A

SYNOPSIS

of

MINOR PROJECT

on

Fake News Detection using NLP

तमसो मा ज्योतिर्गमय्

GITS

Darkness to Light

Submitted by

Karnot Bhagyawardhan Singh

21EGICS053

Project Guide   **Head of Department**
Mr. Fateh Bahadur Kuwar   **Dr. Mayank Patel**

Geetanjali Institute of Technical Studies, Dabok , Udaipur (Raj.)
Department of Computer Science and Engineering
July,2024

Problem Statement:

The rapid proliferation of misinformation and fake news on digital platforms has become a significant challenge, impacting public perception and decision-making. The ability to automatically and accurately detect fake news based on textual content is crucial to mitigate the spread of false information and ensure the integrity of news dissemination.

Brief Description:

This project focuses on developing a system that utilizes Natural Language Processing (NLP) and Machine Learning (ML) to identify fake news articles. By analysing the text content of news articles, the system aims to detect patterns and features indicative of fake news. The project involves data collection, preprocessing, feature extraction, model training, and evaluation to create a reliable and efficient fake news detection tool.

Objective and Scope:

Objectives:
- **Data Collection and Preprocessing:** Gather a comprehensive dataset of news articles label as real or fake and preprocess the text data for analysis.
- **Model Development:** Develop and train various machine learning models using the extracted features to classify news articles as real or fake.
- **Evaluation and Optimization:** Evaluate the performance of the models using appropriate metrics and optimize the models for improved accuracy.
- **Implementation:** Create a user-friendly interface for the fake news detection system, allowing users to input news articles and receive a classification result

Scope:

- **Natural Language Processing:** Techniques for text preprocessing, tokenization, and feature extraction.
- **Machine Learning:** Algorithms for classification, including but not limited to logistic regression, decision trees, random forests, and neural networks.
- **Data Analysis:** Statistical and exploratory analysis of the dataset to understand patterns and correlations.
- **Model Evaluation:** Using metrics such as accuracy, precision, recall, and F1-score to assess model performance.
- **Implementation:** Developing a web-based or standalone application for end-users to utilize the fake news detection system.

Methodology:

In my framework, I am expanding on the current literature by introducing modern techniques with various linguistic feature sets to classify news articles from multiple domains as true or fake. The modern techniques along with Linguistic Inquiry and Word Count feature set used in this research are the novelty of our approach.
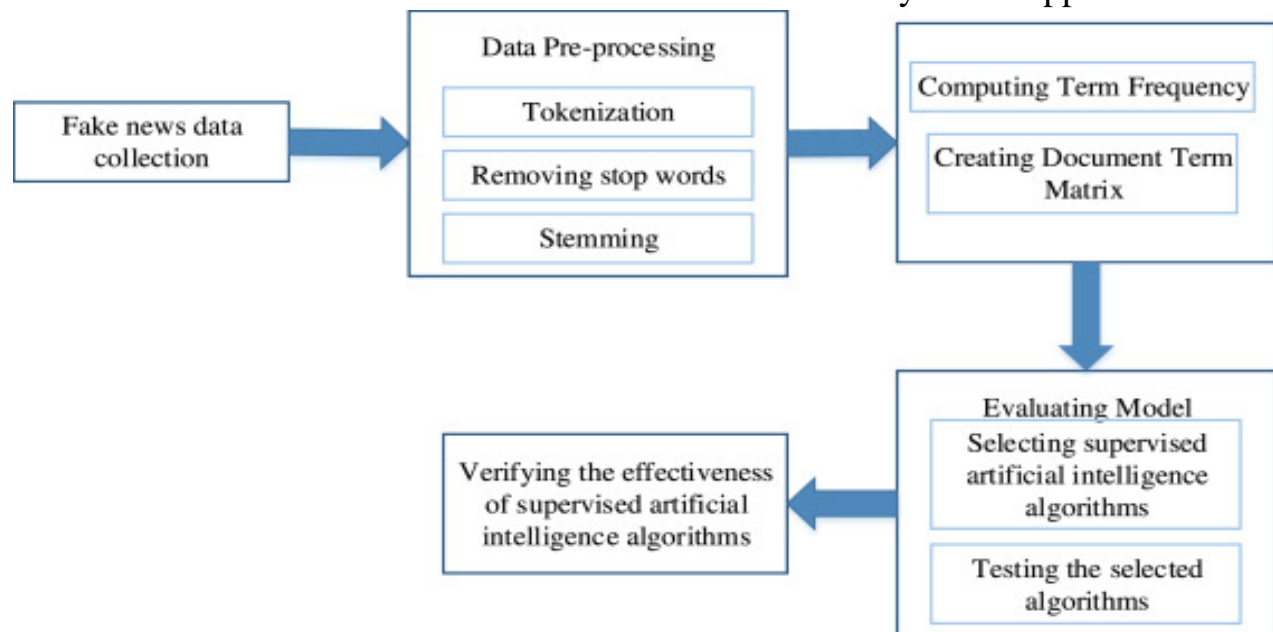
Hardware and Software Requirements:

Hardware Requirements:
- **Processor:** Intel Core i5 or higher
- **RAM:** Minimum 8 GB (16 GB recommended for faster processing)
- **Storage:** Minimum 256 GB SSD (Solid State Drive) for faster read/write speeds
- Internet Connection: Stable internet connection for downloading datasets and libraries.

Software Requirements:
- **Operating System:** Windows 10/11
- **IDE/Code Editor:** Visual Studio Code, PyCharm, or Jupyter Notebook

Technologies:

- **Programming Language:** Python(Used as the primary programming language for development)
- **Natural Language Processing (NLP):**
- **NLTK:** Handles text processing tasks like tokenization, stemming, and stop word removal.
- **spaCy:** Utilized for advanced NLP tasks such as named entity recognition and part-of-speech tagging.
- **Machine Learning:**
- **Scikit-learn:** Implements machine learning algorithms like logistic regression, decision trees, and random forests.
- **TensorFlow/Keras**: Utilized for building and training deep learning models.
- **PyTorch:** An alternative framework for developing and training complex models.

- **Data Manipulation and Analysis:**
- **Pandas:** Used for data manipulation and analysis, providing data structures like dataframes.
- **NumPy:** Employed for numerical computations and handling multi-dimensional arrays.
- **Web Development and Deployment:** Streamlit(A framework for creating and deploying interactive web applications with Python)
- **Version Control:** Git and GitHub/GitLab:
- **Virtual Environment:** Virtualenv

Testing Techniques:

1. Unit Testing:

Unit testing involves testing individual components or functions of the code to ensure that each part works correctly in isolation.
Tools:
unitest: A built-in Python module for writing and running tests.
pytest: A framework that makes it easy to write simple and scalable test cases.

2. Integration Testing:

Integration testing focuses on verifying the interactions between different modules or components of the application.
Tools:
pytest: Can be extended to perform integration tests.
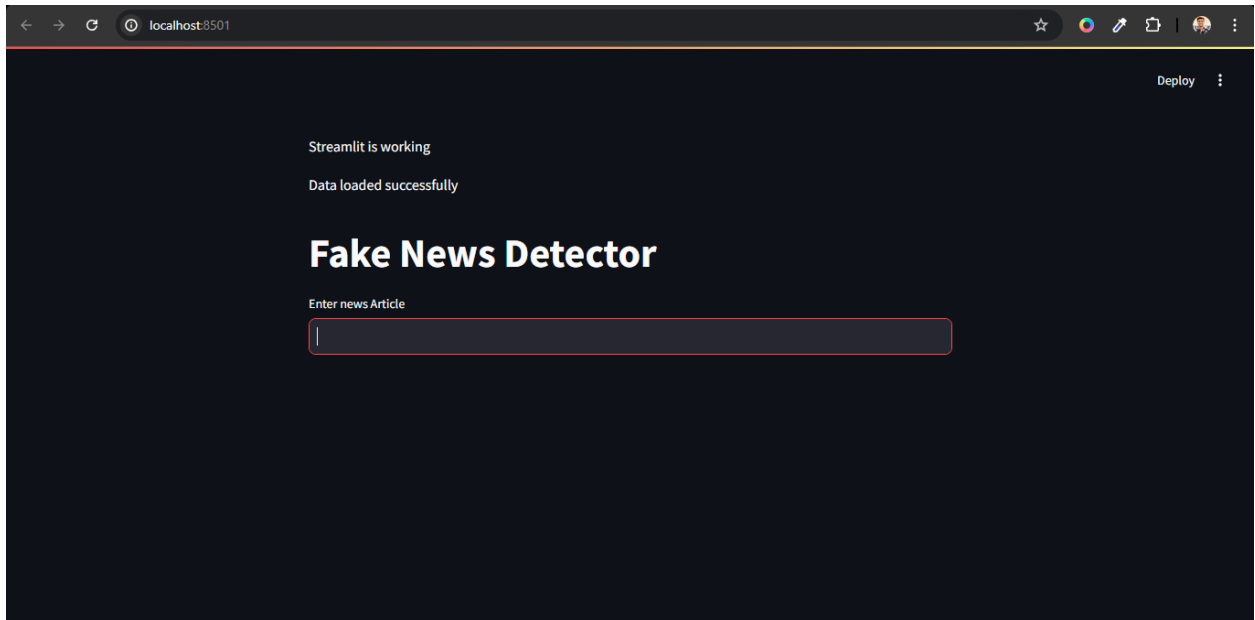
3. Functional Testing:

Functional testing ensures that the application behaves as expected when interacting with it as a user.
Tools:
Streamlit's built-in testing tools: To test the Streamlit app.

Project Screenshots:

1. **User Interface of Streamlit App:**

2. Fake News Detection by using NLP:

Streamlit is working

Data loaded successfully

Fake News Detector

Enter news Article

We Didn't Even See Comey's Letter Until Jason Chaffetz Tweeted It,Darrell Lucus,"House Dem Aide: We D

The News is Fake

3. Real News Detection of an live article on web:

Streamlit is working

Data loaded successfully

Fake News Detector

Enter news Article

America first! We have way too many problems and debt in this country to continue our foreign policies

The News is Real

Project Contribution:

This project significantly advances the field of fake news detection through the application of advanced Natural Language Processing (NLP) techniques. By leveraging state-of-the-art algorithms, it effectively analyses news articles to distinguish between authentic and misleading content. A key contribution lies in the curation and labelling of a comprehensive dataset of news articles, which not only enhances the diversity of resources available but also serves as a valuable benchmark for future research in fake news detection. Integrated with sophisticated Machine Learning models, the project achieves high accuracy in identifying patterns and linguistic features characteristic of fake news. Moreover, the development of a user-friendly web application using Streamlit facilitates easy access and interaction with the detection system, making complex NLP-based analysis accessible to a broader audience.