Phase-1

Student Name: Kamesh.K

Register Number: 410723106016

Institution: Dhanalakshmi college of engineering

Department: Electronics and Communication Engineering

Date of Submission: 28/04/2025

1.Problem Statement

Exposing the truth with advanced fake news detection powered by natural language processing.

In the age of digital media, fake news spreads rapidly across social platforms, influencing public opinion and causing social harm. Detecting and filtering out false information is crucial to maintaining trustworthy news dissemination. This project aims to develop a machine learning-based system to classify news articles as real or fake, helping users and platforms identify misinformation effectively.

2.Objectives of the Project

1. Build a binary classification model to detect whether a given news article is real or fake.
2. Analyse text patterns, keywords, and writing styles that differentiate fake news from real news.
3. Present clear visual insights into the data and model performance.
4. Deliver a prototype tool for users to test or visualize fake news predictions.

3.Scope of the Project

1.Features to Analyse/Build:

1. News title, full text, subject, publication date
2. Natural language features (word frequency, sentiment, etc.)
3. NLP preprocessing (stop words, lemmatization, vectorization)

2.Constraints:

1. Use only publicly available datasets.
2. Limit models to interpretable and widely used NLP algorithms.
3. Final product can be a dashboard or notebook-based tool, not deployed to production-level web service.

3.Model Development:

1. Text classification models (e.g., Support Vector Machines, Decision Trees, Neural Networks)
2. Sentiment analysis techniques
3. Named Entity Recognition to identify key entities in articles.

4.Data Sources

1. Dataset: Fake and Real News Dataset.
2. Source: Kaggle.
3. Type: Public.
4. Nature: static.
5. Link: https://www.kaggle.com/datasets/khushikyad001/fake-news-detection

5.High-Level Methodology

1.Data Collection:

- Gather a diverse and representative dataset of real and fake news articles from reputable sources.
- Include multiple languages, domains (politics, health, finance), and formats (text, headlines, social media posts).

2.Data Preprocessing:

- Clean and normalize text (tokenization, lowercasing, stop word removal, stemming/lemmatization).
- Remove noise such as HTML tags, emojis, or non-textual elements.
- Handle class imbalance through techniques like oversampling or under sampling.

3.Feature Engineering

- Extract linguistic features: n-grams, POS tags, named entities.
- Use semantic features: word embeddings (e.g., Word2Vec, GloVe), contextual embeddings (BERT, RoBERTa).
- Consider metadata: publication date, source credibility, author, social engagement.

4.Model Development

- Train classification models using traditional ML (Logistic Regression, SVM) and deep learning (LSTM, Transformers).
- Fine-tune pre-trained transformer models (like BERT) for fake news classification.
- Explore ensemble methods to improve accuracy.

5.Explainability & Truth Justification

- Use explainable AI (e.g., SHAP, LIME) to interpret model decisions.
- Implement claim verification modules using external fact-checking databases and knowledge graphs.

6.Evaluation

- Use metrics such as Accuracy, Precision, Recall, F1-score, and AUC-ROC.
- Perform cross-validation and test on unseen datasets to evaluate generalizability.

7. Deployment & Monitoring

- Deploy the model in a real-time or batch-processing pipeline.
- Continuously monitor performance and retrain with fresh data to adapt to evolving misinformation tactics

6.Tools and Technologies

- Programming Language: Python
- Notebook/IDE: Google Colab.
- Libraries: pandas, NumPy, matplotlib, seaborn, scikit-learn, nltk, spaCy, Sklearn,feature extraction, text.
- Optional Deployment Tool: Streamlit, Gradio.

7. Team Members and Roles

S.NO	NAMES	ROLES	RESPONSIBILITY
1.	Kamesh.K	Leader	Data collection & cleaning
2.	Kishore kumar.K	Member	Visualization & Interpretation
3.	Monishraj.V	Member	Exploratory data analysis (EDA) Feature engineering
4.	Selvan samuvel.A	Member	Model building,model evaluation