Fake News Detection using Natural Language Processing (NLP)

Overview

This script is designed to detect fake news articles using Natural Language Processing (NLP) techniques. It leverages a dataset containing both real and fake news articles, preprocesses the text data, and trains a machine learning model to make predictions. The goal is to classify news articles as either real (0) or fake (1).

Prerequisites

Before running this script, ensure you have the following prerequisites installed:

- Python (3.x recommended)
- Required Python libraries (pandas, nltk, scikit-learn)

You should also have the following CSV datasets in the same directory as this script:

- 'True.csv': A dataset of real news articles.
- 'False.csv': A dataset of fake news articles.

Script Breakdown

1. **Data Loading and Labeling:**
   - Load the 'True.csv' and 'False.csv' datasets using pandas.
   - Add a 'label' column to indicate real news (0) and fake news (1).
   - Concatenate the two datasets and shuffle the data for randomness.

2. **Data Preprocessing:**
   - Download and use stopwords from NLTK to remove common words.
   - Define a function for text preprocessing, which can include lowercasing and removing stopwords.
   - Apply text preprocessing to the 'text' column of the dataset.

3. **Data Splitting:**
   - Split the preprocessed dataset into training and testing sets.
   - Define features (X) and labels (y) for training and testing.

4. **Text Vectorization:**
   - Create a TF-IDF vectorizer with a limit of 5000 features.
   - Transform the text data into TF-IDF vectors for training and testing sets.

5. **Model Training:**
   - Train a classifier using Multinomial Naive Bayes. You can experiment with other classifiers.

   - Fit the model using the training data.

6. **Model Evaluation:**
   - Make predictions using the test data.
   - Print a classification report to evaluate the model's performance.

Usage

1. Make sure you have the necessary prerequisites installed.

2. Place 'True.csv' and 'False.csv' in the same directory as this script.

3. Run the script using Python:
   ```bash
   python NLP.py
   ```