This Python script demonstrates a basic fake news detection system using machine learning. It classifies news text as either **real** (label 0) or **fake** (label 1) based on a small example dataset.

**Libraries Used**

* **pandas:** For handling structured data as DataFrames.
* **scikit-learn:** Provides tools for splitting data, feature extraction (TF-IDF), machine learning models (Logistic Regression), and evaluation metrics.

### Code Breakdown

#### 1. Import Libraries

* pandas: Load and manipulate data.
* train\_test\_split: Split dataset into training and test sets.
* TfidfVectorizer: Converts text into TF-IDF numerical features.
* LogisticRegression: Classification algorithm.
* accuracy\_score and classification\_report: Evaluate model performance.

#### 2. Prepare Dataset

* Creates a small example dataset of news headlines/articles.
* Each entry has a text field (news content) and a label (0 for real, 1 for fake).

#### 3. Split Data into Training and Test Sets

* Splits data randomly: 70% for training the model, 30% for testing.
* random\_state=42 ensures reproducibility.

#### 4. Text Vectorization with TF-IDF

* Transforms text into numerical features based on Term Frequency-Inverse Document Frequency (TF-IDF).
* stop\_words='english' removes common English words that don't add much meaning.

#### 5. Train the Logistic Regression Model

* Initializes the logistic regression classifier.
* Trains (fit) on the vectorized training data and labels.

#### 6. Make Predictions on Test Data

Predicts labels (real/fake) for the test dataset.

7. Evaluate Model Performance

* Prints overall accuracy score.
* Prints precision, recall, F1-score for each class (real and fake news).

### Notes & Recommendations

* This is a **minimal demo** and **not suitable for real-world use** without a larger dataset.
* Real fake news detection requires extensive datasets, preprocessing, and more advanced models.
* Consider using datasets like Kaggle's Fake News Dataset for practice.
* Experiment with more sophisticated models (SVM, Random Forest, Transformers like BERT).
* Text cleaning steps (removing punctuation, lowercasing, stemming) improve results.
* TF-IDF can be replaced with word embeddings for better context awareness.

### Notes & Recommendations

* This is a **minimal demo** and **not suitable for real-world use** without a larger dataset.
* Real fake news detection requires extensive datasets, preprocessing, and more advanced models.
* Consider using datasets like Kaggle's Fake News Dataset for practice.
* Experiment with more sophisticated models (SVM, Random Forest, Transformers like BERT).
* Text cleaning steps (removing punctuation, lowercasing, stemming) improve results.
* TF-IDF can be replaced with word embeddings for better context awareness.