Twitter Sentiment Prediction Python Package

This project focuses on predicting the sentiment of Twitter posts using Natural Language Processing (NLP) techniques. It is an end-to-end machine learning project that includes data preprocessing, model building, hyperparameter tuning, and deployment.

How I built this package

Download code packaging process.pdf

Features

- **ETL Pipeline**: Extract, transform, and load (ETL) processes implemented using PySpark and SQL.
- **Text Processing**: Cleaned text data using regex, removed special characters, and vectorized text using TF-IDF.
- **Model Building**: Implemented Logistic Regression and Multinomial Naive Bayes for sentiment prediction.
- **Hyperparameter Tuning**: Optimized models using GridSearchCV.
- **Experiment Tracking**: Logged experiments with MLflow for efficient performance comparison.
- **Model Packaging**: Prepared reusable model packages using `sdist` and `wheel`.
- **Deployment Ready**: Plan to host the project on Render for public access.

Installation

1. Clone the repository:
   `bash git clone https://github.com/vijaytakbhate2002/sentiment_prediction_python_package.git`
2. Install the package using pip:
   `bash pip install git+https://github.com/vijaytakbhate2002/sentiment_prediction_python_package.git`

Usage

Prediction Example

To predict the sentiment of a Twitter post, you can use the following Python code:

```
from sentiment_prediction import predict
predict.predictor("this is a negative tweet")
```

This will return:

```
['Negative']
```

Sentiment Classes

The prediction function supports four sentiment classes:
- **Positive**

- **Negative**
- **Neutral**
- **Irrelevant**

Project Structure

```
sentiment_prediction_python_package/
■
■■■ sentiment_prediction/                # Main package directory
■   ■■■ config/
■   ■   ■■■ config.py                     # Configuration file
■   ■
■   ■■■ data_manipulation/               # Model training and prediction scripts
■   ■   ■■■ data_handling.py             # Training pipeline
■   ■   ■■■ data_processing.py           # Prediction pipeline
■   ■   ■■■ text_filer.py                # Text filtering utilities
■   ■
■   ■■■ trained_models/                  # Trained models and metadata
■       ■■■ classifier.pkl               # Classifier model
■       ■■■ vectorizer.pkl               # Vectorizer model
■       ■■■ encoder.pkl                  # Encoder model
■
■■■ dist/                                # Distribution packages (generated)
■
■■■ build/                               # Build files (generated)
■
■■■ sentiment_prediction.egg-info/       # Egg-info metadata (generated)
■
■■■ .gitignore                           # Excluded files and folders
■■■ MANIFEST.in                          # Configuration file with paths and constants
■■■ README.md                            # Project documentation
■■■ requirements.txt                     # Project dependencies
■■■ setup.py                             # Package metadata and configuration
```

Contribution

Feel free to submit issues or pull requests. Contributions are welcome!

License

This project is licensed under the MIT License.

Author

Vijay Dipak Takbhate

GitHub Repository: Twitter Sentiment Prediction Python Package

---