FAKE NEWS CLASSIFICATION

Department of Computer Science

Supervisor: Prof. Dr Usman Ghani

Overview

This project addresses the critical challenge of identifying fake news using advanced natural language processing (NLP) techniques. By analyzing textual content, it classifies news articles as either real or fake, aiming to mitigate the spread of misinformation. The solution leverages a combination of text preprocessing, feature extraction, and machine learning.

Key Components

- **Data Preprocessing**
  - Text normalization
  - Tokenization
  - Stopword removal
  - Lemmatization
  - Handling missing values
  - Removing special characters and numbers
- **Feature Extraction**
  - TF-IDF vectorization
  - N-grams for capturing context
  - Bag-of-Words representation
- **Model Selection and Training**
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Random Forest Classifier
  - Gradient Boosting (e.g., XGBoost, LightGBM)
- **Performance Evaluation**
  - Accuracy, Precision, Recall, F1-Score

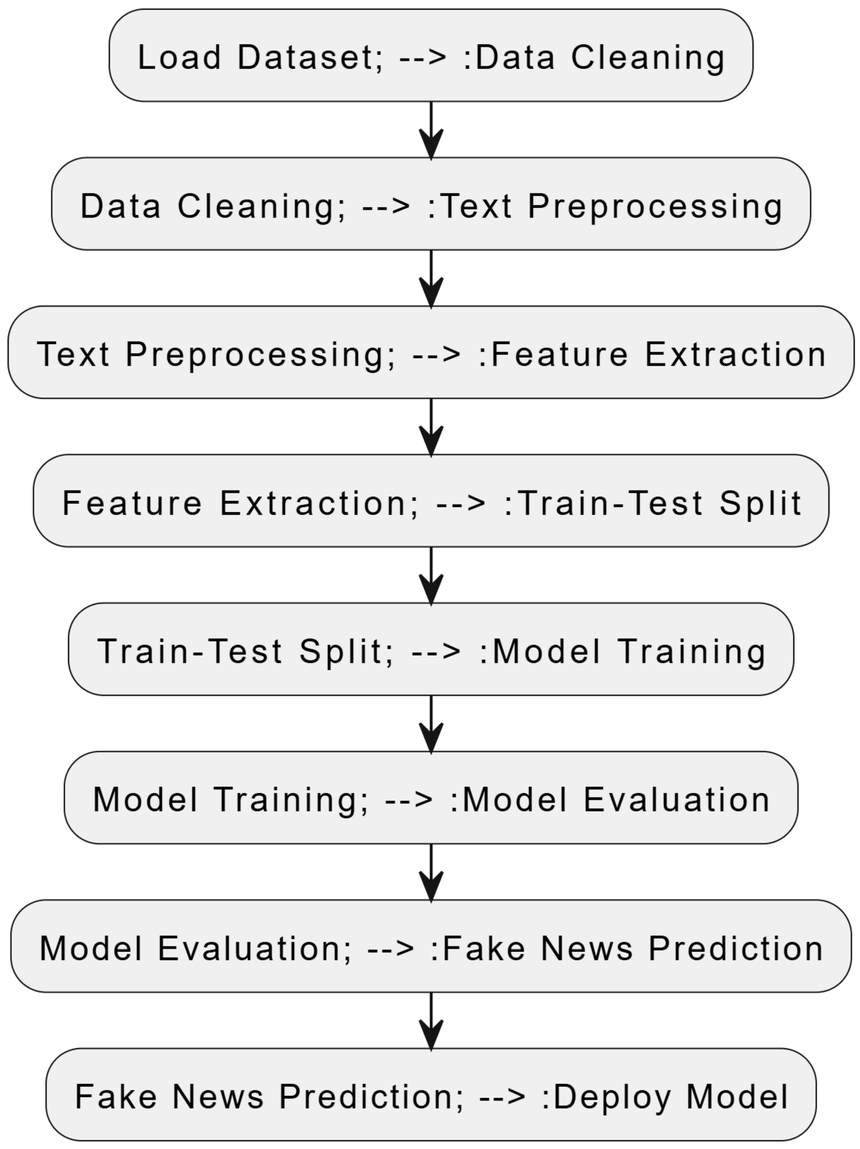
Future Directions

- **Deep Learning Approaches**
  - Experiment with LSTM and Transformer models for better context understanding.
- **Larger and Diverse Datasets**
  - Incorporate multilingual datasets for broader applicability.
- **Integration with Real-Time Systems**
  - Deploy the model for live news verification.

Technical Stack

Python, Scikit-learn, NLTK, Pandas, NumPy, Matplotlib, Seaborn

Flowchart



Performance Metrices

- **Accuracy:** Measures overall prediction correctness.
- **Precision:** Focuses on minimizing false positives.
- **Recall:** Focuses on minimizing false negatives.
- **F1-Score:** Balances precision and recall.

Applications

- **Social Media Platforms**: Flagging fake news posts to reduce misinformation.
- **News Agencies:** Assisting editors in verifying content credibility.
- **Government and NGOs:** Monitoring and controlling misinformation campaigns.
- **Search Engines:** Improving search quality by prioritizing credible news sources.
- **Educational Tools:** Teaching users about misinformation patterns and reliable sources.

Developed By:

Yasir Hassan 2021-CS-28