Phase - 1

Exposing the Truth with Advanced Fake News Deduction Powered by Natural Language Processing

Student Name: KOKILA.V

Register Number: 620123106056

Institution: AVS ENGINEERING COLLEGE

Department: ECE

Date of submission: 30/04/2025

1. Problem Statement

In today's digital age, fake news spreads faster than ever before, influencing public opinion, creating confusion, and sometimes leading to serious social and political consequences. Traditional fact-checking methods are too slow and manual to handle the massive volume of online content. There is an urgent need for an automated, reliable, and scalable system to detect fake news in real-time,

2. Objectives of the Project

- To build an intelligent system that can classify news articles as real or fake using NLP techniques.

- To preprocess, analyze, and extract key features from news content.

- To apply machine learning models that understand linguistic patterns

3. Scope of the Project

- Focuses on text-based news articles and social media posts.

- Targets English-language content.

- Covers major fake news topics like politics, health, technology, and finance.

- The system will classify news articles into binary categories: Fake or Real.

- Future scope includes multi-language support, sarcasm detection, and context-based verification.

4. Data Source

- Kaggle datasets:

- Fake News Detection Dataset (contains labeled real and fake news articles)

- LIAR Dataset (short political statements labeled with six truthfulness levels)

- Social Media: Twitter API for recent posts related to trending topics (with caution).

5. High-Level Methodology

1. **Data Collection**: Gather datasets from trusted sources.

2. **Data Preprocessing**:

    - Text cleaning (removing noise, stop words, punctuation)

- Word embeddings (e.g., Word2Vec, GloVe)

- Named Entity Recognition (NER) features

3. Model Building:

- Classical ML Models: Logistic Regression, Random Forest, SVM.

- Deep Learning Models: LSTM, BERT-based transformers.

4. Training and Validation:

- Train models on labeled datasets.

- Evaluate using accuracy, precision, recall, F1 score.

5. Deployment:

- Build a web application to input articles and classify them.

- Optionally, use APIs to automate news verification.

6.Tools and Technology

- Programming Languages: Python

- NLP: NLTK, spaCy, transformers (HuggingFace)

- Data Handling: pandas, numpy

- Visualization: matplotlib, seaborn

- Other Tools:

- Jupyter Notebook / Google Colab (for development

- GitHub (for version control)

7.Team Members and Roles

S.NO	Members	Roles	Responsibility
1	V.KOKILA	Team Leader	I led the team by planning the project timeline assigning responsibility and ensuring consistent progress.I facilitated communication among members, resolved blockers,and made sure we stayed aligned with our goal of detecting fake news accurately
2	M.S.KOVARTHANA	Data Engineer	I handled the end to end data pipeline collecting,cleaning,and storing the dataset of real and fake news.I also performed feature engineering and ensured the data was in the right format for model training.
3	S.AKALYA	NLP Specialist	I desinged and trained the fake news detection model using NLP techinque and machain learning algorithms.I tested different models like logistic regression ,random forest and BERT to find best best performing solution