Journal of Science, Computing and Engineering Research (JSCER)
Volume-7, Issue-10, October 2024.
DOI: *https://doi.org/10.46379/jscer.2023.07080*

Detecting Fake News Article and Images by Using Machine Learning Algorithm

V. Sriram,R. Subash, D. Sundar Rajan, K. Vimal Raj, R. Nivethitha

[1-4] UG Students, [5]Assistant Professor Computer Science and Engineering, K.L.N. College of Engineering, Sivagangai,India

Article Information

Received : 16 Oct 2024
Revised : 20 Oct 2024
Accepted : 21 Oct 2024
Published : 31 Oct 2024

Corresponding Author:
R. Nivethitha
Email: niverenga30@gmail.com

Abstract— The rampant spread of misinformation on platform like Facebook poses a critical challenge to public discourse. This proposed system focuses on detecting fake news and fake images using machine learning algorithms. To develop a robust machine learning- based system for detecting fake news specifically on Facebook. By leveraging techniques like Logistic Regression and image analysis models, the project aims to identify patterns that distinguish between real and deceptive content. The system processes textual and visual data to flag misinformation, enhancing the reliability of online information. The system utilizes a combination of text-based and image-based machine learning techniques to achieve this goal. For fake news detection, a Logistic Regression algorithm is employedto classify news articles as either real or fake. The model is trained on a label dataset containing various features extracted from the news articles, such as word frequency, sentiment analysis, and metadata. By learning from these features, the model can effectively identify patterns that are common in deceptive or misleading news content. On the image analysis front, the proposed system incorporates advanced image recognition and manipulation detection algorithms. Techniques like Convolutional Neural Networks (CNNs) are used to detect anomalies in images that may indicate forgery or tampering.

Keywords: ***Fake News Detection, Machine Learning, Logistic Regression, Convolutional Neural Networks (CNNs), Error Level Analysis (ELA).***

Citation: V. Sriram,R. Subash, D. Sundar Rajan, K. Vimal Raj, R. Nivethitha, "Detecting Fake News Article and Images by Using Machine Learning Algorithm", Journal of Science, Computing and Engineering Research, 7(10), October 2024.

I. INTRODUCTION

In the digital age, the rapid proliferation of information has fundamentally transformed how users engage with content online. Content-based recommender systems have become a cornerstone of this transformation, designed to curate and deliver personalized information to users based on their previous interactions. These systems utilize algorithms to continuously showcase content in users' feeds that closely resembles previously viewed items or those with which the user has engaged, such as through likes or comments. While these algorithms enhance user satisfaction and retention, they inadvertently present a significant opportunity for the dissemination of fake news [1]. Once a user encounters a piece of misleading information, the recommender system perpetuates the cycle of misinformation by suggesting similar content, regardless of its veracity [2].

Furthermore, these systems often incorporate mechanisms to diminish the visibility of certain posts, leading to an ecosystem where undesirable content is consistently masked rather than eliminated [3]. This raises critical ethical questions about the responsibility of platform providers in managing the information environment. Historically, the academic focus on fake news detection has predominantly centered on linguistic and compositional characteristics. Researchers have sought to identify fake news by examining various parameters, including the presence of identifiable authorship, the credibility of sources, and the overall length of articles [4], [5], [6], [7].

This approach operates under the assumption that clear distinctions exist between the linguistic features of fake and factual news, allowing for the identification of unreliable sources. However, this methodology has significant limitations. It fails to account for the complex characteristics ofusers who engage with or propagate fake news, as well as the intricate

dynamics of the social media networks that facilitate the spread of such information.The rise of sophisticated AI models, such as ChatGPT (Generative Pre-trained Transformer), has further complicated the landscape of fake news detection.

These advanced language models can generate coherent, stylistically appropriate text that closely resembles professional journalism, thereby obscuring the lines between authentic reporting and fabricated narratives [4]. As a result, the traditional reliance on linguistic features for detecting fake news is increasingly inadequate.

The ease with which misleading content can now be generated means that users are often unable to distinguish between news articles produced by AI and those crafted by experienced journalists. This shift necessitates a reevaluation of detection methodologies to effectively combat misinformation. In response to these challenges, this study proposes a comprehensive fake news detection model that moves beyond conventional techniques by integrating an analysis of not only the visual characteristics of content but also the nuanced behaviors of users who generate and disseminate fake news.

The sophistication of fake news generation has escalated, particularly with advancements in AI technology that allow for the rapid creation of articles that closely mimic real news. For instance, AI-driven bots deployed on social media platforms like Twitter can create a vast array of user accounts capable of generating support or opposition for specific narratives, effectively obfuscating the authenticity of the information presented [8]. This manipulation of public discourse complicates the landscape of news consumption and amplifies the need for innovativedetection strategies.

To effectively address the limitations of existing research, this study aims to enhance the predictive performance of fake news detection by considering the characteristics of information recipients and the specific contexts in which misinformation proliferates. By constructing a detection model that incorporates various content features, user characteristics in social media, and the networks through which fake news is propagated, we seek to offer a more holistic approach to understanding and identifying misinformation.

Central to our methodology is the concept of feature selection, which serves as a critical step in optimizing the detection process. We employ Boost (Extreme Gradient Boosting) to identify the most significant explanatory variables for effective fake news detection. The model is developed through the application of five distinct machine learning techniques: Logistic Regression (LR), Neural Networks (NNET), Random Forest (RF), Support Vector Machines (SVM), and Classification and Regression Trees (CART). Each model is rigorously evaluated, allowing us to derive the one with the highest predictive performance rate in accurately identifying fake news.

By advancing the detection of fake news through a multifaceted approach, this study aims to contribute significantly to the field of misinformation research. Our work not only addresses the shortcomings of traditional detection methods but also offers valuable insights into the interplay between content characteristics, user behavior,and social media dynamics. In doing so, we hope to providea framework that enhances the accuracy of fake news detection, ultimately fostering a more informed public discourse.

II. RELATED WORKS

The rise of social media has led to an increase in misinformation, prompting various studies on fake news detection. Previous research primarily focused on linguistic features, employing methods such as keyword frequency analysis and sentiment analysis to differentiate between fake and real news articles [1]. For instance, some studies identified fake news based on the presence of sensational language and emotional appeal [3, 4].

Recent advancements in machine learning have shifted the paradigm toward automated detection. Techniques such as Logistic Regression and Random Forest have been widely used, with models achieving notable accuracy in classifying news content [2, 5]. The introduction of deep learning methods, particularly Convolutional Neural Networks (CNNs), has further improved detection capabilities by allowing modelsto learn intricate patterns in text data [6].

Furthermore, the role of user engagement in the propagation of fake news has gained traction. Studies suggest that user behavior, including likes and shares, can be indicativeof content credibility [7]. This user-centric approach provides insights into how misinformation spreads and how it can be mitigated. Despite these advancements, many existing methodsstill rely heavily on textual features and often overlook the dynamics of social media networks

III. PROPOSED SYSTEM

A comprehensive fake news detection system is proposed to identify and mitigate the spread of misinformation across social media platforms. The system aims to enhance the accuracy of news verification by integrating both text analysis and image recognition techniques, providing a dual-layered approach to content validation. By leveraging advanced machine learning algorithms, the system not only analyzes textual content but also

Journal of Science, Computing and Engineering Research (JSCER)
Volume-7, Issue-10, October 2024.
DOI: *https://doi.org/10.46379/jscer.2023.07080*

evaluates accompanying images, thereby improving the overall reliability of the detection process.

Advantages of the Proposed System:

- Multimodal Analysis: The system combines text and image

Figure 1: Frameworks and Libraries used in Python

a nalysis to provide a more robust assessment of news authenticity.

- Real-time Detection: Users receive immediate feedback on the authenticity of content, enabling quick decisions regarding the information they consume.
- User-Friendly Interface: For Entering Articles and post, Images Data
- Educational Tool: The system provides explanations for detection outcomes, helpingusers understand why a piece of content is classified as real or fake.

A. Hardware Requirements

Computer or laptop with GPU Support:

A computer equipped with a powerful Graphics Processing Unit (GPU) is essential for efficiently running the TensorFlow library, enabling faster model training and inference.

B. Software Requirements

1. Python Environment:

The primary programming language for implementing the fake news detection algorithms.

Essential libraries include:

NumPy: For efficient numerical computations.

Pandas: For data manipulation and analysis.

Scikit-learn: For implementing various machine learning algorithms.

TensorFlow: For building and training deep learning models.

PIL (Pillow): For image processing tasks.

2. Integrated Development Environment (IDE):

Google Colab, short for Colaboratory, is a cloud-based platform by Google that enables users to write, execute, and share Python code within a Jupyter Notebook environment. It's especially popular for tasks like machine learning, data science, and deep learning, as it provides free access to GPUs and TPUs, offering significant computational power without the need for expensive hardware.

Colab integrates seamlessly with Google Drive, making it easy to store and share notebooks, and supports real- time collaboration, allowing multiple users to work on the same notebook simultaneously. It also includes popular libraries like TensorFlow, PyTorch, and pandas pre- installed, making it a powerful tool for bothbeginners and experts in the field.

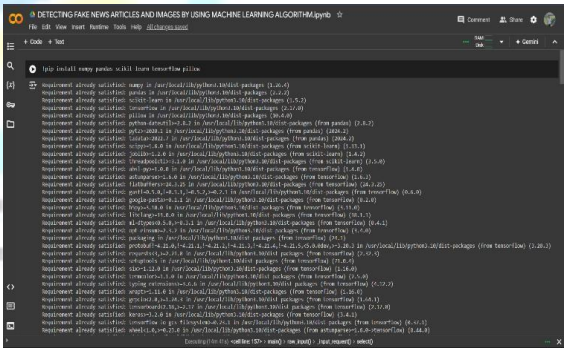
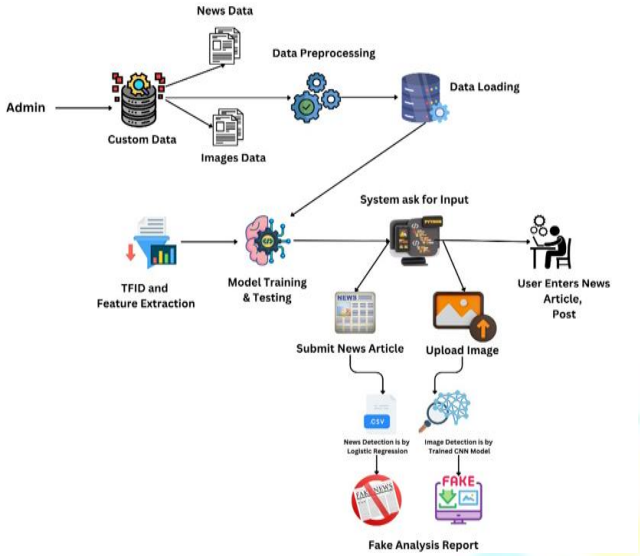


Figure 2: Google Collaboratory - Integrated Development Environment

Journal of Science, Computing and Engineering Research (JSCER)
Volume-7, Issue-10, October 2024.
DOI: *https://doi.org/10.46379/jscer.2023.07080*

3. System Architecture



The proposed system architecture consists of three main components:

Data Collection Input:

Users can input text and upload images for analysis. The system retrieves this data and prepares it for processing.

Figure 3: User Input for News or Article Data

Data Processing Unit:

Text Analysis Module: Utilizes Natural Language Processing (NLP) techniques, such as TF-IDF vectorization and machine learning classifiers (e.g., Logistic Regression, Support Vector Machines), to analyze the textual content for indicators of fake news.

Image Analysis Module: Employs convolutional neural networks (CNNs) to evaluate uploaded images against a database of known fake images, detecting inconsistencies and anomalies.

Figure 5: Output of Fake news and Image

Result Layer:

Provides users with a report indicating whether the content is classified as real or fake. Explanations based on identified keywords, textual features, and visual characteristics are presented for transparency.

D. Workflow

1. The user submits a news article and optionally uploads an image for analysis.

Journal of Science, Computing and Engineering Research (JSCER)
Volume-7, Issue-10, October 2024.
DOI: *https://doi.org/10.46379/jscer.2023.07080*

2. The system processes the text and image through the respective modules, applying algorithms to evaluate authenticity.

3. Results are generated, indicating whether the content is classified as real or fake, along with detailed explanations.

4. Users can continue analyzing other posts or exit the system after reviewing the findings.

IV. IMPLEMENTATION

The implementation of the proposed Fake News Detection System is a multi-layered approach that combines text analysis, image analysis, and machine learning algorithms to accurately detect and mitigate the spread of misinformation. The core implementation can be divided into several key phases: text preprocessing, feature extraction, classification using various machine learning models, and image verification using convolutional neural networks. The first step in thesystem is to process the textual content of the news articles. The system starts with text preprocessing, which cleans and standardizes the raw text for analysis. The news article is broken down into tokens using tokenization, which splits the text into individual words or terms.

Each token is then subjected to lemmatization, where words are reduced to their root forms. For running" becomes "run" and "children" becomes "child." This process eliminates grammatical variations and ensures uniformity. Additionally, stop word removal is performed, where common words like "is," "the," and "and" are removed as they do not contribute to distinguishing between real and fake news.

Once the text is preprocessed, the next step is to transform the cleaned text into a format suitable for machine learning models. TF-IDF (Term Frequency- Inverse Document Frequency) is used for this transformation, a common technique in text mining to convert the text into a matrix of numerical values. The formula for TF-IDF is as follows:

Where:

TF(t,d) represents the term frequency, or the numberof times a term t appears in a document d.

N is the total number of documents in the dataset.

DF(t) is the document frequency, or the number of documents in which the term t appears.

The transformation process begins by applying the Term Frequency-Inverse Document Frequency (TF-IDF) technique,which generates a weighted matrix that encapsulates the significance of individual words across the documents in the dataset.

This method calculates two crucial metrics: Term Frequency (TF), which measures how frequently a term appears in a document relative to the total number of terms in that document, and Inverse Document Frequency (IDF), which assesses how important a term is by considering the number of documents in which it appears: Next, feature extraction is performed on the TF-IDF matrix to identify key characteristics of the text. Key features extracted include:

Sentence Length (L): Fake news often contains longer sentences, designed to convey complex or misleading information. The system calculates the average sentence length in the article.

Exaggerated Words (W): Fake news tends to usesensational or exaggerated language. The frequency of such words is calculated relative to the total word count.

$$W = \frac{\text{Number of exaggerated words}}{\text{Total word count}}$$

Source Credibility (S): This feature flags articles thatlack credible sources. News articles without verifiable sources or with suspicious authorship are flagged for furtheranalysis.

After feature extraction, the system moves to the classification phase. To classify news articles as fake or real, several machine learning models are implemented. The following models are trained on a labeled dataset and are fine-tuned for optimal accuracy:

Logistic Regression (LR):

Logistic regression is a binary classification algorithmthat models the probability that a news article belongs to one of two categories (fake or real). The formula for logistic regression is:

$$P(y = 1|X) = \frac{1}{1 + e^{-z}}$$

Where $z=\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn$, and P(y=1) is the probability that the news article is classified as fake.

Support Vector Machine (SVM):

SVM works by finding the hyperplane that maximizesthe margin between fake and real news in a high- dimensional space. The objective is to minimize classification errors. The decision boundary is defined as:

Journal of Science, Computing and Engineering Research (JSCER)
Volume-7, Issue-10, October 2024.
DOI: *https://doi.org/10.46379/jscer.2023.07080*

$$f(x) = \mathbf{w} \cdot \mathbf{x} - b = 0$$

Comparison of three double tail comparator circuits being done. All the circuits simulated by using cadence design tools 90nm technology with the supply voltage of 0.6 volt. Using the power gating technique, power consumption and delay is reduced in the proposed double tail comparator. The proposed double tail comparator consumes less power and delay is also reduced compare to previous comparator circuits. Due to additional NMOS transistors there is an increase in area. The comparator circuit used in analog to digital converter structures, sense amplifier, operational amplifier and pre defined amplifier.

3. Random Forest (RF):

Random Forest is a method that constructs multiple decision trees and aggregates their outputs. The classification is based on a majority vote from all trees. The system ensures robustness by comparing the results from different models. Based on accuracy, the optimal classifier is selected for final deployment. In addition to text-based analysis, the system incorporates image verification using Convolutional Neural Networks (CNNs) to further validate theauthenticity of news articles. Many fake news articles use misleading images to manipulate perception. The CNN model is trained on a large dataset of real and fake images and is capable of detecting manipulated or misleading visuals by comparing images to a database of verified images. The architecture of the CNN involves multiple layers such as convolutional layers, ReLU activation functions, and pooling layers to extract spatial features from the images. Once the system has processed both the text and the image data, itaggregates the results from both layers to make a final decisionon whether the article is fake or real. If either the text or the image is flagged as suspicious, the article is classified as fake. The final classification result is displayed on a console- based interface, where users can see detailed feedback, including which features contributed most to the classification This comprehensive implementation ensures that fake news is detected based on both content and context, providing a robust tool for mitigating the spread of misinformation on social media.

V. RESULTS

The proposed system provides real-time analysis and classification of various content types, including Facebook posts, news articles, blog posts, and images, directly in the console. For text-based content, such as news articles and blog entries, it identifies potential misinformation by examining linguistic patterns, the credibility of sources, and markers like exaggerated claims or sensationalist headlines. For images, thesystem employs image recognition techniques to detectmanipulation signs, such as pixel inconsistencies or alterations in metadata. It cross-references images with known

VI. CONCLUSION

In today's digital age, the spread of fake news has become a significant challenge, impacting public opinion and trust in media. As misinformation proliferates, it is crucial to develop robust detection systems that can accurately identify and mitigate this issue. Our proposed fake news detection system addresses this need byleveraging advanced text analysis and image recognition techniques. By examining the linguistic features of articles,blog posts, and social media content, along with employing image verification methods, we can effectively classify content as genuine or misleading. This system not only aids in combating misinformation but also promotes a more informed society. With technology at the forefront, our approach provides a proactive solution to an urgent problem, ensuring that users are equipped with the necessary tools to discern the truth in an ever-evolving digital landscape.

VII. FUTURE WORKS

In the future, this project can be expanded to include features such as real-time notification alerts for users regarding the authenticity of news articles and images. Additionally, integrating a browser extension could allow users to receive instant feedback on the credibility of content they encounter online. Implementing a user- friendly dashboard for personalized tracking of misinformation trends will also enhance user engagement. Moreover, collaborating with social media platforms to flag or remove fake content before it spreads cansignificantly impact the overall effectiveness of the system.

REFERENCES

[1]. H. L. Gururaj, H. Lakshmi, B. C. Soundarya, F. Flammini and V. Janhavi, "Machine Learning-Based Approach for Fake News Detection," in Journal of ICT Standardization, vol. 10, no. 4, pp. 509-530, 2022, doi: 10.13052/jicts2245-800X.1042.

[2]. Khalil, A., Jarrah, M. & Aldwairi, M. Hybrid Neural Network Models for Detecting Fake News Articles. Hum-Cent Intell Syst 4, 136–146 (2024). doi.org/10.1007/s44230-023-00055-x

[3]. Abdullah All Tanvir, E. M. Mahir, S. Akhter and M. R. Huq, "Detecting Fake News using Machine Learning and Deep Learning Algorithms," 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, 2019, pp. 1-5, doi: 10.1109/ICSCC.2019.8843612.

[4]. S. Chaturvedi, A. Saxena, A. Gupta and F. Alam, "Detecting Fake News using Machine Learning Algorithms," 2021 3rd

International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida,India,2021,pp.212-216,doi: 10.1109/ICAC3N53548.2021.9725560.

[5]. R. R. Mandical, N. Mamatha, N. Shivakumar, R. Monica andA. N. Krishna, "Identification of Fake News Using Machine Learning," 2020 IEEE International Conference on Electronics, Computing and Communication

[6]. M. Q. Alnabhan and P. Branco, "Fake News Detection Using Deep Learning: A Systematic Literature Review," in IEEE Access, vol. 12, pp. 114435-114459, 2024, doi: 10.1109/ACCESS.2024.3435497.

[7]. D. K. Sharma and S. Garg, "Machine Learning Methods to identify Hindi Fake News within social-media," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-6,

[8]. V. Gupta, R. S. Mathur, T. Bansal and A. Goyal, "Fake News Detection using Machine Learning," 2022 International Conferen ce on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 84-89, doi: 10.1109/COM-IT- CON54601.2022.9850560.

[9]. 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," in IEEE Access, vol. 9, pp. 156151-156170, 2021,

[10]. M. Park and S. Chai, "Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques," in IEEE Access, vol. 11,pp. 71517-71527, 2023,

[11]. Velivela Durga Lakshmi, Ch Sita Kumari. Detection of Fake News using Machine Learning Models. International Journal of Computer Applications. 183, 47 (Jan 2022), 22-27. DOI=10.5120/ijca2022921874J.

[12]. Mandical, Rahul R et al. “Identification of Fake News Using Machine Learning.” 2020 IEEE International Conference on Electronics, Computing and Communication Technologies(CONECCT) (2020): 1-6