Darshan K M
PA2312052010003

Report on Sentiment Analysis Model

Objective

The primary goal of the project was to scrape customer reviews from Amazon, perform sentiment analysis using a fine-tuned BERT-based model, and derive meaningful insights and engagement metrics for A/B testing.

Data Collection

- Customer reviews for Amazon products were scraped using Selenium and BeautifulSoup.
- Reviews were extracted along with metadata such as username, review date, rating, and review text.
- Data was saved to a CSV file **(amazon_reviews.csv)** for preprocessing and analysis.

Model Details

Architecture

- **Base Model:** BERT (bert-base-uncased)
- **Head:** Sequence classification layer with three labels (positive, neutral, negative).

Training Data

- **Dataset:** A cleaned CSV file **(train2.csv)** containing customer review texts and their sentiment labels.
- **Preprocessing:**
    - Removed URLs and mentions using regex.
    - Mapped sentiment labels to numerical values: positive: 2, neutral: 0, negative: 1.
- **Splitting:** Train-test split (80%-20%) for model training and evaluation.

Training Details

- **Tokenizer:** BERT tokenizer with truncation and padding to a max length of 128 tokens.
- **Loss Function:** SparseCategoricalCrossentropy (from logits).
- **Optimizer:** Adam optimizer with a learning rate of 2e-5.
- **Batch Size:** 16
- **Epochs:** 3

Model Performance

Classification Report:

Class	Precision	Recall	F1-Score	Support
Positive	0.86	0.83	0.84	2549
Neutral	0.88	0.86	0.87	1765
Negative	0.84	0.91	0.87	1889

- **Accuracy:** 85.99%
- **Macro Average F1-Score:** 0.86
- **Weighted Average F1-Score:** 0.86

Key Observations:

- The model performs consistently across all classes with F1-scores between 0.84 and 0.87.
- The model exhibits high recall for negative sentiment (~91%), suggesting strong sensitivity in identifying negative reviews.
- Slightly lower recall for positive sentiment (~83%) indicates potential improvement areas.

A/B Testing and Engagement Metrics

Methodology

- Users were split into two groups (A and B) based on their index in the dataset:
    - Group A: Even indices
    - Group B: Odd indices

Engagement Metrics

Group A:

- Average Rating: 3.98
- Review Count: 50
- Positive Reviews: 31
- Negative Reviews: 2
- Neutral Reviews: 17

Group B:

- Average Rating: 4.00
- Review Count: 50
- Positive Reviews: 24
- Negative Reviews: 3

- Neutral Reviews: 23

Statistical Analysis

- **T-Test Results:**
    - **T-Statistic:** -0.3747
    - **P-Value:** 0.7087

Inference:

- The p-value (>0.05) indicates no statistically significant difference between the engagement of Group A and Group B.
- Both groups exhibited similar engagement metrics, suggesting no observable bias or effect from group splitting.

Predictions and Insights

Sentiment Prediction

- The trained model was used to predict sentiments for the scraped reviews.
- Sentiment predictions were mapped back to the original dataset and saved in **updated_reviews.csv.**

Insights Derived

- The proportion of sentiments across reviews provided insights into customer sentiment distribution.
- Average ratings and sentiment proportions helped quantify user satisfaction.

Conclusion

The sentiment analysis model demonstrated robust performance across all sentiment classes, with an overall accuracy of ~86%. The A/B testing analysis revealed no significant differences between groups, indicating consistent engagement. Future work can focus on addressing improvement areas and expanding the dataset for better generalization and scalability.