

News Article Classification (Fake vs Real)

1. Introduction

With the rapid growth of digital media platforms, misinformation and fake news have become significant global concerns. The ability to automatically detect fake news is critical for maintaining information credibility and preventing the spread of misleading content. This project focuses on building a Machine Learning-based system that classifies news articles as Real or Fake using Natural Language Processing (NLP) techniques.

2. Abstract

This project implements a supervised machine learning model to classify news articles into Real or Fake categories. The dataset was preprocessed using text cleaning techniques such as removing special characters and converting text to lowercase. TF-IDF (Term Frequency–Inverse Document Frequency) was used to transform textual data into numerical features. A Logistic Regression classifier was trained on the processed data and evaluated using performance metrics including Accuracy, Precision, Recall, and F1-score. The final model achieved high classification accuracy and was deployed using Streamlit to create an interactive web-based application.

3. Tools Used

- Python – Programming language used for development
- Pandas & NumPy – Data manipulation and numerical operations
- Scikit-learn – Machine Learning model training and evaluation
- TF-IDF Vectorizer – Feature extraction from text
- Logistic Regression – Binary classification algorithm
- Streamlit – Web application deployment
- Joblib – Model serialization and saving

4. Steps Involved in Building the Project

1. Data Collection: Downloaded labeled Fake and Real news datasets. 2. Data Preprocessing: Cleaned text by removing special characters, converting to lowercase, and shuffling the dataset. 3. Feature Extraction: Applied TF-IDF vectorization to convert text into numerical form. 4. Train-Test Split: Divided dataset into training (80%) and testing (20%) sets. 5. Model Training: Trained Logistic Regression classifier on TF-IDF features. 6. Model Evaluation: Evaluated using Accuracy, Precision, Recall, and F1-score. 7. Model Saving: Saved trained model and vectorizer using Joblib. 8. Deployment: Built an interactive web interface using Streamlit for real-time predictions.

5. Conclusion

The News Article Classification system successfully demonstrates the practical implementation of NLP and Machine Learning techniques to address real-world misinformation challenges. The model

achieved high accuracy and provides real-time predictions through a user-friendly web interface. This project highlights the effectiveness of TF-IDF and Logistic Regression for text classification tasks and serves as a strong foundation for future enhancements such as deep learning models, explainable AI integration, and large-scale deployment.