International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 13, Issue 4 (Jul.-Aug. 2024), PP. 01-04

Fake News Detection Using Machine Learning Methods: A Review

Rajan Kumar Tiwari
Assistant Professor, Department of Computer Science & Information Technology, Jharkhand Rai University, Ranchi (Jharkhand)

***Abstract**: In the present age, where the Internet is also omnipresent, everyone uses various online information resources. Along with the increase using social networks such as Facebook, Twitter, etc. news items spread like wildfire among millions of people. Spread Fake news has far-reaching effects, such as generating biased opinions that can influence the outcome of elections in favour of certain individuals or groups or parties or a particular section of the society. Moreover, spammers use enticing news headlines and clickbait ads to generate revenue which is unlawful. Because fake news is so common on social media, the widespread usage of these platforms has had a detrimental effect on our society. Similar to the internet, publishers promoted their interests by disseminating inaccurate and misleading information. Due to social media's low cost, quick access, and quick news propagation, people use it frequently. Fake news has long been known to be detrimental to both individuals and society as a whole. Thus, the problem of identifying bogus news arose. Unsuitable news was released to draw attention and enable the sender to begin spreading rumours. As a result of this, people began to see social media as a source of news as a dubious news agency. This led to inconvenience for offline news as well because when people will too much depend on online platforms that will reduce the offline users.*

This paper reviews the machine learning algorithms that are currently in use for identifying and minimizing bogus news on various social media sites, such as Facebook, Twitter, WhatsApp, and Convolutional Neural Network, LSTM, Neural Network, and Support Vector Machine. This review offers a thorough analysis that includes representative datasheets, assessment measures, and a perspective on data mining. Additionally, an analysis of current state-of-the-art methods is provided, and unresolved issues with false news detection are emphasized. The dearth of high-quality and quantity datasets has significantly impeded research in the field of false news detection. As a result, this evaluation contrasts the current methods for creating models with the anticipated advancements that can be made by combining several machine learning techniques.

***Keywords**: NLP techniques, machine learning algorithms, fake news, e-news, Social Media, classification*

Date of Submission: 08-07-2024 Date of acceptance: 25-07-2024

I. Introduction:

In the current situation, consumers' perspectives are greatly influenced by the stuff that is available online. Online reviews and comments provide e-commerce clients with information about the quality of the products. However, fake news has become a significant type of opinion spam that deceives consumers in recent times. It has always been difficult to stop the spread of fake news because so many websites, including those found on social media platforms like Google Plus, contribute significantly to its dissemination. For the average individual, it can be quite difficult to recognise fake news, and even for someone with extensive understanding in that field, the results might not be reliable. The emergence of machine learning in fake news has made identification easier. Natural language processing is employed in the current work to identify three different kinds of fake news. Analogously, data mining methods such as clustering and classification problems were employed in some of the previous research.

The current studies have an accuracy rate of about 85% when it comes to identifying fake news. However, more study in this area is required.

Because false news identification may be a relatively recent development in relation to the public's concern and interest, research on the topic is still being conducted in its early stages. We usually review a wide range of published publications.

Experts are of the opinion that false information generally comes in three forms. The first is the outright false information, or news that the writers of the piece have made up and is completely false. Making derogatory remarks is the second category of fake news. The information is made up to make the reader laugh. The third category comprises news articles with a substantial percentage of trash content, some genuine news, but bad writing.

Previous studies on detection of fake news by Machine Learning methods

(Probierz et al, 2021) aims to develop a new model to quickly identify fake news based on the news title alone without analyzing the entire text. The model uses NLP techniques, machine learning algorithms, single classifiers, and ensemble methods for prediction. Experiments confirmed the hypothesis and the model yielded good results after initial analysis of news titles. However, the running time of the SVM classifier was a problem, but the random forest algorithm was a good trade-off. Future work may involve developing a multi-criteria model for a two-step analysis, focusing on the quality of classification and running time to identify true or fake news quickly and efficiently.

(Nagaraja et al, 2021) reveals that users tend to read e-news through unofficial or fake sources, affecting both clients and society. This paper uses algorithms to identify if a news is fake or not, including three algorithms for calculations and analysis: Natural Language Processing (NLP), Naïve Bayes, and Support Vector Machine (SVM). Semantic examination is used in both procedures, with Naïve Bayes providing the least accuracy and SVM providing the best precision. This approach allows users to identify if a news is fake or real, potentially helping users to protect themselves and society.

(Mishra et al, 2022) explores the concept of fake news detection using deep learning techniques. Despite the success of probability latent semantic analysis, the constantly changing nature of false news on social media makes classification difficult. The study compares traditional machine learning and deep learning techniques on fake news, revealing that deep learning outperforms traditional methods. Bi-LSTM (Long Short-Term Memory,) achieved the best detection rate of 95% accuracy and F1 score. It is worthwhile to mention here that Long Short-Term Memory is an improved version of recurrent neural network. LSTM excels in sequence prediction tasks, capturing long-term dependencies. This is ideal for time series, machine translation, and speech recognition. The research can be useful for police departments and society as a whole, but its limitations lie in its focus on textual data. Future research could expand to image data.

(Reis et al, 2019) aims to detect fake news by analyzing recent and related works. It proposes novel features, such as those related to the source domain, which appear five times more often in the best models. The study reveals that detecting fake news is challenging, with only 2.2% of models achieving an AUC higher than 0.85. The findings suggest that certain types of fake news are identified by models with specific combinations of features, highlighting the complexity of the problem and the difficulty of tackling all forms of fake news stories. Future work plans to categorize fake news stories as a strategy for constructing effective and robust ensembles of classifiers. The study also shows different models of clusters made of random combinations of features, suggesting that ensemble techniques combining models from different clusters are a promising avenue for investigation.

(Baarir et al, 2021) proposes a machine learning system for fake news detection, utilizing TF-IDF and SVM techniques, and a dataset of fake and true news, demonstrating its efficiency. The paper presents a method for detecting fake news using support vector machine, focusing on identifying the best features and techniques. The solution uses a preprocessed news dataset, extracting features, and applying the algorithm for classification. This study found that text, author, source, date, and sentiment are the best features for detecting fake news, with a 100% recognition rate. Sentiment analysis is useful for opinion mining. The N-gram method is better for bulky datasets and large texts. Support vector machine is the best algorithm for detecting fake news, with parameters like Cost C, gamma, and epsilon influencing its performance.

(Sahoo et al, 2021) presents a fake news detection method for Facebook users using machine learning and deep learning classifiers in a chrome environment. The method analyzes user profile and news content features. A chrome extension is developed using crawled data extracted by a crawler. A deep learning algorithm called Long Short-Term Memory is used to enhance the performance. The method achieved 99.4% accuracy compared to machine learning algorithms. Experimental results show the method successfully detects fake news in real-time on the user side. Future work could include analyzing features using other deep learning algorithms, such as bidirectional LSTM, bidirectional GRU, or hybrid approach-based deep learning classifiers, for better decision-making with more datasets.

(Khan et al, 2021) analyzes nineteen machine learning approaches on three datasets, including traditional learning models, deep learning models, and advanced pre-trained language models like BERT. BERT-based models outperform all other models on all datasets, and pre-trained BERT-based models perform better on small sample sizes. Naive Bayes with n-gram can achieve similar results to neural network-based models when the dataset size is sufficient. LSTM-based models' performance depends on the dataset length and the information provided in a news article. With adequate information, LSTM-based models have a higher probability of

overcoming overfitting. The author claimed that the results could help organizations like online news portals and social media choose the most suitable model for detecting fake news, particularly during the COVID-19 pandemic.

This paper examines the detection of fake news in social media through two stages: characterization and disclosure. The first stage highlights the basic concepts and principles of fake news, while the discovery stage reviews current methods for detecting fake news using different supervised learning algorithms. The paper uses Naive Bayes classifier to detect fake news from various sources with an accuracy of 74%. However, it has been found that the results are not accurate for untruthful sources.

The researchers used data from Kaggle with an average accuracy of 74.5%. They also used Naive Bayes algorithms to detect Twitter spam senders with an accuracy rating of 70% to 71.2%. Three common methods used in their research were Naïve Bayes, Neural Network, and Support Vector Machine (SVM). Naïve Bayes had an accuracy of 96.08%, while neural networks and SVM reached an accuracy of 99.9%.

The researchers examined the 2012 Dutch elections fake news on Twitter and concluded that the decision tree algorithm worked best for the dataset. A counterfeit detection model using N-gram analysis achieved the highest accuracy of 92%.

The research summary and system analysis concluded that most research papers used Naïve Bayes algorithm, with prediction precision between 70-76%. The authors propose adding POS textual analysis to these methodologies, which depends on adding numeric statistical values as features. The proposed features include total words (tokens), total unique words (types), Type/Token Ratio (TTR), number of sentences, Average sentence length (ASL), Number of characters, Average word length (AWL), nouns, prepositions, and adjectives.

(Sharma et al, 2020) aims to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning**.** Most tasks in the twenty-first century are completed online. Online news items and social media apps like Facebook and Twitter are replacing newspapers that were once preferred in printed form. Another important source is Whatsapp's forwards. The growing issue of false news only serves to aggravate matters and tries to influence public opinion and attitudes on the usage of digital technology. When someone is tricked by fake news, two things may occur: first, they begin to think that their preconceived notions about a certain subject are accurate. Therefore, in an effort to stop the situation, the authors have created a system called phoney News Detection that allows users to submit content and determines whether it is phoney or true. Several NLP and machine learning techniques must be applied in order to do this. An adequate dataset is used to train the model, and a variety of performance metrics are used to assess its performance. The news headlines or articles are classified using the best model, or the model with the highest accuracy. Their best model, with an accuracy of 65% for static search, turned out to be logistic regression. Consequently, they employed grid search parameter optimization to boost the logistic regression's performance, resulting in a 75% accuracy rate. Therefore, they state that there is a 75% likelihood that a user will classify a news story or its headline accurately if they feed it into our model.
The customer has the option to verify the website's legitimacy in addition to checking the news article or keywords online. With a dynamic system, the accuracy is 93% and gets better with each iteration. The authors' goal is to create a dataset of their own that is updated based on the most recent information. Using an online database and web crawler, all of the most recent information and current news get stored in one place.

II. Conclusion:

The review study discusses pioneering existing work in the field of false news detection. Machine learning-based classification algorithms play a very important role in the detection of fake news or rumors from social media, which is a very complicated and difficult process due to the diverse political, social and economic, and many other related factors. This review discusses various machine learning approaches such as NLP, Linear Regression, KNN, SVM, LSTM, Artificial Neural Networking, and many more.

References:

[1]. Probierz, B., Stefański, P., & Kozak, J. (2021). Rapid detection of fake news based on machine learning methods. Procedia Computer Science, 192, 2893-2902.
[2]. Nagaraja, A., KN, S., Sinha, A., RAJENDRA KUMAR, J. V., & Nayak, P. (2021, April). Fake news detection using machine learning methods. In International Conference on Data Science, E-learning and Information Systems 2021 (pp. 185-192).
[3]. Mishra, S., Shukla, P., & Agarwal, R. (2022). Analyzing machine learning enabled fake news detection techniques for diversified datasets. Wireless Communications and Mobile Computing, 2022(1), 1575365.
[4]. Reis, J. C., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019, June). Explainable machine learning for fake news detection. In Proceedings of the 10th ACM conference on web science (pp. 17-26).

[5]. Baarir, N. F., & Djeffal, A. (2021, February). Fake news detection using machine learning. In 2020 2nd International workshop on human-centric smart environments for health and well-being (IHSH) (pp. 125-130). IEEE.
[6]. Sahoo, S. R., & Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing, 100, 106983.
[7]. Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. Machine Learning with Applications, 4, 100032.
[8]. Khanam, Z., Alwasel, B. N., Sirafi, H., & Rashid, M. (2021, March). Fake news detection using machine learning approaches. In IOP conference series: materials science and engineering (Vol. 1099, No. 1, p. 012040). IOP Publishing.
[9]. Sharma, U., Saran, S., & Patil, S. M. (2020). Fake news detection using machine learning algorithms. International Journal of creative research thoughts (IJCRT), 8(6), 509-518.