Microsoft Elevate

CAPSTONE PROJECT

FAKE-NEWS-DETECTION

PRESENTED BY

STUDENT NAME: SUBHAM KUMAR

COLLEGE NAME: ARCADE BUSINESS COLLEGE

DEPARTMENT: CS (BCA)

EMAIL ID: badalku8102@gmail.com

Microsoft Elevate

OUTLINE:

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

Microsoft Elevate

PROBLEM STATEMENT: Fake News Detection System

- In today's digital world, news spreads rapidly through social media, websites, and online platforms. While this enables quick information sharing, it also increases the spread of **fake news**, misinformation, and misleading content. Fake news can create panic, influence public opinion, damage reputations, and even affect elections and financial markets.
- Currently, it is difficult for users to manually verify whether a news article is real or fake due to the large volume of content available online. There is a need for an automated system that can analyze news text and classify it as **Real** or **Fake** accurately and efficiently.
- This project aims to develop a **Machine Learning-based Fake News Detection System** that:
- Takes news text as input from the user
- Preprocesses the text data
- Uses Natural Language Processing (NLP) techniques
- Applies a trained Machine Learning model
- Predicts whether the news is Real or Fake
- The system helps users quickly verify news authenticity and reduces the impact of misinformation in society.

Microsoft Elevate

PROPOSED SOLUTION:

- The system uses machine learning and NLP techniques to automatically classify news articles as Real or Fake based on their content.
- **Data Collection:**
  - The dataset consists of real and fake news articles collected from CSV files (True.csv and Fake.csv).
- **Data Preprocessing:**
  - The text data is cleaned by removing stopwords, punctuation, and converting it into a numerical format using TF-IDF.
- **Machine Learning Algorithm:**
  - Logistic Regression is used to train the model to classify news as Real or Fake.
  - **Deployment:**
  - The trained model is deployed using a Flask web application where users can enter news text and get instant predictions.
  - **Evaluation:**
  - The model performance is evaluated using accuracy and other classification metrics.

  - Result: The system successfully predicts whether the given news text is Real or Fake through a web interface.

SYSTEM APPROACH:

The system approach describes the overall methodology used to develop and implement the Fake News Detection System using Machine Learning and Natural Language Processing techniques.

- **System requirements :**
  - **Operating System:** Windows / Linux
  - **Software:** Python 3.x, VS Code
  - **Web Browser:** Google Chrome or any modern browser
  - **Hardware:** Minimum 4GB RAM, Basic Processor
  - **Dataset:** True.csv and Fake.csv (news articles dataset)

- **Library required to build the model :**
  - **Pandas** – For data handling and preprocessing
  - **NumPy** – For numerical operations
  - **NLTK** – For text cleaning and stopword removal
  - **Scikit-learn** – For TF-IDF vectorization and Logistic Regression model
  - **Pickle** – For saving and loading the trained model
  - **Flask** – For building and deploying the web application

ALGORITHM & DEPLOYMENT:

- Algorithm Selection:
  - Logistic Regression is used as the classification algorithm because it is efficient, simple, and performs well for binary classification problems like detecting Real and Fake news.
  - Data Input:
  - The input to the model is news article text collected from True.csv and Fake.csv datasets. The text is converted into numerical features using TF-IDF vectorization.
  - Training Process:
  - The dataset is cleaned and split into training and testing sets. The Logistic Regression model is trained on the training data, and performance is evaluated using accuracy and classification metrics.
  - Prediction Process:
  - When a user enters news text in the web application, the text is preprocessed and transformed using the saved TF-IDF vectorizer. The trained model then predicts whether the news is Real or Fake and displays the result in the browser.

RESULT:

The model achieved good accuracy in classifying news as Real or Fake using test data. The results show that the system can effectively detect misinformation. The prediction is displayed instantly through the web application interface.

Microsoft Elevate

Fake News Detection

The Indian government announced new measures to boost digital infrastructure across rural areas.

Analyze

Analysis Result

✓ REAL NEWS

Microsoft Elevate

CONCLUSION:

The Fake News Detection System successfully classifies news articles as Real or Fake using Machine Learning and NLP techniques. The model demonstrates good accuracy and shows that automated systems can help reduce the spread of misinformation. The web-based implementation makes the system simple, practical, and user-friendly for real-time news verification.

FUTURE SCOPE:

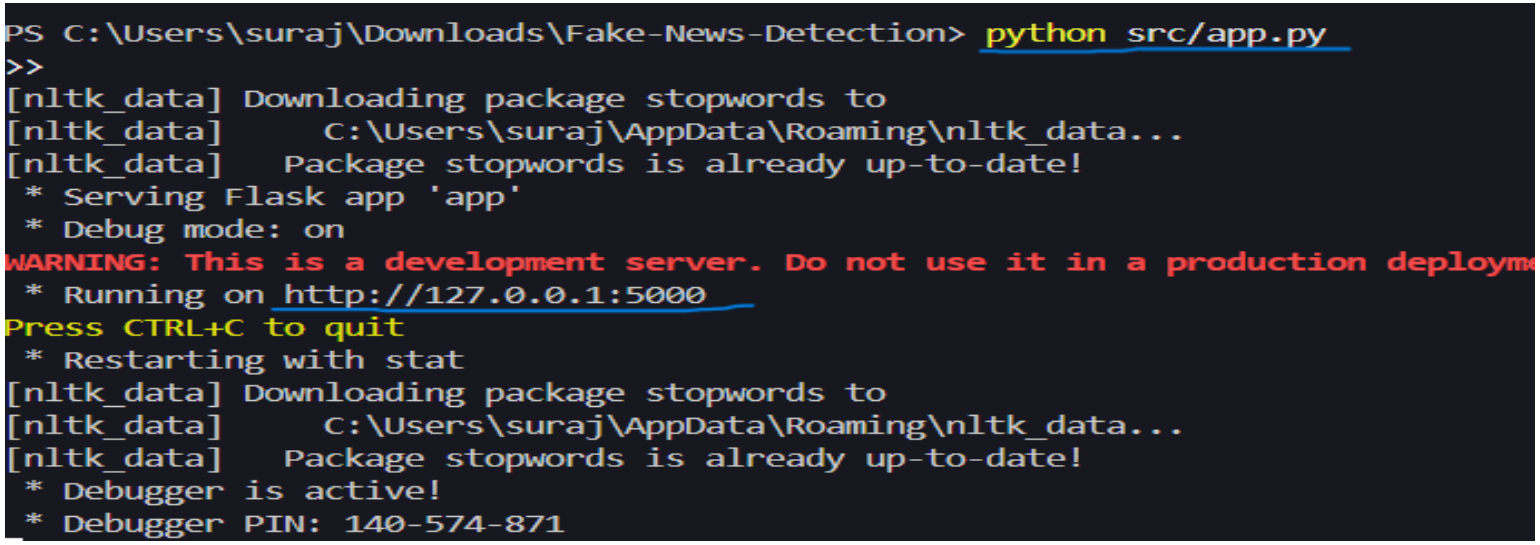
The system can be improved by training the model on a larger and more diverse dataset to increase accuracy and reliability. Advanced deep learning models such as LSTM, BERT, or Transformer-based models can be implemented for better text understanding. The project can also be deployed on cloud platforms and integrated with social media platforms for real-time fake news detection.

REFERENCES:

List and cite relevant sources, research papers, and articles that were instrumental in developing the proposed solution. This could include academic papers on bike demand prediction, machine learning algorithms, and best practices in data preprocessing and model evaluation.

GitHub Link: https://github.com/Subhamkumar2003/ss-

- **How to run the file**



Microsoft Elevate

Thank You