of data preprocessing techniques, which are used for the reformation of the data in a suitable format for the application of machine learning techniques. Further, TF-IDF vectorization is used for the conversion of the data into a suitable format for the application of machine learning techniques. Such a conversion is commonly used in the application of machine learning techniques in the existing literature on the classification of text data [1], [4].

LITERATURE REVIEW

Traditional Machine Learning Techniques

The classification of text data is a characteristic example of a high-dimensional and sparse feature space problem, primarily because of the TF-IDF transformation used in the traditional machine learning techniques used for the classification of text data. The TF-IDF transformation considers words in the data set as features of the data set [1], [4]. Therefore, in the context of traditional machine learning techniques used for the classification of text data, the number of features is in the range of thousands for a classification problem. The sparse features of the data set create a number of computational and algorithmic problems. Nevertheless, traditional machine learning techniques such as Naïve Bayes, Support Vector Machines (SVM), and boosting techniques are effective in the application of text mining and the classification of fake news [4], [5], [6].

Of the methods explored, the effectiveness of the Support Vector Machines (SVM) in pattern recognition, thereby enabling the separation of classes when the data is linearly separable in a high- dimensional TF-IDF feature space, is noteworthy. A support vector machine classifier works by determining the optimal hyperplane that maximizes the margin, where the margin is the distance between the data points of the respective classes and the hyperplane. Additionally, support vector machines have been found to be marginally immune to

overfitting compared to other learning algorithms that make use of the entire set of training data, an aspect that is important when the data is scarce.

Probabilistic Models

The Naïve Bayes classifier, based on Bayes' theorem, provides a probabilistic approach for classification problems. This method uses the information provided by each feature individually to calculate the posterior probabilities of whether a specific instance belongs to a certain class. Naïve Bayes assumes conditional independence between features. Although this assumption is not valid for natural language processing, it has been noted that this assumption does not affect the performance of Naïve Bayes for text classification problems [4], [7].

Dataset Description

The current research uses a corpus of 44,898 news articles, out of which 23,481 articles are categorized as Fake and 21,417 as True. This demonstrates the actual occurrence of fake and true news articles in the given corpus. The existing corpora, LIAR and FEVER, have been widely used in fake news- related research for the development of various supervised learning-based approaches for fake news detection, as discussed in [12] and [13]. In the case of binary classification, fake news articles are represented as 0 and true news articles as 1.

For the performance evaluation of the proposed model, the entire dataset was divided equally into a training set of 35,918 articles and a test set of 8,980 articles.

Ensemble Methods

Random Forest is an ensemble learning technique that combines the results of multiple predictions made by decision trees, where the trees work independently and the results are combined through a voting process that is facilitated by bagging. The main reason for the application of ensemble learning is to improve the stability of results by reducing the variance compared to that of a single decision tree, as explained in references [5] and [6]. In this context, the results of individual trees in the ensemble learning technique are combined to form an ensemble that has minimal variance, reducing the effect of random fluctuations that may be experienced in individual trees. One of the major benefits of the Random Forest algorithm is its ability to deal with non-linear problems, which include textual features and class labels. The algorithm is robust and can achieve a high level of accuracy even with noisy data, but this comes at the cost of increased complexity.

Deep Learning Techniques

From the current trends, it is evident that significant improvement has been made in the design of deep learning architectures, with more focus on the implementation of the transformer architectures as recommended by OpenAI, Google, among other organizations. The most important advantage of the transformer models is the improved ability to learn contextual information associated with words. Despite the design of the transformer architectures, some problems have been noted in the application of the models, which may limit their use [2], [9].

Gaps in the Research

Despite the significant amount of research and literature that has been conducted so far in the context of fake news detection, there are a number of issues that need to be addressed in the context of the topic. Firstly, it is important to understand that the high dimensionality of TF-IDF features results in high computational costs, and it is also important to understand that the best mapping between the features and the classification results may not always be achieved, especially in the context of overfitting. A balanced model that can achieve high accuracy, scalability, interpretability, and efficiency is an important direction for future research.

Stratified sampling has been used for the data partitioning process, ensuring that the proportion of fake and real news is preserved in the training and testing sets.

FEATURE ENGINEERING

The TF-IDF transformation is utilized for measuring the level of significance of a particular term in a given document in comparison to the total corpus size. The TF-IDF transformation of a term t in a document d is given by:

TF-IDF(t, d) = TF(t, d) * log(N/DF(t))

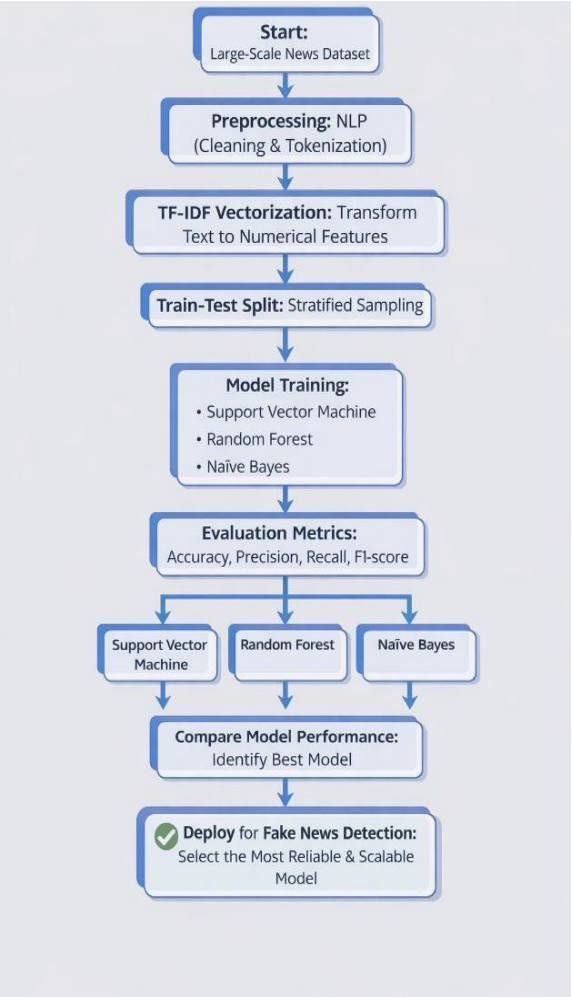
where TF(t, d) refers to the term frequency of term t in document d, and DF(t) refers to the document frequency of term t. It has been established that a direct relationship exists between term frequency and the overall significance of term t in document

d. It has been recognized that DF(t) plays a vital role in determining the overall significance of term t in a given document.

If N refers to the total number of documents in the set of all documents, it has been established that the logarithmic factor in the TF-IDF transformation log(N/DF(t)) has a substantial impact in reducing the overall significance of term t in document d when DF(t) is

sufficiently large. The TF-IDF transformation combines term frequency and term rarity and is therefore a highly effective transformation for fake news detection through the use of traditional machine learning approaches.

MODEL ARCHITECTURES



1. Naïve Bayes

An accuracy of 93.16% was attained by the Naïve Bayes classifier, which shows that it performed very well. Precision values were found to range from 93% to 94%, while recall values ranged from 92% to 94%. These values show that there is a good balance in classification performance. Bayes' theorem is the foundation of the Naïve Bayes classifier, which is given by:

$$P(C|X) = [P(X|C)P(C)]/P(X)$$

Here, it is possible to find the probability of class C given the set of features X. It is known that the Naïve Bayes classifier is computationally efficient and is best used for high-dimensional text classification [4], [6]. However, it is found that the main disadvantage of the Naïve Bayes classifier is that it assumes independence of features, which is not possible in natural language processing because of the relationships that exist among words.

2. Random Forest

It was found that the accuracy of the classifier was maximum at 99.77%. Precision, recall, and F1 score were found to be 100%. The prediction is given by the following equation:

ŷ = majority_vote(T1, T2, ..., Tn)

It is seen that the result of prediction is derived by performing majority voting over multiple decision trees. Such a method is able to handle nonlinear relationships and is highly robust with noisy data [5], [6]. However, it is seen that there is a high computational cost of performing the voting operation over multiple decision trees and high- dimensional features [5].

3. Support Vector Machine

The accuracy of the Support Vector Machine (SVM) classifier was determined by the following expression:

Accuracy = (4679 + 4262) / 8980 = 8980 / (4679 +
4262) = 0.9956 = 99.56%.

The optimization problem for the SVM model is formulated by the following expressions:

Minimize (1/2) ||w||^2 + C ∑ ξ_i subject to:

y_i (w · x_i + b) ≥ 1 − ξ_i, for all i.

The ability of the SVM classifier to deal with high- dimensional and sparse text classification problems is proved in [4], [5]. This classifier has been chosen to be the final one because of its potential to provide the highest margin and improve the generalization performance, as well as its relatively low probability of overfitting compared to other tree- based models.

COMPARATIVE ANALYSIS

MOD EL	ACCUR ACY	PRECISI ON	RECA LL	F1-SCORE
Naïve Bayes	93.16%	93%	93%	93%
SVM	99.56%	99%	99%	99%
Rand om Forest	99.77%	99%	99%	99%

The table presented above is a summary of the performance of three different machine learning models in the classification of deceptive news articles. The Naïve Bayes model had an accuracy of 93.16%. Precision, recall, and F1-score had similar values, which implies an average level of effectiveness in the classification of deceptive news articles using the Naïve Bayes model. The Support Vector Machine model had a high level of effectiveness in the classification of deceptive news articles compared to the Naïve Bayes model. The SVM model had an accuracy of 99.56%. Precision,

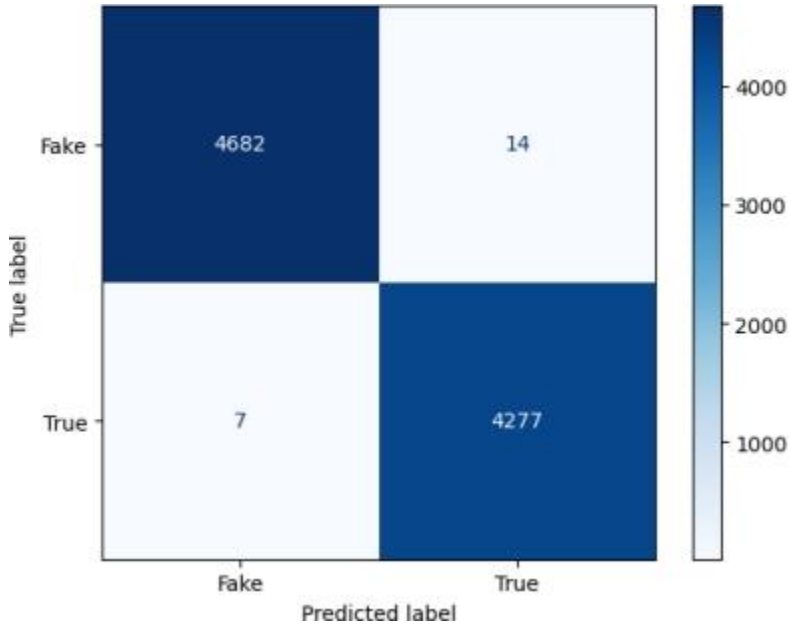
recall, and F1-score had similar values, which implies a high level of effectiveness in the classification of deceptive news articles using the SVM model. The efficiency level of the Random Forest model in the classification of fake news is high. The model had an accuracy of 99.77%. Precision, recall, and F1-score had similar values, which implies a high level of effectiveness in the classification of deceptive news articles using the Random Forest model. The results obtained in the research are in line with the literature, which implies that traditional machine learning models are effective in the classification of fake news when an effective feature engineering technique such as TF- IDF is used [2], [5].

METHODOLOGICAL INSIGHTS

The TF-IDF technique was instrumental in transforming the information represented in the news articles into numerical form, which improved its compatibility with machine learning algorithms [1], [4]. In the resulting feature space, the terms in the entire dataset represented unique features, resulting in an extremely high-dimensional space with more than 10,000 features. Nevertheless, the features utilized in the representation of the data in each individual document remained few, resulting in a sparse feature space with a majority of zeros. Such a sparse feature space was instrumental in selecting the model since it alleviated the problems related to handling high-dimensional data. High- dimensional data sparsity is a recognized characteristic of the feature space in text classification problems, as cited in the literature [1], [2].

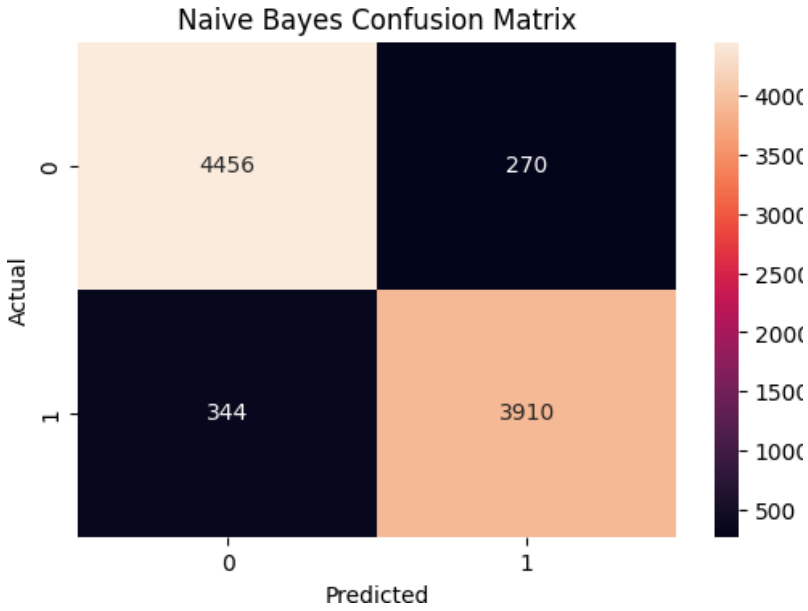
The resulting feature space was completely linearly separable following the application of the TF-IDF technique, such that a linear decision boundary was adequate in differentiating the real and fake news.

The performance of support vector machines (SVMs) was better than the other models, as they are designed to achieve an optimal separation of classes with maximum margin, as well as being computationally efficient [5], [6]. The sparsity also supports the appropriateness of SVMs for this purpose.

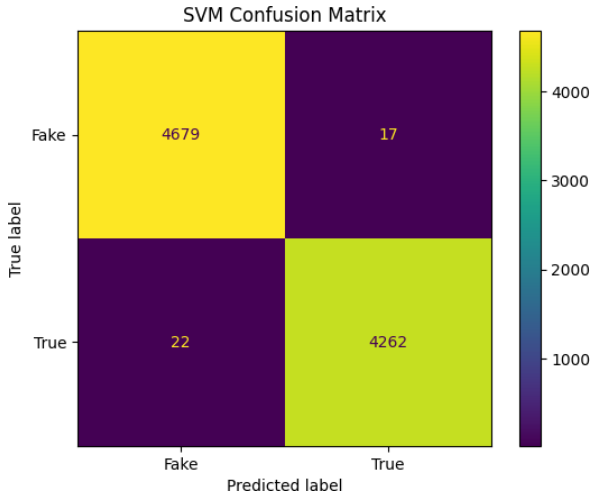


The performance of the random forests model was near-perfect, which may be due to its ability to handle complex relationships between individual words. However, this came at the cost of considerable computational complexity, especially for large-dimensional feature spaces, which made the computations more complex [5].

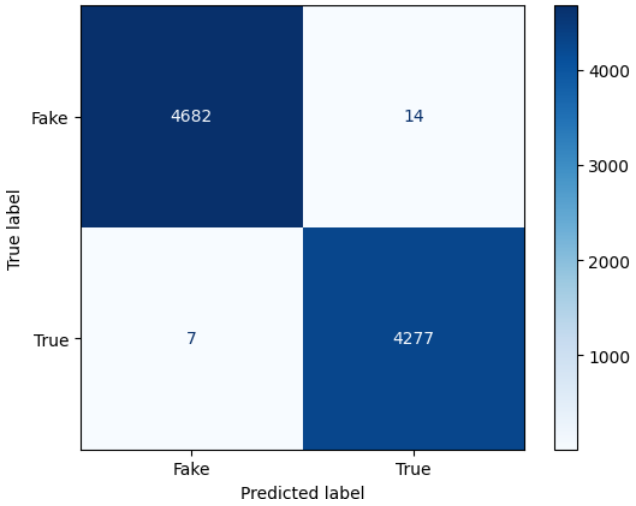
The performance of the naive Bayes classifier was slightly compromised compared to the other two models, namely SVM and random forests, as it assumes independence between words in the dataset, which fails to consider the possible relationships or dependencies between groups of words, as seen in natural language.



Confusion Matrix of Naïve Bayes



Confusion Matrix of SVM



Confusion Matrix of RF

Aspect	**Key Observation**	**Impact on Model Performance**
TF-IDF Representatio n	It is observed that the data is represented in the form of high-dimensional sparse vectors [1], [2]	This gives rise to the development of models that perform efficient operations on sparse high-dimensional data.
Linear Decision Boundary	It is observed that the data points are linearly separable due to the TF-IDF transformatio n [4]	This gives rise to high performance with low computationa l and training complexity.
SVM Performance	It is observed that the margin maximization process is facilitated by effectively handling sparse data [5]	High accuracy is achieved by the proposed method, with a feature set that is significantly larger than the number

		of data samples. Moreover, low computationa lcomplexity is achieved.

CONCLUSION

The results obtained in the present study show that traditional machine learning methods can provide positive results in terms of the level of accuracy in the identification of a large quantity of misinformation compared to modern artificial intelligence methods, provided that intelligent preprocessing is used in combination with the extraction of features by an efficient TF-IDF method. It is also anticipated that the statistical learning process from unstructured news texts to structured numerical data will help to identify false information with an acceptable level of accuracy. The transformation of unstructured news texts into structured numerical data is also anticipated to help statistical learning methods to achieve stable performance in text classification tasks, as reported in the literature [2], [5].

Thus, the results obtained in the present study show that, despite the exploratory nature of the application of multiple deep learning methods, the
In the framework of Support Vector Machine (SVM) methodology, an accuracy of 99.56% is anticipated with high generalization performance. Due to mathematical properties, Support Vector Machines (SVMs) are capable of maximizing margins for mitigating overfitting phenomena while being effective in high-dimensional sparse spaces that are generated through TF-IDF. Therefore, it can be concluded that Support Vector Machines are highly suitable for large-scale implementations. These characteristics make Support Vector Machines highly suitable for large-scale text classification systems [5].

In the case of Naïve Bayes methodology, an accuracy of 93.16% has been achieved. Considering the accuracy achieved through Naïve Bayes, it is evident that Naïve Bayes is one of the computationally most efficient methodologies that could be used for these analyses.

REFERENCES

1. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. “Fake News Detection on Social Media: A Data Mining Perspective.” ACM SIGKDD Explorations Newsletter 2017, 19(1), 22–36.
2. Zhou, X.; Zafarani, R. “A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities.” ACM Computing Surveys 2020, 53(5), 1–40.

3. Bondielli, A.; Marcelloni, F. “A Survey on Fake News and Rumor Detection Techniques.” Information Sciences 2019, 497, 38–55.
4. Ahmed, H.; Traore, I.; Saad, S. “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, 2018, 127–138.
5. Sahoo, S. R.; Gupta, B. B. “Multiple Features Based Approach for Automatic Fake News Detection on Social Networks Using Machine Learning.” Applied Soft Computing 2021, 100.
6. Dalbhanjan, M.; Mehta, S.; Kumar, P. “Fake News Detection Using Machine Learning Approaches.” Procedia Computer Science 2021, 165, 696–703.
7. Rubin, V. L.; Chen, Y.; Conroy, N. J. “Deception Detection for News: Three Types of Fakes.” In Proceedings of the Association for Information Science and Technology (ASIST), 2015, 52(1).
8. Monti, F.; Frasca, F.; Eynard, D.; Mannion, D.; Bronstein, M. M. “Fake News Detection on Social Media Using Geometric Deep Learning.” IEEE Transactions on Signal Processing 2021, 69, 259–273.