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 aims to build an automated Natural Language Processing (NLP) model that classifies customer reviews into positive, neutral, or negative sentiments to support faster, data-driven decision-making.

2. Data Overview

The project uses the Amazon US Customer Reviews dataset from Kaggle. Sentiment labels were derived from star ratings, with preprocessing steps applied to clean noisy text, remove missing values, and address class imbalance.

3. Methodology

Exploratory Data Analysis was conducted to understand class distribution, review lengths, and noise patterns. Text preprocessing included normalization, tokenization, and vectorization using TF–IDF. Multiple models were trained, including Logistic Regression and transformer-based approaches, and evaluated using accuracy and macro F1-score.

4. Model Evaluation & Results

The final selected model achieved strong performance across sentiment classes, meeting the project success criteria of balanced classification. Evaluation metrics demonstrated the model's ability to generalize across positive, neutral, and negative reviews while maintaining efficiency.

5. Business Insights

Analysis revealed that negative reviews often focus on delivery issues, product quality concerns, and customer service problems. Positive reviews emphasized value for money and product reliability. These insights enable targeted improvements across teams.

6. Recommendations

1. Integrate sentiment monitoring dashboards for real-time customer feedback tracking.
2. Route negative sentiment reviews directly to customer support teams for rapid response.
3. Use aggregated sentiment trends to inform product improvement and marketing strategies.

7. Limitations & Future Work

The model may struggle with sarcasm, slang, or domain-specific language. Future work could include fine-tuning transformer models, topic modeling on negative reviews, and extending the approach to multilingual datasets.

8. Conclusion

This project demonstrates how NLP-driven sentiment classification can transform unstructured customer feedback into actionable business insights. The final model provides a scalable solution for organizations seeking to improve customer satisfaction and operational efficiency.