

## **Predicting Numerical Ratings from E-commerce Reviews Using Advanced Machine Learning Models**

**Abstract:** This study investigates the application of advanced machine learning techniques for sentiment analysis, specifically aimed at predicting numerical customer ratings from e-commerce reviews. Utilizing the "Amazon Product Review Dataset" from the University of California, San Diego, I explore the performance of Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and Transformer models. The data preprocessing included text cleaning to remove non-textual elements and handling missing values, followed by model training and systematic evaluation using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). My findings indicate that the LSTM model outperforms other models in capturing long-range dependencies in textual data, providing more accurate sentiment predictions. The CNN model also demonstrated strong performance by effectively identifying local patterns in text. Although the Transformer model showed potential, it did not perform as well as expected, highlighting the need for further tuning. Additionally, this study delves into the analysis of specific words and features that significantly influence customer ratings, offering valuable insights into consumer satisfaction and preferences. For instance, words like "comfort" and "price" were found to have a substantial impact on review scores. This research contributes to the field by demonstrating the capabilities and limitations of each model, paving the way for more nuanced and actionable sentiment analysis tools that can help businesses better understand and enhance consumer satisfaction. Future work will focus on refining these models, exploring additional features, and incorporating more sophisticated attention mechanisms to further improve prediction accuracy and model interpretability.