Phase-1 Submission

Student Name: Bharathidhasan M

Register Number: 712523104007

Institution: PPG INSTITUTE OF TECHNOLOGY

Department: BE-CSE

Date of Submission: 27.04.2025

1.Problem Statement

The rapid spread of fake news on digital platforms poses a significant threat to societal trust, public safety, and democratic processes. This project aims to build an NLP-based system that can automatically detect fake news, helping users and organizations make informed decisions and combat misinformation effectively.

2.Objectives of the Project

- *interface. Build a machine learning model that accurately classifies news as real or fake.*
- *Analyze linguistic patterns in fake vs real news using NLP techniques.*

3.Scope of the Project

- *Text preprocessing and NLP-based feature extraction*
- *Multiple classification models (Logistic Regression, Random Forest, etc.)*

Limitations/Constraints:

- *Model trained on static datasets*
- *Focus limited to English-language news.*

4.Data Sources

- *Fake and real news dataset from Kaggle (e.g., "Fake News Detection" dataset)*
- *Static dataset downloaded once.*

Data Source Link: https://www.kaggle.com/datasets/mahdimashayekhi/fake-new

5.High-Level Methodology

- *Data Collection: Download dataset from Kaggle.*
- *Data Cleaning: Remove duplicates, handle missing values, standardize text formatting.*
- *EDA: Word clouds, frequency plots, sentiment analysis.*
- *Feature Engineering: TF-IDF, n-grams, stop word removal, POS tagging.*
- *Model Building: Logistic Regression, Naive Bayes, Random Forest, XG Boost.*
- *Model Evaluation: Accuracy, Precision, Recall, F1-score, ROC-AU*
- *Data Collection: Download dataset from Kaggle.*
- *Data Cleaning: Remove duplicates, handle missing values, standardize text formatting.*
- *EDA: Word clouds, frequency plots, sentiment analysis.*
- *Feature Engineering: TF-IDF, n-grams, stop word removal, POS tagging.*
- *Model Building: Logistic Regression, Naive Bayes, Random Forest, XG Boost.*

6.Tools and Technologie

- ***Programming Language:*** *Python.*
- ***Notebook/IDE:*** *Google Colab, Jupyter Notebook*
- ***Libraries:*** *pandas, numpy, matplotlib, seaborn, scikit-learn, nltk, spaCy, TensorFlow/Keras*
- ***Deployment Tools (Optional):*** *Streamlit, Gradio, Flask.*

7.Team Members and Roles;

NAME	ROLE	WORK
MURALIDHARAN.K	NLP Engineer	*Model development and evaluation.*
GOWTHAM.P	Marketing & Outreach Lead	*Data collection, cleaning, and preprocessing.*
PUGAZHENTHI.G	Document and presentation	*Visualization and interpretation of results.*
BHARATHIDHASAN.M	Testing and deployment	*(Optional) Deployment and UI design.*
JESLIN SAJAN	Data Scientist	*Materials and coordination among the team members*