

editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE
RESEARCH IN ENGINEERING MANAGEMENT
AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)e-ISSN :
2583-1062Vol. 04, Issue 12, Decembaer 2024, pp : 1347-13517.001

A REVIEW OF MACHINE LEARNING AND NLP FOR FAKE NEWS DETECTION

Bhargavi Jakkula¹, Nuthana Bellamkonda², M. Geetha³, Remalli Rohan⁴

¹M. Sc-Data Science - PG Scholar, Department of Computer Science, Telangana Social Welfare Residential Degree College for Women, Jagathgirigutta, Hyderabad, India.

^{2,3}Degree lecturer, Department of Computer Science, Telangana Social Welfare Residential Degree College for Women, Jagathgirigutta, Hyderabad, India.

⁴Researcher, Computer Science Educator, Hyderabad, Telangana, India.

DOI: https://www.doi.org/10.58257/IJPREMS37826

ABSTRACT

Fake news identification has become one of the most important problems of the twenty-first century due to the sharp rise in social media use and the falling cost of internet access. The dissemination of rumors and false information has been greatly accelerated by this accessibility, leading to a rise in problems associated with fake news that range from online arguments to violent hate crimes. Despite its dependability, traditional fact-checking techniques are unable to handle the enormous amount of web content produced every day. Therefore, it is crucial to create an effective, scalable, and empirical approach to identify bogus news. In order to solve this issue, this study highlights machine learning (ML) and natural language processing (NLP) as crucial strategies. Real-time, automated news analysis and credibility-based classification are made possible by these technologies. This survey attempts to review previous research in order to compile knowledge, analyze current approaches, and determine their efficacy and correctness. Models' advantages, disadvantages, and areas for development are highlighted through comparison with industry standards. With accuracy as the key to success, the ultimate goal is to provide a template for a more successful false news detecting system. This study offers insightful information about how to use ML and NLP techniques to lessen the effects of false information and promote trust in online communication.

Keywords: Fake News Detection, Machine Learning (ML), Natural Language Processing (NLP), Misinformation and Digital Communication.

1. INTRODUCTION

Fake news identification has advanced significantly between 2019 and 2024, especially with the use of machine learning (ML) and deep learning techniques. With models reaching accuracy levels as high as 95% to 98% when applied to wellknown datasets like those from fig1, the accuracy of fake news detection systems has significantly increased during this time. More sophisticated algorithms like BERT (Bidirectional Encoder Representations from Transformers) have shown to be far more effective than more conventional techniques like Random Forest and Logistic Regression, especially when deep learning models became more popular in 2021. Moreover, the capacity to detect in real time has gained significant attention, particularly with the growth of social media platforms that produce enormous volumes of data every day. These developments highlight the growing importance of ML and deep learning in combating misinformation, as they offer scalable, adaptable, and more precise solutions to address the increasing spread of fake news across digital platforms.



Figure 1: History of Fake News Detection

@International Journal Of Progressive Research In Engineering Management And Science

LIPREMS	INTERNATIONAL JOURNAL OF PROGRESSIVE	e-ISSN :
	RESEARCH IN ENGINEERING MANAGEMENT	2583-1062
	AND SCIENCE (IJPREMS)	Impact
www.ijprems.com	(Int Peer Reviewed Journal)	Factor :
editor@ijprems.com	Vol. 04, Issue 12, Decembaer 2024, pp : 1347-1351	7.001

2. LITERATURE SURVEY

The increasing proliferation of false information on digital platforms and social media has made the detection of fake news a crucial field of study. To address this problem, numerous studies have looked into different machine learning and deep learning strategies. To classify news as "real" or "fake," for example, Tahura Nikhath et al. [1] and Ranga Rao et al. [2] used Random Forest classifiers on the Kaggle Fake News Dataset, highlighting the shortcomings of straightforward word-count-based models and the necessity of context-aware methods. Similar to this, P. S. Surendra et al. [3] addressed the mixing of bogus and legitimate news on social media by combining natural language processing (NLP) methods with classifiers like Logistic Regression, Naïve Bayes, and SVM to improve detection accuracy. Feature extraction and optimization to enhance detection were the main topics of several studies. The usage of n-grams, TF-IDF, and dimensionality reduction methods like PCA in conjunction with algorithms like Random Forest was investigated by Yash Shukla et al. [4] and Ch. Sita Kumari et al. [11] in order to improve performance. To develop reliable detection systems, Tejaswi Gaikwad et al. [12] suggested frameworks that make use of feature selection techniques and the Bag of Words. It has also been investigated how user involvement and behavior affect the detection of bogus news. Madre Mario Perez et al. [9] emphasized the incorporation of user engagement data to improve model accuracy, while Anu Shrestha et al. [6] investigated the role of personality factors in disseminating false information. Furthermore, utilizing content, creator, and context-based features, Fátima Leal et al. [20] presented a novel strategy for real-time fake news detection that combines supervised and unsupervised techniques. This field has seen an increase in the use of deep learning and advanced machine learning models. The higher performance of neural networks and contemporary frameworks like BERT over conventional techniques was demonstrated by studies such as those conducted by Tushar Rane et al. [15] and Ali Raza et al. [7]. Similarly, in order to enhance early detection capabilities, Yuta Yanagi et al. [8] suggested utilizing neural networks to generate artificial remarks. The versatility of false news detection algorithms to different settings was shown by domain-specific research such those by Pop Mihai Ionut et al. [5], which concentrated on economic sectors, and Nagul Cooharojananone et al. [19], which targeted Thai health articles. These initiatives demonstrate how tactics are changing and how crucial it is to use linguistic, contextual, and interactionbased elements in order to successfully handle the difficulties presented by false information.

Author Name(s)	Dataset Used	Feature Extraction Method(s)	Algorithm(s)	Results
Tahura Nikhath et al. [1]	Kaggle Fake News Dataset	Count Vectorizer, TF-IDF	Random Forest	Developed a model to classify articles into real or fake news using a Random Forest classifier, noting limitations due to lack of word order/context
L Sai Venkaya Ranga Rao et al. [2]	Kaggle Fake News Dataset	Count Vectorizer, TF-IDF	Random Forest	Applied a Random Forest classifier for binary classification, focusing on detection of fake news in social media contexts like Facebook.
Yash Shukla et al. [3]	Custom Dataset	n-grams, TF-IDF	Various algorithms compared	Compared multiple algorithms to find the best fit for fake news prediction, analyzing various feature extraction methods and their effectiveness.
Ms. Ch. Uma Devi et al. [4]	Social Media Feeds	NLP Techniques	NLP, Logistic Regression, Naïve Bayes, SVM	Applied multiple classifiers to detect fake news, focusing on social media misinformation and its impact on public perception.

Table 1. Summary of	Methodologies
---------------------	---------------



www.ijprems.com

editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

(Int Peer Reviewed Journal)

e-ISSN : 2583-1062 Impact Factor :

7.001

Achieved 0.73 accuracy for Binary English and 0.77 for Spanish, PAN at Writing style, n-Anu Shrestha et Classification with finding personality features CLEF 2020 grams, BERT al. [5] BERT, Sentiment significantly impact fake news Dataset embeddings, LIWC Analysis spreader detection on social media. Reviewed current algorithms, Facebook, emphasizing NLP and neural Natural Language Ali Raza et al. Twitter, Naïve Bayes, NLP, networks, for detecting fake Processing, Part-of-Neural Networks [6] Weibo news on social networks, Speech Tagging Datasets relevant for ethical and political issues. Showed that generating fake News Social context comments improved detection, Yuta Yanagi et Articles with Neural Network, generation for fake supporting early fake news Generated al. [7] Deep Learning detection when limited social comments Comments context is available. Developed a scoring model for Custom Scoring model detecting fake news in the Mihai-Ionuț Pop Dataset on Custom Dataset on tailored to economic energy field, aiming to combat et al. [8] Economic **Economic Sectors** data misinformation affecting public Sectors perception of economic sectors. Deep Learning, Logistic Demonstrated that including user Regression, SVM, Mario Perez Fake News User engagement, engagement data improves Madre et al. [9] Net dataset linguistic features Random Forest, accuracy over content-only Light GBM, XG models. Boost Signal Bi-grams, Explored limitations of machine Media for Probabilistic Priyanshi Goyal learning for fake news detection; TF-IDF, PCFG et al. [10] OpenSource Context-Free highlighted potential for deep Grammar (PCFG learning enhancements. s.co Random Forest, Random Forest with Count News Ch Sita Kumari TF-IDF, Count SVM, Logistic Vectorizer achieved the best Articles et al. [11] Vectorizer, PCA Regression, Naïve performance based on accuracy, Dataset precision, and recall **Bayes** Social Voting classifiers improved Tejaswi Gaikwad TF-IDF, Bag of Naïve Bayes, SVM, Media News accuracy for distinguishing real Words et al. [12] Logistic Regression from fake news Dataset Large-Scale Prof. Shweta News and Random Forest, Focused on Random Forest and ROC curves, other Kahurke et al. Social Logistic LSTM with ROC curve evaluation metrics Regression, LSTM [13] Media evaluation Dataset



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

(Int Peer Reviewed Journal)

e-ISSN : 2583-1062 Impact

Factor :

7.001

www.ijprems.com editor@ijprems.com

Vol. 04, Issue 12, Decembaer 2024, pp : 1347-1351

Shagun Kingaonkar et al. [14]	Labeled News Articles	Normalization	Support Vector Machine (SVM)	Aimed to achieve high accuracy with SVM, noting normalization's importance in data cleansing.
Adwait Banda, Tushar Rane et al. [15]	Fake News Dataset, Twitter Datasets	Dimensionality Reduction (PCA, Chi-Square), NLP	CNN, LSTM, PCA, Chi-Square, Decision Tree, Logistic Regression, KNN, Random Forest, SVM, Naive Bayes, RNN, LSTM	Deep learning methods like Attention-based models and BERT outperformed traditional methods for accuracy.
E. V. Nagalakshmi et al. [16]	Labeled News Articles	Pattern Learning	KNN, Decision Tree, Logistic Regression	Developed a tool aiming for accurate classification of real and fake news.
Mohammed Obaid et al. [17]	Twitter Dataset, Social Media Posts	Twitter API, Sentiment Analysis	SVM, Naive Bayes, Logistic Regression, RNN	Found SVM and Naive Bayes to outperform other models.
Pshko Rasul Mohammed Amin et al. [18]	Labeled News Articles, Social Media Data	Classification	Decision Tree, SVM	SVM and Decision Tree showed high performance
Nagul Cooharojananon e et al. [19]	582 reliable and 435 unreliable Thai health and medical articles from 8 websites	Textual, sentiment- based, and lexicon features including specific terms like herbs, fruits, and vegetables to identify article objectives	XG Boost with Lasso Feature Selection	97.76% accuracy, 7.16% improvement over previous model
Fátima Leal et al. [20]	Real-time Twitter dataset	Creator-, content-, and context-based features using Natural Language Processing; explainable classification displayed in a dashboard	Combination of unsupervised and supervised ML approaches with online lexica	80% accuracy and macro-level validation

3. CONCLUSION AND FUTURE SCOPE

In conclusion, machine learning-based techniques, which have been shown to be the most successful in predicting and recognizing fake news, are used in the majority of research on fake news detection (about 95%). Because it affects how the algorithm represents and interprets text data, the word embedding selection has a big impact on improving the detection models' accuracy. Furthermore, deep learning models—in particular, neural networks—have become the most

IJPREMS	INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT	e-ISSN : 2583-1062
	AND SCIENCE (IJPREMS)	Impact
www.ijprems.com	(Int Peer Reviewed Journal)	Factor :
editor@ijprems.com	Vol. 04, Issue 12, Decembaer 2024, pp : 1347-1351	7.001

widely used approach, accounting for around 90% of the research in this field and continuously producing remarkable accuracy rates. Future studies should concentrate on strengthening these models' resilience by investigating new word embedding strategies, integrating various machine learning approaches, and tackling the problems of multilingual content and real-time detection. Additionally, in order to better comprehend deep learning models' decision-making process and guarantee transparency in fake news detection systems, efforts should be focused on improving the interpretability of these models.

4. REFERENCES

- [1] Ehzaam, M., Fareed, S., Nikhath, T., & Anitha, T. Fake News Detection Using Kaggle Dataset and Machine Learning Algorithm. International Journal of Research, Vol 6(10), Pg.890-895,2019.
- [2] Ranga Rao, L. S., Sreenath, T. S., & Sasidhar, K. Fake News Detection in Facebook Using Machine Learning. International Journal of Research, Vol 6(8), Pg.263-270, 2019.
- [3] Devi, C. U., Priyanka, R., Surendra, P., Priyanka, B. S., & Nikhila, C. N. D. L. Fake News Detection Using Machine Learning. Journal of Emerging Technologies and Innovative Research (JETIR), Vol 6(4), Pg.104-112, 2019.
- [4] Shukla, Y., Yadav, N., & Hari, A. A Unique Approach for Detection of Fake News Using Machine Learning. International Journal for Research in Applied Science & Engineering Technology IJRASET Vol 7(6), Pg.1-6, 2020.
- [5] Pop, M.-I. Scoring Model for the Detection of Fake News. Bucharest University of Economic Studies. Published March 2020.
- [6] Shrestha, A., Spezzano, F., & Joy, A. Detecting Fake News Spreaders in Social Networks via Linguistic and Personality Features. Notebook for PAN at CLEF 2020. Boise State University.
- [7] Raza, A., Shafiq, S., Bibi, M., Rehan, M., & Anwar, H. Automatic Fake News Detection: Issues and Solutions. Department of Software Engineering, University of Sialkot, Punjab, Pakistan. Published in April 2020.
- [8] Yanagi, Y., Orihara, R., Sei, Y., Tahara, Y., & Ohsuga, A. Fake News Detection with Generated Comments for News Articles. EasyChair Preprints, 2020
- [9] Mario Perez Madre. Fake News Detection using News Content and User Engagement. Dissertation submitted in fulfillment of the degree of Master in Data Engineering and Science. Director: Alvaro Rodrigo Yuste, Universidad Nacional de Educación a Distancia. June 2021.
- [10] Priyanshi Goyal, Dr. Swapnesh Taterh, Mr. Ankit Saxena. Fake News Detection using Machine Learning: A Review. International Journal of Scientific Research in Computer Science, Engineering, and Information Technology,2021
- [11] Velivela Durga Lakshmi, Ch. Sita Kumari. Detection of Fake News using Machine Learning Models. International Journal of Computer Applications, Vol 183, Pg. 47. January 2022.
- [12] Tejaswi Gaikwad, Bhaskar Rajale, Prasad Bhosale, Swapnil Vedpathak, Mrs. S.S. Adagale. Detection of Fake News using Machine Learning Algorithms. International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol 10, Issue V, May 2022.
- [13] Prof. Shweta Kahurke, Jyoti S.Haldar, Ashwini D. Bhagat, Rasika S. Andeo, Parag M. Dahikar, Ajinkya K.Kinhekar, Suraj S. Yelore. Fight Misinformation and Detect Fake News Using Machine Learning Algorithm. International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol 10, Issue VIII, August 2022.
- [14] Shagun Kingaonkar et al.: "Fake News Detection using Machine Learning," International Journal of Scientific Research in Engineering and Management (IJSREM), Vol 07, Issue 03, March 2023.
- [15] Adwait Banda, Tushar Rane : "A Review on Fake News Detection," International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol 11, Issue V, May 2023.
- [16] E. V. Nagalakshmi et al.: "Fake News Detection using Machine Learning A Working Model of Fake News Detection," International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol 11, Issue V, May 2023.
- [17] Mohammed Obaid et al. :"Detecting Fake News Using Machine Learning," International Journal of Innovative Research in Modern Physics and Science (IJIRMPS), Vol 11, Issue 1, January-February 2023.
- [18] Pshko Rasul Mohammed Amin : "Fake News Detection Using Machine Learning," Indonesian Journal of Computer Science, Vol 12, Pg. 4, 2023.
- [19] Kanokwan Atchariyachanvanich, Chotipong Saengkhunthod, Parischaya Kerdnoonwong, Hutchatai Chanlekha, and Nagul Cooharojananone :"Improvement of a Machine Learning Model Using a Sentiment Analysis Algorithm to Detect Fake News: A Case Study of Health and Medical Articles on Thai Language Websites.", 2024
- [20] Francisco de Arriba-Pérez, Silvia García-Méndez, Fátima Leal, Benedita Malheiro, and Juan Carlos Burguillo: "Exposing and Explaining Fake, 2024.