INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD
ISSN(O): 2455-0620 [Impact Factor: 9.47]
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value : 86.87
Volume - 12, Issue - 2, February - 2026

DOIs:10.2015/IJIRMF/202602003 --:-- Research Paper / Article / Review

Fake Detection Using Machine Learning and Deep Learning

[1] Pillalamarri Madhavi, [2] Lakshmi Muthavarapu, [3] Kotti Bhuvana Teja
[1]Assistant Professor, Department of Electrical and Electronics Engineering, Hyderabad Institute of Technology and Management, Hyderabad, India
[2,3]UG Scholar , Department of Data Science, Hyderabad Institute of Technology and Management, Hyderabad, India
Email -[1]madhavipillalamarry@gmail.com, [2]lakshmimuthavarapu123@gmail.com, [3]bhuvana16teja@gmail.com

Abstract: *The fast spread of information on social media has made detecting fake news a major challenge, as false content can sway public opinion and lower trust in online information sources. Traditional machine learning methods have been commonly used for classifying fake news; however, they often fail to fully understand the meaning of complex text. To overcome this issue, this paper introduces a hybrid model for fake news detection that combines both deep learning and machine learning approaches. The model uses Bidirectional Encoder Representations from Transformers (BERT) to get detailed semantic information from news articles, which is then used for classification with the XGBoost algorithm. The model is tested on a dataset of labeled fake news using metrics like accuracy, precision, recall, F1-score, and confusion matrix. The results show that the BERT + XGBoost model performs better in terms of accuracy and overall effectiveness compared to traditional machine learning models and standalone BERT classifiers. The study shows that combining contextual feature extraction with a strong gradient boosting classifier improves the effectiveness and reliability of fake news detection.*

Key Words: *Fake News Detection, BERT, XGBoost, Transformer Models, Machine Learning, Text Classification*

1. INTRODUCTION:

The rapid growth of social media platforms such as Twitter, Facebook, and online news portals has transformed the way information is created and consumed. Users can share news instantly, allowing information to reach millions of people within a short time. While this has improved communication and access to information, it has also increased the spread of fake and misleading news. Fake news refers to deliberately fabricated or false information presented as legitimate news, often created to mislead readers or manipulate opinions. The widespread circulation of fake news can have serious consequences, including public misinformation, damage to reputations, social unrest, and political manipulation. During critical events such as elections, natural disasters, and health emergencies, the presence of fake news can cause panic and influence decision-making. Due to the enormous volume of content generated on social media every day, manually verifying news for authenticity is time-consuming and impractical. Therefore, automatic fake news detection has become a significant research challenge in the fields of machine learning and natural language processing. Early research in fake news detection mainly relied on traditional machine learning techniques such as Naive Bayes, Support Vector Machines (SVM), Logistic Regression, and Random Forest. These approaches use manually engineered features such as word frequency, n-grams, and syntactic patterns. Although these methods provide reasonable results, their performance is limited because they struggle to capture the contextual and semantic meaning of complex text, especially when dealing with ambiguous or long-form news articles.

To overcome these limitations, deep learning techniques have gained attention due to their ability to automatically learn meaningful representations from raw text. Models such as Convolution Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have been successfully applied to fake news detection by learning spatial and sequential patterns in text. More recently, transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) have shown superior performance by understanding contextual relationships between words in a sentence. BERT captures bidirectional context, enabling a deeper understanding of language compared to earlier models. Despite the success of deep learning models, using them alone for classification may not always yield optimal performance. Machine learning classifiers such as XGBoost are known for their robustness, efficiency, and strong generalization ability. Combining deep learning models for feature extraction with powerful machine learning classifiers can further improve classification accuracy.

In this work, a hybrid fake news detection model is proposed that integrates BERT for contextual feature extraction and XGBoost for final classification. The proposed approach leverages the strengths of both deep learning and machine learning techniques to achieve improved detection performance. Experimental results demonstrate that the hybrid model outperforms traditional machine learning methods and standalone deep learning models, making it a reliable solution for fake news detection on social media platforms.

2. RESEARCH METHOD:

This section describes the dataset, preprocessing steps, feature extraction process, classification model, and evaluation metrics used in the proposed fake news detection system. The overall workflow of the proposed methodology is illustrated as a pipeline consisting of data preprocessing, feature extraction using BERT, and classification using XGBoost.

2.1 Dataset Description: A publicly available labeled fake news dataset is used for experimentation. The dataset consists of news articles labeled as either fake or real. Each record contains the news text along with its corresponding class label. The dataset is divided into training and testing sets to evaluate the performance of the proposed model. Balanced class distribution is maintained to avoid bias in classification results.

2.2 Data Preprocessing: Before model training, the raw news text is preprocessed to improve model performance. The preprocessing steps include removal of special characters, URLs, punctuation, and extra whitespace. Text is converted into a standardized format suitable for tokenization. Since BERT uses its own tokenizer, minimal preprocessing is applied to preserve contextual information. The cleaned text is then tokenized using the BERT tokenizer, and input IDs and attention masks are generated for each news article.

2.3 Feature Extraction Using BERT: Bidirectional Encoder Representations from Transformers (BERT) is employed to extract contextual semantic features from the news text. BERT is a transformer-based language model pre-trained on large-scale text corpora, enabling it to understand word meanings based on surrounding context. In this work, the pre-trained BERT model is used as a feature extractor rather than a standalone classifier.
For each news article, the output embedding corresponding to the [CLS] token is extracted, as it represents the overall semantic representation of the input text. These embeddings capture deep contextual information and serve as high-quality feature vectors for classification.

2.4 Classification Using XGBoost: XGBoost is used as the final classifier in the proposed hybrid model. It is a gradient boosting algorithm known for its high accuracy, efficiency, and ability to handle complex decision boundaries. The BERT-generated feature vectors are provided as input to the XGBoost classifier, which learns to distinguish between fake and real news. By combining BERT's contextual understanding with XGBoost's strong classification capability, the proposed model achieves improved performance.

2.5 Model Training and Evaluation: The dataset is split into training and testing sets. The BERT model is used to generate embeddings for both sets, while XGBoost is trained only on the training embeddings. The trained model is then evaluated on the test set. Performance is measured using standard evaluation metrics including accuracy, precision, recall, F1-score, and confusion matrix. These metrics provide a comprehensive assessment of the model's effectiveness in fake news detection.

3. BLOCK DIAGRAM:

The block diagram of the proposed fake news detection system illustrates the step-by-step flow of data from input to final classification. The system is designed as a hybrid architecture combining deep learning–based feature extraction and machine learning–based classification.

1. Input News Text: The process begins with raw news text collected from a labeled fake news dataset. Each news article serves as an input to the system

2. Text Preprocessing: The input text undergoes basic preprocessing to remove noise such as special characters, URLs, punctuation, and extra spaces. Minimal preprocessing is applied to preserve contextual information required by the BERT model.

3. BERT Tokenization: The preprocessed text is tokenized using the BERT tokenizer. Input tokens are converted into token IDs, and attention masks are generated to indicate valid tokens.

4. BERT Feature Extraction: The tokenized text is passed through the pre-trained BERT model to extract contextual embeddings. The embedding corresponding to the [CLS] token is selected as it represents the overall semantic meaning of the news article.
5. Feature Vector Generation: The extracted BERT embeddings are transformed into fixed-length numerical feature vectors suitable for machine learning classification.
6. XGBoost Classification: The feature vectors are fed into the XGBoost classifier, which learns to classify the news articles as fake or real based on learned patterns.
7. Output Prediction: The final output of the system is a binary prediction indicating whether the given news article is fake or real.

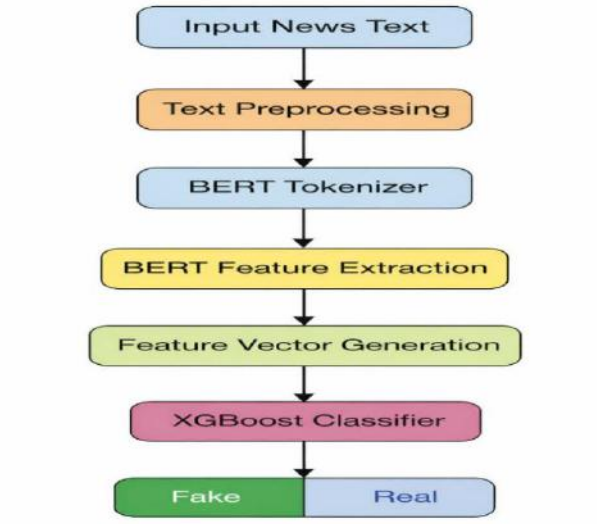

Figure 1: Block Diagram

4. RESULTS:

This section presents the experimental results of the proposed hybrid fake news detection model using BERT for feature extraction and XGBoost for classification. The performance of the proposed approach is evaluated and compared with traditional machine learning and standalone deep learning models using standard evaluation metrics.

4.1 Sample Prediction Output

To verify the real-time applicability of the proposed system, a test input containing a genuine news statement was provided to the trained model. The system successfully classified the input as REAL NEWS, demonstrating the model's capability to correctly identify authentic news content.

```
# Test with a true news example
true_news = "Indian Space Research Organisation successfully launched a new satellite into orbit."

cleaned = clean_text(true_news)
true_emb = get_bert_embeddings(pd.Series([cleaned]))

prediction = model.predict(true_emb)

print("Input:", true_news)
if prediction[0] == 1:
    print("Output: REAL NEWS")
else:
    print("Output: FAKE NEWS")
```

```
Input: Indian Space Research Organisation successfully launched a new satellite into orbit.
Output: REAL NEWS
```

Fig. 4.1: Sample Output of Real News Prediction

4.2 Confusion Matrix Analysis:

The confusion matrix of the proposed BERT + XGBoost model is shown in Fig. 4.2. It provides a detailed view of correct and incorrect classifications.

- True Negatives (TN): 503
- False Positives (FP): 21
- False Negatives (FN): 30
- True Positives (TP): 446

The high number of correctly classified instances indicates that the proposed model performs reliably for both fake and real news categories. The low false positive and false negative rates demonstrate the robustness of the hybrid approach.

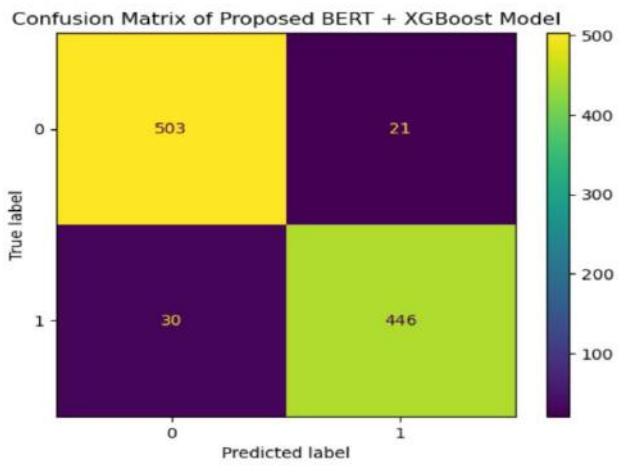

Fig. 4.2: Confusion Matrix of Proposed BERT + XGBoost Model

4.3 Accuracy Comparison with Existing Models:

To highlight the effectiveness of the proposed model, its accuracy is compared with several baseline models, including Naive Bayes, Support Vector Machine (SVM), Random Forest, and standalone BERT. The comparison is illustrated in Fig. 4.3.

The proposed BERT + XGBoost model achieves the highest accuracy of approximately **94.9%**, outperforming all other models. This improvement is due to BERT's contextual feature extraction combined with XGBoost's powerful gradient boosting classification capability.

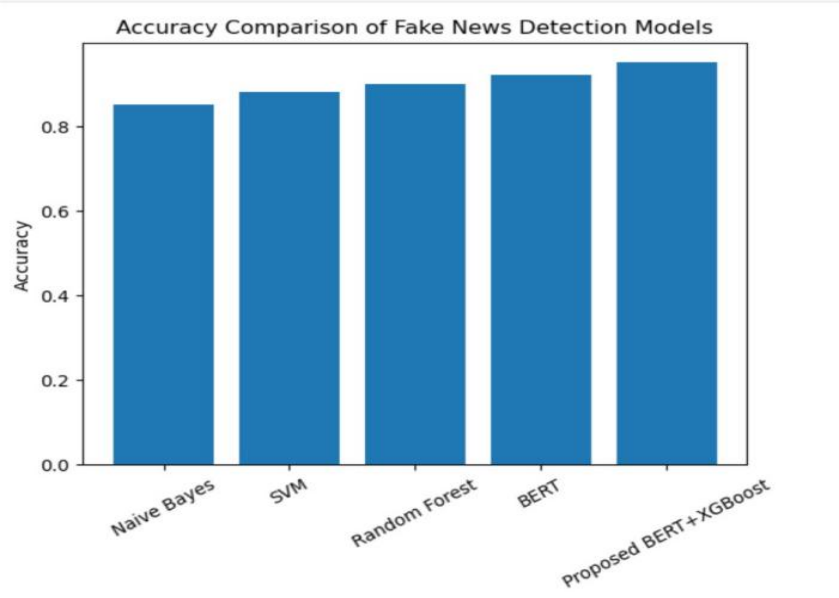

Fig. 4.3: Accuracy Comparison of Fake News Detection Models

4.4 Classification Report Analysis

The detailed performance metrics of the proposed model are summarized using precision, recall, and F1-score, as shown in Table 4.1.

- The model achieves high precision and recall for both fake and real news classes.
- The overall accuracy of **94.9%** indicates strong generalization capability.
- The balanced macro and weighted averages confirm that the model performs consistently across classes.

	precision	recall	f1-score	support
0	0.943715	0.959924	0.951750	524.000
1	0.955032	0.936975	0.945917	476.000
accuracy	0.949000	0.949000	0.949000	0.949
macro avg	0.949373	0.948449	0.948834	1000.000
weighted avg	0.949102	0.949000	0.948974	1000.000

Table 4.1: Classification Report of Proposed Model

5. DISCUSSION:

The experimental results clearly show that the proposed hybrid approach outperforms traditional machine learning models and standalone deep learning models. Traditional classifiers rely heavily on handcrafted features, limiting their ability to understand complex semantic patterns. In contrast, BERT captures deep contextual information from text, while XGBoost effectively learns non-linear decision boundaries. The integration of these two techniques results in improved accuracy, reduced misclassification, and enhanced reliability for fake news detection.

6. CONCLUSION:

This study presents and evaluates a hybrid fake news detection framework that integrates BERT for contextual feature extraction with XGBoost for effective classification. By leveraging BERT's ability to capture deep semantic relationships within news text and XGBoost's strong learning capability, the proposed model demonstrates improved performance over conventional machine learning approaches and standalone deep learning models, achieving an overall accuracy of 94.9%. The evaluation results, including the confusion matrix and classification metrics, indicate a significant reduction in misclassification while maintaining high precision and recall across both fake and real news classes. These findings highlight the advantage of combining transformer-based language models with gradient boosting classifiers to enhance the reliability and effectiveness of fake news detection systems. Future research can explore multilingual extensions, integration of social and contextual signals, and real-time deployment through web or mobile platforms to further address the challenges of misinformation spread.

REFERENCES:

1. N. F. Baarir and A. Djeffal, "Machine learning based fake news detection using TF-IDF and SVM," IEEE Conference on Smart Environments, 2021.
2. R. Shaikh and R. Patil, "Fake news detection using machine learning classifiers with TF-IDF features," IEEE International Conference on Sustainable Energy and Signal Processing, 2020.
3. U. Sharma, S. Saran, and S. M. Patil, "Fake news detection using artificial intelligence and NLP techniques," International Journal of Creative Research Thoughts, 2020.
4. C. K. Hiramath and G. C. Deshpande, "Fake news detection using machine learning and deep learning models," International Conference on Advances in Computing, 2019.

5. S. Kumar, R. Asthana, and S. Upadhyay, "Deep learning based fake news detection using CNN and Bi-LSTM with attention," IEEE Transactions on Computational Social Systems, 2020.
6. L. Waikhom and R. S. Goswami, "Ensemble learning for fake news detection using LIAR dataset," International Conference on Advancements in Computing, 2019.
7. S. I. Manzoor and J. Singla, "Fake news detection using SVM and NLP techniques," IEEE Conference on Trends in Electronics and Informatics, 2019.
8. M. F. Mridha, A. J. Keya, M. A. Hamid, and M. M. Monowar, "Headline and article stance detection using deep neural networks," IEEE Access, 2021.
9. A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Deep learning approaches for fake news detection," SMU Data Science Review, 2018.
10. D. H. Lee, Y. R. Kim, H. J. Kim, and S. M. Park, "Twitter fake news detection using machine learning classifiers," Journal of Information Processing Systems, 2019.