

Phase-1 Submission Template

Student Name: Durga Renuka J

Register Number: 712523106005

Institution: PPG Institute Of Technology

Department : Electronics and Communication Engineering

Date of Submission: 25/04/2025

1.Problem Statement

Design and develop an advanced fake news detection system leveraging natural language processing (NLP) techniques to accurately identify and expose false information, mitigating the spread of misinformation and promoting a more informed public discourse.

2.Objectives of the Project

- 1. Fake News Detection: Develop an NLP-powered system to detect fake news articles, social media posts, and other online content.*
- 2. Accuracy and Reliability: Ensure high accuracy and reliability in detecting fake news, minimizing false positives and negatives.*
- 3. Real-time Analysis: Enable real-time analysis and detection of fake news, allowing for swift action to mitigate its spread.*

3.Scope of the Project

- 1. Fake News Detection System Development: Design and development of an NLP-powered system to detect fake news articles and social media posts.*

2. *Dataset Collection and Annotation: Collection and annotation of a large dataset of labeled fake and real news articles.*
3. *NLP Technique Implementation: Implementation of NLP techniques such as text classification, sentiment analysis, and named entity recognition.*
4. *System Deployment: Deployment of the system in a cloud-based environment for real-time analysis.*

4.Data Sources

Dataset: Fake and real news dataset from Kaggle

- Source: Kaggle (<https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset>)
- Public dataset, static in nature
- Contains labeled examples of fake and real news articles

5.High-Level Methodology

- **Data Collection :**

Real News: Verified media outlets (e.g., BBC, Reuters, The Guardian)

Fake News: Fact-checking sites (e.g., PolitiFact, Snopes, FakeNewsNet)

- **Data Cleaning :**

Noise Reduction: Removing irrelevant elements such as HTML tags, advertisements, special characters, URLs, and excessive white spaces. This step ensures a cleaner text corpus for NLP processing.

- **Exploratory Data Analysis (EDA)** – *Perform EDA to understand data distribution, identify patterns, and extract insights using techniques like text analysis, visualization, and correlation analysis to inform fake news detection model development..*

- **Model Building** – *Build and train NLP models using techniques like text classification, sentiment analysis, and deep learning (CNNs, LSTMs, BERT) to detect fake news, and evaluate performance using metrics like accuracy, precision, and F1-score.*
- **Model Evaluation** – *Evaluate the fake news detection model's performance using metrics like accuracy, precision, recall, F1-score, and ROC-AUC, and perform cross-validation to ensure robustness and generalizability.*
- **Visualization & Interpretation** – *Visualize model results using techniques like confusion matrices, ROC curves, and feature importance plots to interpret performance, identify biases, and inform model refinement for advanced fake news detection.*
- **Deployment** – *Deploy the fake news detection model in a cloud-based environment, integrating with APIs for real-time analysis, and ensuring scalability, security, and monitoring for optimal performance and impact.*

6.Tools and Technologies

- **Programming Language** – *Utilize Python as the primary programming language, leveraging popular NLP libraries like NLTK, spaCy, and TensorFlow/Keras for advanced fake news detection model development and deployment.*
- **Notebook/IDE** – *Utilize Jupyter Notebook or Google Colab as the primary development environment, and PyCharm or VS Code for larger-scale development and deployment of the fake news detection model(Python extension, Jupyter Lab)*
- **Libraries** – *Utilize popular NLP libraries like NLTK, spaCy (e.g., for tokenization, entity recognition)*

- **Optional Tools for Deployment** – Utilize optional deployment tools like Docker, Flask, Django, AWS, Google Cloud, and Kubernetes to support scalable, secure, and efficient deployment of the fake news detection model.

7.Team Members and Roles

MEMBERS	ROLE	DESCRIPTION
KISHORE S	<i>Data Collection & Cleaning</i>	<i>Responsible for data collection and cleaning. He will ensure the dataset is gathered from reliable sources and properly preprocessed to remove inconsistencies, duplicates, and irrelevant content.</i>
SUBAGAANTHAN B	<i>EDA & Feature Engineering</i>	<i>Handles exploratory data analysis (EDA) and feature engineering. He will visualize trends and patterns in the dataset and extract meaningful features using techniques like TF-IDF and CountVectorizer.</i>
DURGA RENUGA J	<i>Model Building & Evaluation</i>	<i>3. Focuses on model building and evaluation. She will experiment with various machine learning models, fine-tune them, and evaluate performance using metrics such as accuracy, precision, and recall.</i>
NITHYA S	<i>Visualization & Deployment</i>	<i>In charge of visualization and deployment. She will present the final results</i>

		<i>using graphs and dashboards, and optionally deploy the model using a framework like Streamlit or Flask for user interaction</i>
--	--	--