README File

Sentiment Analysis of Reviews using Python

Project Description

This project aims to analyze customer reviews to understand sentiment and quality perception of products based on user-generated content. By identifying patterns and trends, classifying sentiments, and providing actionable insights, businesses can improve customer satisfaction and product quality.

Objectives

1. **Detailed Analysis of Reviews Data**
    - Trend Analysis
    - Pattern Recognition
    - Rating and Helpfulness Analysis
2. **Sentiment Classification**
    - Development of a Sentiment Analysis Model
    - Model Evaluation and Selection
    - Integration of Sentiment Analysis
3. **Business Application of Insights**
    - Feedback Loop for Product Improvement
    - Customer Service Enhancement
    - Market Strategy Adaptation

Data Description

The dataset can be found at the link : https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews

The dataset consists of 5,68,411 rows and 10 columns, including unique identifiers for reviews, products, and users, as well as textual data for reviews and summaries. The columns are:

- **Id:** Unique identifier for each review.
- **ProductId:** Unique identifier for the product being reviewed.
- **UserId:** Identifier for the user who wrote the review.
- **ProfileName:** Name of the user profile.
- **HelpfulnessNumerator:** Number of users who found the review helpful.
- **HelpfulnessDenominator:** Number of users who indicated whether they found the review helpful or not.
- **Score:** Rating given to the product by the reviewer.
- **Time:** Timestamp when the review was posted.
- **Summary:** Summary of the review.
- **Text:** Full text of the review.

Installation Instructions

Ensure you have the following libraries installed. You can install them using pip:

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

Installation Links for Required Packages

Below are the links for details and commands (if required) to install the necessary Python packages:

- pandas: Go to https://pypi.org/project/pandas/ or use command: pip install pandas
- numpy: Go to https://pypi.org/project/numpy/ or use command: pip install numpy
- matplotlib: Go to https://pypi.org/project/matplotlib/ or use command: pip install matplotlib
- seaborn: Go to https://pypi.org/project/seaborn/ or use command: pip install seaborn
- scikit-learn: Go to https://pypi.org/project/scikit-learn/ or use command : pip install scikit-learn
- nltk: Go to https://pypi.org/project/nltk/ or use command: pip install nltk

Usage

Clone the repository and navigate to the project directory:

```
git clone <repository_url>
cd <repository_directory>
```

Run the Jupyter Notebook to perform the analysis:

```
jupyter notebook final_project_reviews.ipynb
```

Insights and Analysis

The project involves the following steps:

1. **Data Preprocessing**
    - Loading and inspecting data
    - Cleaning data by handling missing values and removing duplicates
    - Feature engineering to create useful metrics
2. **Exploratory Data Analysis (EDA)**
    - Visualizing distributions and patterns
    - Generating word clouds for positive and negative reviews
3. **Text Preprocessing**
    - Tokenization, stop words removal, lemmatization
    - Vectorization using Count Vectorizer and TF-IDF
4. **Model Training and Evaluation**
    - Training models like Logistic Regression and Naive Bayes
    - Evaluating models using accuracy, precision, recall, and F1-score
5. **Business Applications**
    - Providing actionable insights for product improvement and customer service enhancement

Conclusion

The sentiment analysis project effectively classifies customer reviews, providing valuable insights into customer satisfaction. The Logistic Regression model with Count Vectorizer stands out for its balanced performance, making it suitable for practical deployment in sentiment analysis tasks.

Contact

For any questions or suggestions, please contact:

Tharakh George Chacko
Email: tharakhgeorge@yahoo.co.in
LinkedIn: www.linkedin.com/in/tharakh-george-chacko-27a50051