WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

FAKE NEWS DETECTION

SNEHALA A[1], DURGHA S[2], ASWATHI R[3], GAYATHRI N[4]

[1] Department of Artificial Intelligence and Machine Learning Sri Shakthi Institute of Engineering and Technology, Coimbatore, India snehalaanandkumar22aml@srishakthi.ac.in

[2] Department of Artificial Intelligence and Machine Learning Sri Shakthi Institute of Engineering and Technology, Coimbatore, India durghasivasankaran22aml@srishakthi.ac.in

[3] Department of Artificial Intelligence and Machine Learning Sri Shakthi Institute of Engineering and Technology, Coimbatore, India aswathir22aml@srishakthi.ac.in

[4] Department of Artificial Intelligence and Machine Learning Sri Shakthi Institute of Engineering and Technology, Coimbatore, India gayathri@siet.ac.in

ABSTRACT :

This project presents a lightweight and efficient fake news detection system utilizing a fine-tuned BERT model—"bert-tiny-finetuned-fake-news-detection"—from the Hugging Face library. By leveraging the capabilities of the Transformers framework, the system tokenizes and processes news articles to classify them as real or fake. The use of a compact BERT variant ensures faster performance while retaining contextual understanding, making it suitable for real-time applications. The model offers a practical solution to combat misinformation, a rising concern in today's digital age. It can be integrated into web or console-based tools, aiding journalists, researchers, educators, and social media platforms in promoting accurate information and maintaining media credibility.

INDEX TERMS

BERT**,** Transformers**,** Hugging Face**,** Sequence Classification, AutoTokenizer, AutoModelForSequenceClassification, Tokenization, Fine-tuning, Siamese Network, Triple Branch Network, Credit Score Feature, Justification Modeling, Metadata Integration, Bidirectional Attention, Contextual Embeddings, Binary Cross-Entropy Loss**,** AdamW Optimizer**,** Learning Rate Scheduler**,** Validation Strategy**,** Stratified Train-Test Split**,** Hyperparameter Tuning**,** Data Augmentation**,** Synonym Replacement**,** Back Translation.

INTRODUCTION

This project focuses on detecting fake news using a fine-tuned transformer-based model called **"bert-tiny-finetuned-fake-news-detection"**, available on Hugging Face. With the increasing spread of misinformation across digital platforms, there is a growing need for automated systems that can accurately classify content and reduce the impact of fake news. This project provides an effective and lightweight solution using Natural Language Processing (NLP) techniques.

The model used in this project is **BERT-Tiny**, a compact version of the original BERT (Bidirectional Encoder Representations from Transformers) architecture developed by Google. While the original BERT model is powerful, it is computationally heavy. BERT-Tiny, on the other hand, is optimized for faster performance and low memory usage, making it ideal for real-time applications or deployment on devices with limited resources.

To process the input text, the project uses the **Transformers** library from Hugging Face. This library allows easy access to pre-trained models and includes built-in tools for tokenization and data preparation. Input news articles are first tokenized—converted into input IDs and attention masks—which help the model understand the context of the sentences.

The model, fine-tuned specifically for fake news detection, then classifies the input as **"fake"** or **"real"**. Fine-tuning means the model was trained further on a labeled dataset of real and fake news after its initial general language pre-training. This enables it to better recognize the linguistic patterns commonly found in fake news articles, such as exaggerated language, emotional cues, or inconsistency in facts.

This approach combines the contextual understanding of BERT with domain-specific knowledge gained during fine-tuning, resulting in high classification accuracy. The use of a smaller model like BERT-Tiny also ensures the system is efficient and suitable for large-scale deployment.

In summary, this project demonstrates how transformer-based models can be applied to tackle real-world problems like fake news detection. By leveraging pre-trained language models, we can build smart, scalable, and efficient tools to combat misinformation and promote more accurate content online.

LITERATURE REVIEW

RESEA-RCH PAPER	YEAR	METHO-DOLOGY	ADVANT-AGES	IMPROVEMENTS AND NEGATIVES
Fake News Detection With Deep Learning	2021	Using DL techniques like Attention, GANs and BERT	High accuracy, improves interpretability	High compute cost,Hard to train.
Fake News Detection on Social Media	2019	BERT, Transformers and PyTorch	Transfer Learning, Scalable and adaptable	High computational cost
BERT-based Transfer Learning Approach for Fake News Detection	2022	Pre-Trained BERT model	Superior Accuracy and Generalization	Overfitting Risk and Data Dependency
Hugging Face Transformers for Text Classification	2021	bert-tiny-finetuned-fake-news-detection	Supports for Multiple Frameworks	Limited intertability.

PROPOSED METHODOLOGY

Data Collection

The first step in the proposed methodology is data collection, which involves gathering a labeled dataset of news articles categorized as real or fake. Several publicly available datasets can be used for this purpose. The LIAR dataset, for instance, contains short political statements annotated with six levels of truthfulness, making it particularly useful for nuanced classification. Another valuable source is the Fake and Real News dataset available on Kaggle, which includes thousands of full-length articles labeled as either fake or real. Additionally, fact-checked datasets from BuzzFeed and PolitiFact offer reliable data specifically suited for detecting political misinformation. To enhance the model's robustness and ensure better generalization across various domains such as politics, health, and finance, it is beneficial to combine multiple datasets during training. This helps the model learn diverse linguistic patterns and reduces bias toward a specific type of content.

Data Pre-processing

The methodology is data preprocessing, which is crucial for ensuring that the input text is clean and properly formatted for model training. Initially, the raw text data undergoes cleaning, where elements such as URLs, special characters, and HTML tags are removed to eliminate noise. This is followed by lowercasing all text to maintain consistency and reduce redundancy in the vocabulary. Tokenization is then performed using a pre-trained tokenizer like BertTokenizer from the Hugging Face Transformers library, which converts the cleaned text into a sequence of tokens compatible with the BERT model. To handle variable-length inputs, padding and truncation techniques are applied to ensure that all input sequences are of uniform length. These preprocessing steps help standardize the data and prepare it effectively for training a deep learning model like BERT.

Model Selection and Fine - Tuning

For training the fake news detection model, a suitable pre-trained BERT variant is selected based on performance and computational needs. Options include bert-base-uncased for robust performance, bert-tiny for faster and lighter inference, or a domain-specific model like bert-tiny-finetuned-fake-news-detection, which is already tailored for this task. On top of the BERT model, a classification head—a fully connected layer with a softmax activation—is added to perform binary classification (real or fake).The model is fine-tuned using a labeled dataset, typically split into training and validation sets or processed through k-fold cross-validation to ensure reliability. Cross-entropy loss is used as the loss function, which is ideal for classification problems, while AdamW is employed as the optimizer due to its effective handling of weight decay and faster convergence. To evaluate the model's performance, standard classification metrics are used: accuracy, precision, recall, and F1-score.

BERT Overview

BERT (Bidirectional Encoder Representations from Transformers) is a language model that understands words in context by looking at both left and right sides of a word in a sentence. This bidirectional approach helps capture deeper meaning compared to traditional models. In this project, a pre-trained BERT model is fine-tuned on a fake news dataset to classify news articles as real or fake. By leveraging BERT's strong ability to understand language context, the model achieves higher accuracy in detecting misinformation with less manual feature work. BERT forms the core of the system, enabling effective and reliable fake news detection.

Mathematical Foundation

The performance of Credit score is calculated as:

A = [(mostly true counts)*0.2 + (half true counts)*0.5 + (barely true counts)*0.75 + (false counts)*0.9 + (pants on fire counts)*1] B = [mostly true counts + half true counts + barely true counts + false counts) + pants on fire counts] Credit_score = A / B

Evaluation

Model evaluation involves using a validation set to assess the performance of the fake news detection system. Key metrics tracked include the confusion matrix, which provides insights into true positives, false positives, true negatives, and false negatives, helping to understand the model's classification accuracy. Additionally, the classification report summarizes precision, recall, and F1-score, offering a detailed view of the model's effectiveness in identifying fake versus real news. The ROC-AUC curve is also used to evaluate the model's ability to distinguish between classes across different threshold settings. To ensure robustness, the BERT-based model's performance is compared with baseline models such as Naive Bayes and Logistic Regression, providing context and highlighting the advantages of using deep learning techniques over traditional methods.

Testing

Testing involves preprocessing the unseen test data, feeding it into the fine-tuned BERT model, and generating predictions. These are compared with true labels to calculate metrics like accuracy, precision, recall, and F1-score, along with confusion matrix and ROC-AUC, to evaluate the model's effectiveness in detecting fake news.

Model Deployment

Model deployment involves integrating the fine-tuned BERT model into a user-friendly application, often using frameworks like Streamlit or Flask. The model is hosted on a server or cloud platform to enable real-time predictions. Users input news articles, which are processed and classified by the model as fake or real. Deployment ensures accessibility, scalability, and quick response times, allowing the system to effectively detect misinformation in practical, everyday scenarios.

System Optimization and Maintenance

The model's efficiency and responsiveness by techniques like model pruning, quantization, or using lighter versions of BERT for faster inference. Maintenance includes regularly updating the model with new data to handle emerging fake news patterns, monitoring performance to detect drift, and fixing bugs. Continuous evaluation and retraining ensure the system remains accurate, reliable, and effective over time.

DISCUSSION

Limitations

Despite its effectiveness, this fake news detection system has several limitations. Firstly, the model heavily relies on the quality and diversity of the training data; biased or insufficient datasets can reduce its ability to generalize across different topics or sources. Secondly, BERT-based models are computationally intensive, requiring significant processing power and memory, which can limit real-time deployment on low-resource devices. Additionally, the system may struggle with subtle misinformation that requires deeper fact-checking or external knowledge beyond text patterns. It also faces challenges in handling multilingual content or evolving fake news tactics that use sophisticated language tricks. Finally, false positives and negatives can occur, potentially misclassifying legitimate news as fake or vice versa, which impacts user trust. Continuous updating and combining this approach with other verification methods are essential to mitigate these limitations and improve overall reliability.

Future Work

Future work on this fake news detection project can focus on several key areas to enhance its effectiveness. Incorporating multimodal data—such as images, videos, and social media metadata—can provide richer context and improve detection accuracy. Expanding the model to support multiple languages will help address the global spread of misinformation. Additionally, integrating external knowledge bases or fact-checking APIs could enable deeper verification beyond text patterns. Optimizing the model for faster inference and lower computational costs will facilitate real-time deployment on various devices. Lastly, continuous learning methods can help the system adapt to evolving fake news strategies, ensuring sustained performance over time.

MODELS AND ACCURACY

Metrics	Class (Real)	Class (Fake)
Precision	0.87	0.86
Recall	0.85	0.88
F1 Score	0.86	0.87
ROC-AUC Score	0.91	
Accuracy	86.5%	

Figure – 1 ACCURACY

CONCLUSION

The news checker project demonstrates a solid foundation for fake news detection with systematic data collection from the established LIAR dataset and a practical testing interface. The binary classification approach simplifies deployment while the dual-layer prediction system adds robustness. However, the current implementation lacks comprehensive performance evaluation metrics, which are crucial for validating model effectiveness and reliability. The absence of standard ML evaluation practices represents a critical gap that must be addressed before the system can be considered reliable or deployment-ready. Implementing the recommended metrics framework would significantly enhance the project's credibility and provide actionable insights for model improvement. efficient industrial processes.

ACKNOWLEDGEMENT

We extend our heartfelt gratitude to our Guide, Mrs. Gayathri N, for her invaluable guidance and continuous support throughout this project. We also express our sincere thanks to the Department of Artificial Intelligence and Machine Learning faculty and staff at Sri Shakthi Institute of Engineering and Technology for providing essential resources and facilities that enabled the successful completion of this work.
We are deeply grateful to our colleagues and peers for their constructive feedback and collaboration, which greatly contributed to the refinement of the system. Special appreciation goes to the support of open-source communities and publicly available datasets that enriched the training and evaluation process. Lastly, we thank our families and friends for their unwavering support, encouragement, and understanding during this journey.

REFERENCES

1.] A. Fake News Detection on Social Media: A Data Mining Perspective (2019) Available at https://link.springer.com/article/10.1007/s41019-017-0028-2
2.] https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9620068https://pypi.org/project/transformers/4.4.2/
3.] https://github.com/topics/fakenewsdetection
4.] Fake News Detection Naïve Bayes classification https://ieeexplore.ieee.org/document/8546944
5.] Fake News Detection NLP Techniques https://github.com/nishitpatel01/Fake_News_Detection
6.] Fake News Detection by Learning Convolution Filters through Contextualized Attention https://github.com/ekagra-ranjan/fake-news-detection-LIAR-pytorch