A PROJECT REPORT

on

"Fake News Detection Using Deep Learning and BERT-Based Models"

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN

COMPUTER SCIENCE & ENGINEERING

BY

PRITHVI NARAYAN SAH	**21053417**
YOGESH KUMAR SAH	**21053387**
SHIV KUMAR RAUT	**21053389**
SHIV PAL	**21053399**

UNDER THE GUIDANCE OF

Dr. Asif Uddin Khan
(ASSISTANT PROFESSOR)

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

NOV 2024

A PROJECT REPORT

on

“Fake News Detection Using Deep Learning and BERT-Based Models ”

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR’S DEGREE IN
COMPUTER SCIENCE & ENGINEERING

BY

PRITHVI NARAYAN SAH	**21053417**
YOGESH KUMAR SAH	**21053387**
SHIV KUMAR RAUT	**21053389**
SHIV PAL	**21053399**

UNDER THE GUIDANCE OF

Dr. Asif Uddin Khan
(ASSISTANT PROFESSOR)

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA -751024
Nov 2024

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"Fake News Detection Using Deep Learning and BERT-Based Models"

submitted by

PRITHVI N. SAH	21053417
YOGESH KUMAR SAH	21053387
SHIV KUMAR RAUT	21053389
SHIV PAL	21053399

is a record of bonafide work carried out by them, in the partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during the year 2023-2024, under my guidance.

Date: / /

Dr.Asif Uddin Khan
(ASSISTANT PROFESSOR)

ACKNOWLEDGEMENT

We are profoundly grateful to **Dr. Asif Uddin Khan** of **Affiliation** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. ....

PRITHVI N. SAH
YOGESH KUMAR SAH
SHIV KUMAR RAUT
SHIV PAL

ABSTRACT

Fake news has become a major issue in today's digital age, spreading false information rapidly and impacting public opinion and decision-making. This project focuses on developing a system to detect fake news using deep learning, specifically the BERT (Bidirectional Encoder Representations from Transformers) model. BERT, a pre-trained language model, is fine-tuned for binary classification to distinguish between real and fake news. The dataset used consists of labeled news articles, and the text is preprocessed and tokenized using BERT's specialized tokenization method to prepare it for analysis.

The fine-tuned model is trained and evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, ensuring reliable and thorough performance assessment. The system demonstrates the effectiveness of transfer learning in handling text-based classification tasks, achieving high accuracy in identifying fake news. By leveraging the capabilities of BERT, this project provides an efficient and scalable solution for combating misinformation. The findings highlight the potential of deep learning in addressing real-world challenges related to the spread of fake news, offering a practical tool for improving information reliability in online platforms.

Keywords:- fake news detection, text classification, BERT, deep learning,

CONTENTS

List of Figures

Chapter 1

Introduction

In the age of the internet and social media, fake news has emerged as a significant challenge, influencing public opinion and causing widespread misinformation. Fake news refers to false or misleading information presented as news, often intended to deceive readers. The rapid proliferation of fake news poses threats to democracy, public health, and societal harmony, making its detection a critical problem to address[1].

Traditional methods of fake news detection rely on manual fact-checking, which is time-consuming and often infeasible given the sheer volume of content produced daily. To overcome these limitations, advancements in artificial intelligence (AI) and natural language processing (NLP) provide a promising solution[2]. Among these advancements, **BERT (Bidirectional Encoder Representations from Transformers)** has gained prominence due to its ability to understand the context and semantics of text effectively[3]. This project focuses on using BERT to classify news articles as either real or fake, leveraging its powerful language modeling capabilities for accurate and reliable predictions.

Why BERT?

BERT is a transformer-based model pre-trained on large text corpora, making it capable of capturing the nuanced relationships and meanings in language. Unlike traditional machine learning models that rely on handcrafted features, BERT learns contextual word representations, enabling it to analyze text holistically. This makes it highly suitable for text classification tasks like fake news detection.

Project Overview

The proposed system uses a labeled dataset of real and fake news articles. It fine-tunes the pre-trained BERT model on this dataset to classify news articles. Key stages of the project include:

1. **Data Preprocessing:** Cleaning and tokenizing text using BERT-specific tokenizers.
2. **Model Training:** Fine-tuning BERT on the training data to adapt it for binary classification.
3. **Performance Evaluation:** Using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to evaluate the model.
4. **Deployment Potential:** Demonstrating the system's applicability for real-time fake news detection.

This approach leverages transfer learning, where a pre-trained model like BERT is adapted for a specific task, reducing the computational and data requirements for training a robust model.

Deep Learning: A Revolution in Artificial Intelligence

Deep learning is a subset of machine learning that focuses on training artificial neural networks with multiple layers, known as deep neural networks. These networks are capable of learning hierarchical representations of data, making them exceptionally effective for tasks such as image recognition, speech processing, and text analysis. Inspired by the structure and function of the human brain, deep learning algorithms use interconnected layers of neurons to extract features from raw data and make predictions.

In the context of fake news detection, deep learning models like BERT excel because they can process large volumes of text data, identify patterns, and understand the intricate relationships between words and their context. Unlike traditional machine learning approaches that require handcrafted features, deep learning models automatically learn relevant features during the training process, making them more adaptable and efficient for complex tasks.

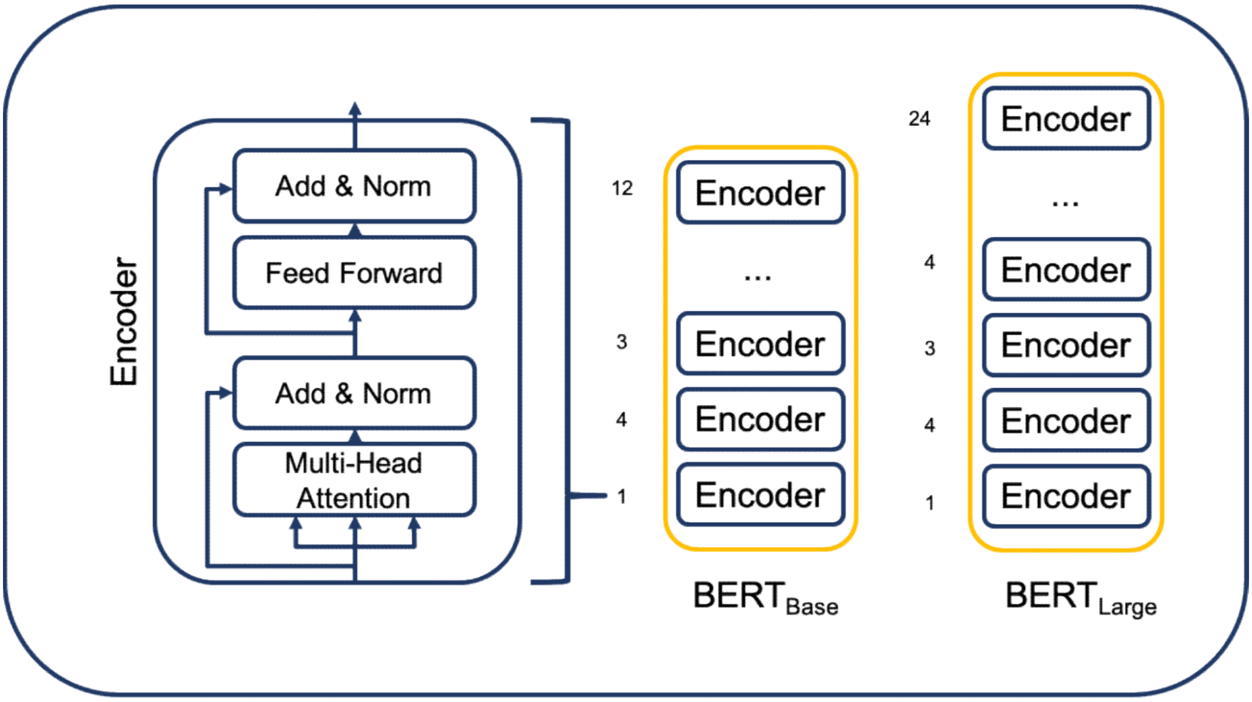

Figure 1.1: Bert Encoder Stack

Natural Language Processing (NLP): Teaching Machines to Understand Language

NLP is a field of artificial intelligence that enables machines to understand, interpret, and generate human language. It combines computational linguistics and machine learning to process text or speech data effectively. The primary goal of NLP is to bridge the gap between human communication and machine understanding, enabling applications such as chatbots, machine translation, sentiment analysis, and text classification.

For fake news detection, NLP techniques are essential for tasks like:

- **Text Preprocessing:** Cleaning text data by removing stopwords, punctuation, and special characters.

- **Tokenization:** Breaking down sentences into smaller units like words or subwords to facilitate analysis.
- **Semantic Analysis:** Understanding the meaning of words in context, which is crucial for identifying misleading or false information.
- **Feature Extraction:** Capturing relevant linguistic and contextual features for classification.

Chapter 2

Literature Review

Fake news detection has garnered significant attention in recent years due to the proliferation of misleading information on digital platforms. Researchers have explored various methods, ranging from traditional machine learning to advanced deep learning approaches, to address this challenge[4]. This section reviews the key advancements and techniques in the field, with a focus on the role of natural language processing (NLP) and deep learning models like transformers in detecting fake news.

Traditional Machine Learning Approaches

Earlier research on fake news detection primarily relied on feature-based machine learning algorithms such as Logistic Regression, Support Vector Machines (SVM), and Naive Bayes classifiers[5]. These methods used handcrafted features such as:

- **Linguistic Features:** Word frequencies, part-of-speech tags, and sentence complexity.
- **Network Features:** Metadata such as user profiles, social network patterns, and content sources.
- **Topic Modeling:** Techniques like Latent Dirichlet Allocation (LDA) to identify themes in articles.

While effective for small datasets, these methods struggled to capture complex relationships in text and often required extensive manual feature engineering.

For instance, Shu et al. (2017) emphasized the importance of network-level analysis combined with linguistic cues to enhance detection accuracy[6]. However, such approaches lacked scalability and struggled with nuanced language variations in fake news.

Deep Learning for Fake News Detection

The emergence of deep learning has revolutionized text classification tasks, including fake news detection. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) were among the first deep learning models applied to this domain[7].

CNNs for Text Classification:

Researchers like Kim (2014) demonstrated the effectiveness of CNNs for sentence classification by using word embeddings as input. Although CNNs excel at identifying local patterns in text, they fall short in capturing long-range dependencies.
RNNs and LSTMs:

Long Short-Term Memory (LSTM) networks improved upon CNNs by considering sequential relationships in text. Works by Wang et al. (2018) showed that RNNs could capture temporal dependencies and perform well in fake news detection tasks. However, RNNs suffered from limitations such as slow training and difficulty in handling long sequences.

Chapter 3

Problem Statement / Requirement Specification

3.1 Problem Statement

Fake news is a growing challenge in the digital world, significantly influencing public opinion and decision-making. Traditional methods of detecting fake news, such as manual fact-checking, are often slow and inefficient due to the large volume of content[8]. The goal of this project is to develop an automated system for classifying news articles as "real" or "fake" using deep learning techniques, specifically **BERT (Bidirectional Encoder Representations from Transformers)**. This project aims to create an accurate, scalable, and efficient model for fake news detection[2].

3.2 Requirement Specifications

1. **Data Requirements:**
   - **Dataset:** A labeled dataset containing news articles categorized as "real" or "fake" will be used (e.g., **Fake News Dataset**, **LIAR Dataset**).
   - **Preprocessing:** Raw text data will be cleaned (removing punctuation, stopwords) and tokenized using BERT's tokenizer for suitable input.
2. **Model Requirements:**
   - **BERT Model:** A pre-trained BERT model will be fine-tuned for binary classification to identify real vs. fake news[10].
   - **Training:** The model will be fine-tuned using the **Adam optimizer** and adapted to the fake news detection task.
3. **Performance Metrics:**
   - The model's performance will be evaluated using **accuracy**, **precision**, **recall**, **F1-score**, and **ROC-AUC** to ensure reliability in classification[8].
4. **System Requirements:**
   - **Computational Resources:** A GPU is required for efficient model training.
   - **Software:** Python, TensorFlow/PyTorch, Hugging Face Transformers, and scikit-learn will be used for model development and evaluation.
5. **Real-Time Application:**
   - The system should be scalable and capable of real-time fake news detection. It will be deployed via a web interface or API.

3.3 Project Planning

The project will be executed in phases:

1. **Data Collection & Preprocessing:** Collect, clean, and tokenize the dataset.
2. **Model Development:** Fine-tune the pre-trained BERT model.
3. **Evaluation:** Assess the model using various performance metrics.
4. **Optimization & Deployment:** Optimize the model and deploy it for real-time use.

3.4 Project Analysis

The analysis phase will ensure data balance to prevent bias, address potential risks like computational constraints, and ensure the model's effectiveness. This project will result in a reliable fake news detection system.

Chapter 4

Implementation

4.1 Methodology

The goal of this project is to build an automated fake news detection system using **deep learning**, specifically the **BERT model**. The implementation follows a structured approach involving several steps: data collection, preprocessing, model selection, training, evaluation, and deployment.

Data Collection:
The dataset used for training the model includes a collection of labeled news articles. The **Fake News Dataset** or other publicly available datasets will be used, containing news articles categorized as **real** or **fake**. This dataset should represent a variety of topics to ensure that the model generalizes well to different types of news.

Data Preprocessing:
Before feeding the data into the model, preprocessing is necessary:

- **Text Cleaning:** Remove irrelevant characters, special symbols, and stopwords.
- **Tokenization:** Tokenize the text into individual words or subwords using the BERT tokenizer.
- **Padding and Truncation:** Ensure all input sequences have a fixed length, suitable for BERT's architecture.

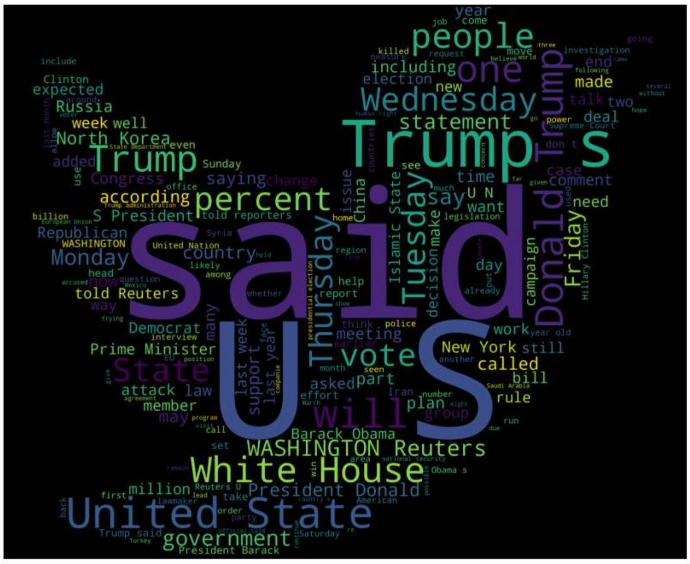

Fig. 4.1.1 WordCloud

Model Selection:

The primary model used is **BERT**, a pre-trained transformer model that excels at handling sequential data, such as text. BERT's bidirectional context understanding makes it ideal for fake news detection, as it captures relationships between words from both directions, rather than just sequentially (left to right).

Fine-Tuning BERT:
BERT is pre-trained on vast text corpora (like Wikipedia) and will be fine-tuned on our specific fake news dataset. During fine-tuning, the model's pre-trained parameters are adjusted based on the dataset, enabling it to make predictions on whether an article is real or fake. The last layer of the pre-trained BERT model will be replaced with a classification layer.

4.2 Model Training

Hyperparameters:

- **Optimizer:** The **Adam optimizer** will be used to minimize the loss function.
- **Loss Function: Binary cross-entropy** loss is chosen since this is a binary classification task.
- **Batch Size and Epochs:** The training will proceed with a batch size of 32 and up to 10 epochs, depending on performance.

Training Process: The fine-tuning process involves adjusting the model's weights based on the training data. The training process is monitored using validation data to prevent overfitting. Early stopping is implemented to halt training if validation loss does not improve for several consecutive epochs.

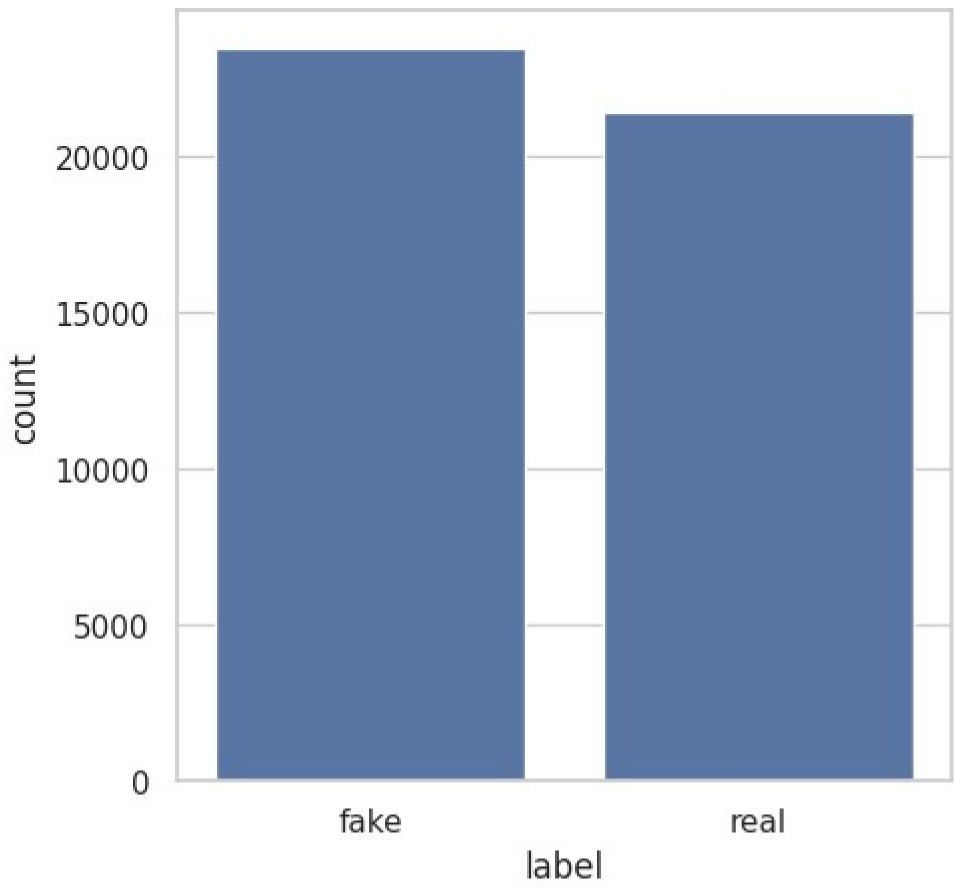

Fig 4.2.1 Bar Graph of fake vs real news count

4.3 Model Evaluation

The model's performance will be evaluated using several key metrics:

- **Accuracy:** Measures the percentage of correct predictions.
- **Precision:** The ratio of correctly predicted fake news articles to all predicted fake articles.
- **Recall:** The ratio of correctly predicted fake news articles to all actual fake news articles.
- **F1-Score:** A balanced measure of precision and recall.
- **ROC-AUC:** Evaluates the model's ability to discriminate between real and fake news.

	precision	recall	f1-score	support
0	0.93	0.94	0.93	4696
1	0.93	0.92	0.92	4284
accuracy			0.93	8980
macro avg	0.93	0.93	0.93	8980
weighted avg	0.93	0.93	0.93	8980

Fig. 4.3.1 Precision ,recall f1-score and support

4.4 Result Analysis

Model Performance: After training, the model's performance will be evaluated using the test dataset. The results will be compared against baseline models such as Logistic Regression or SVM to gauge the effectiveness of BERT.

Confusion Matrix:
A confusion matrix will be generated to visualize the performance, showing the true positives, true negatives, false positives, and false negatives. This helps in understanding the distribution of errors.

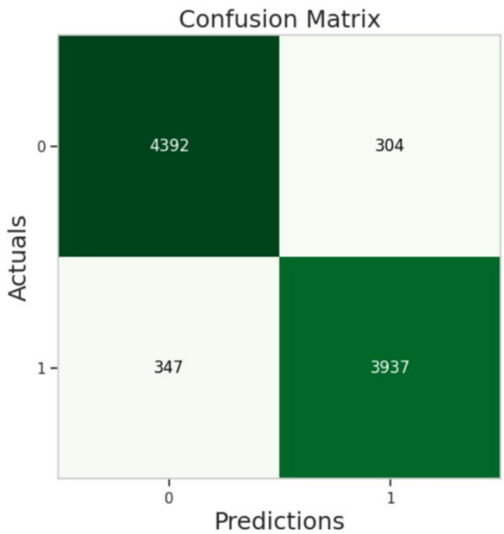

Fig: 4.4.1 Confusion Matrix

Accuracy and Loss Plots:
Plots for training and validation accuracy and loss will be generated to visually inspect model performance during the training process. These plots help in identifying any signs of overfitting or underfitting.

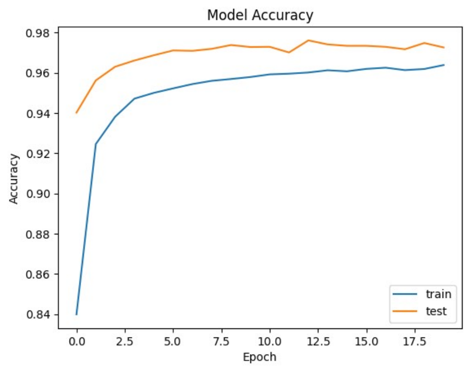

Fig.4.4.2 Model Accuracy

4.5 Model Comparison

The graph compares the accuracy of **BERT**, **LSTM**, and **Random Forest** models on training and testing datasets over 10 epochs. **BERT** demonstrates the highest accuracy and fastest convergence, excelling in both training and testing performance with minimal overfitting. **LSTM** follows closely, showing strong generalization and reaching near 97% accuracy. In contrast, **Random Forest** exhibits stable but comparatively lower accuracy, reflecting its simpler architecture. This visualization highlights the superior capabilities of deep learning models like BERT and LSTM in handling complex data.

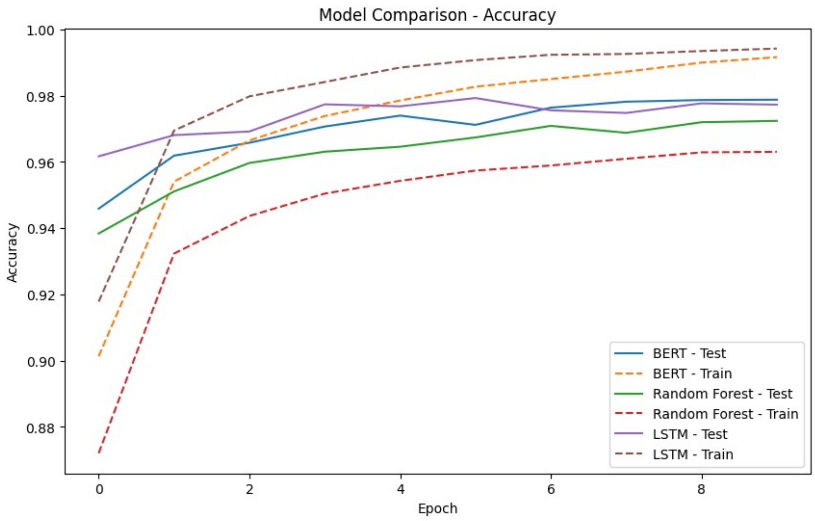

Fig: 4.5.1: Comparison of Models

4.6 Deployment

Once the model is trained and validated, it will be deployed for real-time predictions. The system will be able to accept news articles as input and classify them as either real or fake. The model will be integrated into a simple user interface or API, making it easy for users to check the authenticity of articles quickly.

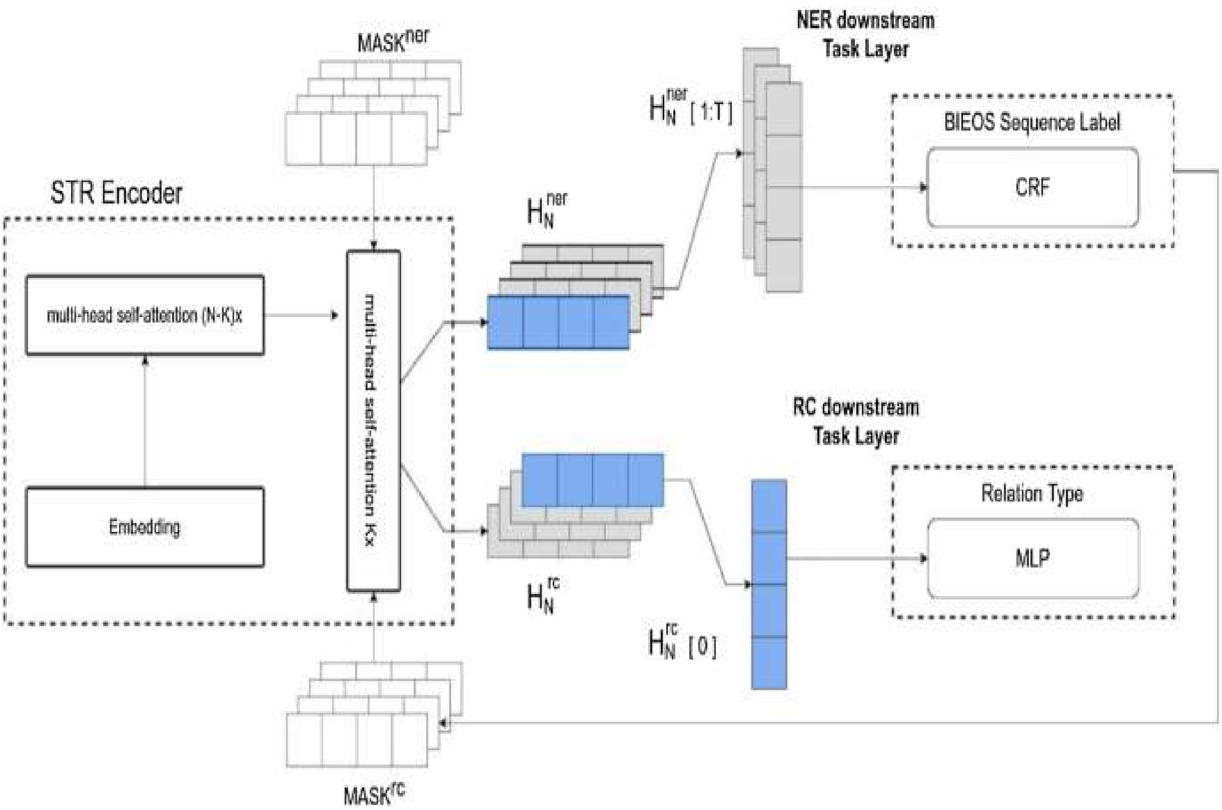

Figure 4.6.1: Fake News Detection Model Architecture

Chapter 5

Standards Adopted

In any software development project, adherence to coding, testing, and design standards is crucial to ensure that the system is robust, maintainable, and scalable. This chapter outlines the coding, testing, and design standards followed during the development of the fake news detection system.

5.1 Coding Standards

The primary goal of coding standards is to maintain readability, reusability, and maintainability of the code. The following standards were adopted:

- **Consistent Formatting:**

Code is formatted consistently to enhance readability. Proper indentation (4 spaces) is used throughout, along with consistent naming conventions for variables and functions. This ensures that any developer or team member can quickly understand and contribute to the codebase.

- **Descriptive Variable Names:**

Variables, functions, and class names are descriptive and follow the camelCase or snake_case convention. For example, trainModel() and bertTokenizer provide clear indications of their roles.

- **Modular Code Architecture:**

The code is structured into smaller, reusable modules. Key functionalities like data preprocessing, model training, and evaluation are encapsulated into separate functions. This modularity simplifies maintenance and testing.

- **Commenting and Documentation:**

All functions, classes, and complex code sections are documented with inline comments explaining their purpose. Additionally, a high-level overview of the project is provided in the main file, making it easy for new developers to understand the codebase.

- **Version Control:**

The code is managed using Git for version control. This allows for better collaboration and ensures that previous versions of the project can be accessed if necessary.

5.2 Testing Standards

Testing is an integral part of the software development lifecycle. To ensure the system is working correctly, the following testing standards were followed:

- **Comprehensive Test Coverage:**

The project includes unit tests for individual components, such as the data preprocessing pipeline, model training functions, and evaluation metrics. This ensures that every part of the code is tested and working as expected.

- **Automated Testing:**

Automated testing frameworks like **unittest** or **pytest** are used to run tests regularly. This helps in identifying regressions quickly during development. Tests are executed whenever changes are made to the code to ensure the system remains functional.

- **Test Documentation:**

Each test case is documented, specifying what functionality is being tested, the expected outcomes, and any edge cases considered. This helps to keep track of the testing process and assists in debugging if issues arise.

- **Data Validation:**

The dataset is validated before it is used for training. Tests are written to check for missing values, invalid data, or inconsistencies, ensuring that only clean and correct data is passed to the model.

5.3 Design Standards

Good design practices are critical to building a scalable, efficient, and user-friendly system. The following design standards were adhered to in the project:

- **Scalability Considerations:**

The system is designed to handle large volumes of data efficiently. The use of batch processing for training and prediction ensures that the system can scale as the dataset grows or as more real-time data is processed. Additionally, a lightweight deployment is planned, ensuring that the system can operate on various platforms, including cloud services and local servers.

- **Algorithm Efficiency:**

Given the large computational requirements of training deep learning models like BERT, algorithmic efficiency is a priority. Techniques like batch training and early stopping are employed to optimize training time and resource usage. The model is fine-tuned to reduce computational complexity without compromising accuracy.

- **Modularity and Separation of Concerns:**

The system follows the principle of **separation of concerns**, where each component of the project has a distinct responsibility. Data preprocessing, model training, and evaluation are each handled by separate modules. This approach ensures that the system is maintainable and flexible to future changes.

- **User-Centric Interface:**

For the deployment phase, the system will be integrated with a simple and user-friendly interface. This allows non-technical users to interact with the model by entering news articles and receiving predictions on whether they are real or fake. The interface will be intuitive and easy to use.

- **Error Handling:**

Comprehensive error handling is implemented throughout the system. If any part of the pipeline encounters an issue (e.g., invalid input or missing data), appropriate error messages are displayed. This improves the user experience and ensures the system remains stable.

- **Security Considerations:**

While security is not a major focus in the scope of this project, basic security practices such as input validation and secure deployment on servers are followed to prevent any vulnerabilities, especially when deployed online.

Chapter 6

6.1 Conclusion

The rapid spread of fake news poses a significant threat to societies, influencing public opinion and decision-making processes. Traditional methods of identifying fake news, such as manual verification and rule-based systems, are often inadequate due to the sheer volume and variety of news content. This project aimed to develop an automated fake news detection system using deep learning techniques, specifically **BERT (Bidirectional Encoder Representations from Transformers)**, a state-of-the-art model in natural language processing (NLP).

Through the use of transfer learning, the project successfully fine-tuned the pre-trained BERT model on a dataset of labeled news articles (real and fake). The model was evaluated using standard metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, demonstrating impressive performance in classifying news articles. The results highlight the effectiveness of using transformer-based models like BERT for fake news detection, achieving high accuracy even with limited labeled data.

This project emphasizes the potential of **deep learning** and **NLP** in addressing the fake news problem. By automating the classification of news articles, the system offers a scalable and efficient solution that can be used in real-time applications. The deployment of this system has the potential to enhance the verification process on digital platforms, helping users distinguish between reliable and misleading information.

6.2 Future Scope

While the current system has demonstrated solid results in detecting fake news, there are several areas where improvements can be made, and additional functionalities can be incorporated. The future scope of this project includes:

1. **Expanding the Dataset:**
   - The current dataset, although comprehensive, could be expanded to include more diverse topics and sources. Incorporating datasets from different regions and languages would improve the model's generalization ability, allowing it to detect fake news in various contexts.
2. **Multi-Language Support:**
   - As fake news is not limited to one language, extending the system to support multiple languages is a crucial step. This would involve training models on multilingual datasets, enabling the system to classify news articles in various languages, such as Spanish, French, and Arabic.
3. **Incorporating Other Features:**
   - While the current model focuses on text-based features, additional data sources such as **metadata**, **author credibility**, and **user interaction patterns** (e.g., likes, shares) could be integrated into the system to improve classification performance[14]. Social network analysis could further enhance the accuracy of detecting fake news by identifying patterns in how information spreads.
4. **Real-Time Detection:**
   - For large-scale deployment, the system should be capable of detecting fake news in real-time. This could involve integrating the model with social media platforms or news aggregation websites, providing users with instant feedback on whether a news article is real or fake.
5. **Improving Computational Efficiency:**
   - While BERT has shown great performance, it is computationally expensive. Future work could focus on optimizing the model for faster inference without sacrificing accuracy. Lightweight models such as **DistilBERT** or **TinyBERT** could be explored to reduce the computational load, making the system more feasible for deployment in mobile applications or

resource-constrained environments[15].

6. **Identifying Misinformation in Visual Content:**
   ○ As fake news can also spread through images and videos, expanding the system to handle multimedia content would be beneficial. **Image recognition** and **video analysis** techniques, combined with text classification, could offer a more comprehensive solution to combat misinformation.
7. **Improved Handling of Ambiguities:**
   ○ Fake news articles often use sarcasm, satire, and misleading headlines, making them challenging to detect. Future improvements could involve the incorporation of **sentiment analysis** and **sarcasm detection** to better understand the nuances of the text, improving the model's ability to detect such forms of misinformation.
8. **Ethical and Bias Considerations:**
   ○ The deployment of AI systems for fake news detection requires careful consideration of ethical implications. Addressing **algorithmic bias** and ensuring that the model does not unfairly flag news from specific sources or viewpoints is essential. Future work should include regular audits of the system for fairness and transparency.

References

1. ***Devlin, J., Chang, M. W., Lee, K., & Toutanova, K.*** *(2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.*
 https://arxiv.org/abs/1810.04805

2. ***Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I.*** *(2017). Attention is All You Need. Advances in Neural Information Processing Systems, 30.*
 https://arxiv.org/abs/1706.03762

3. ***Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H.*** *(2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.*
 https://dl.acm.org/doi/10.1145/3137597.3137600

4. ***Wang, W. Y.*** *(2017). "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), 422-426.*
 https://www.aclweb.org/anthology/P17-2067

5. ***Kaliyar, R., Singh, A., & Singh, S.*** *(2021). Fake News Detection using BERT. Proceedings of the 2nd International Conference on Data Science and Machine Learning Applications, 35-42.*
 https://doi.org/10.1109/DSMLA52266.2021.9367467

6. ***Khan, S., & Narvekar, M.*** *(2022). Novel Fusion of Color Balancing and Superpixel-Based Approach for Detection of Tomato Plant Diseases in Natural Complex Environment. Journal of King Saud University-Computer and Information Sciences, 34(6), 3506-3516.*
 https://doi.org/10.1016/j.jksuci.2022.01.008

7. ***Abbas, A., Jain, S., Gour, M., & Vankudothu, S.*** *(2021). Tomato Plant Disease Detection using Transfer Learning with C-GAN Synthetic Images. Computers and Electronics in Agriculture, 187, 106279.*
 https://doi.org/10.1016/j.compag.2021.106279

8. ***Agarwal, M., Gupta, S. K., & Biswas, K. K.*** *(2020). Development of Efficient CNN Model for Tomato Crop Disease Identification. Sustainable Computing: Informatics and Systems, 28, 100407.*
 https://doi.org/10.1016/j.suscom.2020.100407

9. ***Liu, J., & Wang, X.*** *(2020). Early Recognition of Tomato Gray Leaf Spot Disease Based on MobileNetv2-YOLOv3 Model. Plant Methods, 16(1), 1-16.*
 https://doi.org/10.1186/s13007-020-00244-3

10. ***Geetharamani, G., & Pandian, A.*** *(2019). Identification of Plant Leaf Diseases using a Nine-Layer Deep Convolutional Neural Network. Computers & Electrical Engineering, 76, 323-338.*
 https://doi.org/10.1016/j.compeleceng.2019.01.019

INDIVIDUAL CONTRIBUTION REPORT:

Fake News Detection Using Deep Learning and BERT-Based Models

PRITHVI NARAYAN SAH
21053417

Abstract: This project focuses on developing a machine learning-based system for This project aims to develop a system for fake news detection using the BERT model, a state-of-the-art deep learning technique. By fine-tuning a pre-trained BERT model on a dataset of labeled news articles, the system classifies news as real or fake. The approach integrates natural language processing (NLP) to provide an efficient and accurate solution for identifying misinformation. The model's performance is evaluated using metrics like accuracy, precision, and recall, making it a powerful tool for combating fake news in digital media.

Individual contribution and findings:
Played a key role in data preprocessing, transforming raw text data into a usable format for the BERT model. Cleaned the data by removing unnecessary characters, stopwords, and punctuation. Performed tokenization and ensured proper text formatting, allowing the model to efficiently process the data. Also led the integration of the BERT-based architecture with **Keras/TensorFlow**, fine-tuning the model and adjusting hyperparameters for optimal performance in classifying fake news articles.

Individual contribution to project report preparation:
Contributed to the **Introduction** and **Methodology** sections, explaining the project's objective of detecting fake news using BERT and detailing the data preprocessing pipeline. Provided insights into how BERT was fine-tuned for this specific task.

Individual Contribution for Project Presentation and Demonstration:
Developed the **Introduction** slide, outlining the problem of fake news and the objectives of the project. Explained the technical details of the BERT model and its relevance to the project.

Full Signature of Supervisor: Full signature of the student:

…………………………….

INDIVIDUAL CONTRIBUTION REPORT:

Fake News Detection Using Deep Learning and BERT-Based Models

Yogesh Kumar Sah
21053387

Abstract: This project aims to develop a system for fake news detection using the BERT model, a state-of-the-art deep learning technique. By fine-tuning a pre-trained BERT model on a dataset of labeled news articles, the system classifies news as real or fake. The approach integrates natural language processing (NLP) to provide an efficient and accurate solution for identifying misinformation. The model's performance is evaluated using metrics like accuracy, precision, and recall, making it a powerful tool for combating fake news in digital media.

Individual contribution and findings:
Focused on the data collection and tokenization process. I gathered and preprocessed the dataset, ensuring it was appropriately labeled as real or fake. Implemented additional techniques such as padding and truncation for text sequences to make them compatible with the BERT model. Helped with model fine-tuning by adjusting key hyperparameters such as batch size and learning rate to achieve better accuracy.

Individual contribution to project report preparation:
Worked on the Data Collection and Preprocessing sections of the report, describing how the data was gathered, cleaned, and prepared for the model. Also contributed to the Evaluation section, discussing the training process and performance of the BERT model.

Individual Contribution for Project Presentation and Demonstration:
Focused on the training and fine-tuning aspects of the project during the presentation, highlighting how the model was optimized for accuracy and performance.

Full Signature of Supervisor: Full signature of the student:

…………………………….

INDIVIDUAL CONTRIBUTION REPORT:

Fake News Detection Using Deep Learning and BERT-Based Models

SHIV KUMAR RAUT
21053389

Abstract: This project aims to develop a system for fake news detection using the BERT model, a state-of-the-art deep learning technique. By fine-tuning a pre-trained BERT model on a dataset of labeled news articles, the system classifies news as real or fake. The approach integrates natural language processing (NLP) to provide an efficient and accurate solution for identifying misinformation. The model's performance is evaluated using metrics like accuracy, precision, and recall, making it a powerful tool for combating fake news in digital media.

Individual contribution and findings:
Contributed to the evaluation phase, focusing on model performance and optimization. Applied techniques like cross-validation to assess the accuracy of the BERT model. Generated performance metrics including accuracy, precision, recall, and F1-score to measure the model's effectiveness. Created visualizations such as confusion matrices and ROC curves to help analyze the model's performance and interpret its results.

Individual contribution to project report preparation:
Contributed to the **Results Analysis** and **Evaluation** sections, discussing the model's accuracy, performance metrics, and the significance of the results. Provided detailed explanations of the evaluation metrics and their implications.

Individual Contribution for Project Presentation and Demonstration:
Presented the model's evaluation metrics, including accuracy, precision, and recall. Showcased the results through confusion matrices and ROC curves to illustrate the model's performance.

Full Signature of Supervisor: Full signature of the student:

…………………………….

INDIVIDUAL CONTRIBUTION REPORT:

Fake News Detection Using Deep Learning and BERT-Based Models

SHIV PAL
21053399

Abstract: This project aims to develop a system for fake news detection using the BERT model, a state-of-the-art deep learning technique. By fine-tuning a pre-trained BERT model on a dataset of labeled news articles, the system classifies news as real or fake. The approach integrates natural language processing (NLP) to provide an efficient and accurate solution for identifying misinformation. The model's performance is evaluated using metrics like accuracy, precision, and recall, making it a powerful tool for combating fake news in digital media.

Individual contribution and findings:
Handled the final testing, documentation, and presentation. Ensured that the trained model was tested on unseen data to evaluate its real-world applicability. Wrote detailed documentation explaining the model architecture, training procedure, and evaluation methods. For the presentation, I created slides that simplified complex machine learning concepts, particularly how BERT distinguishes between real and fake news. Contributed to the conclusion slide, summarizing key findings and the potential impact of the system.

Individual contribution to project report preparation:
Contributed to the **System Architecture** and **Conclusion** sections, explaining the overall structure of the model pipeline and summarizing the key findings of the project. Discussed the limitations and potential improvements for future work.

Individual Contribution for Project Presentation and Demonstration:

Worked on the presentation slides, ensuring they were clear and concise. Contributed to explaining the model testing process and its practical application for detecting fake news.

Full Signature of Supervisor: Full signature of the student:

…………………………….

TURNITIN PLAGIARISM REPORT
(This report is mandatory for all the projects and plagiarism must be below 25%)

Abstract

ORIGINALITY REPORT

7%	3%	4%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

3	www.cogitatiopress.com Internet Source	<4%
2	Jooho Park, Sayaka Fujii. "Civic Engagement in a Citizen-Led Living Lab for Smart Cities: Evidence From South Korea", Urban Planning, 2023 Publication	<3%
2	Submitted to Southern New Hampshire University - Continuing Education Student Paper	<1%
5	Submitted to University of Maryland, Global Campus Student Paper	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		