**Project name : Exposing the Truth with Advanced Fake News Detection Powered By Natural Language Processing**

**Phase-1**

**Student Name:** Akila S

**Register Number**: 623523106003

**Institution**: AVS college of technology

**Department**: ECE

**Date of Submission:** 14.05.2025

**1.Problem Statement**

In the age of digital communication, the rapid spread of fake news across social media and online platforms has become a serious global issue. False or misleading information can easily reach millions of people within minutes, influencing public opinion, promoting misinformation, and even affecting political and social outcomes. Traditional manual fact-checking methods are too slow and resource-intensive to keep up with the volume of content generated every day.

**2.Objectives of the Project**

* To develop an accurate fake news detection model using Natural Language Processing (NLP) techniques.
* To analyze linguistic patterns and features that differentiate fake news from real news.
* To evaluate multiple machine learning and deep learning models (e.g., Logistic Regression, Naive Bayes, LSTM, BERT) for performance comparison.
* To create a user-friendly tool or web application that allows users to input news content and receive a fake/real prediction.
* To provide actionable insights into the common traits of misinformation to support awareness and digital literacy.

**3.Scope of the Project**

**Features to analysis:**

* Use of NLP techniques for fake news detection
* Text preprocessing and feature extraction (e.g., TF-IDF, Word2Vec)
* - Model building using ML/DL algorithms (Logistic Regression, LSTM, BERT)
* Creating a simple web app for user input and prediction.

**Out of Scope /Limitations:**

* No real-time news scraping or updates
* Only English-language news content
* Limited to static, publicly available datasets
* Basic local/web deployment only (no large-scale production)

**4.Data Sources**

* Dataset Used: Fake and Real News Dataset
* Source: Kaggle (public dataset)
* Type: Static – downloaded once and used locally
* Content: Contains labeled news articles (real vs. fake) with fields like title, text, subject, and date
* Usage: Used for training, testing, and evaluating the fake news detection model

*.* **5.High-Level Methodology**

**Data Collection**

* **Source:** We will use publicly available datasets such as the Fake and Real News Dataset from Kaggle.
* **Method**: The dataset will be downloaded manually and used as a static dataset. No real-time scraping or API access will be involved in this phase.

**Data Cleaning**

* **Missing Values:** Check for null or missing entries in fields like title or text. Drop rows with insufficient text content.
* **Duplicates:** Identify and remove duplicate news articles to prevent data leakage.
* **Inconsistent Formats**: Standardize text (e.g., convert to lowercase, remove special characters, HTML tags, and extra whitespace).
* **Stopwords and Noise Removal:** Use NLP preprocessing to remove common stopwords, punctuation, and irrelevant words that do not add value to model training.

**Exploratory Data Analysis (EDA)**

* Use word clouds, frequency plots, and histograms to visualize common terms in fake vs. real news.
* Analyze text lengths, class distributions, and word usage differences between the two categories.
* Use correlation matrices or bar graphs to explore relationships between categorical features (e.g., subject/topic vs. truthfulness).

**Feature Engineering**

* Text Vectorization using methods like:
* TF-IDF (Term Frequency-Inverse Document Frequency)
* CountVectorizer (Bag of Words)
* Word Embeddings like Word2Vec or BERT embeddings
* Create derived features such as:
* Average sentence length
* Number of punctuation marks or uppercase words
* Sentiment polarity using tools like TextBlob

**Model Building**

* Logistic Regression – Simple, effective baseline for binary classification
* Naive Bayes – Performs well on text classification problems
* Random Forest – Handles complex patterns and avoids overfitting
* LSTM (Long Short-Term Memory) – Captures sequential context in news text
* BERT – A powerful transformer-based model for deep text understanding

**Model Evaluation**

* Accuracy – Measures the overall correctness of the model.
* Precision – Indicates how many predicted fake news articles were actually fake.
* Recall – Shows how many actual fake news articles were correctly identified.
* F1-Score – Balances precision and recall for better overall performance.
* Confusion Matrix – Helps visualize true positives, false positives, etc.

**Visualization & Interpretation**

* Word Clouds – To highlight frequently used words in fake vs. real news.
* Bar Charts & Pie Charts – For class distribution, model performance comparison.
* Confusion Matrix – To visualize prediction accuracy and errors.
* Line/Bar Plots – To compare metrics like precision, recall, and F1-score across models..

**Deployment**

The project will be deployed as a web application using Streamlit.This will allow users to input a news article or headline and receive an instant prediction (Real or Fake).We chose Streamlit for its simplicity, quick setup, and interactive UI for NLP applications.

**6.Tools and Technologies**

**Programming Language**:

* Python – Chosen for its strong support in NLP, machine learning, and data analysis.

**Notebook / IDE:**

* Google Colab – For cloud-based coding with GPU support.
* Jupyter Notebook – For local development and visualization.

**Libraries:**

* Data Processing: pandas, numpy, re
* NLP: nltk, spaCy, TextBlob, transformers
* Modeling: scikit-learn, TensorFlow, Keras, XGBoost
* Visualization: matplotlib, seaborn, wordcloud
* Evaluation: sklearn.metrics (for accuracy, precision, recall, etc.)

**Deployment Tools (Optional):**

* Streamlit – For creating an interactive web app.
* Flask – For backend API deployment (optional alternative).

**7.Team Members and Roles**

* **S.Harini** - Team Leader / NLP & Model Development Lead
* **S.Akila**- Data Analyst & Visualization Specialist
* **P.Abinaya** - Backend Developer
* **S.Janani**- Deployment Engineer
* **S.Atchaya keerthi -** Documentation & Presentation Coordinator