Phase-2 Submission Template

Student Name: DEEPAK T

Register Number: 622423121013

Institution: Salem College of Engineering and Technology

Department: BioMedical Engineering

Date of Submission: 08-05-2025

Github Repository Link:
https://github.com/deepaktk697/Deepak.T-phase-2-project.git

TITLE: Exposing the Truth: Advanced Fake News Detection Powered by Natural Language Processing

1. Problem Statement

The rapid spread of fake news poses significant threats to public trust, democracy, and societal harmony. Despite the availability of information, distinguishing fact from misinformation remains a challenge for users. The goal of this project is to detect and classify fake news articles using state-of-the-art Natural Language Processing (NLP) techniques.

This project expands on the Phase-1 baseline classifier by incorporating transformer-based models, semantic understanding, and explainable AI. The problem type is binary classification, where the objective is to label news as either fake or real based on textual and source-based features.

Addressing this problem helps social media platforms, journalists, and readers ensure the integrity of the information they consume and share.

2. Project Objectives

- **Develop an Intelligent NLP-Based Classifier**
Build a robust machine learning model capable of accurately distinguishing fake news from real news using state-of-the-art Natural Language Processing techniques.

- **Leverage Transformer-Based Models**
Implement advanced transformer architectures such as BERT to enhance semantic understanding and improve classification performance over traditional models.

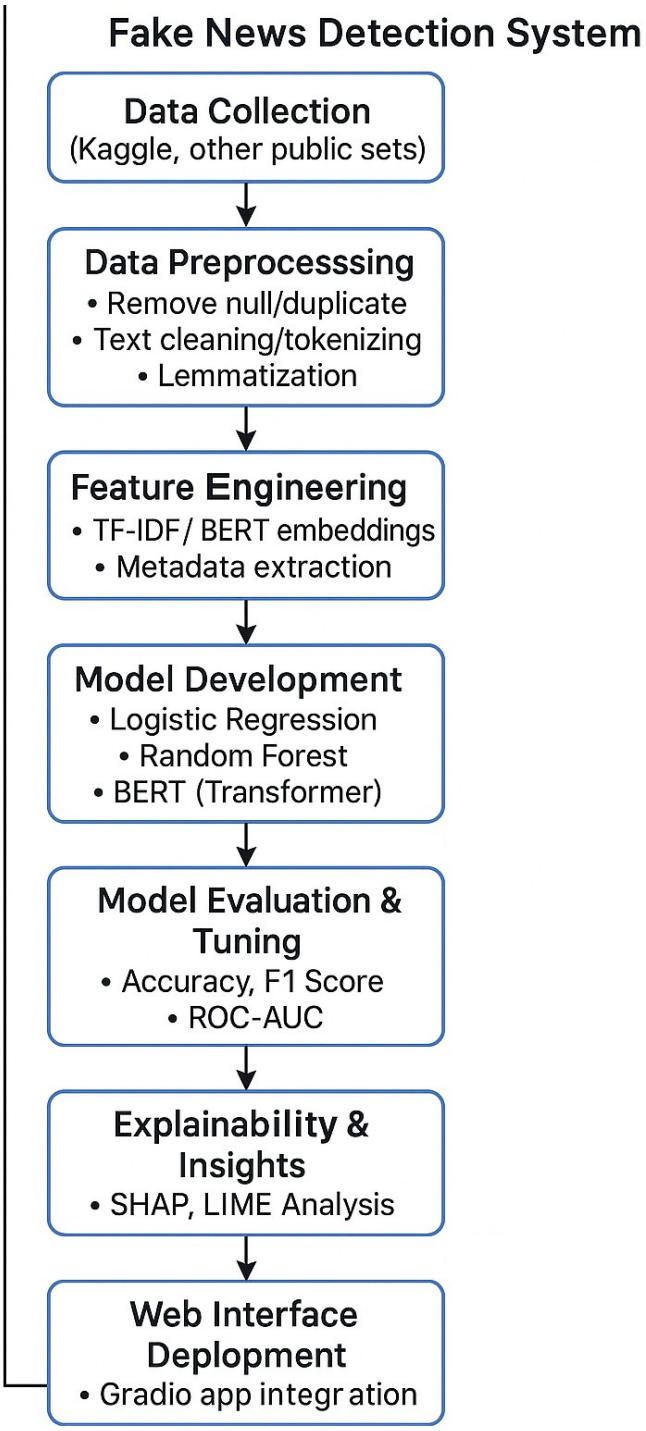
- **Enable Model Explainability**
Integrate explainable AI tools (like LIME and SHAP) to provide transparency and interpretability of model predictions, fostering user trust and accountability.

- **Analyze Linguistic and Contextual Features**
Identify and rank the most influential textual and contextual elements—such as keywords, sentiment, and source credibility—used in fake news detection.

- **Develop a User-Friendly Web Interface**
Create an interactive and accessible web interface using Gradio, allowing users to input news articles and receive real-time predictions with interpretive insights.

- **Promote Public Awareness and Media Literacy**
Provide insights and tools that help users, journalists, and media platforms critically evaluate the authenticity of digital news content. .

3. Flowchart of the Project Workflow



4. Data Description

- ***Dataset Name:*** *Fake and Real News Dataset*
- ***Source:*** *Kaggle / Other public datasets*
- ***Type of Data:*** *Structured tabular data (text + metadata)*
- ***Records and Features:*** *~50,000 news articles, including title, text, subject, and source*
- ***Target Variable:*** *Label (Fake = 1, Real = 0)*
- ***Static or Dynamic:*** *Static dataset*
- ***Attributes Covered:*** *Article title, full text, publication date, subject category, source*

5. Data Preprocessing

- ***Removed*** *null or incomplete articles.*
- ***Merged*** *duplicate entries and standardized column names.*
- ***Text normalization:*** *lowercasing, punctuation removal, stopword filtering.*
- ***Tokenization*** *and lemmatization using SpaCy.*
- ***Vectorization*** *using TF-IDF and embeddings (BERT for Phase-2).*
- ***Handled*** *class imbalance using resampling techniques like SMOTE or class weights.*

6. Exploratory Data Analysis (EDA)

- ***Univariate Analysis:***
    - *Distribution of article lengths, word counts, and label frequency*
- ***Bivariate & Multivariate Analysis:***
    - *Word cloud and keyword analysis for fake vs real articles*
    - *Frequency of certain biased phrases and sources*
    - *Sentiment analysis to detect tone shifts in fake vs real news*

Key Insights:

- *Fake news tends to use more emotionally charged and sensational language.*
- *Certain sources are consistently associated with fake content.*
- *Real news articles are generally longer and use more formal language*

7. Feature Engineering

Extracted linguistic features: POS counts, sentiment scores, readability scores

Built embeddings using BERT for semantic representation

Derived metadata features (e.g., publishing frequency of sources)

Removed multicollinearity through feature selection and variance thresholding

8. Model Building

Algorithms Used:

- Logistic Regression (baseline)
- Random Forest Classifier
- BERT (Bidirectional Encoder Representations from Transformers)

Model Selection Rationale:

- Logistic Regression: fast and interpretable
- Random Forest: handles non-linear data well and shows feature importance
- BERT: state-of-the-art language model for understanding context and semantics

Train-Test Split:

- 80% training, 20% testing
- Stratified sampling to maintain label distribution

Evaluation Metrics:

- Accuracy
- Precision, Recall, F1 Score
- ROC-AUC Curve

9. Visualization of Results & Model Insights

Feature Importance:

- Highlighted top keywords and phrases most predictive of fake news

Model Comparison:

- BERT achieved the highest F1 score, significantly outperforming traditional models

Explainability:

- Used LIME and SHAP to interpret individual predictions

User Testing:

- Integrated model into Gradio app for testing articles and viewing explanations

10. Tools and Technologies Used

- ***Programming Language:*** *Python 3*
- ***Notebook Environment:*** *Google Colab*
- ***Key Libraries:***
  - ***pandas, numpy –*** *data handling*
  - ***nltk, spacy, transformers –*** *NLP processing*
  - ***sklearn, xgboost –*** *machine learning*
  - ***Gradio –*** *interface deployment*
  - ***LIME, SHAP –*** *model explainability*

11. Team Members and Contributions

- **T. Deepak**

- Data Cleaning
- Feature Engineering
- Documentation and Reporting

- **S. Deepak**

- Exploratory Data Analysis (EDA)
- Model Development

- **S.Deepak**

- Interface & Deployment
- Model Development