**Fake News Detection Project**

**Project Overview**

This project focuses on building a **Fake News Detection** model using a dataset of news articles. The model classifies news articles as either *real* or *fake* using **Natural Language Processing (NLP)** techniques and a **Logistic Regression** classifier. The dataset contains the title, author, and label of each news article.

**Table of Contents**



**Dependencies**

To run this project, you need the following libraries:

- numpy
- pandas
- re (regular expressions)
- nltk (Natural Language Toolkit)
- sklearn (scikit-learn)

You can install the necessary packages using:

```
pip install numpy pandas nltk scikit-learn
```

**Dataset**

The dataset (train copy.csv) consists of the following columns:

- author: Name of the author of the news article
- title: Title of the news article
- label: 0 indicates real news, and 1 indicates fake news

**Data Preprocessing**

1. **Handling Missing Values**: Missing values in the dataset are filled with empty strings.
2. **Text Merging**: The author and title columns are merged to form a content column.
3. **Text Cleaning & Stemming**:
    - All non-alphabetical characters are removed.
    - The text is converted to lowercase.

    - Stopwords are removed, and words are stemmed using the PorterStemmer from NLTK.
4. **Feature Extraction**:
    - The cleaned text data is converted into numerical features using **TF-IDF Vectorization**.

**Model Building**

- The project uses **Logistic Regression** from scikit-learn for classification.
- The dataset is split into training (80%) and testing (20%) sets.

**Evaluation**

The model is evaluated using **accuracy score** on both the training and testing data.

**Usage**

1. Ensure that the required libraries are installed.
2. Run the fake_news_detection.py script:

   python fake_news_detection.py
3. The script outputs the accuracy on the training and test datasets.

**Results**

- The model outputs the accuracy score for both the training and test sets.
- A simple predictive system is implemented to classify a single news article from the test set:
    - If the output is 0, the news is classified as **real**.
    - If the output is 1, the news is classified as **fake**.

**Notes**

- The script contains a line to download NLTK stopwords. Uncomment and run it if necessary:

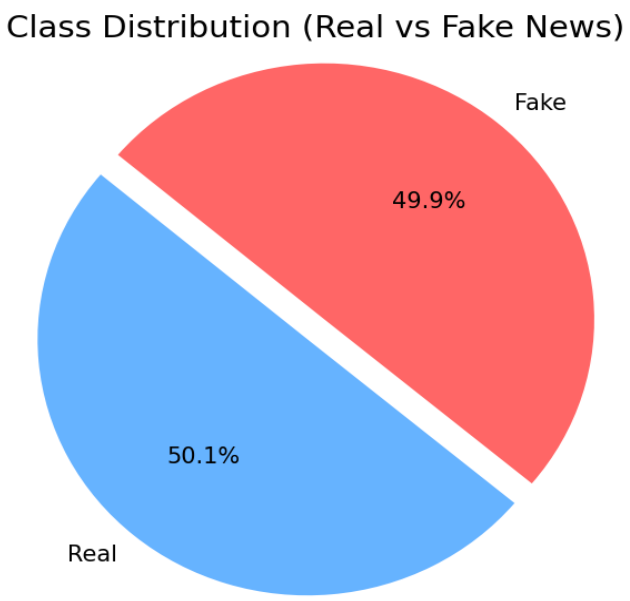
  nltk.download('stopwords')
- Further improvements can be made by experimenting with different models and feature extraction techniques.

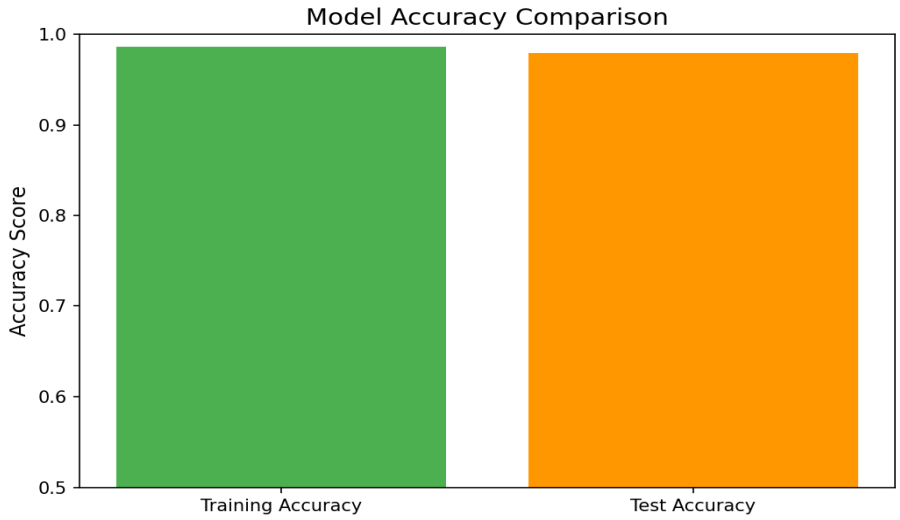
**References**

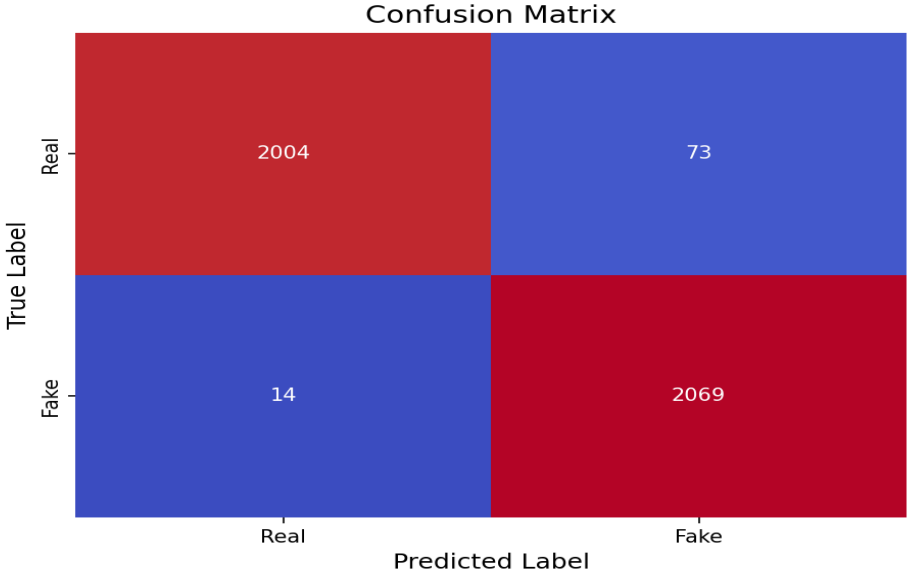
- Dataset: The dataset used in this project is assumed to be a custom or publicly available CSV file.
- NLTK: Natural Language Toolkit Documentation
- scikit-learn: scikit-learn Documentation

This project serves as an introductory example of NLP for text classification. Feel free to modify and extend the project for better performance and additional features.

Charts Prepared by the the program :



Most Common Words in the News Dataset