

Sentiment Analysis Project

September 8, 2024

1 Project Overview

This project demonstrates the use of various sentiment analysis techniques on a dataset of product reviews. It explores both classical and modern NLP models to analyze the sentiment of customer reviews.

1.1 Data Collection and Basic Analysis

Loaded and performed basic operations on the dataset.

1.2 Exploratory Data Analysis

Visualized various aspects of the data to uncover insights.

1.3 Sentiment Analysis

- **NLTK (VADER):** Performed sentiment scoring using VADER.
- **Transformer (RoBERTa):** Applied a pre-trained transformer model (RoBERTa) for sentiment scoring.

1.4 Model Comparison

Compared the results from VADER and RoBERTa models.

1.5 Case Studies

Reviewed examples of specific sentiment scores, including discrepancies (e.g., negative sentiment for a 5-star review).

2 Installation

To run this notebook, you'll need the following Python packages installed:

```
!pip install numpy
!pip install pandas
!pip install matplotlib
!pip install seaborn
!pip install nltk
!pip install tqdm
!pip install transformers
!pip install scipy
!pip install torch
```

3 How to Run

1. Clone this repository:

```
git clone https://github.com/your-username/sentiment-analysis.git
```

2. Install the dependencies:

```
pip install -r requirements.txt
```

3. Open the `sentiment.ipynb` notebook and run the cells in sequence.

4 Dataset

The dataset used for this project is **Amazon Fine Food Reviews**, which can be found on Kaggle. Due to its large size, it is not included in this repository, but it can be accessed and downloaded from this link: <https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews>.

Load the dataset using the following command:

```
review = pd.read_csv('Reviews.csv')
```

5 Key Sections

- **VADER Sentiment Scoring:** This section uses VADER from the NLTK library to score the reviews based on sentiment.
- **RoBERTa Pre-trained Model:** This section applies the RoBERTa model to compute sentiment scores using the Hugging Face Transformers library.
- **Comparison of Models:** Sentiment scores from both models are compared and visualized.

6 Results

Visualizations and analyses reveal differences in how the models score sentiments for the same reviews, with a detailed comparison of results for specific examples.

7 Libraries Used

- Numpy
- Pandas
- Matplotlib
- Seaborn
- NLTK
- Transformers (Hugging Face)
- TQDM
- Torch
- Scipy

8 Conclusion

This project showcases the use of both rule-based and deep learning-based sentiment analysis models. It compares the results of each approach and explores specific cases where they differ.