**Phase-1 Submission Template**

**Student Name:**EZHILARASAN S

**Register Number:** 422623106032

**Institution:** University College of Engineering, Panruti

**Department:** Electronics & Communication Engineering

**Date of Submission:** 30/04/2025

**1.Problem Statement:**

Fake news has become a pervasive issue with significant social, political, and economic consequences. With the rapid spread of misinformation via social media, the public can be easily misled. Detecting and mitigating the spread of fake news is essential to ensuring access to factual information and supporting informed decision-making.

**2.Objectives of the Project:**

- Develop an automated system to classify news articles as fake or real.  
- Leverage Natural Language Processing (NLP) techniques and machine learning algorithms.  
- Provide insights through visualizations and data profiling.  
- Ensure the system is accurate, scalable, and suitable for real-world applications.

**3.Scope of the Project:**

- Analyze and clean text data from news articles.  
- Apply feature engineering techniques like TF-IDF.  
- Train and evaluate a logistic regression model.  
- Limitations: Focused on text classification only; uses static datasets; no deep learning methods implemented.

**4.Data Sources:**

- **Dataset:** Fake and real news dataset sourced from public repositories (e.g., Kaggle).  
- **Nature:** Public and static.  
- **Format:** Text data with labels indicating whether the news is fake or real.

**5.High-Level Methodology:**

- **Data Collection -** Downloaded from a public dataset repository in Kaggle.  
- **Data Cleaning -** Performed stemming, removal of stopwords, and basic preprocessing.  
- **Exploratory Data Analysis (EDA)** - Utilized profiling tools like ydata-profiling, Sweetviz, and D-Tale.  
- **Feature Engineering -** Implemented TF-IDF vectorization to transform text into numeric format.  
- **Model Building -** Trained a Logistic Regression classifier.  
- **Model Evaluation -** Evaluated using accuracy, confusion matrix, and classification report.  
- **Visualization & Interpretation -** Used EDA tools to generate dashboards and statistical summaries.  
- **Deployment -** Model saved for future reuse, not deployed in current phase.

**6.Tools and Technologies:**

- **Programming Language:** Python.  
- **Notebook/IDE:** Google Collab.  
- **Libraries:** NLTK, scikit-learn, pandas, numpy, ydata-profiling, Sweetviz.  
- **Optional Tools for Deployment:** Not deployed; model is saved locally for potential future use.

**7.Team Members and Roles:**

- **N.M.Baalaji:** Data Collection and Sourcing.  
- **J.Kishore:** Data Cleaning and Preprocessing.  
- **Yagnesh Kumaar Sakthivel Mohana:** Exploratory Data Analysis and Visualization.  
- **S.Ezhilarasan:** Model Building & Model Evaluation.  
- **K.Yogaraj:** Deployment.