Grenze International Journal of Engineering and Technology, January Issue

Fake News Detection using BERT Transformer Model

Bibhas Das[1], Suraj Majumder[2], Prima Giri[3] and Ritushna Roy[4]
[1-4]Adamas University Department of Computer Science and Engineering, Kolkata, India
Email: {bd.bibhas, surajmajumder212, primagiri001, royritushna}@gmail.com

***Abstract*—The prevalence of fake news, with its deleterious effects on societal trust and cohesion, underscores the pressing need for innovative approaches that combine technological advancements with human judgment. Our research work occupies the forefront of this endeavour, integrating state-of-the-art machine learning techniques, notably the BERT Transformer model, with advanced natural language processing (NLP) algorithms to effectively combat misinformation. Through the development of a user-friendly web application, accessibility and usability are paramount, ensuring that our tools cater to a diverse audience. Beyond mere identification and mitigation, our broader mission is to cultivate a society that values media literacy and restores confidence in information sources.**

***Index Terms*— Fake news, Machine learning, BERT Transformer model, Natural Language Processing (NLP), Web application.**

I. INTRODUCTION

In today's digital age, the widespread dissemination of false or deceptive information, known as fake news, has become a critical issue. This deliberate spread of misleading content, masked as genuine news, significantly impacts public perception, shapes political discourse, and can incite societal tensions. The ease of creating and sharing content on social media and online platforms exacerbates the problem, highlighting the urgent need for effective solutions to mitigate misinformation.

The rapid circulation of misleading information online threatens societal trust and discourse. Addressing this challenge requires innovative approaches and robust mechanisms to discern fact from fiction amidst the vast amount of daily shared information. Current methods for detecting misinformation lack scalability and real-time capabilities, making users vulnerable to deceptive content.

The primary goal of this research work is to develop a reliable system for effectively identifying fake news. Utilizing advanced technologies such as machine learning, natural language processing, and data analytics, the research work aims to create an algorithm which is capable of distinguishing between genuine and fabricated news articles. By analyzing textual content, source reliability, linguistic structures, and other relevant features, the system will enable real-time detection and empower users to verify the legitimacy of news articles confidently. Ultimately, this initiative seeks to combat the spread of misinformation and restore trust in digital information sources.

II. LITERATURE SURVEY

A. Machine Learning Approach to Fake News Detection:

Numerous studies have delved into supervised and unsupervised machine learning algorithms for detecting

Grenze ID: 01.GIJET.11.1.403

deceptive content. A seminal study **[1]** employed Support Vector Machines (SVM), Random Forest, Naive Bayes, and neural networks to categorize news articles based on linguistic features, sentiment analysis, and network-based analysis. These algorithms leverage labelled datasets to learn patterns indicative of fake news, enabling them to classify news articles accurately. Moreover, advancements in deep learning have enabled the development of more complex neural network architectures capable of capturing intricate relationships within textual data, further enhancing the discrimination between genuine and fabricated news.

B. Linguistic and Semantic Analysis

Research efforts have focused on analyzing linguistic and semantic features to differentiate between genuine and fake news. Notably, investigations in linguistic analysis **[2]** have scrutinized linguistic patterns, stylistic variances, sentiment analysis, and lexical characteristics to enhance the accuracy of classification. By dissecting the textual content of news articles, researchers aim to uncover subtle cues indicative of misinformation. Moreover, the integration of semantic analysis techniques, such as word embeddings and semantic similarity measures, enables the detection of semantic inconsistencies and anomalies in fake news narratives, contributing to more robust detection models.

C. Social Network Analysis and Propagation Patterns

Understanding how fake news propagates through social networks is crucial. Research in this realm **[3]** has utilized network-based analysis techniques, including information diffusion models and user behaviors analysis, to understand the dissemination and virality of misleading information across online platforms. By studying the dynamics of information flow within social networks, researchers gain insights into the mechanisms driving the spread of fake news. Additionally, the identification of influential nodes and communities within these networks facilitates targeted interventions to mitigate the amplification of misinformation.

D. Fact-Checking and Dataset Creation

The creation of annotated datasets and fact-checking frameworks is pivotal for training and evaluating fake news detection models. Studies emphasize the importance of high-quality, diverse datasets with labelled instances of both real and fake news articles **[4]**, facilitating the development of reliable detection algorithms. Moreover, collaborative efforts between researchers, journalists, and fact checking organizations contribute to the continuous enrichment of these datasets, ensuring their relevance and representativeness across different contexts and languages. Furthermore, the development of automated fact-checking tools powered by natural language processing enables scalable verification of news articles, augmenting the capabilities of human fact-checkers.

E. Ethical and Societal Implications

Scholarly discussions have explored the ethical and societal ramifications of fake news beyond technical considerations. Research in this area [5] addresses its impact on public trust, democratic processes, and the role of media literacy in combating misinformation. The dissemination of fake news not only erodes trust in traditional media sources but also undermines democratic institutions and fosters societal polarization. Consequently, interventions aimed at promoting media literacy and critical thinking skills are imperative to empower individuals to discern between reliable and misleading information.

F. Challenges and Future Directions

Persistent challenges in fake news detection include the evolution of deceptive tactics, the necessity for real-time detection, and ethical dilemmas regarding content moderation. Future research directions involve integrating diverse data sources, enhancing the explanation ability of AI models, and employing adaptive learning techniques to bolster detection systems. Additionally, interdisciplinary collaborations between researchers from diverse fields, including computer science, social sciences, and ethics, are essential to address the multifaceted nature of the fake news phenomenon and develop holistic solutions that mitigate its adverse effects on society.

In conclusion, the battle against fake news necessitates a comprehensive approach integrating advanced technologies with ethical considerations and societal ramifications. Through the exploration of machine learning algorithms, linguistic analysis, social network dynamics, and the creation of reliable datasets, researchers are advancing the field of fake news detection. However, persistent challenges, such as the dynamic nature of deceptive tactics and the urgency for real-time detection mechanisms, underscore the ongoing need for vigilance and innovation. As we confront these challenges, it is crucial to maintain a proactive stance and foster collaboration across disciplines. By leveraging insights from research and promoting collective efforts, we can strive towards a more informed and resilient society, where trust in information sources is strengthened, and the integrity of public discourse is upheld.

III. PROPOSED METHODOLOGY

The methodology for the fake news detection project involves a systematic approach encompassing data collection, pre-processing, exploratory data analysis (EDA), feature engineering, model selection, and training and evaluation.

A. Data Collection

Datasets were gathered from sources like Kaggle [6] and Papers with Code [7], resulting in six key datasets: cleaned_fake_and_real_news.xlsx, Fake_Real_News_Data.xlsx, WELFAKE.xlsx, news (1).xlsx, train.xlsx, and True.xlsx. These datasets were processed, reducing 110,000 data samples to 85,520 after cleaning and pre-processing.

B. Data Pre-processing

Pre-processing involved dropping unnecessary columns, standardizing text by replacing non-alphabetic characters and lowercasing, removing stop words, punctuation, and URLs, and assigning labels (0 for real news, 1 for fake news). Tokenization was performed to prepare the dataset for analysis.

C. Exploratory Data Analysis (EDA)

EDA involved examining the dataset's structure, checking for missing values, and analyzing the distribution of real and fake news, ensuring the dataset was balanced. Word frequencies for real and fake news were visualized using bar charts.

D. Feature Engineering

The BERT-based-uncased tokenizer was used for feature engineering, converting text into numerical representations. Parameters were fine-tuned for maximum token length, and distinct encoding strategies were adopted for training, validation, and testing datasets to ensure consistent processing.

E. Model Selection

Various models were evaluated, with the BERT Transformer model chosen for its exceptional ability to capture contextual information and semantic nuances in text. Despite testing other models like LSTM, BERT outperformed them, making it the cornerstone of the fake news detection system.

F. Training and Evaluation

The dataset was split into training, validation, and testing sets, with model performance evaluated using metrics such as accuracy, precision, recall, and F1-score. The BERT Transformer consistently outperformed LSTM networks, demonstrating superior accuracy and robustness in detecting fake news. This led to its selection as the preferred model for the project.

G. Creating Web Application

After training and testing the fake news detection model, in Fig.1 a user-friendly web application was developed using React for the front end and Node.js for the back end. The application features real-time data collection via the NewsData.io API, refreshing every 12 hours to ensure up-to-date news. A robust backend database using MongoDB securely stores user credentials, facilitating efficient account management and secure authentication. The Flask framework was employed to integrate the front end, back end, and database components smoothly. This comprehensive web application combines advanced machine learning algorithms with real-time data integration and secure user management, providing users with effective tools to detect and combat misinformation.

Our research work involves the development of a web application shown in Fig. 2 that combines secure authentication with a news verification system. When users input their credentials, the application sends a One-Time Password (OTP) to their registered email address to ensure secure access. Upon successful login and OTP verification, users gain access to a dashboard. Within this dashboard, users can enter news articles or text to be evaluated. The system uses advanced machine-learning algorithms to analyze the input and determine its authenticity, classifying the news as either true as shown in Fig. 3 or false as shown in Fig. 4. This functionality aims to provide a reliable tool for users to identify and combat misinformation.

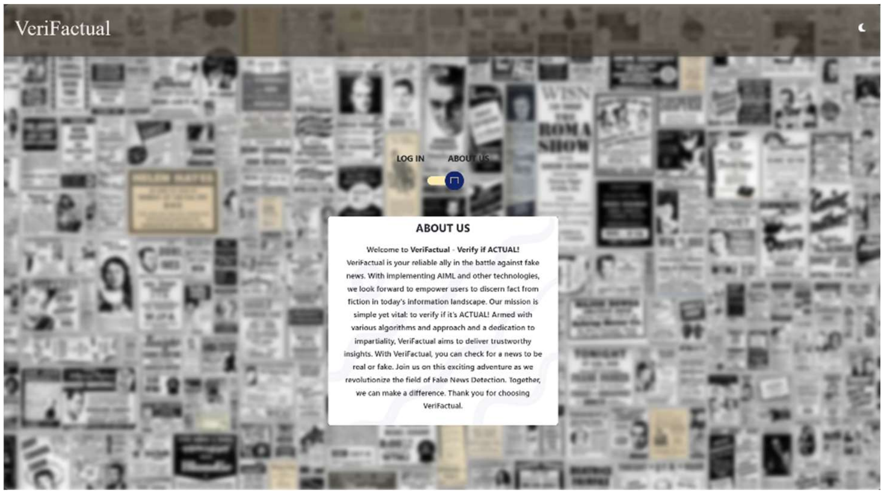

Fig. 1: Interface of the web application

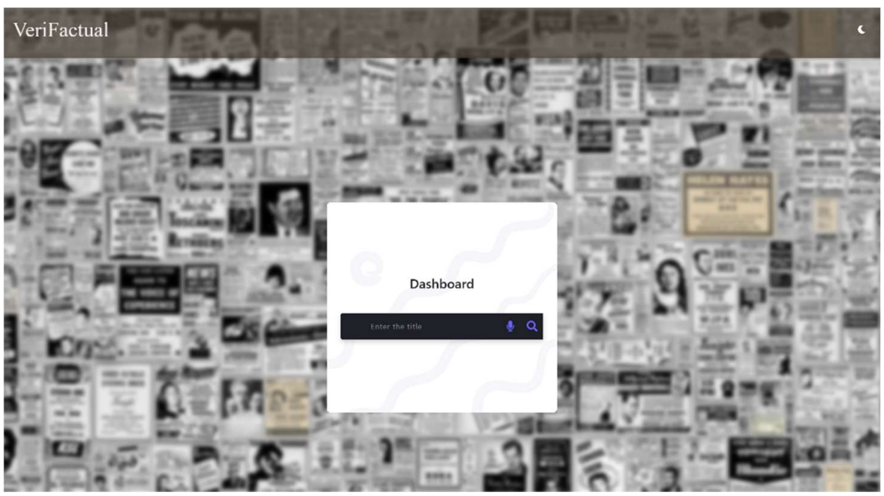

Fig. 2: Dashboard of deployed web application

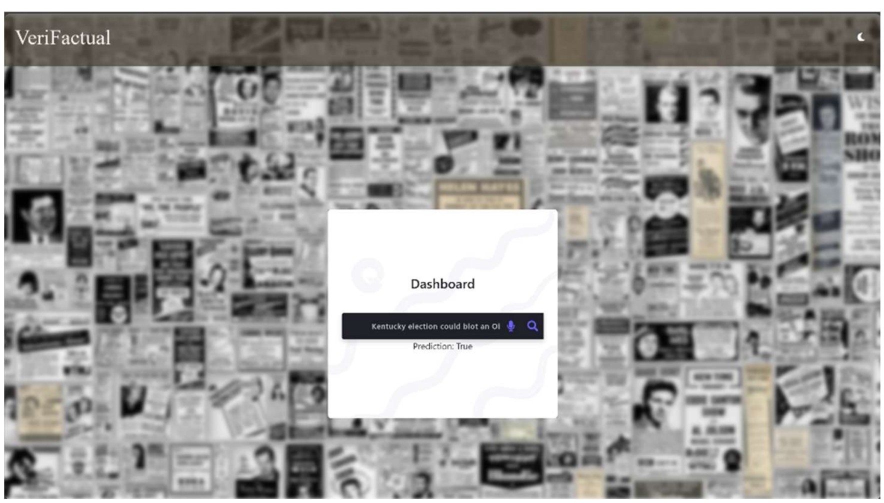

Fig. 3: News predicted as True

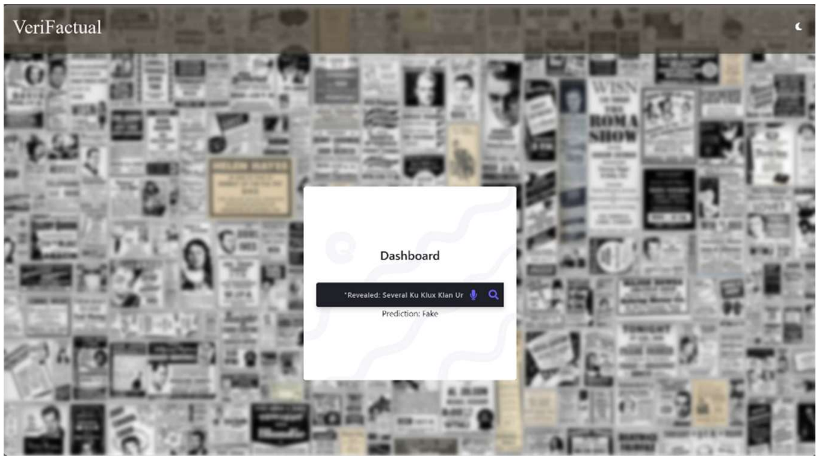

Fig. 4: News predicted as Fake

IV. Result and Discussion

A. For the LSTM model, we have got

Average Training Accuracy: 0.9273454546928406
Average Validation Accuracy: 0.8987371365229289
Testing Accuracy: 0.8854069113731384
The above results are based on 15 epochs.
The LSTM model displays strong performance in fake news detection as shown in Fig.5, with an average training accuracy of approximately 92.7% and an average validation accuracy of around 89.9%. However, the testing accuracy of about 88.5% indicates a slight drop in performance when applied to unseen data.
Overall, while the model shows promising results, further refinement may be necessary to address this discrepancy and enhance its performance.

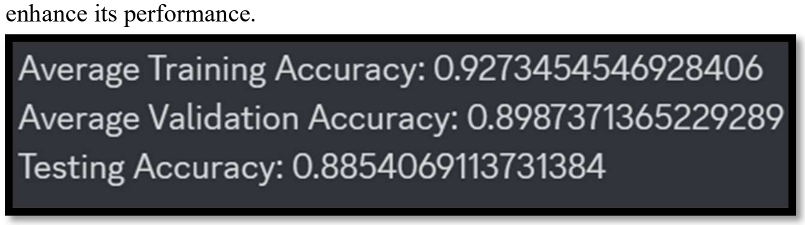

Fig. 5: LSTM model accuracy results

B. For the BERT Transformer model, we have got

Average Training Accuracy: 0.9829
Average Validation Accuracy: 0.9845
Testing Accuracy: 0.9881
The BERT Transformer model demonstrates outstanding performance in fake news detection mentioned in Fig. 6, achieving remarkable accuracies across all evaluation metrics. With an average training accuracy of 98.29% and an average validation accuracy of 98.45%, the model exhibits robust learning capabilities and effectively generalizes to unseen data. Additionally, the testing accuracy of 98.81% further confirms the model's high accuracy and reliability in correctly classifying news articles as real or fake.
Overall, the BERT Transformer model displays superior performance compared to the LSTM model, consistently delivering high accuracies across training, validation, and testing datasets. Its exceptional accuracy underscores its efficacy in capturing contextual information and semantic nuances within textual data, making it the preferred choice for fake news detection in our research work.

III. Conclusions

The fake news detection research work was a dedicated effort to combat digital misinformation by developing a

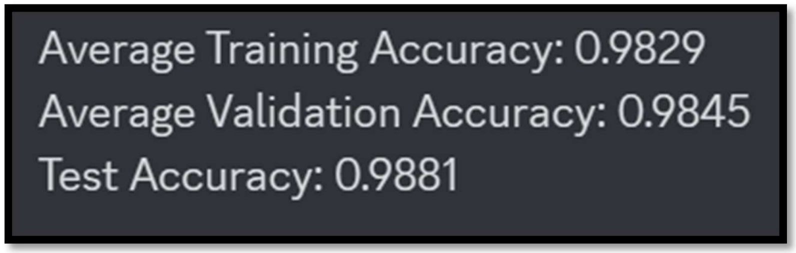

Fig. 6: BERT Transformer model accuracy results

robust system capable of accurately identifying fake news articles. Utilizing meticulous data collection, model training, and evaluation techniques, the team achieved exceptional performance with advanced machine learning, particularly leveraging the BERT Transformer model. Despite encountering challenges such as hardware limitations and restricted API access, the team's perseverance led to innovative strategies and optimization techniques, ensuring the delivery of a reliable solution.

Looking forward, our research work paves the way for further advancements in fake news detection through continuous refinement of models, optimization of hardware resources, and expanded access to comprehensive news data. Emphasizing collaboration and interdisciplinary approaches underscores the importance of addressing complex societal issues like misinformation. Ultimately, the project's success demonstrates technology's pivotal role in promoting truth and integrity in the digital landscape, with ongoing dedication and innovation poised to foster a more informed and resilient society.

References

[1] I.G. Hosea et al., “A machine learning approach to fake news detection using support vector machine (SVM) and unsupervised learning model,” Advances in Multidisciplinary and scientific Research Journal Publication, vol. 11, no. 1, pp. 11–18, 2023. doi:10.22624/aims/csean smart2023p1 N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

[2] C. Wang, I. Castellón, and E. Comelles, “Linguistic analysis of datasets for semantic textual similarity,” Digital Scholarship in the Humanities, vol. 35, no. 2, pp. 471–484, 2019. doi:10.1093/llc/fqy076.

[3] “Introduction to social network analysis,” Social Network Analysis, pp. 1–6, 2020. doi:10.4135/9781506389332.n4 M. Inamdar, N. Mehendale, “Real-Time Face Mask Identification Using Facemasknet Deep Learning Network”, SSRN Electron. J. (july 29, 2020).

[4] Vlachos and S. Riedel, “Fact checking: Task definition and dataset construction,” Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science, 2014. doi:10.3115/v1/w14-2508 R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 142-158, 1 Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.

[5] K. Santosh and C. Wall, “AI and ethical issues,” AI, Ethical Issues and Explainability— Applied Biometrics, pp. 1–20, 2022. doi:10.1007/978-981-19-3935-8_1 C.Y. Fu, W. Liu, A.Ranga, A. Tyagi, AC Berg, “Dssd: Deconvolutional single shot detector”, arXiv: 1701.06659 (2017).

[6] “Find open datasets and Machine Learning Projects,” Kaggle, https://www.kaggle.com/datasets.

[7] “Papers with code - machine learning datasets,” Machine Learning Datasets | Papers with Code, https://paperswithcode.com/datasets.