Fake News Detection ML-Model

Submitted By - Chahit Kumar Gaware

1. **Overview :-**
   This project implements a machine learning pipeline to detect fake news based on the FakeNewsNet dataset. The dataset contains fake and real news articles, and the model is trained using the `Logistic Regression` algorithm. The goal of this project is to classify a given news article as either Fake or Real using its title.

2. **Dataset :-**
   The dataset consists of news from two sources:
   A. Politifact: Contains real and fake news labels.
   B. GossipCop: Another source of real and fake news labels.

   The dataset is split into two categories:
   a. Fake news: Labeled as **`0`**
   b. Real news: Labeled as **`1`**

3. **Prerequisites :-**
   The following Python libraries are required to run the project:
   - Pandas
   - Scikit-learn
   - joblib

To install these libraries, you can use the following command:
- pip install pandas scikit-learn joblib

4. File Descriptions :-

a. **load_fakenewsnet_data()**
   This function loads and concatenates the CSV files from the Politifact and GossipCop datasets, returning a combined DataFrame with titles and labels (0 for fake, 1 for real).

b. **Model Training**
   The pipeline processes the dataset, vectorizes the text using TF-IDF, and trains a Logistic Regression** classifier. The steps are:
   - Train/Test Split: The dataset is split into training (90%) and test (10%) sets.
   - TF-IDF Vectorization: Text data is converted to numerical features using TF-IDF.
   - Logistic Regression: A logistic regression model is trained on the TF-IDF features.

c. Model Evaluation

The performance of the model is evaluated using:

- Accuracy Score
- Confusion Matrix
- Classification Report (precision, recall, F1-score)

d. Saving the Model

The trained Logistic Regression model and the TF-IDF vectorizer are saved as `.pkl` files using `joblib` for future use.

e. Loading the Model

To predict new data, the saved model and vectorizer are loaded using `joblib`.

5. Usage :-

a. Model Training

To train the model, ensure that the FakeNewsNet dataset is stored in the appropriate directory and run the script:

base_path = '/path/to/FakeNewsNet/dataset'
df = load_fakenewsnet_data(base_path)
Continue with train-test split, training, and evaluation...

b. Model Prediction

Once the model is trained and saved, you can predict whether a piece of news is fake or real by loading the saved model and vectorizer and passing the news title to the `predict_news()` function:

sample_news = "Vishnu Deo Sai is chhattisgarh's new cm."
prediction = predict_news(sample_news)
print(f"The news is predicted to be: {prediction}")

6. Files

- **fake_news_detection_model.pkl**: The saved logistic regression model.
- **tfidf_vectorizer.pkl**: The saved TF-IDF vectorizer.
  (Note - pkl is the extension to save the trained model)

Trained model we have used for the discrimination that the news is fake or real.
In the Discrimination we have to give the headline of the news then it will Discriminate by their observation.

7. Conclusion

This project demonstrates a simple and effective way to detect fake news using machine learning. The trained Logistic Regression model, combined with TF-IDF vectorization, achieves reasonable performance in classifying news titles as fake or real.