Fake News Detection: Project Documentation

Created by: Dulhara Lakshan (2025/03/15)

1. Introduction

Project Overview

This project focuses on building a **fake news detection system** using **machine learning** techniques. The goal is to classify news articles as either **FAKE** or **REAL** based on their content.

Objective

- Develop a **text classification model** to identify fake news.
- Implement **text preprocessing** and **feature extraction** techniques.
- Evaluate **model performance** using appropriate metrics.
- Explore **potential improvements** for more advanced fake news detection.

2. Dataset

Dataset Description

- The dataset consists of news articles with columns:
    - **Title**: The headline of the news article.
    - **Text**: The body content of the article.
    - **Label**: The classification (FAKE or REAL).

Preprocessing Steps

- **Removed Index Column**: The dataset contained an index column with unnecessary values.
- **Converted Labels**: Mapped FAKE to **0** and REAL to **1**.
- **Merged Title & Text**: Created a new column called **"content"** by combining the title and text.
- **Text Cleaning**:
    - Lowercased text.
    - Removed **brackets**, **URLs**, **punctuation**.
    - Optionally removed numbers (not used in this case).
- **Reordered Columns**: Organized dataset as title, text, content, label.

3. Feature Extraction

- **TF-IDF Vectorization**:
    - Converted text into numerical format using **TfidfVectorizer**.
    - Removed English **stop words**.
    - Used **5000 most important features** for classification.
    - Transformed textual data into a sparse matrix representation to improve computational efficiency.

4. Model Training

- **Split Data**:
    - **80%** training set
    - **20%** testing set
- **Model Used**: Logistic Regression
- **Training Process**:
    - Trained the model on the transformed **TF-IDF vectors**.
    - Predicted labels for the test set.
    - Optimized hyperparameters using **default settings**, but potential for fine-tuning.

5. Model Performance

Evaluation Metrics

Metric	**Score**
Accuracy	91.95%
Precision (Fake News - 0)	91%
Recall (Fake News - 0)	93%
Precision (Real News - 1)	93%
Recall (Real News - 1)	91%
F1-score (Overall)	92%

Confusion Matrix

A **heatmap** visualization of the confusion matrix was generated to analyze **false positives** and **false negatives**.

- The confusion matrix indicates that the model correctly classifies most fake and real news articles, with minor misclassifications.
- **False positives** (real news misclassified as fake) and **false negatives** (fake news misclassified as real) remain a challenge.
- Further **data augmentation** or **ensemble learning** techniques could help reduce these misclassifications.

6. Findings & Insights

- The model performed **well (91.95% accuracy)** but can be **improved** with:
    - **More advanced NLP techniques** (Word Embeddings, Transformers like BERT).
    - **Deep Learning models** (LSTM, BiLSTM for sequential text learning).
    - **Ensemble Methods** (combining multiple models for better results).
    - **Fact-checking databases** (incorporating external fact-based sources).
    - **Additional linguistic analysis** to detect sentiment and misleading phrases.

7. Future Improvements

- Implement **BERT/RoBERTa** for deeper context understanding.
- Use **Graph Neural Networks (GNNs)** to analyze **news propagation patterns**.
- Integrate **explainability techniques** to understand why the model flags news as fake.
- Apply **semi-supervised learning** to improve classification accuracy on unseen data.
- Investigate **adversarial attacks** to test model robustness against deceptive fake news techniques.

8. Conclusion

This project successfully built a **fake news classifier** using **Logistic Regression** and **TF-IDF vectorization**, achieving **high accuracy (91.95%)**. Future work can enhance detection using **deep learning and network-based approaches**. The findings suggest that integrating **more sophisticated NLP models and external verification sources** could lead to **greater detection accuracy** and **real-world applicability**.