Sentiment Analysis on Customer Feedback

Group 23

Group Members

21903488 Emrehan Hoşver, 21901798 Gökberk Altıparmak, 21802480 Tuğberk Dikmen, 21901418 Efser Efe Kantik

Description of the Datasets

Customer Feedback Dataset: This dataset is sourced from Kaggle, consisting of unconstructed text data for customer feedback and reviews from an e-commerce platform, accompanied by sentiment labels, making it suitable for sentiment analysis. The goal is to classify the customer feedback into sentiment categories such as positive, negative, or neutral. This dataset will serve as our primary source for training and evaluation [1].

Brazilian E-commerce Dataset: Another Kaggle dataset, it includes customer reviews and comments pertaining to e-commerce transactions in Brazil. These reviews can provide supplementary data for sentiment analysis and help gain insights into customer sentiment within a specific geographic context [2].

Jigsaw Toxic Comment Classification Challenge: While initially designed for toxic comment classification, this Kaggle dataset comprises text data with associated labels indicating toxicity. We can adapt this dataset to perform sentiment analysis by redefining the sentiment labels. This versatile dataset allows us to explore sentiment in a different context, beyond standard customer feedback [3].

Problem Definition

The central question this project aims to address is: "How can sentiment analysis of customer feedback enhance e-commerce operations and customer satisfaction?" We will leverage machine learning techniques to classify customer feedback into sentiment categories and identify patterns and factors influencing sentiment. The goal is to provide actionable insights for improving customer satisfaction and making data-driven business decisions.

Plans for the Project

For dataset processing, we will perform data cleaning to remove irrelevant entries and apply text preprocessing techniques, including tokenization, stop-word removal, and label mapping to ensure a consistent sentiment classification format.Regarding model selection, we plan to explore a combination of machine learning and deep learning models: Naive Bayes and Support Vector Machine (SVM) for traditional methods, and Transformer-based Models (BERT) and Recurrent Neural Network (RNN) to capture contextual and sequential aspects of the text. Model performance will be evaluated using standard classification metrics.

Future Work

Real-time Sentiment Analysis: Developing a real-time sentiment analysis system to process incoming customer feedback and generate instant insights for timely interventions.

Continual Model Improvement: Regularly updating and improving the sentiment analysis model to adapt to evolving customer feedback trends and linguistic changes.

References

[1] V.Salodkar, “Customer Feedback Dataset, ” Kaggle. [Online]. Available: https://www.kaggle.com/datasets/vishweshsalodkar/customer-feedback-dataset [Accessed: Oct 19, 2023].

[2] Olist, “Brazilian E-Commerce Public Dataset by Olist, ” Kaggle. [Online]. Available: https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce [Accessed: Oct 19, 2023].

[3]J. Peller, “jigsaw-toxic-comment-classification-challenge, ” Kaggle. [Online]. Available:https://www.kaggle.com/datasets/julian3833/jigsaw-toxic-comment-classification-challenge/code [Accessed: Oct 19, 2023].