Fake News Detection

1. Introduction

Fake news has become a significant issue in today's digital world, influencing public opinion and causing misinformation. This project aims to develop a machine learning-based system to detect and classify fake news articles accurately.

2. Objectives

• To analyze textual data and identify fake news patterns.

• To apply machine learning techniques for fake news classification.

• To evaluate model performance for accurate prediction.

3. Dataset Description

The dataset used in this project contains:

• News headlines and articles.

• Labels indicating whether the news is "Fake" or "Real".

• Additional metadata such as source and publication date.

4. Methodology

The following steps were followed to build the fake news detection system:

4.1 Data Preprocessing

• Removal of stopwords and punctuation.

• Tokenization and stemming.

• Text vectorization using TF-IDF.

4.2 Exploratory Data Analysis (EDA)

• Visualization of word frequency.

• Analysis of fake vs. real news word distributions.

4.3 Machine Learning Models Applied

• Logistic Regression: A baseline classification model.

• Naïve Bayes: Suitable for text classification problems.

• Random Forest: An ensemble learning approach.

• Deep Learning (LSTM/Transformers): Advanced techniques for improved accuracy.

5. Results and Analysis

• Performance evaluation using accuracy, precision, recall, and F1-score.
• The best-performing model was determined based on validation results.
• Visualization of classification results and misclassified examples.

6. Conclusion

The project successfully classified fake and real news using machine learning techniques. The best-performing model can help mitigate the spread of misinformation by identifying unreliable news sources.

7. Future Work

• Implementing real-time news classification.

• Enhancing the model using larger datasets.

• Using advanced NLP models like BERT or GPT for improved detection.

Prepared by: Abeer Fatima
Domain: AI/ML

DHC-3603