JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Automated Fake News Detection on Social Media Platforms Using Machine Learning

[1]Annu, [2] Rahul

[1]Dept. of CS&E, Ganga Institute of Technology & Management, MDU, Haryana
[2]Dept. of CS&E, Ganga Institute of Technology & Management, MDU, Haryana

Abstract: The extensive dissemination of misinformation on social media has emerged as a significant concern, affecting public opinion and decision-making. The capacity to autonomously detect false information is essential in alleviating its detrimental impacts. This study introduces a machine learning methodology for identifying false news through the analysis of textual data utilizing many classification models, such as Logistic Regression, Naïve Bayes, and K-Nearest Neighbors (KNN). This study's dataset includes both authentic and fabricated news stories, which are subjected to preprocessing methods like tokenization, stopword elimination, and TF-IDF vectorization to transform textual content into structured numerical data. The efficacy of these models is assessed using critical metrics including accuracy, precision, recall, and F1-score, facilitating a comparison analysis to identify the most successful classification algorithm. The findings indicate that machine learning can markedly improve the precision of fake news detection, providing a dependable and automated method to address the proliferation of disinformation. This research advances the field of digital content verification and underscores the necessity of incorporating artificial intelligence-driven methodologies to enhance the reliability of online information.

Keywords**:** Fake News Detection, Social Media, Machine Learning, Text Classification, Logistic Regression, Naïve Bayes, K-Nearest Neighbors, Natural Language Processing, Misinformation, TF-IDF Vectorization.

I. INTRODUCTION

The emergence of social media has significantly altered news consumption, providing rapid access to information while also exacerbating the dissemination of disinformation. Fabricated news, encompassing intentionally false or misleading information, has emerged as a major issue because to its capacity to sway public opinion, distort political results, and erode confidence in credible news outlets (Al-Tarawneh, 2024). Identifying false information is notably difficult as it frequently emulates the style and format of authentic news stories, rendering human verification arduous and time-intensive. As a result, researchers have employed machine learning (ML) techniques to automate false news identification, utilizing text analysis, sentiment assessment, and classification models **(**Hamed**,** S. K., 2023**).**

In recent years, diverse machine learning techniques, such as logistic regression, support vector machines, and deep learning methodologies, have been utilized to enhance the accuracy of fake news detection. Through the examination of linguistic patterns, word embeddings, and user engagement metrics, these algorithms can distinguish between genuine and fraudulent information with considerable accuracy (Roumeliotis, 2025). Nonetheless, difficulties include dataset bias, new misinformation strategies, and adversarial content development persist as significant impediments. This paper examines the function of machine learning in identifying false information, emphasizing current approaches, obstacles, and possible remedies.

1.1 Machine Learning Techniques for Fake News Detection

A variety of machine learning methodologies have been suggested for the detection of false information. Conventional models, including Naïve Bayes and decision trees, depend on statistical text features, but sophisticated deep learning methodologies employ neural networks to identify intricate patterns in news stories. Moreover, hybrid models that integrate natural language processing (NLP) with attitude and emotion analysis have shown enhanced precision in detecting misleading content (Mohsen, F.,2024)**.**

1.2 Challenges in Fake News Detection

Notwithstanding progress, the detection of fake news encounters numerous problems, such as the evolution of disinformation tactics, the scarcity of labeled datasets, and adversarial assaults aimed at circumventing detection systems (Roumeliotis, 2025). Moreover, discrepancies in writing styles and sources hinder models' capacity to generalize across various forms of fake news information. Researchers are enhancing detection systems through the integration of explainable AI and transfer learning methodologies.

1.3 Future Directions and Ethical Considerations

As the technology for detecting false information advances, ethical issues related to censorship, bias, and privacy require attention. Future research must concentrate on creating resilient, transparent models that reconcile precision with ethical accountability (Shen, Y. 2023)**.** Furthermore, incorporating fact-checking protocols with AI-based detection systems could improve dependability and reduce the likelihood of false positives.

2. LITERATURE REVIEW

1."Cinelli et al. (2020) - The Influence of Social Media on Misinformation in the context of COVID-19"

Cinelli et al. (2020) investigated the influence of social media on the propagation of disinformation during the COVID-19 epidemic, emphasizing how platforms such as Facebook, Twitter, and YouTube expedited the distribution of erroneous information. Provocative headlines garnered greater engagement than substantiated facts, while automated accounts exacerbated the spread of misinformation. This resulted in public bewilderment, detrimental actions, and interruptions in health initiatives. The study encouraged social media businesses to enhance algorithm transparency and fortify content control to reduce the dissemination of disinformation and alleviate its adverse effects on public health.

2."Donepudi (2019) - The Role of Artificial Intelligence in Fake News Detection"

Donepudi (2019) investigated the function of artificial intelligence in identifying fake news, highlighting the capability of machine learning to automate the detection of disinformation. The study addressed the obstacles encountered by AI models, such as the absence of labeled datasets, biases in training data, and the dynamic nature of misinformation, which complicate the preservation of accuracy. The study emphasized that imbalanced datasets may result in biased predictions, undermining the reliability of detection systems. Moreover, as misinformation perpetually transforms, AI models must adjust to novel manifestations of deceit. Notwithstanding these constraints, the study acknowledged AI as a viable instrument for addressing misinformation and emphasized the necessity for enhancements in dataset quality, bias reduction, and algorithmic flexibility to augment its efficacy.

3."Shu et al. (2017) - Techniques for Fake News Detection on Social Media"

Shu et al. (2017) examined data mining methodologies for identifying false information on social media, assessing machine learning models including decision trees, support vector machines, and deep learning. Their research highlighted the need of integrating textual analysis with user-related data, such as credibility scores and engagement patterns, to enhance detection accuracy. They emphasized the significance of content-based and social context-based attributes, such as shares and comments, in detecting misinformation. The study emphasized the necessity for adaptive models that progress alongside new false content, promoting integrated detection systems that utilize various data sources to improve reliability and counteract misinformation dissemination.

4."Vosoughi, Roy, and Aral (2018) - The Spread of True vs. False News"

Vosoughi, Roy, and Aral (2018) examined the dissemination of real and false news on Twitter, demonstrating that disinformation propagates more rapidly, reaches a broader audience, and elicits greater engagement than factual information. The swift dissemination is chiefly propelled by the emotional allure of misinformation, which elicits intense responses such as astonishment and revulsion, hence increasing the likelihood of user sharing. The research highlighted that human psychology, rather than solely algorithms, is essential in this occurrence, as individuals are inherently predisposed to disseminate emotionally charged content.

The researchers proposed incorporating emotional reaction analysis into false news detection algorithms to enhance their efficacy. They advocated for the promotion of digital literacy and the encouragement of critical thinking to assist users in assessing misleading content prior to dissemination, emphasizing the necessity of an integrated technological and psychological strategy to counter disinformation.

5. "Wang (2017) - Machine Learning Framework for Fake News Detection"

Wang (2017) proposed a machine learning framework for identifying false news through the examination of linguistic characteristics and social environment. The research assessed deep learning models such as CNNs and RNNs for feature extraction, emphasizing syntax, semantics, and writing style to distinguish between fake and authentic news. It emphasized the significance of integrating user involvement, source reliability, and metadata to enhance detection accuracy. The study revealed that hybrid models integrating deep learning with conventional machine learning outperformed singular-method strategies. Wang underscored the necessity for ongoing enhancements in detection systems to accommodate the increasing strategies of misinformation.

3. METHODOLOGY

3.1 Data Collection

The dataset for this study consists of labeled fake and actual news stories obtained from publicly accessible datasets, including the Fake News Dataset and the True News Dataset. These databases comprise textual information sourced from several online channels, including news websites and social media outlets. Each news article is meticulously categorized as either false (1) or authentic (0) to enable supervised learning.

3.2 Data Preprocessing

Raw text data contains noise that can impact the accuracy of machine learning models. Therefore, preprocessing techniques were applied to clean and standardize the text. The steps include:

- **Text Cleaning:** Removing special characters, punctuation, numbers, and unnecessary white spaces.
- **Tokenization:** Splitting the text into individual words for further processing.
- **Stopword Removal:** Eliminating commonly used words (e.g., "the," "is," "and") that do not contribute meaningfully to classification.
- **Lemmatization:** Converting words to their base form to maintain uniformity in representation.
- **Feature Extraction using TF-IDF:** The Term Frequency-Inverse Document Frequency (TF-IDF) technique was applied to convert the textual data into numerical form. TF-IDF assigns importance scores to words based on their occurrence in the document and across the dataset. This transformation is crucial for effective machine learning classification.

3.3 Dataset Splitting

To assess the efficacy of several models, the dataset was partitioned into:

- **80% Training Set:** Utilized for the training of machine learning models.
- **20% Testing Set:** Utilized to assess model efficacy on unobserved data. The train-test split guarantees that models generalize effectively to novel data and mitigates overfitting.

3.4 Machine Learning Models and Implementation

Three commonly utilized machine learning models were employed for classification:

3.4.1 Logistic Regression

- A statistical model employed for binary classification.
- It assesses the likelihood of an article being authentic or fraudulent based on its textual characteristics.
- Appropriate for high-dimensional data, including text-based datasets.

3.4.2 Naïve Bayes (Multinomial Naïve Bayes)

- A probabilistic model founded on Bayes' theorem.
- Presumes that characteristics (words) are independent, hence enhancing computing efficiency.
- Commonly employed in text classification tasks, such as spam detection and false news categorization.

3.4.3 K-Nearest Neighbors (KNN)

- A distance-based categorization technique that designates a label according to the predominant K-nearest neighbors.
- Employed in this work for comparative analysis to assess performance disparities between logistic regression and Naïve Bayes.

Each model was trained on the processed dataset utilizing Python-based machine learning frameworks, including Scikit-learn.

3.5 Model Evaluation and Performance Metrics

Various evaluation indicators were employed to evaluate the efficacy of each model.

- **Accuracy Score:** The proportion of accurately classified news articles.
- **Precision, Recall, and F1-score:** These measures assess the efficacy of each model in distinguishing between fake and real news.
- **Confusion Matrix:** A visualization instrument that displays accurate and erroneous predictions for both fabricated and authentic news categories.
- **Comparison of Model Performance:** A bar chart was employed to compare the accuracy of Logistic Regression, Naïve Bayes, and KNN.

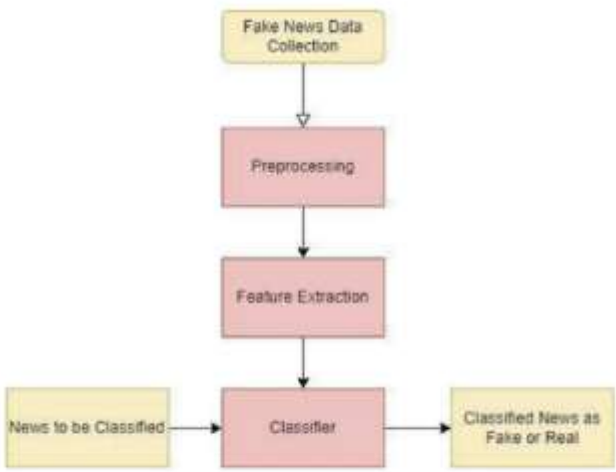

Figure 3.1 Flowchart of a Fake News Detection System

3.6 Fake News Prediction System

To extend the usability of this study, a prediction function was developed that allows users to input any news article and classify it as either fake or real. The function follows these steps:

1. **Preprocess the input text** using the same cleaning techniques as applied to the dataset.
2. **Convert the text into numerical format** using TF-IDF vectorization.
3. **Use the trained model (Logistic Regression)** to predict whether the input news is fake or real.
4. **Output the prediction result** (0 = Real News, 1 = Fake News).

This feature enables real-time fake news detection using machine learning.

4. RESULT AND ANALYSIS

4.1 Data Overview and Preprocessing

This study employs a dataset of news articles categorized as either authentic or fabricated. Authentic news consists of verifiable content, while fabricated news contains misleading or false information. Machine learning models are trained to differentiate between the two.

To enhance model performance, textual data undergoes preprocessing, which includes:

- **Text Cleaning –** Removal of URLs, HTML tags, numbers, special characters, and punctuation to retain meaningful words.
- **Tokenization –** Splitting text into individual words for analysis and classification.
- **Stopword Removal –** Eliminating common words (e.g., "the," "is," "and") to reduce noise and improve efficiency.
- **Text Normalization –** Converting text to lowercase for consistency and applying lemmatization to retain base word forms.
- **Feature Extraction (TF-IDF) –** Converting text into numerical representations to highlight word significance for model training.

These steps refine the dataset, improving the accuracy of fake news detection models by ensuring structured and relevant input for learning algorithms.

4.2 Class Distribution

The Class Distribution of Fake & Real News graph illustrates the ratio of genuine (0) to fabricated (1) news articles within the dataset. A balanced dataset mitigates model bias, facilitating precise classification. Given the nearly equal representation of both categories, no supplementary methods such as oversampling or undersampling are required, rendering the dataset optimal for training an effective false news detection model.

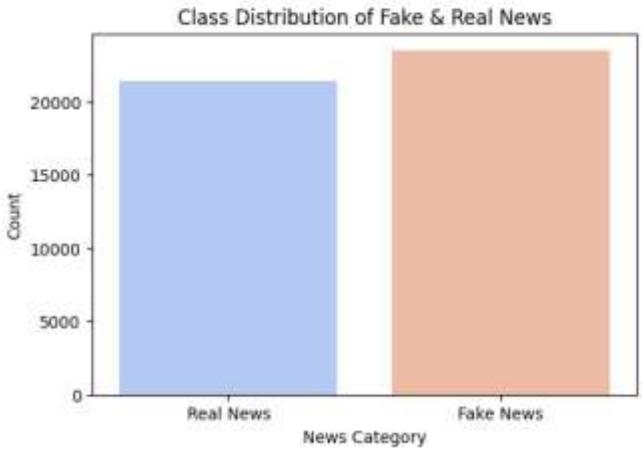

Figure 4.1 Class Distribution of Fake and Real News

4.3 Text Length Distribution

The Text Length Distribution graph illustrates the fluctuation in article word counts. Concise writings may be devoid of context, whilst lengthy ones may introduce superfluous information. The histogram facilitates the evaluation of distribution, guaranteeing equitable preprocessing for enhanced model correctness.

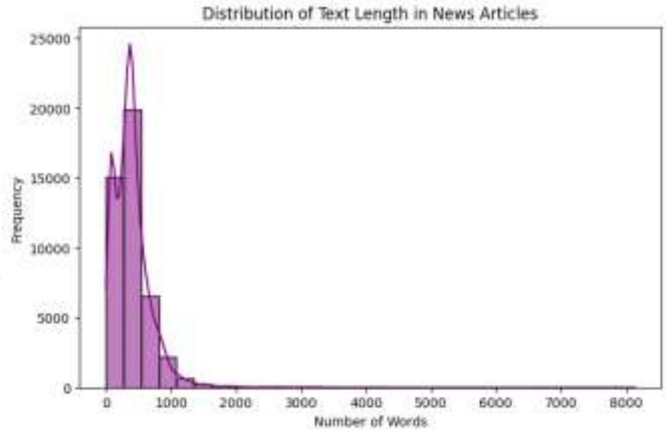

FIGURE 4.2 TEXT LENGTH DISTRIBUTION

4.4 Word Cloud for Fake and Real News

A Word Cloud visually depicts the frequency of words in both fake and authentic news stories, with larger words indicating higher usage. This facilitates the identification of essential terminology, trends, and language patterns distinctive to each category of news.

Word Cloud for Fake News

The Word Cloud for Fake News emphasizes often utilized terms in fraudulent stories, uncovering prevalent themes, sensationalist language, and misleading phrasing employed to deceive readers. This aids in recognizing linguistic patterns that differentiate false information from credible sources.

Figure 4.3 Fake News Word Cloud

Word Cloud for Real News

The Word Cloud for Real News emphasizes frequently utilized terms in reputable articles, indicating genuine reporting, authenticated sources, and impartial language. Analyzing the fake news word cloud facilitates comprehension of the structure of reliable news.

Figure 4.4 Real News Word Cloud

Analyzing these word clouds reveals significant vocabulary disparities, aiding in the identification of patterns that enhance the efficacy of false news detection models.

4.5 Most Common Words in Real and Fake News

A bar graph illustrating the most prevalent words in fake and real news underscores linguistic disparities. Examining word frequencies uncovers patterns, as misinformation frequently employs deceptive terminology, whereas authentic news adheres to a factual and impartial lexicon.

Most Common Words in Fake News

The bar graph of prevalent terms in fabricated news underscores linguistic trends frequently associated with sensationalism and disinformation. Recognizing these prevalent phrases enhances detection models and augments the identification of deceptive information.

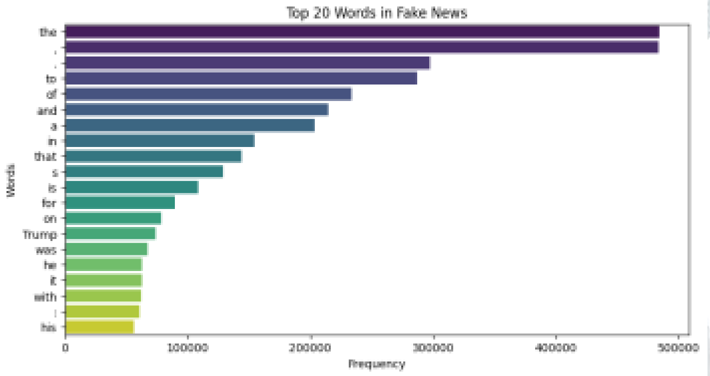

Figure 4.5 Top 20 words in Fake News

Most Common Words in Real News

The bar graph of prevalent terms in authentic news emphasizes words linked to genuine reporting and reliable sources. Examining these terms aids in differentiating genuine news from fabricated news, underscoring the significance of precise information.

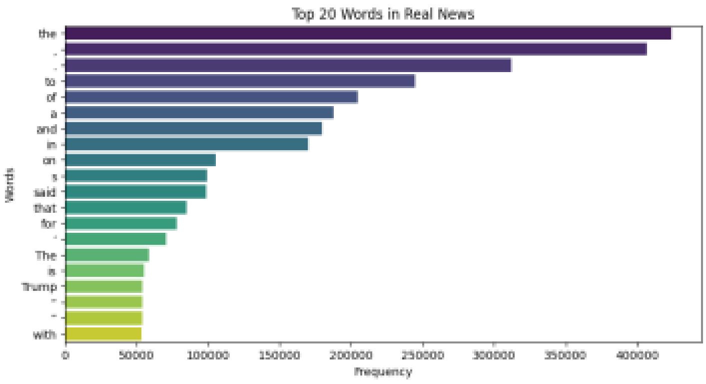

Figure 4.6 Top 20 words in Real News

4.6 Model Training and Evaluation

Model Training

Three machine learning models—Logistic Regression, Naive Bayes, and KNN—were trained for fake news detection. The dataset was preprocessed through text cleaning, tokenization, stopword removal, and TF-IDF feature extraction. It was then divided into 80% training and 20% testing data to ensure effective model validation.

Model Evaluation

The models were assessed based on accuracy, precision, recall, F1-score, and a confusion matrix to measure classification performance. The results are summarized in the table below:

✧ *Table 4.1: Model Performance Comparison*

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.9869	0.99	0.99	0.99
Naive Bayes	0.9248	0.92	0.94	0.93
K-Nearest Neighbors (KNN)	0.7073	0.65	0.97	0.78

4.7 Model Comparison

A bar graph illustrates the efficacy of Logistic Regression, Naive Bayes, and KNN in detecting fake news. Logistic Regression exhibited the highest performance, succeeded by Naive Bayes, however KNN demonstrated the lowest accuracy attributable to its computational complexity. This validates Logistic Regression as the most efficient model for this assignment.

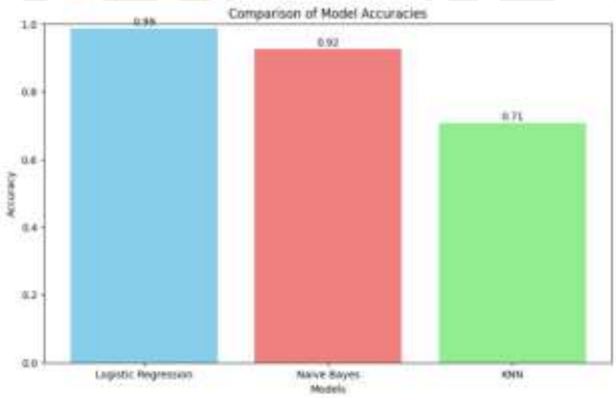

Figure 4.7 Model Comparison

5. CONCLUSION

The growing prevalence of disinformation on social media requires effective and efficient detection systems. This study examined the function of machine learning in detecting false information by utilizing classification models like Logistic Regression, Naïve Bayes, and K-Nearest Neighbors. The research effectively converted unstructured news content into significant numerical representations using rigorous data preprocessing, which encompassed text cleaning, tokenization, and TF-IDF vectorization. The assessment of these models through accuracy, precision, recall, and F1-score revealed that machine learning may substantially improve the precision of fake news identification.

Notwithstanding progress in automated false news detection, obstacles such as the evolution of disinformation strategies, dataset bias, and adversarial assaults persist. Addressing these difficulties necessitates ongoing enhancement of detection methodologies, the integration of real-time validation systems, and the adoption of explainable AI frameworks to augment transparency. Future study may concentrate on integrating various machine learning models with fact-checking databases and deep learning methodologies to improve detection precision and flexibility.

This research offers an automated, scalable method for identifying misinformation, thereby enhancing digital content verification initiatives and alleviating the adverse societal effects of fake news. Going forward, cooperation among technological specialists, legislators, and media entities will be crucial in formulating comprehensive methods to guarantee the dependability of online information.

REFERENCES:

[1]. Collins, B., Hoang, D.T., Nguyen, N.T., & Hwang, D. (2021). Trends in Combating Fake News On Social Media–a Survey. Journal of Information and Telecommunication, 5,2:247-266.

[2]. Yu, J., Huang, Q., Zhou, X., & Sha, Y. (2020). Iarnet: An Information Aggregating and Reasoning Network Over Heterogeneous Graph for Fake News Detection. Proceedings of 2020 International Joint Conference On Neural Networks (IJCNN), Glasgow, UK, Pp. 1-9.

[3]. R. Kozik, S. Kula, M. Choraś, And M. Woźniak, "Technical Solution to Counter Potential Crime: Text Analysis to Detect Fake News and Disinformation," J. Comput. Sci., Vol. 60, Apr. 2022, Art. No. 101576.

[4]. M. Ravinder, A. Jaiswal and S. Gulati, "Deep Learning-based Object Detection in Diverse Weather Conditions," International Journal Of Intelligent Information Technologies, Vol. 18, P. 1–14, March 2022.

[5]. S. R. Sahoo and B. B. Gupta, "Multiple Features Based Approach for Automatic Fake News Detection On Social Networks Using Deep Learning," Applied Soft Computing, Vol. 100, P. 106983, March 2021.

[6]. Arqoub, O.A.; Elega, A.A.; Özad, B.E.; Dwikat, H.; Oloyede, F.A. Mapping The Scholarship Of Fake News Research: A Systematic Review. J. Pract. 2022, 16, 56–86.

[7]. Lin, S.Y.; Kung, Y.C.; Leu, F.Y. Predictive Intelligence In Harmful News Identification By Bert-based Ensemble Learning Model With Text Sentiment Analysis. Inf. Process. Manag. 2022, 59, 102872.

[8]. García S. A., García G. G., Prieto M. S., Guerrero A. J. M., And Jiménez C. R., The Impact of Term Fake News On The Scientific Community Scientific Performance And Mapping In Web Of Science, Social Sciences. (2020) 9, No. 5.

[9]. MK Singh, J Ahmed, MA Alam, KK Raghuvanshi… - Multimedia Tools and …, 2024 – Springer

[10]. Raponi S, Khalifa Z, Oligeri G, Di Pietro R (2022) Fake News Propagation: A Review of Epidemic Models Datasets and Insights. ACM Trans Web 16(3):1–34

[11]. Andry Alamsyah and Asla Sonia. 2021. Information Cascade Mechanism and Measurement Of Indonesian Fake News. In 2021 9th International Conference On Information and Communication Technology (Icoict'21). 566–570.

[12]. E. Aïmeur, S. Amri, And G. Brassard, "Fake News, Disinformation and Misinformation in Social Media: A Review," Soc. Netw. Anal. Min., Vol. 13, No. 1, 2023, Doi: 10.1007/S13278-023-01028-5.

[13]. S. Tufchi, A. Yadav, And T. Ahmed, "A Comprehensive Survey Of Multimodal Fake News Detection Techniques: Advances, Challenges, And Opportunities," Int. J. Multimed. Inf. Retr., Vol. 12, No. 2, 2023, Doi: 10.1007/S13735-023-00296-3. DOI: Https://Doi.Org/10.1007/S13735-023-00296-3.

[14]. M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, And M. S. Rahman, "A Comprehensive Review On Fake News Detection with Deep Learning," IEEE Access, Vol. 9, 2021.

[15]. Y. Papanastasiou, "Fake News Propagation and Detection: A Sequential Model," Manage. Sci., Vol. 66, No. 5, 2020, Doi: 10.1287/Mnsc.2019.3295.

[16]. H. Thakar and B. Bhatt, "Fake News Detection: Recent Trends and Challenges," Soc. Netw. Anal. Min., Vol. 14, No. 1, P. 176, 2024.

[17]. Akilandeswari, J., Jothi, G., Dhanasekaran, K., Kousalya, K., & Sathiyamoorthi, V. (2022). Hybrid Firefly-ontology-based Clustering Algorithm For Analyzing Tweets To Extract Causal Factors. [Ijswis]. International Journal On Semantic Web and Information Systems, 18(1), 1–27.

[18]. Mani, S., & Annadurai, S. (2022). An Improved Structural-based Ontology Matching Approach Using Similarity Spreading. [Ijswis]. International Journal On Semantic Web and Information Systems, 18(1), 1–17.

[19]. Xia, B., Liu, W., He, Q., Liu, F., Pang, J., Yang, R., Yin, J. B., & Ge, Y. (2023). Binary Vulnerability Similarity Detection Based On Function Parameter Dependency. [Ijswis]. International Journal On Semantic Web and Information Systems, 19(1), 1–16.

[20]. Jones, M., Li, W.: Enhancing Fake News Detection with Graph-based Propagation Models. In: Proceedings of The 30th ACM SIGKDD Conference, Pp.

[21]. Ibrahim, N., Noor, M., Ibrahim, M.: Detecting Fake News In Low-resource Languages Using Transfer Learning. Inf. Process. Manag. 61, 102578 (2024).

[22]. Singh, R., Sharma, M.: Early Fake News Detection Using Big Data Analytics and Deep Learning Techniques. Expert Systems with Applications 213,118733 (2023)

[23]. Rashid, M., Asif, M., Khan, T., Mahmood, A.: A Hybrid Approach for Fake News Detection Using Machine Learning. IEEE Access 8, 21974–21985(2020).