

Introduction

We have analysed the *Amazon Reviews*²³¹ dataset providing 571.54M reviews from May 1996 - September 2023 across 34 categories (including one “unknown”) to answer the following 4 questions: (1) Are reviews for some categories of product on Amazon overall more positive than for other Categories? (2) Are reviews more subjective for some classes of products than for others? (3) Which aspects of different classes of products are the most important in the reviews? (4) Can one predict the star rating from the review text?

Insights / Results

(1) The analysis shows that certain product categories consistently receive more positive sentiment in reviews than others. Digital Music, Handmade Products, Gift Cards, and Books lead with the highest positive sentiment scores, while Beauty and Personal Care and Automotive categories show the lowest sentiment scores. However, the differences between categories are relatively modest, with most clustering near the global median sentiment score of 0.74 [Fig. 1]. (2) Furthermore we saw that Home/Kitchen products, Digital Music, and Gift Cards show the highest levels of subjective language in reviews, while Electronics, Tools, and Musical Instruments demonstrate more objective review content. Notably, all categories show consistently high subjectivity scores. [Fig. 2]

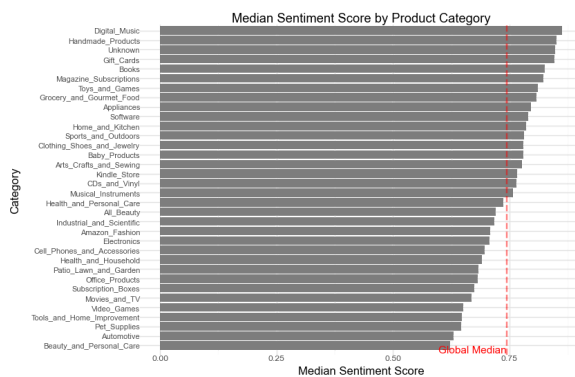


Fig. 1: Sentiment Analysis Results

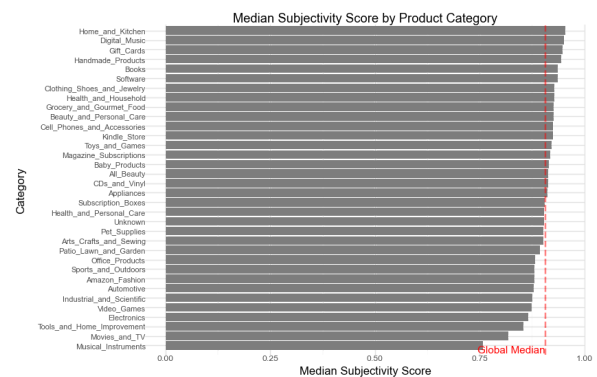


Fig. 2: Subjectivity Analysis Results

(3) Thirdly, we failed to extract any meaningful product characteristics that reviews primarily focus on. Instead aspect extraction surfaced generic descriptors across all categories both globally as well as under category specific rankings. [Fig. 3]

¹ <https://huggingface.co/datasets/McAuley-Lab/Amazon-Reviews-2023>

(4) Finally, using another discrete sentiment classification model we were able to predict product ratings from review text with approximately 68% accuracy. The model performs best with 5-star ratings but struggles more with 1-star and 4-star ratings. Those results need further validation to outrule any potential on overfitting as the underlying model has been fine tuned on another dataset² of Amazon reviews. [Fig. 4]

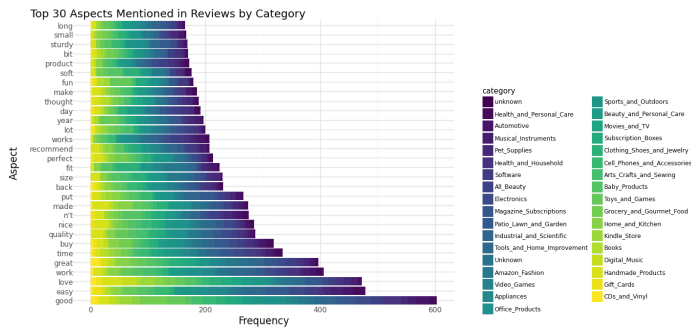


Fig. 3: Keyword Extraction Results

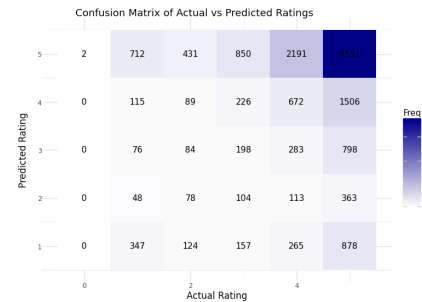


Fig. 4: Rating Prediction Results

Methodology We relied on various pretrained and fintuned BERT-models for sentiment analysis³, subjectivity analysis⁴ and rating prediction⁵ tasks. Due to their computational requirements we needed to randomly sample 1000 reviews per category for our analysis - Additionally we ran it on a Google Colab T4 instance. For textual analysis we removed HTML tags from text and title fields, stripped whitespaces, eliminated reviews with empty text or titles and dropped all columns besides the numerical rating, the reviews text and title as well as its category. No further processing has been applied.

Conclusion We successfully answered 3 out of 4 research questions using a combination of NLP techniques and data analysis. The results provided valuable insights into review sentiment, subjectivity and star rating prediction, highlighting the complexities and nuances of consumer feedback on Amazon.

² <https://www.kaggle.com/datasets/cynthiarempel/amazon-us-customer-reviews-dataset>

³ <https://huggingface.co/LiYuan/amazon-review-sentiment-analysis>

⁴ <https://huggingface.co/cfll/bert-base-styleclassification-subjective-neutral>

⁵ <https://huggingface.co/LiYuan/amazon-review-sentiment-analysis>