# AI- BASED SYSTEM FOR DETECTING AND COMBATTING FAKE NEWS PROPAGATION


**PRESENTED BY**

MALARVIZHI.R (711021205013)

POOJASHREE.R (711021205017)

DINESH BALAJI.R(711021205302)

**UNDER THE GUIDENCE OF**

Mr. M. NAGARASAN , AP/CSE

ASSISTANT PROFESSOR,

DEPARTMENT OF COMPUTER SCIENCE OF ENGINEERING.

# TABLE OF CONTENT

# ABSTRACT

The rapid spread of misinformation on digital platforms has become a critical concern in today's interconnected world. This project proposes an AI-based system to detect and mitigate fake news propagation using machine learning and natural language processing (NLP) techniques. The system leverages logistic regression, fact-checking algorithms, and user credibility analysis to identify and suppress false information.

# INTRODUCTION

**Communication Challenges for Individuals with Disabilities:**

- ✓ With the advent of social media and online news platforms, the dissemination of information has become instantaneous.
- ✓ However, this also presents an opportunity for the rapid spread of false or misleading content.
- ✓ Fake news can manipulate public opinion, impact elections, and incite violence. Traditional fact-checking methods are inefficient against the sheer volume of misinformation.
- ✓ AI-driven solutions offer a scalable and efficient way to detect and combat fake news by analyzing textual patterns, sources, and dissemination behaviors.

# OBJECTIVES

- ✓ To develop an AI-driven system for detecting fake news using NLP and machine learning techniques.
- ✓ To integrate fact-checking mechanisms with credible sources for verification.
- ✓ To analyze and classify fake news based on linguistic patterns and source credibility.
- ✓ To create a real-time monitoring system that alerts users and authorities about misinformation.
- ✓ To raise awareness about misinformation and provide a tool to validate news credibility.

# PROBLEM IDENTIFICATION

- ✓ Language Barrier:
  Primarily designed for English, lacking multilingual support.

- ✓ No Real-Time Detection:
  None of the papers address real-time fake news detection on social media.

- ✓ Lack of Explainability:
  Models are often black boxes, making it hard to trust their decisions.

- ✓ Influence:
  Fake news can influence public opinion, elections, and create social unrest.

# RESEARCH GAPS

- ✓ User-Friendly Alerts: Provide easy-to-understand fake news alerts for end-users.
- ✓ Social Impact Analysis: Measure how fake news influence's public opinion.
- ✓ Probability: Measures the probability between 0-1 with accuracy.
- ✓ Contextual Understanding: Improve the system’s ability to understand context and sarcasm in fake news.
- ✓ Inadequate real-time detection capabilities for fast-spreading misinformation.

# LITERATURE REVIEW

| S.NO | YEAR | PROJECT TITLE | AUTHOR NAME | MERITS | DEMERITS |
|---|---|---|---|---|---|
| 01 | 2023 | Big Data ML-Based Fake News Detection Using Distributed Learning | Alaa Altheneyan and Aseel Alhadlaq | The paper uses distributed machine learning with Apache Spark, making it highly scalable for handling large datasets, which is crucial for analyzing the massive volume of data on social media. | The model relies on batch processing, meaning it cannot perform real-time detection of fake news as it spreads on social media platforms. |

| | | | | | |
|---|---|---|---|---|---|
| 02 | 2024 | Advancing Fake News Detection: Hybrid Deep Learning With FastText and Explainable AI | Ehtesham Hashmi, Sule Yildirim Yayilgan, Muhammad Mudassar Yamin | The paper proposes a hybrid deep learning model (CNN-LSTM) combined with FastText embeddings, which effectively captures both local features and sequential dependencies in text data, achieving high accuracy in fake news detection. | the model is limited to text-based fake news and does not address multimedia content (images, videos), which are increasingly used in fake news propagation. |

| | | | | | |
|---|---|---|---|---|---|
| 03 | 2024 | Systematic Review of Fake News, Propaganda, and Disinformation: Examining Authors, Content, and Social Impact Through Machine Learning. | Darius Plikyns, Ieva Rizgeline | The paper provides a comprehensive review of fake news detection techniques, offering valuable insights into authorship, content, and social impact, which helps identify key challenges and gaps in the field. | The paper is theoretical and lacks practical implementation or experimental results, making it less useful for developing real-world solutions.<br>. |

# ALGORITHM

**NATURAL LANGUAGE PROCESSING (NLP) + LOGISTIC REGRESSION**

- ✓ **NLP (Natural Language Processing)**: Converts raw text into numerical data for machine learning.
- ✓ **Logistic Regression**: A machine learning algorithm for binary classification (Fake vs. Real News).

**How You Are Using Them:**

- ✓ **NLP Processing** → Clean text (remove stopwords, stemming).
- ✓ **Feature Extraction** → Convert text into numbers using **TF-IDF (TEAM FREQUENCY INVERSE DOCUMENT FREQUENCY)**.
- ✓ **Logistic Regression** → Classifies news as **Fake or Real** based on extracted features.

# NATURE LANGUAGE PROCESSING

## NATURE LANGUAGE PROCESSING

Natural Language Processing (NLP) enables computers to comprehend and interact using human language. Its primary purpose is to bridge the communication gap between humans and machines, facilitating seamless interaction.

- ✓ Automating repetitive tasks.
- ✓ Enhancing data analysis.
- ✓ Improving search functionalities
- ✓ Generating content.

# LOGISTIC REGRESSION

## LOGISTIC REGRESSION

- ✓ Logistic Regression is a statistical method used for analyzing datasets where the outcome variable is binary, meaning it has two possible outcomes such as "yes" or "no", "success" or "failure".
- ✓ The primary goal is to model the probability that a given input point belongs to a particular category.

## DECISION RULE FOR CLASSIFICATION

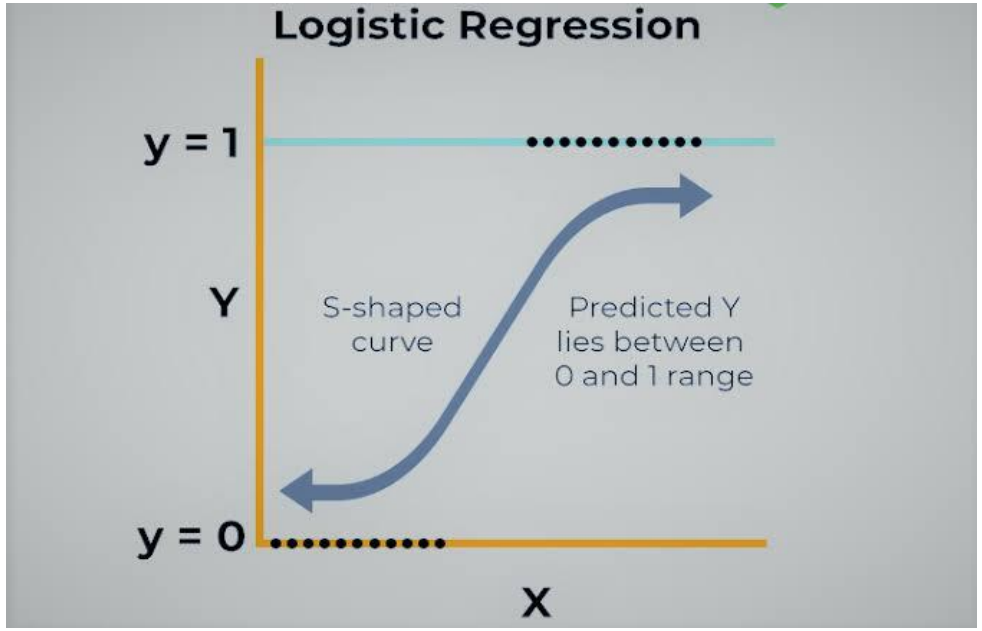
After calculating the probability $P(Y=1 \mid X)$:

If $P \geq 0.5$,

classify the news as real.

If $P < 0.5$,

classify the news as fake.



# WORK FLOW OF LOGISTIC REGRESSION

✓ **Data Collection**:

Gather a dataset with relevant features and a binary outcome variable.

✓ **Data Preprocessing**:

Clean the data by handling missing values, encoding categorical variables, and normalizing numerical features if necessary.

✓ **Model Training:**

Split the data into training and testing sets. Use the training set to fit the Logistic Regression model by estimating the coefficients that best predict the outcome variable.

✓ **Model Evaluation**:

Assess the model's performance using metrics such as accuracy, precision.

✓ **Prediction**:

Use the trained model to predict outcomes on new data instances by calculating the probability of the outcome variable and assigning class labels based on a chosen threshold.

# BENEFITS OF LOGISTIC REGRESSION

✓ **Simplicity and Efficiency**:

Logistic Regression is computationally less intensive compared to more complex models, making it suitable for large datasets.

✓ **Interpretability**:

The coefficients in Logistic Regression represent the relationship between each independent variable and the dependent variable, allowing for clear insights into how input features affect the outcome.

✓ **Probability Estimation**:

Unlike some classification methods that only provide a class label, Logistic Regression estimates probabilities, offering a measure of confidence in the predictions.

✓ **Flexibility**:

It can handle various types of relationships by applying appropriate transformations to the predictor variables.

# METHODOLOGY

- ✓ **DATA COLLECTION**

  Use labelled datasets (e.g., Kaggle’s "Fake News Detection").
- ✓ **PROCESSING**

  Clean text (lowercase, remove stop words, tokenize, lemmatize) and apply TF-IDF.
- ✓ **FEATURE EXTRACTION**

  Extract TF-IDF scores, word count, sentence length, etc.
- ✓ **MODAL TRAINING**

  Split data (80/20), train Logistic Regression, optimize hyperparameters.
- ✓ **CLASSIFICATION**

  Preprocess input text, extract features, predict using the model.
- ✓ **FEEDBACK AND DEPLOYMENT**

  Allow user feedback for improvement.

# MODULES

- ✓DATA SELECTION
- ✓DATA PROCESSING
- ✓STEMMING AND STOPWORDS REMOVAL
- ✓FEATURE EXTRACTION
- ✓CLASSIFIER
- ✓EXPERIMENTAL RESULTS
- ✓FAKE NEWS PREDICTION

# MODULES

## MODULE 1 : DATA SELECTION

This is the first step where we select only the relevant data needed for fake news detection. The dataset is chosen based on the objective of identifying misleading or false information. We ensured that the data includes both real and fake news articles for balanced analysis.

## MODULE 2 : DATA PREPROCESSING

We clean the raw data by removing unwanted characters, handling missing values, and performing text normalization techniques such as tokenization. Exploratory Data Analysis (EDA) is also carried out to understand data quality and class distribution.

## MODULE 3 : STOPWORD REMOVAL AND STEMMING

Focuses on refining the text by eliminating common, non-informative words (stopwords) and reducing words to their root form using stemming. This helps unify similar words and reduce noise in the data.

## MODULE 4 : FEATURE EXTRACTION

We convert the processed text into numerical form using techniques like Bag of Words and TF-IDF (Term Frequency-Inverse Document Frequency). These methods help highlight important terms in the data and form the basis for training machine learning models using scikit-learn.

## MODULE 4 : CLASSIFIER

**W**e implemented a Logistic Regression model using the features extracted from the previous step. This model was trained to differentiate between real and fake news articles.

## MODULE 5 : EXPERIMENTAL RESULTS

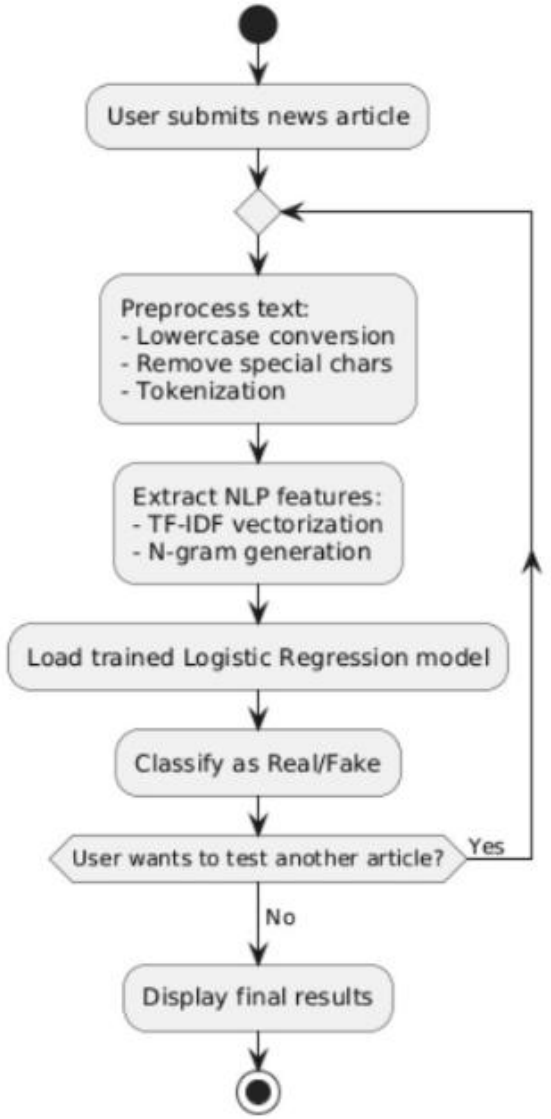
We analyzed the accuracy of the model on both the **training and test datasets**, helping us assess how well the model performs on known and unknown data.

## MODULE 6 : FAKE NEWS PREDICTION

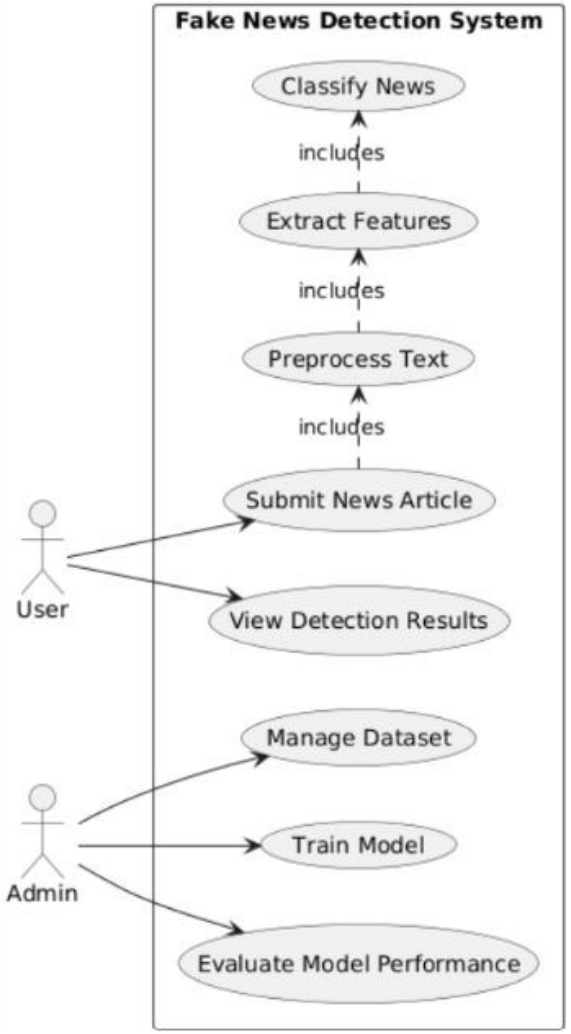
The best-performing model was used to predict whether an input news article is real or fake.

The model provides a **probability score**, indicating the likelihood of the news being true or false, allowing for more informed decisions.

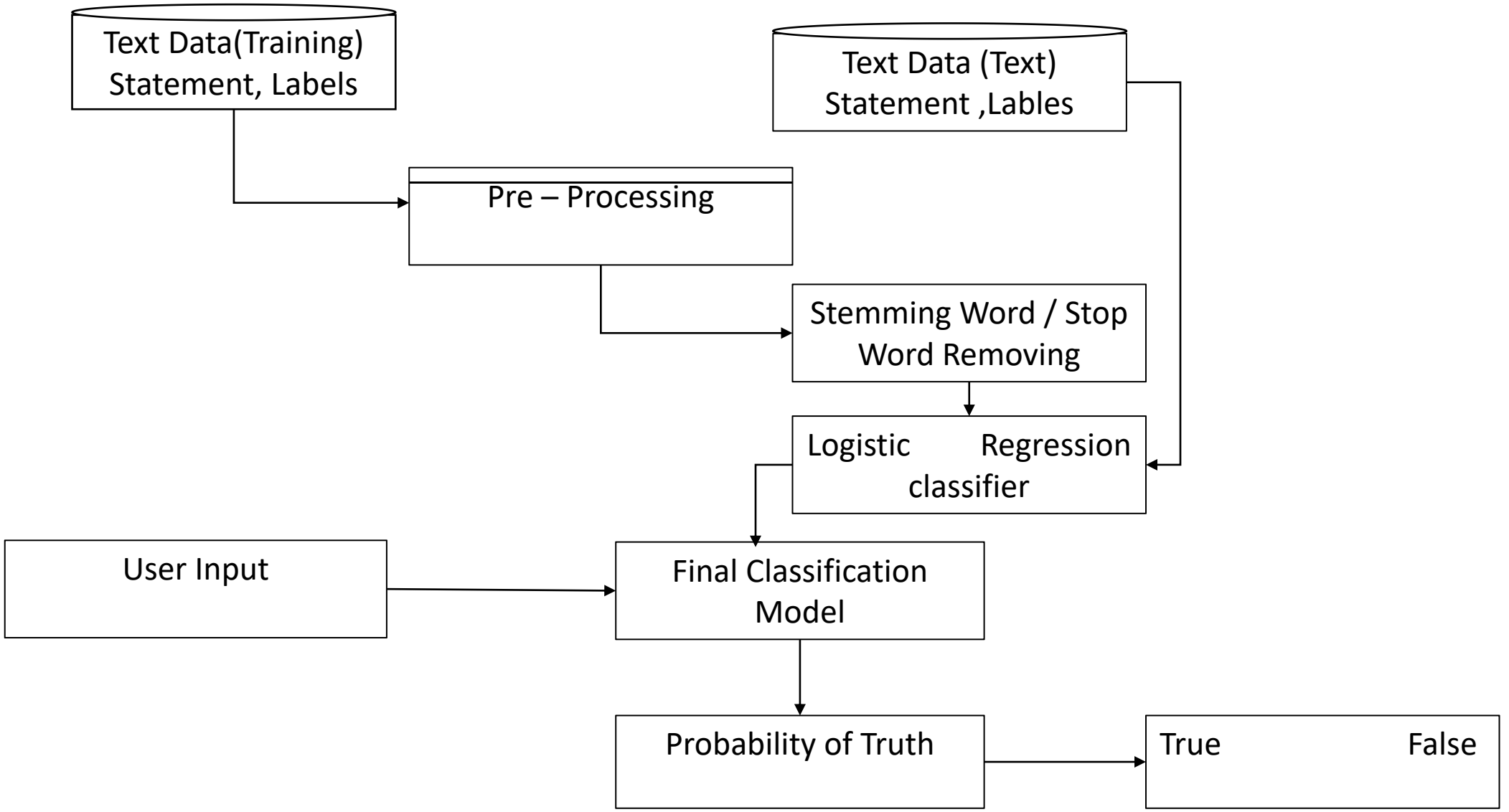
# UML DIAGRAMS:ACTIVITY DIAGRAM



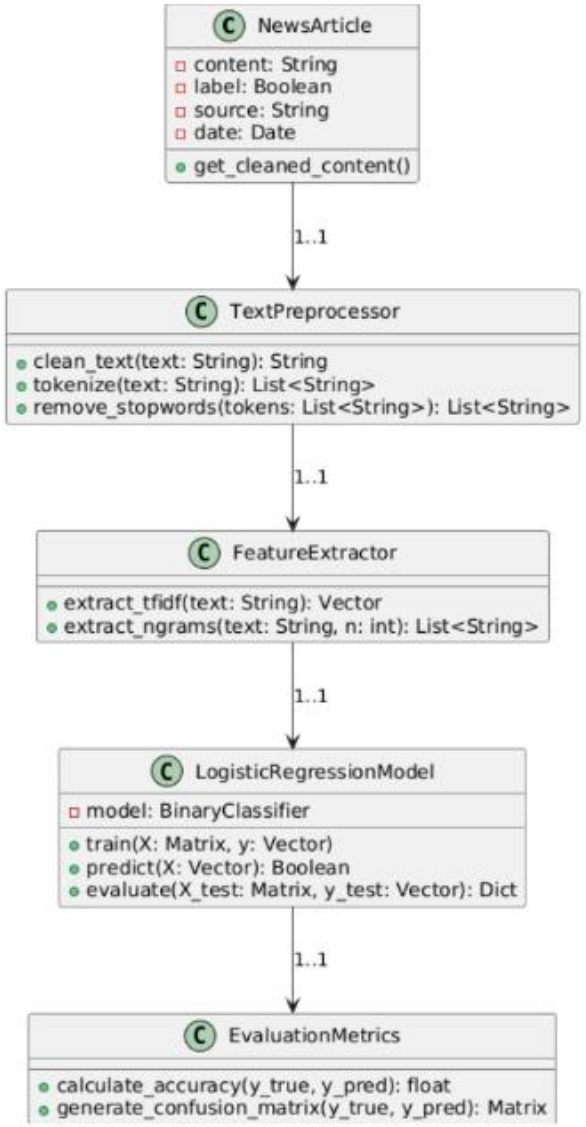
# USE CASE DIAGRAM



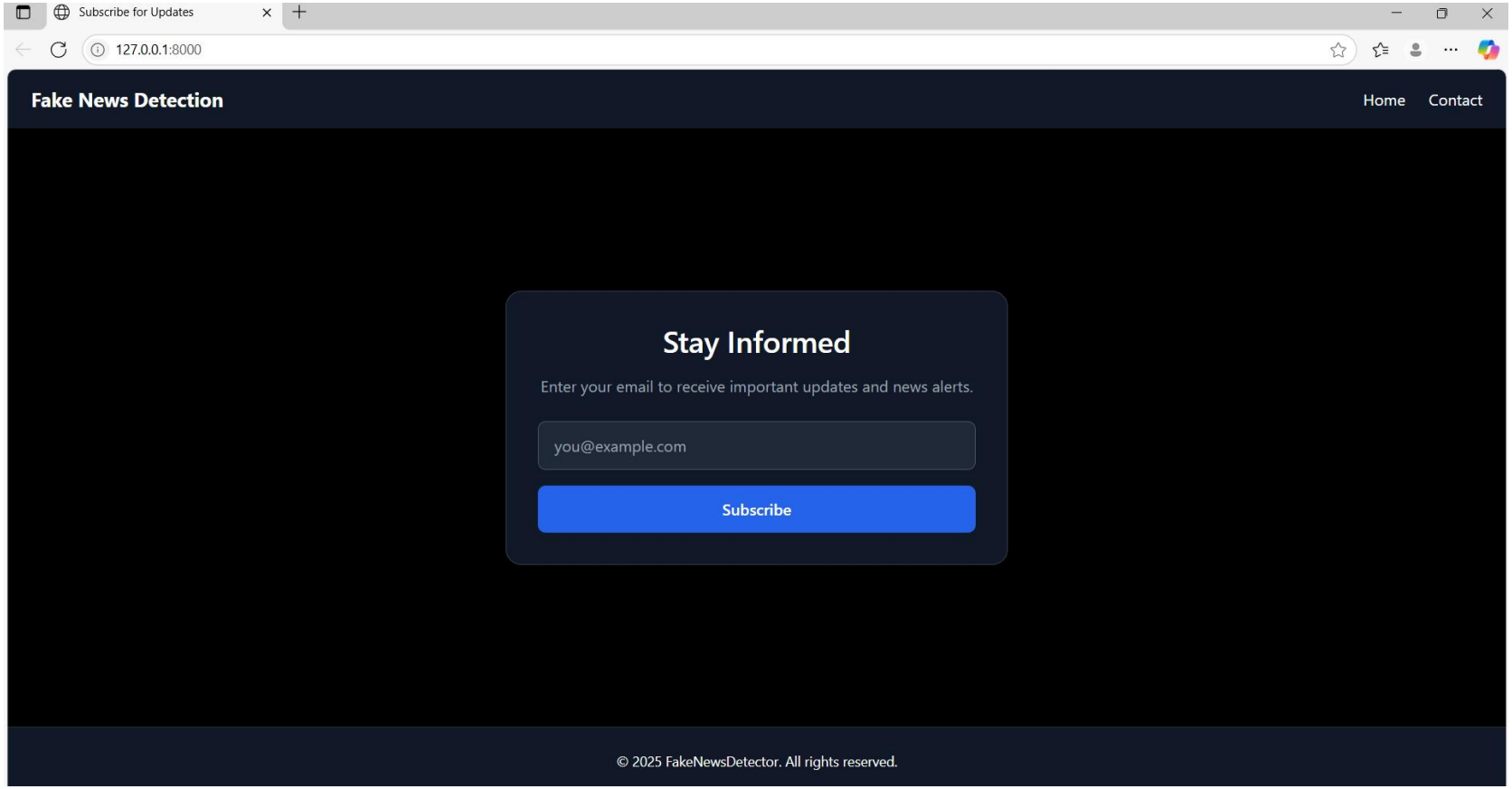
# DATA FLOW DIAGRAM



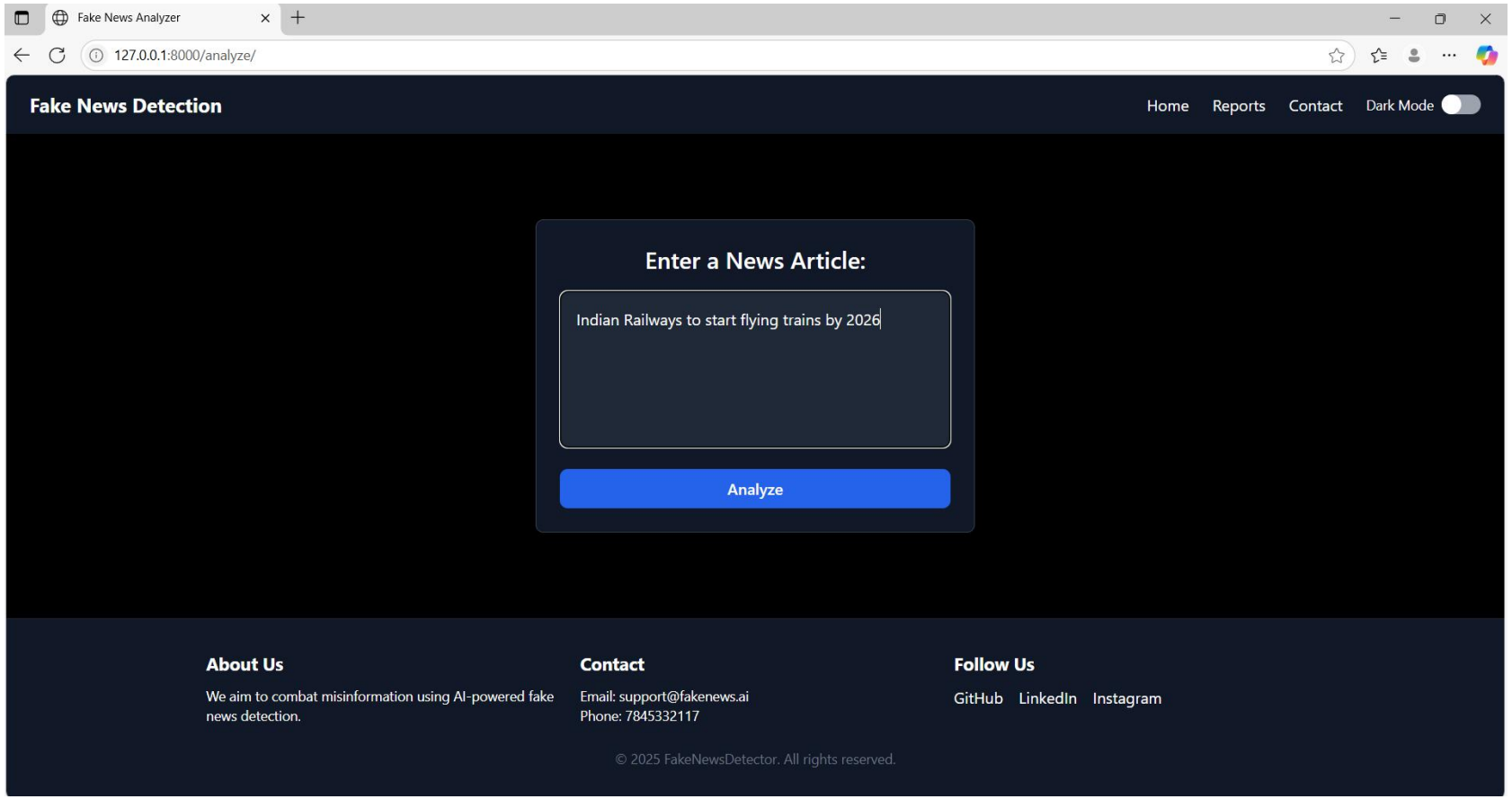
# CLASS DIAGRAM

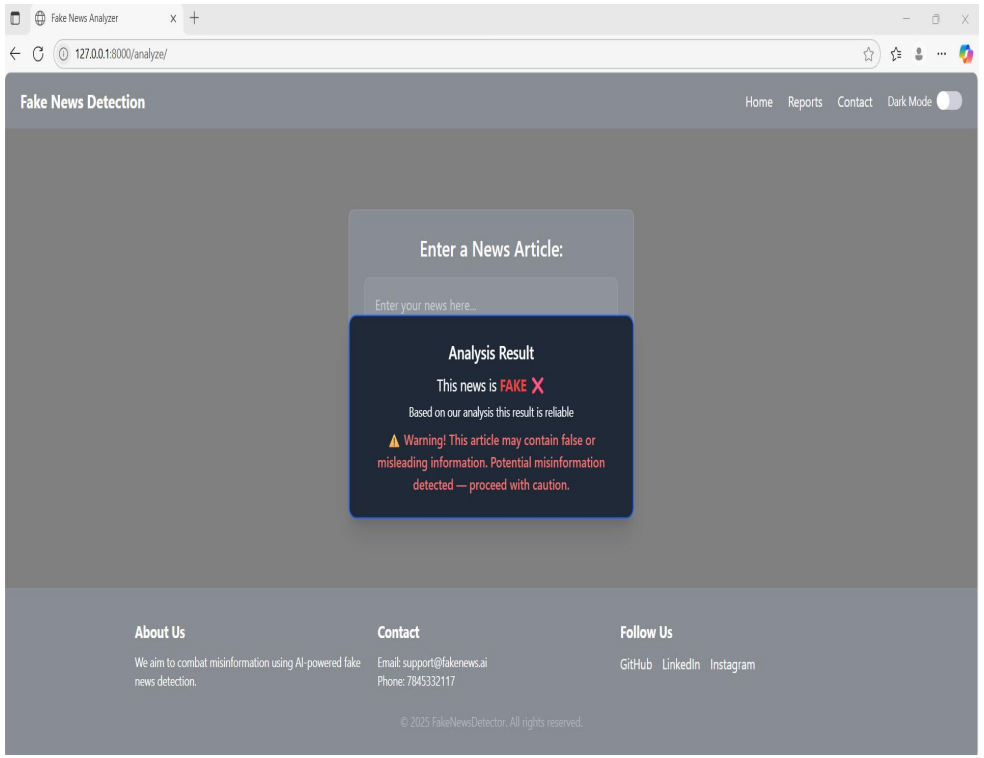


# OUTPUT

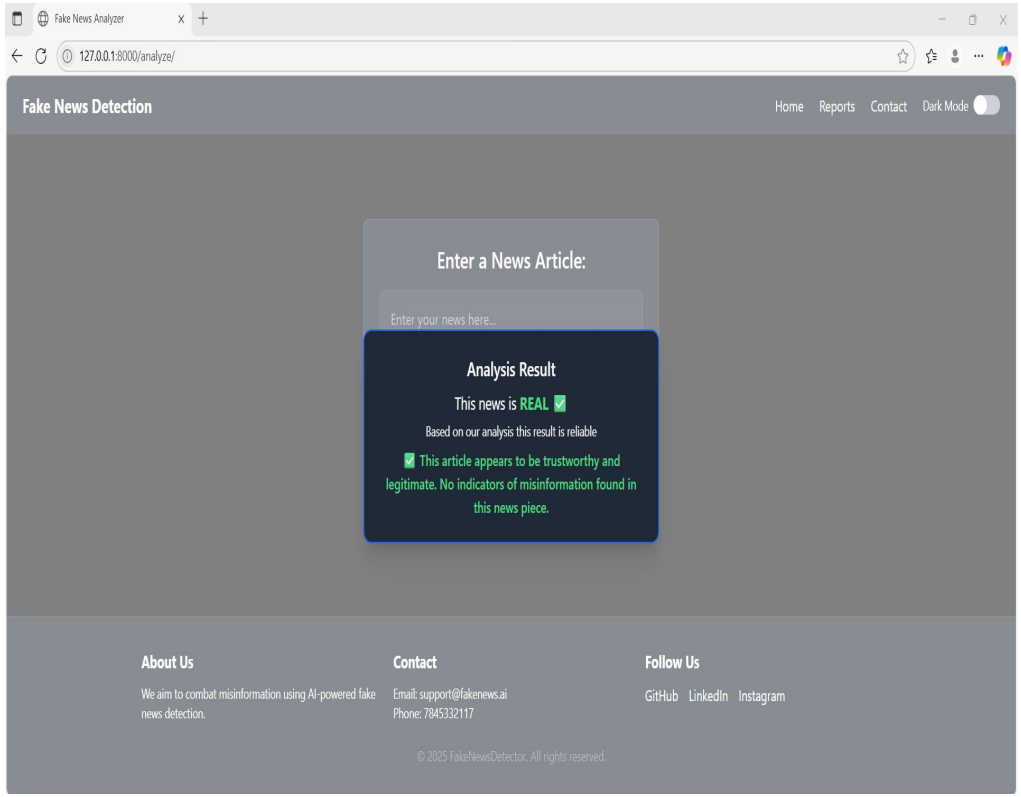

Subscribe for Updates
127.0.0.1:8000
Fake News Detection
Home
Contact
Stay Informed
Enter your email to receive important updates and news alerts.
you@example.com
Subscribe
© 2025 FakeNewsDetector. All rights reserved.

USER INPUT:

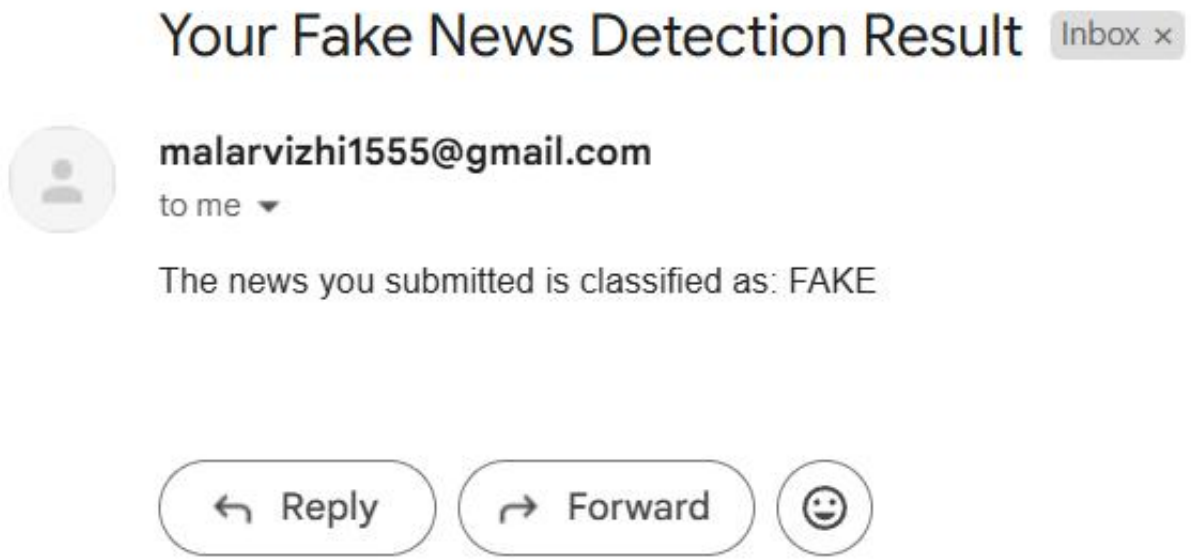

Fake News Analyzer
127.0.0.1:8000/analyze/
Fake News Detection
Home
Reports
Contact
Dark Mode
Enter a News Article:
Indian Railways to start flying trains by 2026
Analyze
About Us
We aim to combat misinformation using AI-powered fake news detection.
Contact
Email: support@fakenews.ai
Phone: 7845332117
Follow Us
GitHub
LinkedIn
Instagram
© 2025 FakeNewsDetector. All rights reserved.




RESULT: REAL OR FAKE



Integrated with the Email registered:



# CONCLUSION

- The Fake News Detection system successfully identifies fake and real news using machine learning and NLP techniques.
- Logistic Regression provides a simple yet effective approach for binary classification of news content.
- The project highlights the potential of automated tools to assist in reducing the spread of misinformation.
- The web application makes the system accessible, allowing users to test news headlines or content instantly.
- This project serves as a foundation for more advanced fake news detection solutions in the future.

# FUTURE ENHANCEMENTS

- Integrate support for multiple languages to detect fake news in regional and global contexts.
- Add image and video content analysis using deep learning (e.g., CNNs for fake image detection).
- Implement a browser extension for on-the-go verification of online news.
- Incorporate a credibility score or confidence level with detailed explanation.
- Enable real-time scraping and analysis of trending news articles from trusted and non-trusted sources.
- Add user feedback and reporting system to improve model accuracy over time

# THANK YOU