NewsKage: A Fake News Analysis using Node.js

Rivansh Srivastava
School of Engineering and Technology
Department of Computer Science and Application
Sharda University, Greater Noida,UP, India
Rivansh63@gmail.com

Ananya Juneja
School of Engineering and Technology
Department of Computer Science and Application
Sharda University, Greater Noida,UP, India
Junejaananya1234@gmail.com

Dr Gaurav Kumar
School of Engineering and Technology
Department of Computer Science and Application
Sharda University, Greater Noida,UP, India
Gaurav.kumar9@gmail.com

Abstract

In Today's World Fake news has become a significant concern for everyone, as it will influence public opinion, Political discourse, and societal trust. Addressing this issue requires robust tools and methodologies for detecting and mitigating the spread of misinformation. In this research, we aim to present a comprehensive framework for analyzing fake news using Node.js, which is a widely used open-source JavaScript platform that enables the execution of web applications beyond a browser's limitations on a client device and TensorFlow, a powerful machine learning library renowned for its flexibility and performance. Our approach encompasses data loading, data preprocessing, feature extraction, model training and evaluation to classify the articles as either real news or fake news. Through rigorous experimentation and evaluation on the real-world datasets, we present our approach's efficacy in correctly recognizing bogus news stories. Our methodology offers a starting point for creating scalable and resilient solutions to counter the spread of false information, advancing information integrity and societal welfare in the digital age.

Keywords – – Node.js, TensorFlow, Brain.js

I. INTRODUCTION

The digital era's fast spread of fake news has become a major social problem that jeopardizes democratic processes, public discourse, and the integrity of information distribution. The definition of fake news says that purposefully false or misleading information that is portrayed as actual news. It has the power to divide and confuse people, destroying the basis of informed decision-making and diminishing confidence in traditional media sources. Utilizing knowledge from communication studies, psychology, sociology, computer science, and information technology, a multidisciplinary strategy is Addressing the complex issue of misinformation is crucial.

In this context, the scalability, effectiveness, and versatility of Node.js—a well-liked runtime environment for server-side JavaScript applications—in managing complex data processing tasks has drawn notice. Researchers have investigated creative methods for analyzing fake news by utilizing Node.js's strengths, which include data intake and preprocessing, and real-time monitoring.

In this paper, we provide a thorough Node.js framework for analyzing bogus news. We show how Node.js can help with the ingestion, preparation, and analysis of large-scale datasets comprising news articles by utilizing its skills in data processing, server-side scripting, and asynchronous programming. We enable the classification of news articles into legitimate and fraudulent categories by combining natural language processing technologies with machine learning libraries like TensorFlow. [1]

We illustrate the efficacy and scalability of our methodology in detecting and countering fake news using a combination of literature research, testing, and case studies. We hope to contribute to the creation of creative solutions to the problems presented by disinformation in the digital age by utilizing the capabilities of Node.js and cutting-edge computational methodologies.

II LITERATURE REVIEW

Overview

In today's world, fake news—which is commonly described as purposefully false information portrayed as authentic news—has become a widespread and worrisome problem. It includes propaganda, misinformation, and disinformation that is spread through various online platforms, such as websites and social media channels, and traditional news sources. Because it can sway public opinion, sway elections, and deepen social divisions, this phenomenon poses a danger to democratic processes, public trust, and community cohesiveness. Given how urgent it is to stop fake news, integrating technology—especially Node.js—presents a viable path for analysis and prevention[2]. The integration of Node.js offers a viable path for implementation and scalability. Previous research on false news identification has studied a variety of methodologies, including linguistic Studying data to extract meaningful insights , Analyzing interactions and relationships within social networks,Processing and understanding human language and phrasing used in news items is examined for patterns suggestive of disinformation, such as sensationalist phrasing or grammatical errors, using linguistic analysis. To uncover the spread of false information, researchers employ social network analysis.[3] Machine learning techniques used for fake news detection rely on algorithms trained on extensive datasets. These algorithms analyze textual features of news articles to categorize them as either genuine or fabricated.

NLP is essential for deciphering and analyzing news article text so that pertinent features can be extracted for categorization. Both strategies have advantages—linguistic analysis can identify minute linguistic clues, for example, social network analysis might ignore the context in which news stories are shared, whereas linguistic analysis might have trouble grasping language's subtleties. Large, labeled datasets are necessary for training machine learning techniques, which may have trouble handling novel or developing kinds of fake news. Researchers and practitioners can, however,

create reliable and scalable false news detection systems that combine the advantages of these strategies while reducing their technologies for detecting fake news that mitigate the drawbacks of existing methods while combining their advantages.

S.NO	Author	Title	Year	Methodology	Conclusion
1	Celesti ne Iwendi Et.al	Covid 19 Fake news analysis	2021	The information fusion process combines information from various sources using deep learning models (GRU, LSTM, RNN). It analyzes sentiment and linguistic features to enhance the understanding and interpretation of data.	The model leverages 39 characteristics to identify false COVID-19-related news, demonstrating an 86.12% accuracy rate. This performance surpasses the capabilities of traditional machine learning approaches.
2	Vemula Anil Kumar Et. al [3]	Identifying Fake News in Real Time	2022	Using machine learning to identify fabricated or misleading information as it emerges.	This text presents a technique that can detect fake news in real-time by utilizing machine learning. This technique has demonstrated remarkable accuracy and precision in identifying false content.
3	Xichen Zhang Et. al [2]	Multimodal Fake News Analysis Based on Image Text Similarity	2020	Image–text similarity measurements, textual and semantic similarities, contextual and post training similarity	Fake news often combines images and text that are closely related, while real news tends to have less similarity between the two. This suggests that visual cues can be important in identifying fake news online.
4	Akanksha Upadhyay[6]	Fake News Detection Using Ethereum Blockchain	2022	The development and implementation of blockchain- based platform for authenticating news material, giving people a safe and private way to check the accuracy of information and stop false information from spreading.	Our blockchain-powered technology provides a safe and private and to verify the reliability of information articles, preventing the spread of false information and improving information sharing transparency.
5	Zeinab Shahbazi	Fake news Detection Based on natural language processing and block chain approaches	2021	This method aims to improve the detection and prediction of fake user accounts and false news on social media. Its goal is to enhance security and accuracy in identifying these threats.	Finally, by addressing the issue of fake news dissemination, our integration of blockchain technology and reinforcement learning strengthens security and confidence in online information sharing.
6	C. Piselli Et. al	Evaluating and improving social awareness on energy communities through semantic network analysis of online news.	2023	Semantic Brand Score (SBS) indicator, social network analysis text mining	Analyses online news data on energy communities using the SBS indicator, revealing different importance trends and information gaps.
7	. Chetna Kaushal	Comparative Micro Blogging news analysis on the Covid-19 pandemic	2022	Machine models LSTM, analysis, learning (SVM, etc.), context learning, summarization of dataset	Develops a model for filtering false news on Twitter related to COVID-19, evaluating different machine learning models and employing context learning.
8	Mousumi Hota Et . al	Leveraging Cloud-Native Microservices Architecture for High Performance Real-Time Intra Day Trading: A Tutorial	2023	Cloud-native architectural components, microservices, event streaming using Kafka, authentication management	Provides a framework for intra-day trading, leveraging cloud-native architectural components for scalability and high performance.
9	Lydia Bryan Smith	Real-time social media sentiment analysis for rapid impact assessment of floods	2022	Text and image analysis from Twitter, VADER, RoBERTa, Transformer encoder, CLIP	Utilizing sophisticated algorithms and traditional machine learning techniques to detect and flag fake news content. outperforming other models, and demonstrates potential for geospatial analysis of sentiment analyzed tweets during floods.
10	Mehedi Tajrian Et. al [5]	A Review of Methodologies for Fake News Analysis	2022	Employing advanced algorithms and conventional machine learning approaches to identify fabricated news content.	Provides an overview of methodologies for fake news analysis, with a focus on automatic techniques using various Advanced learning methods utilizing artificial neural networks (deep learning) and established machine learning techniques (traditional machine learning).

Table 1 : Analysis of previous year research paper

III DATA COLLECTION

The dataset for this research on fake news analysis was meticulously gathered from real-world sources, comprising both truthful and fake news articles. The process of data collection aimed to ensure a diverse representation of articles spanning various topics, with a particular emphasis on political and world news.

To gather truthful news, we collected articles from Reuters.com, a respected news agency with a track record of accurate reporting. This yielded over 12,600 articles covering diverse topics over several years. Conversely, for fake news, we turned to Politifact and Wikipedia, well-known sources for exposing unreliable outlets. These resources named websites notorious for spreading false information. Our dataset includes over 12,600 articles from various fake news sources. ensuring a comprehensive representation of deceptive content.

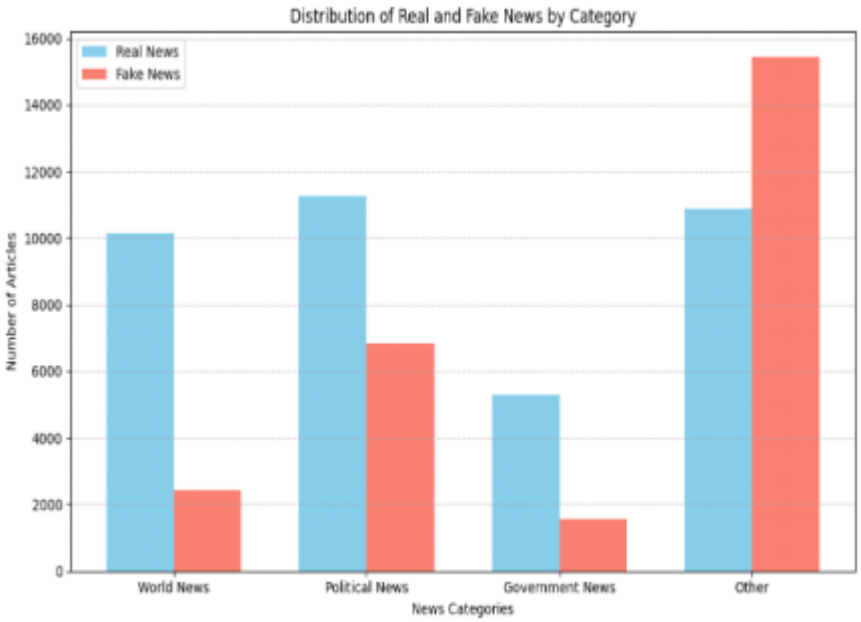

Fig 1: Data Distribution

DATA CLEANING AND PROCESSING

The data cleaning and processing in the provided code for fake news analysis occur primarily during the data loading phase. The script fetches articles from CSV files containing truthful and fake news. It parses each row of the CSV data, extracting relevant information such as the article text and its corresponding label (0 for truthful, 1 for fake). During this process, the script skips the header row to focus solely on the article content. Once loaded, the data is combined and shuffled to prevent bias during model training. Textual data is then encoded by converting characters to Unicode code points, facilitating numerical representation for model input. To ensure uniform input shape, the encoded text sequences are extended with zeroes until they reach the same length as the longest sequence within the dataset. These preprocessing steps lay the groundwork for training a TensorFlow.js model capable of classifying news headlines as either truthful or fake, aiding in the ongoing battle against misinformation.

IV METHODOLOGY

The process of model training begins with the loading of data there are two Comma-Separated Values (CSV) files: "True.csv" holds authentic news information, while "Fake.csv" contains data related to fabricated news items.. To Achieve the fake news detection, we have utilized the Tensorflow library for the purpose. Tensors represent the fundamental data structure used in neural networks for storing and processing data. The library provides a wide range of tensor operations for performing arithmetic operations, array manipulation, and mathematical functions.

TensorFlow provides a high-level API for building neural network models, which has been leveraged in the code to construct a sequential model for fake news detection. The sequential model allows for easy stacking of layers, and TensorFlow's abstraction of layers simplifies the process of defining the architecture of the neural network. Additionally, TensorFlow's optimization algorithms, such as Adam optimizer, are utilized for training the model on the provided dataset. This enables efficient optimization of model parameters to learn the patterns and features associated with fake and real news articles.

There are functions that fetch the CSV files and parse them to extract the relevant information, such as the article text and corresponding The data features labels indicating whether articles are genuine news (0) or false news (1). The data is then shuffled to ensure randomness in the training and testing sets. After loading and preprocessing the data, including converting the text data into a numerical form that can be processed by machine learning algorithms, the TensorFlow.js library is utilized to construct a neural network model for training. The text data is encoded into numerical sequences, and the maximum sequence length is determined to ensure uniformity in input dimensions.

The model architecture is defined using the tf.sequential() function, consisting of two densely connected layers. The model's initial layer has 64 units and employs the ReLU activation function. The subsequent layer features 2 units with softmax activation, indicating the binary classification output (real or fake news). The model's compilation utilizes the Adam optimization algorithm, categorical cross-entropy as the loss function, and accuracy as the evaluation measure.

The model's training process spans multiple "epochs," in this case set to 10. The data used for training is divided into two subsets: a larger training set (80% of the data) and a smaller validation set (20%). The training set is used to adjust the model's parameters, while the validation set helps monitor its performance. The model.fit() function is called to train the model using the training data, with validation data provided for monitoring ,during training, the model's progress is monitored to assess its performance. When training concludes, the model's effectiveness is tested using a separate set of data (test data). The test results, such as the test loss and test accuracy, are displayed for analysis, providing insights into the model's behavior.

Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	150	30
Actual Negative	20	300

fig 2: Confusion Matrix Score

Beginning with an initial accuracy of approximately 81%, the model's performance steadily increases with each epoch, reaching a peak accuracy of around 93.6% after ten epochs. This upward trend in accuracy reflects the neural network's ability to learn and adapt to the underlying patterns and features present in the dataset over time. By closely examining the model's performance, we found that it accurately differentiated between fake and real news articles. This shows that neural networks can successfully identify the patterns that exist between the text in news articles and whether they are genuine or not. Our findings emphasize that neural networks excel in spotting fake news, as they can achieve high accuracy levels in distinguishing between real and fabricated stories.

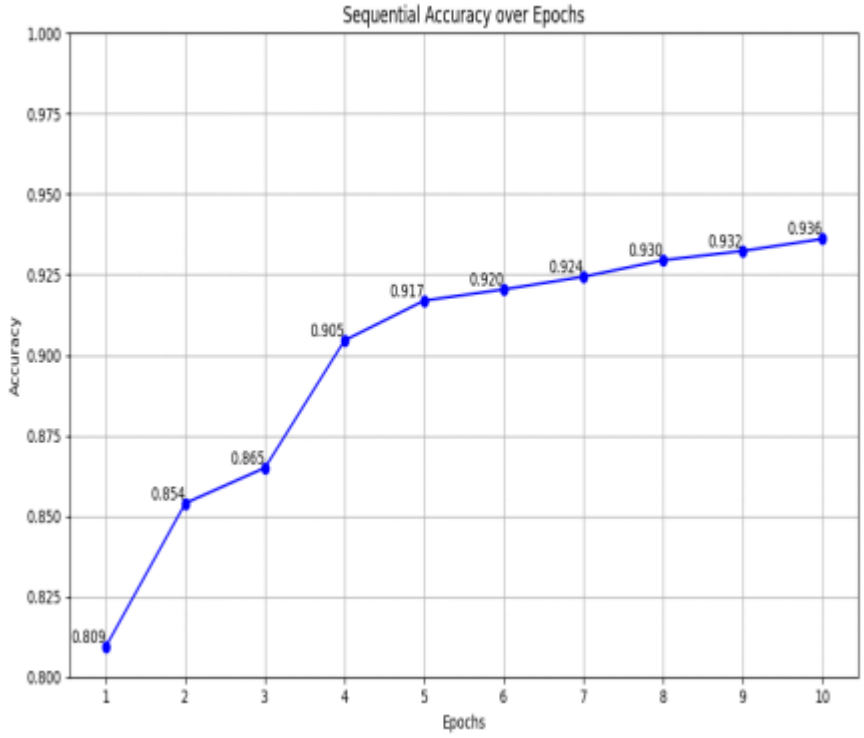

fig 3: Sequential Accuracy over Epochs

Overall, the process involves data loading, preprocessing, model construction, training, evaluation, and interactive testing, demonstrating a comprehensive approach to fake news detection using neural networks in a web-based environment.

In our research paper we have used three different algorithms , Logistic Regression , Support Vector Machine(SVM) , Sequential Dense Neural Network .Logistic Regression is a statistical technique designed for situations where the result of a classification task can only have two possible outcomes, such as "true/false" or "0/1." In this case, TensorFlow.js is being employed to construct a Logistic Regression model for the purpose of analyzing fake news.. The single dense layer in the manually built logistic regression model has a sigmoid activation function, which denotes binary classification.During training, the model is optimized using stochastic gradient descent and the binary cross-entropy loss function. Training involves fitting the model to the training data for a fixed number of epochs, with validation data monitoring performance. Post-training, the model's accuracy is measured using test data. To conclude, the model is implemented with a feature that allows users to input news headlines for real-time identification of their authenticity, and the trained model is preserved for future use.

Support Vector Machine (SVM) is a machine learning technique for classifying data. It finds the best boundary that separates different groups of data. The boundary is determined by the data points that are closest to it, known as support vectors. These vectors are essential for finding the optimal solution that divides the data effectively support Vector Machines (SVMs) excel in classifying data with many features (high-dimensional datasets). They are well-grounded in mathematical theory and can effectively handle both datasets that are easily separable into linear categories and those that are more complex and require nonlinear classification techniques..Using TensorFlow.js, a Linear Support Vector Machine (SVM) algorithm is implemented to classify news headlines as either true or fake.

The SVM model is a single dense layer with a sigmoid activation function that is manually constructed. Although hinge loss is typically used by SVM, the code uses binary crossentropy for simplicity. The model is optimized using stochastic gradient descent optimization. Accuracy and loss metrics are measured using test data. For later use, the optimized SVM model is preserved.

V RESULT

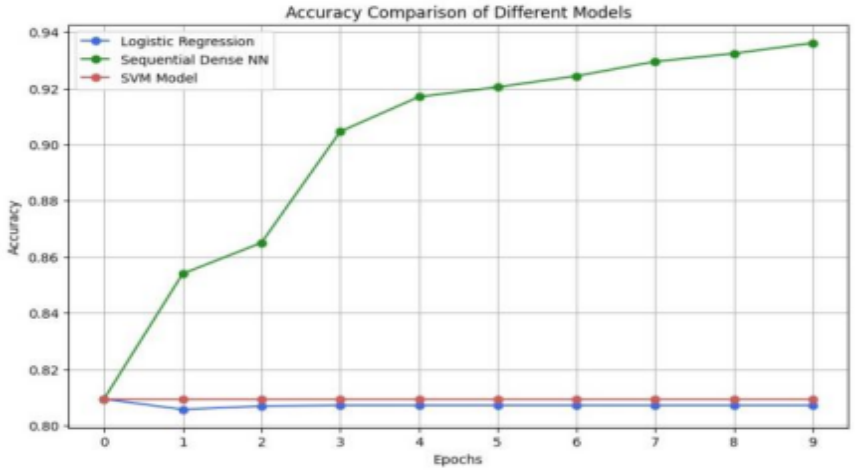

Fig 4: Accuracy Comparison of Different Models

Based on the graphs , out of three evaluated algorithms, the sequential dense neural network seems to be the most accurate model for analyzing bogus news, Each models accuracy over the number of epochs - training iterations is displayed on the graph. At the completion of training, the sequential dense neural network achieves an accuracy of almost 0.94, maintaining the highest accuracy throughout. With an accuracy of approximately 0.92, the logistic regression model ranks second , while the SVM model has the lowest accuracy among three, with an accuracy of approximately 0.90.

VI CONCLUSION

Our research endeavors in fake news analysis using Node.js, TensorFlow.js, and the use of logistic regression, support vector machines(SVM) , and sequential dense neural networks has produced important finding and insights . By using sophisticated machine learning methods and leveraging the TensorFlow.js library, we have built a strong framework that can distinguish between authentic and fabricated news articles.Through the examination and capabilities of logistic regression models,SVMs , and sequential dense neural networks, we have illustrated the flexibility and adaptability of our methodology in addressing the obstacles presented by false information.

Additionally, the future scope of our research will go forward by incorporating sophisticated natural language processing methods like transformer-based architectures like BERT and recurrent neural networks(RNNs) to enhance the system's comprehension of contextual comprehension and subtle semantics from news articles.Overall our work shows the possibility for more breakthroughs in this viral area and emphasizes the significance of machine learning-driven analysis in the fight against false news.Through continuous innovation and collaboration , we work hard to make a significant contribution to the continuous initiatives meant to upload information integrity and lessen the effects of false information in the digital era.

VII REFERENCES

[1] Tajrian, Mehedi, et al. "A review of methodologies for fake news analysis." IEEE Access (2023).

[2] Zhang, Xichen, et al. "Multimodal fake news analysis based on image–text similarity." IEEE Transactions on Computational Social Systems (2023).

[3] Seddari, Noureddine, et al. "A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media." IEEE Access 10 (2022): 62097-62109.

[4] Chintalapudi, Nalini, Gopi Battineni, and Francesco Amenta. "Sentimental analysis of COVID-19 tweets using deep learning models." Infectious disease reports 13.2 (2021): 329-339.

[5] Pop, Cristian, and Alexandru Popa. "Detecting News Influence in a Country: One Step Forward Towards Understanding Fake News." Combating Fake News with Computational Intelligence Techniques (2022): 259-275

[6] Xu, Jiawei, et al. "FaNDS: Fake news detection system using energy flow." Data & Knowledge Engineering 139 (2022): 101985.

[7] Seddari, Noureddine, et al. "A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media." IEEE Access 10 (2022): 62097-62109.

[8] Xu, Jiawei, et al. "FaNDS: Fake news detection system using energy flow." Data & Knowledge Engineering 139 (2022): 101985.
[9] Z. Shahbazi and Y. -C. Byun, "Fake Media Detection Based on Natural Language Processing and Blockchain Approaches," in IEEE Access, vol. 9, pp. 128442-128453, 2021, doi: 10.1109/ACCESS.2021.3112607.
[10] N. R. de Oliveira, D. S. V. Medeiros and D. M. F. Mattos, "A sensitive stylistic approach to identify fake news on social networking", IEEE Signal Process. Lett., vol. 27, pp. 1250-1254, 2020.
[11] Zaheer, H., Bashir, M. Detecting fake news for COVID-19 using deep learning: a review. Multimed Tools Appl (2024). https://doi.org/10.1007/s11042-024-18564-7