Customer Review Sentiment Classification
Springboard Data Science Capstone Project

Author: Abhinaya Gyawali
Program: Springboard Data Science Career Track

1. Problem Statement

Businesses receive large volumes of unstructured customer reviews, making manual sentiment analysis slow, inconsistent, and impractical. This project develops an automated NLP-based sentiment classification system that categorizes reviews into positive, neutral, and negative sentiments to support scalable, data-driven decision-making.

2. Data Overview

The Amazon US Customer Reviews dataset was used for this project. Review star ratings were converted into sentiment labels, and extensive preprocessing was applied to clean noisy text, handle missing values, and manage class imbalance.

3. Methodology

Exploratory Data Analysis examined class distributions, review lengths, and text characteristics. Text preprocessing included normalization, tokenization, and TF–IDF vectorization. Three classical machine learning models were trained and compared: Logistic Regression, Support Vector Machine (SVM), and Multinomial Naive Bayes. Model selection focused on balanced performance across sentiment classes using cross-validated accuracy and macro F1-score.

4. Model Evaluation & Final Model Selection

Among the evaluated models, **Logistic Regression with TF–IDF features** was selected as the final model. It provided the best balance of classification performance, interpretability, and computational efficiency, with stable cross-validation results across sentiment classes.

5. Business Insights

Analysis of model outputs shows a clear relationship between predicted sentiment and customer satisfaction levels. Negative sentiment reviews are concentrated among lower star ratings, providing scalable indicators of dissatisfaction. Positive sentiment reviews dominate the dataset, while neutral sentiment serves as a useful early warning signal when monitored over time.

6. Recommendations

1. Deploy the Logistic Regression sentiment classifier to tag incoming reviews at scale.
2. Prioritize negative sentiment reviews for customer support follow-up.
3. Monitor sentiment trends over time to inform product and marketing decisions.

7. Limitations & Future Work

The TF–IDF–based Logistic Regression model does not capture full linguistic context, limiting performance on nuanced language such as sarcasm or negation. Sentiment labels derived from star ratings introduce some label noise, and neutral sentiment remains challenging due to linguistic overlap with positive and negative reviews. Future work includes fine-tuning transformer-based models and applying topic modeling to deepen insight from negative reviews.

Conclusion

This project demonstrates how NLP-driven sentiment classification can transform large volumes of unstructured customer feedback into reliable, scalable sentiment signals. The final Logistic Regression model balances strong performance with interpretability, making it suitable for real-world deployment.