

# FAKE NEWS DETECTION

**Name:** Abdul Rafay

**Reg ID:** 233679

**Department:** BS Software Engineering (3<sup>rd</sup>)

**Subject:** Introduction to Artificial Intelligence

**Submitted to:** Sir Saad Ur Rehman

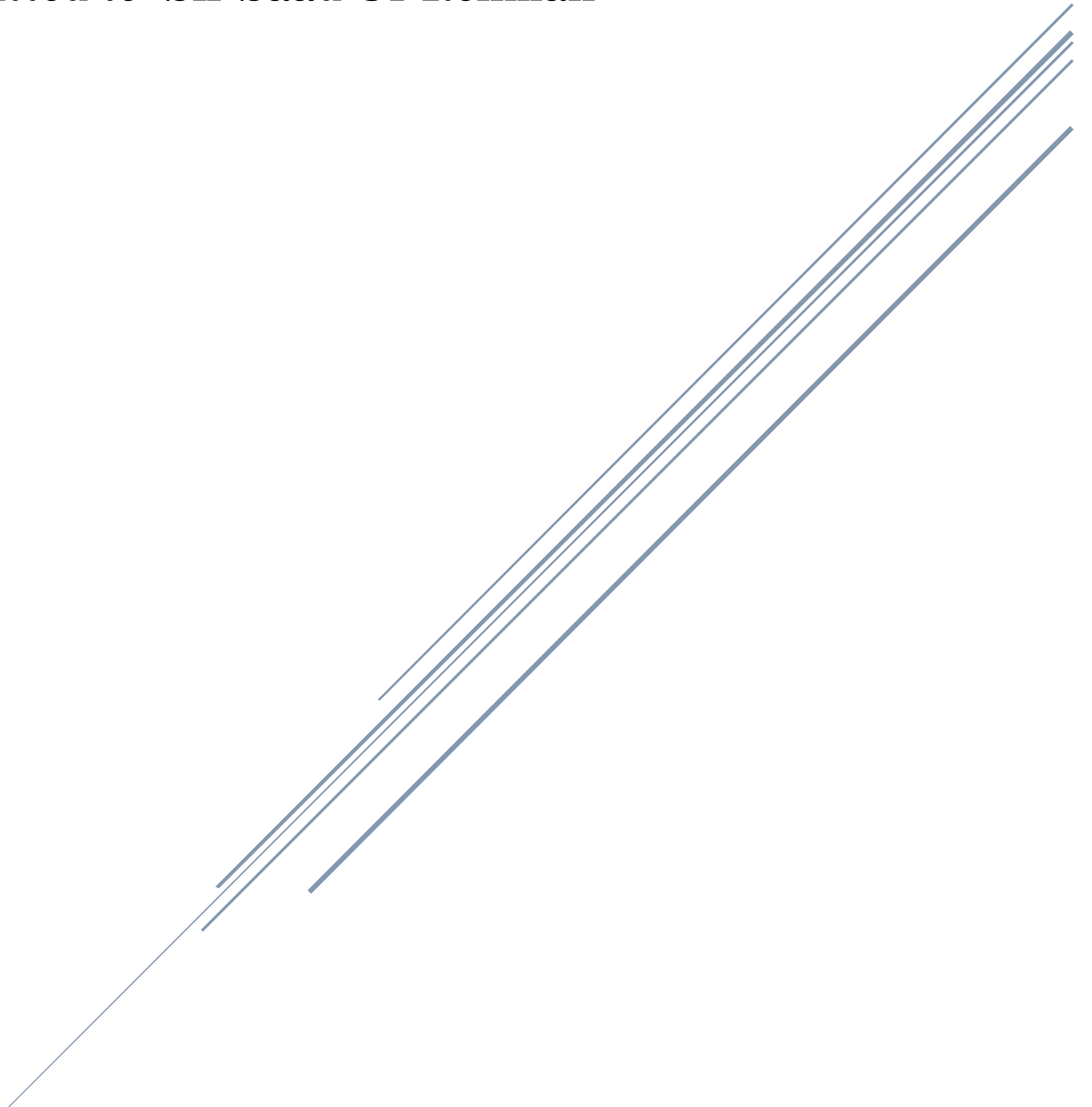


Table of Contents

- 1. Introduction ..... 2
- 2. Main Functionalities..... 2
- 3. Dataset ..... 2
- 4. AI Techniques ..... 3
  - 1. Natural Language Processing: ..... 3
  - 2. Machine Learning Models: ..... 3
  - 3. Performance Metrics:..... 3
- 5. Tools and Technologies ..... 3
  - 1. Programming Language: ..... 3
  - 2. Libraries: ..... 3
  - 3. Environment: ..... 4
  - 4. Deployment: ..... 4
- 6. Project Diagram ..... 4

## 1. Introduction

Fake news has become a significant issue in today's digital age, leading to misinformation and societal consequences. This project focuses on leveraging Machine Learning techniques to detect fake news articles efficiently and accurately. By analyzing text data from credible datasets, the project aims to differentiate between true and fake news articles using natural language processing and classification models. The final solution will be user-friendly and adaptable for real-world deployment.

## 2. Main Functionalities

- **Dataset Loading and Cleaning:**  
Ability to load Fake.csv and True.csv, remove null values, and clean the text for further processing.
- **Text Preprocessing:**  
Tokenization, removal of stop words, stemming/lemmatization, and vectorization to convert text into numerical features.
- **Model Training and Prediction:**  
Training machine learning models to classify news as either "Fake" or "True."
- **Model Evaluation:**  
Assessing model performance using metrics such as accuracy, precision, recall, and F1-score.
- **Interactive Interface:**  
A user-friendly interface, developed using Streamlit, allows users to input news articles and receive predictions on their authenticity.
- **Deployment:**  
Hosting the model and interface online using Streamlit for public access.

## 3. Dataset

- **Fake.csv:** Contains labeled fake news articles. Each entry includes the title, text, and labels for classification.

## Fake\_News\_Detection

- **True.csv:** Contains labeled true news articles. Like Fake.csv, it includes text data for model training.'
- **Key Details**
  - **Source:** Credible dataset repository (Kaggle or a public research source).
  - **Data Format:** CSV format with fields for title, text, and label.
  - **Volume:** The combined size of both datasets ensures robust training and evaluation.

## 4. AI Techniques

### 1. Natural Language Processing:

- **Text Preprocessing:** Tokenization, stemming, and stop word removal.
- **Vectorization:** Converting text into numerical features using techniques like TF-IDF or Count Vectorizer.

### 2. Machine Learning Models:

- **Logistic Regression:** For binary classification.
- **Random Forest:** For robust and interpretable classification.
- **Support Vector Machines (SVM):** For high-margin classification of fake vs. true news.

### 3. Performance Metrics:

- Accuracy
- Precision
- Recall
- F1-Score

## 5. Tools and Technologies

### 1. Programming Language:

- Python

### 2. Libraries:

- **NLP:** NLTK, spaCy
- **Modeling:** Scikit-learn, TensorFlow/Keras (if neural networks are used)
- **Data Analysis:** Pandas, NumPy

## Fake\_News\_Detection

- **Visualization:** Matplotlib, Seaborn'
- **Deployment:** Streamlit for web interface and deployment

### 3. Environment:

- Jupyter Notebook

### 4. Deployment:

- Streamlit for backend and user-friendly web interface.

## 6. Project Diagram

