

PROJECT TITLE: FAKE NEWS DETECTOR

INTRODUCTION

In the contemporary digital landscape, the rapid spread of misinformation poses a major threat to societal trust, public awareness, and democratic values. With the exponential growth of online content, the need for automated systems capable of verifying the authenticity of information has become increasingly important. Fake news has been used to manipulate opinions, influence elections, and cause widespread panic, making detection technologies essential for ensuring credible communication. The Fake News Detector project is designed to classify news content as genuine or fabricated using Natural Language Processing (NLP) and Machine Learning (ML) techniques. By analyzing linguistic patterns, vocabulary, and contextual features, the model predicts whether a piece of news is authentic. This project demonstrates the use of computational intelligence to address one of the most pressing challenges in the digital information age.

ABSTRACT

The objective of this project is to build an intelligent system that detects fake news headlines using supervised machine learning. A dataset consisting of labeled real and fake news entries was utilized. The data underwent a preprocessing phase involving text cleaning, tokenization, and stop-word removal. The processed text was then converted into numerical vectors through the Term Frequency–Inverse Document Frequency (TF-IDF) technique. A Logistic Regression model was selected for classification due to its efficiency in handling sparse text data and interpretability of results. Model performance was evaluated based on accuracy, precision, recall, and F1-score, achieving an overall accuracy of 95%. Additionally, a Streamlit web application was developed to allow users to input a news headline or upload a CSV file for batch testing. This integration of ML and a user interface provides both functionality and accessibility for practical use.

TOOLS USED

1. Python
2. Pandas
3. Scikit-learn
4. Vectorizer

5. Streamlit

STEPS INVOLVED IN BUILDING THE PROJECT

1. **Data Acquisition:** The dataset containing real and fake news headlines was imported for analysis.
2. **Data Cleaning:** Removed unnecessary characters, duplicates, and stop words to enhance data quality.
3. **Feature Extraction:** Text was converted into numerical representations using the TF-IDF vectorizer.
4. **Model Development:** Logistic Regression and SVM algorithms were tested, with Logistic Regression performing best.
5. **Evaluation:** Accuracy, precision, recall, and F1-score were computed to assess model reliability.
6. **Interface Development:** Implemented an interactive Streamlit web app for real-time and batch predictions.
7. **Deployment:** Model and vectorizer were saved for future deployment and scalability.

CONCLUSION

The Fake News Detector project effectively showcases the integration of machine learning and natural language processing to solve real-world problems. By achieving a high accuracy rate of 95%, it validates the effectiveness of TF-IDF and Logistic Regression in detecting misinformation. The inclusion of a Streamlit-based interface enhances usability, allowing both technical and non-technical users to verify news authenticity. In future developments, more complex neural architectures such as BERT and RoBERTa can be implemented to improve contextual understanding. Expanding the dataset across multiple languages and news domains can further increase robustness and generalization. Overall, the project contributes toward building a more informed and trustworthy digital information ecosystem.