ADScAI
Applied Data Science & Artificial Intelligence Symposium

BERT Unleashed: A Robust Bidirectional Shield Against Proliferating Fake News

[†*]Vadivel Abishethvarman
Department of Computing & Information Systems, Faculty of Computing
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
abishethvarman@gmail.com

[*]Gobihanath Balasubramaniam
Department of Computing & Information Systems,
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
bgobihanath@std.appsc.sab.ac.lk

Senthan Prashanth,
Faculty of Engineering and Applied Science, Memorial University of New Foundland
St Johns, New Foundland & Labrodar, Canada
sprasanth@appsc.sab.ac.lk

Banujan Kuhaneswaran
Faculty of Science and Engineering,
Southern Cross University, Gold Coast, Australia
bhakuha@gmail.com

Prof. B.T.G.S Kumara
Department of Data Science, Faculty of Computing
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
kumara@appsc.sab.ac.lk

Keywords— Fake News, BERT, Natural Language Processing, Proliferating

I. INTRODUCTION

Fake news is defined as purposefully misleading material masquerading as legitimate news, which is frequently broadcast through multiple media channels with the objective of fooling the audience. This false information can appear as completely made-up stories, edited images, or facts that are twisted [1]. Detecting fake news is difficult because it constantly changes and takes different forms. Social media platforms increase the reach of fake news, allowing it to spread quickly and reach a large audience. Addressing the challenges posed by fake news is crucial for preserving the integrity of information[2] and fostering a more informed and resilient society.

Unlike standard algorithms[3] that analyze text in a single direction, BERT evaluates the full phrase, using its surrounding context for a more sophisticated interpretation. By pre-training on large volumes of text data, BERT learns to construct embeddings that incorporate deep language links and semantic interpretations. In the context of machine learning-based fake news detection, BERT's bidirectional capability is crucial since it understands the intricate structures of language, allowing for more accurate and context-aware identification of false information. This research investigates the use of BERT in the field of false news detection, highlighting its functions and showing how it improves the performance of machine learning models.

The use of BERT in this research highlights its impact on fake news detection, utilizing its pre-trained knowledge to enhance the model's ability to differentiate between real and false news articles. By leveraging BERT's bidirectional capabilities, this study demonstrates how advanced NLP models can significantly improve the performance of machine learning-based detection systems in the complex landscape of fake news. As authors, we utilize BERT's pre-trained knowledge and contextual awareness, positioning it as a crucial element in our approach. This research contributes to the ongoing exploration of advanced NLP models for addressing the challenge of fake news proliferation.

II. IMPLEMENTATION

A. Literature Review

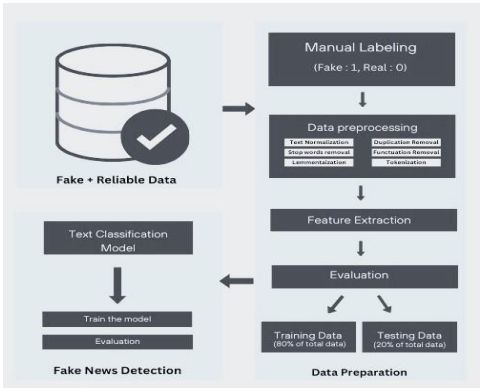
Fake news detection has gained significant attention in recent years, with researchers exploring various machine learning and natural language processing techniques to improve accuracy. Early approaches relied on linguistic cues and handcrafted features, but these methods struggled with generalization across different news sources [4]. Deep learning models, particularly transformer-based architectures like BERT, have demonstrated superior performance in detecting misinformation due to their ability to capture complex contextual relationships. Moreover, advancements in transfer learning have facilitated fine-tuning BERT for domain- specific fake news detection, yielding higher precision and recall compared to conventional classifiers [5]. Researchers have also explored explainability techniques such as attention visualization to enhance trust in BERT's decision-making process. Despite these advancements, challenges remain in addressing adversarial attacks and the dynamic nature of misinformation, necessitating continuous refinement of BERT-based detection frameworks [6]. This study builds upon existing research by implementing BERT for fake news classification, leveraging its bidirectional contextual understanding to enhance detection accuracy and robustness.

B. Materials and Methods

This study employs a BERT-based model for fake news detection, leveraging its bidirectional contextual understanding to enhance classification accuracy. The dataset includes both fake and reliable news articles, which are manually labeled (Fake: 1, Real: 0) to enable supervised

https://doi.org/10.31705/ADScAI.2025.49

learning. The data undergoes preprocessing steps, including text normalization, stop-word removal, lemmatization, duplication and punctuation removal, and tokenization using BERT's tokenization technique. Additionally, GloVe (Global Vectors for Word Representation) is used during preprocessing to generate word embeddings. After preprocessing, the dataset is split into training (80%) and testing (20%) sets to ensure effective model evaluation. The pre-trained 'bert-base-uncased' model from the Hugging Face Transformers library is fine-tuned using transfer learning to adapt to dataset-specific features. Tokenization and padding are applied to standardize input lengths before training the model using PyTorch or TensorFlow, with binary cross-entropy loss and the Adam optimizer. The model's performance is assessed using accuracy, precision, recall, F1-score, and a confusion matrix, providing valuable insights into its classification effectiveness and error distribution.



C. Results and Discussion

Table 1 highlights the evaluation of BERT-based models against traditional machine learning models highlights the superior performance of BERT in fake news detection. Among the BERT models, Bert-Base-Multilingual-Cased achieved the highest accuracy (99.91%), with a precision of 99.97% and recall of 99.86%, demonstrating its strong capability in distinguishing fake news. Bert-Base-Cased and Bert-Base-Uncased also performed well, with accuracies of 99.84% and 98.37%, respectively.

	Accuracy	Precision	Recall	F1Score
RNN LSTM	94.99%	91.06%	99.95%	95.30%
Decision Tree	99.67%	99.74%	99.63%	99.69%
Random Forest	99.41%	99.50%	99.50%	99.00%
SVM	72.65%	72.84%	72.70%	72.07%
Bert-Base-Uncased	98.37%	98.97%	97.90%	98.43%
Bert-Base-Cased	99.84%	99.91%	99.77%	99.84%
Bert-Base-Multilingual-Uncased	96.91%	95.50%	98.75%	97.10%
Bert-Base-Multilingual-Cased	99.91%	99.97%	99.86%	99.91%

Table 1: Performance Comparison

In comparison, traditional models like Random Forest and Decision Tree exhibited high accuracy, indicating their effectiveness in structured datasets. However, models such as SVM performed poorly, achieving only 72.65% accuracy, highlighting its limitations. BERT-based models demonstrated superior consistency across precision, recall, and F1-score.

D. Future Work

Future efforts will focus on enhancing the Semantic Text classification on Fake news in various domains. And domain-specific fine-tuning to handle nuanced situations, dynamic learning for real-time adaptability, and integrating multimodal input for a more comprehensive understanding.

E. Conclusion

The study's findings show that the BERT-based algorithm is very accurate and reliable in detecting false news. By integrating bidirectional contextual knowledge, the model outperforms standard techniques, obtaining considerable precision and recall scores. The attention processes improve interpretability by offering insights into the essential textual characteristics that influence predictions.

III. REFERENCES

[1] V. Abishethvarman, K. Banujan, B. Kumara, and N. Ravikumar, "Unmasking Fake News with Emotional Cues: A Comparative Study of Sentiment Reasoning Models," in *2024 4th International Conference on Advanced Research in Computing (ICARC)*, 2024, pp. 195-200.

[2] V. Abishethvarman, K. Banujan, R. Nirubikaa, and B. Kumara, "Uncovering the multifaceted nature of fake news: Latent dirichlet allocation topic modelling approach," 2024.

[3] V. Abishethvarman, *Fake News Detection: A Comparative Analysis of RNN LSTM and Decision Tree Algorithms for Nonlinear Relationships*, 2023.

[4] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "defend: Explainable fake news detection," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 395-405.

[5] M. Dehghani, R. Boghrati, K. Man, J. Hoover, S. I. Gimbel, A. Vaswani*, et al.*, "Decoding the neural representation of story meanings across languages," *Human brain mapping,* vol. 38, pp. 6096-6106, 2017.

[6] J. Alghamdi, Y. Lin, and S. Luo, "A comparative study of machine learning and deep learning techniques for fake news detection," *Information,* vol. 13, p. 576, 2022.