

Customer Sentiment and Review Helpfulness Analysis in E-Commerce Product Reviews

Mr. Deep Vinodbhai Darji

Student, Department of MSc. IT,

Nagindas Khandwala College, Mumbai, Maharashtra, India

deepdarji425@gmail.com

Abstract: *Understanding customer opinions expressed through online product reviews is critical for evaluating product quality, predicting user ratings, and improving decision-making in e-commerce platforms. This study presents a comprehensive analysis of customer review data by integrating sentiment analysis, linguistic feature extraction, and machine learning techniques within a Google Colab based analytical framework. The research primarily investigates how linguistic and sentiment-driven characteristics influence rating behavior, while also examining the impact of product attributes, reviewer patterns, and temporal dynamics on review helpfulness and overall review quality. The dataset is systematically cleaned and preprocessed to remove noise, handle missing values, and standardize textual content. Natural language processing techniques are employed to extract sentiment and subjectivity scores, along with linguistic features such as word count, positive and negative term frequency, and review length. Numerical ratings are transformed into categorical rating classes to facilitate supervised classification. Feature engineering is performed to enhance representational strength, enabling the models to capture both emotional intensity and informational richness embedded within customer reviews.*

Multiple machine learning classification models, including Logistic Regression, Naïve Bayes, Support Vector Machine, Decision Tree, and Random Forest, are trained and evaluated to assess their effectiveness in predicting rating classes and identifying high-quality reviews. Model performance is measured using accuracy, precision, recall, F1-score, and confusion matrix analysis. Visual analytics techniques are further applied to explore rating distributions, sentiment variations, and reviewer behavior patterns over time. The experimental results demonstrate that sentiment polarity and linguistic richness strongly influence rating behavior, with positive sentiment and detailed textual content closely associated with higher ratings. Additionally, review helpfulness is significantly affected by review length, temporal positioning, and consistency in reviewer behavior rather than sentiment alone. Ensemble-based models consistently outperform linear classifiers, highlighting their ability to capture complex, non-linear relationships within review data..

Keywords: *Customer Reviews, Sentiment Analysis, Rating Prediction, Review Helpfulness, Natural Language Processing, Machine Learning, E-commerce Analytics, Linguistic Features*

